

Jetzek, Thorhildur Hansdottir

**Doctoral Thesis**

## The Sustainable Value of Open Government Data: Uncovering the Generative Mechanisms of Open Data through a Mixed Methods Approach

PhD Series, No. 24.2015

**Provided in Cooperation with:**

Copenhagen Business School (CBS)

*Suggested Citation:* Jetzek, Thorhildur Hansdottir (2015) : The Sustainable Value of Open Government Data: Uncovering the Generative Mechanisms of Open Data through a Mixed Methods Approach, PhD Series, No. 24.2015, ISBN 9788793339316, Copenhagen Business School (CBS), Frederiksberg, <https://hdl.handle.net/10398/9183>

This Version is available at:

<https://hdl.handle.net/10419/208941>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc-nd/3.0/>

**COPENHAGEN BUSINESS SCHOOL**  
SOLBJERG PLADS 3  
DK-2000 FREDERIKSBERG  
DANMARK

**WWW.CBS.DK**

**ISSN 0906-6934**

**Print ISBN: 978-87-93339-30-9**  
**Online ISBN: 978-87-93339-31-6**

THE SUSTAINABLE VALUE OF OPEN GOVERNMENT DATA

Thorhildur Hansdottir Jetzek

# THE SUSTAINABLE VALUE OF OPEN GOVERNMENT DATA

**UNCOVERING THE GENERATIVE MECHANISMS OF OPEN  
DATA THROUGH A MIXED METHODS APPROACH**

The PhD School of LIMAC

PhD Series 24.2015

**CBS**  COPENHAGEN BUSINESS SCHOOL  
HANDELSHØJSKOLEN

PhD Series 24-2015

# **The Sustainable Value of Open Government Data**

**Uncovering the Generative Mechanisms of Open Data  
through a Mixed Methods Approach**

**Thorhildur Hansdottir Jetzek**

**Main supervisor: Niels Bjørn-Andersen**

**Secondary supervisor: Michel Avital**

Submitted for the completion of a PhD degree 30. June 2015 at

**LIMAC PhD School  
Department of IT Managememe  
Copenhagen Business School**

Thorhildur Hansdottir Jetzek  
*The Sustainable Value of Open Government Data*  
*Uncovering the Generative Mechanisms of Open Data through a Mixed Methods Approach*

1st edition 2015  
PhD Series 24.2015

© Thorhildur Hansdottir Jetzek

ISSN 0906-6934

Print ISBN: 978-87-93339-30-9

Online ISBN: 978-87-93339-31-6

LIMAC PhD School is a cross disciplinary PhD School connected to research communities within the areas of Languages, Law, Informatics, Operations Management, Accounting, Communication and Cultural Studies.

All rights reserved.

No parts of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without permission in writing from the publisher.

## **Preface**

In June 2011, I stood in Kastrup Airport with my three daughters, coming to Denmark to stay. My husband had already arrived; my dog was on the way. Both my husband and I had left good careers and a great life in order to allow me to fulfill my dream of pursuing a PhD. I have to say, I was very apprehensive at this moment. My (then) 16-year-old daughter had decided to stay with her father and continue her education in Iceland, and the prospect of being separated from her was next to unbearable. Moreover, I had absolutely no idea how to get into a PhD program. Prior to coming to Denmark, I had sent a couple of emails and gotten quite unenthusiastic responses. Fortunately, a few days before leaving I happened to run into a former colleague, Agnar Hanson. Agnar offered to introduce me to a professor he had collaborated with at CBS, Niels Bjørn-Andersen.

To get an opportunity to meet with Niels Bjørn was truly a turning event for me. Without his kind nature, willingness to help, great network and clear insight into interesting research areas I suppose I would never even have started. After meeting him, it only took a couple of meetings before we had ensured the financial support of a large IT vendor in Denmark, KMD. We resolved that I would study a very interesting e-Government program in Denmark, the Basic Data Program (BDP). The BDP is an Open Data Initiative, led by the Danish Agency for Digitization, which also agreed to participate as a third party. After “only” a couple of hundreds of emails, everything was in place, and we sent an application to the Industrial PhD fund. The (second) application was approved in June 2012, exactly a year after my arrival in Denmark. July 1<sup>st</sup> 2012, I started to work on my dissertation called The Sustainable Value of Open Government Data.

These last three years have been the most challenging but also most rewarding years of my carrier. It is a bit like running an intellectual marathon, and I did hit some walls on the way. Suffice to say, this is not a tale of one woman’s journey. I have so many people to thank who have helped me on the way, and all of them have provided something unique to this project. Thus, I will go through my list of acknowledgements in no particular order.

The only one who will get a special place is Professor Niels Bjørn-Andersen, as without him this project would never have been born. He spent an incredible amount of energy and time to get me to the starting line, which is not a very rewarding process. I hope he knows that I have not, and never will, forget his help and his kindness. He also

took me on as a main supervisor in spite of my subject not being a part of his main research area. Niels's great experience and knowledge in the field of IS has proven to be an invaluable asset to me personally and this dissertation in particular. A bit later in the process, I struck gold again, when I got the chance to get Professor Michel Avital as my secondary supervisor. Michel has this amazing ability of being able to identify the main structural elements of every model and every topic – an ability I can only hope to acquire later in my career. As Niels, Michel has a wide and deep IS knowledge, from which I have benefitted immensely. I want to express my sincerest gratitude to both of them for their help and support and their generous knowledge sharing.

At KMD I have had two company supervisor, although not simultaneously. Morten Binderup was my mentor throughout the application process and for the first half of my dissertation. He is an extremely likable and organized manager who helped me get all the practical elements in place (a surprisingly difficult but necessary part of the PhD). During my PhD course, KMD decided to strengthen their focus on data related solutions, and a new department was born: Grunddata. The manager of Grunddata, Ruth Wisborg, also became my new supervisor. She is a great leader, and I have learned very much from her open and collaborative approach. I also want to thank Ole Jensen, director at KMD and a member of my steering committee, for his valuable support. For the last year, I have had the pleasure of sitting next to the inspiring Nicolas Lemcke Horst who has been developing a new data strategy at KMD. He has achieved so much in less than a year, and I am proud of having been, if only a small, contributor to this strategy. I have also enjoyed working with Anne Juel Jørgensen on an EU Horizon 2020 open data proposal, a huge learning experience. Finally, I must mention my KMD swimming buddies, especially Hanne Vallerbæk Johne who introduced me to triathlons. I believe the training has contributed greatly to my sanity 😊

At the Agency for Digitization, Lars Frelle Petersen made a huge difference by supporting the project. He is one of those visionary leaders who are transforming digital governance in Denmark. I also had third-party supervisor from the Agency, Jens Krieger-Røyen. What an amazing person! Not only is he extremely nice and has always cheered me on and made me feel appreciated, but he is also a gifted leader who has managed to drive forth the program I am writing about, The Basic Data Program. I have observed him and his teammates through the good times and the bad. There have been some major challenges on the way, but he always keeps his smile, his optimism and his unblinking belief that the basic data they are transforming into a strategic resource will someday be a fundamental foundation of Danish society. Unfortunately, I

cannot mention by name all the people that have helped me in the Agency for Digitization; Geodata Agency; Ministry for Housing, Urban and Rural Affairs; The Danish Business Authority and other places where I have gone for interviews and other sources of data. Very driven group of people, and I admire them all for their structured approach and their willingness to collaborate and share.

I must extend my gratitude to my colleagues at the Department of IT Management, especially head of department Jan Damsgaard, who is running a world-class IS department in small Denmark. My fellow doctoral students, who have given me so much needed emotional support. The ITM department secretariat, headed by Bodil Sponholz who are always ready to help, and the LIMAC people, especially Annie Olsen. I am very grateful to Chee-Wee Tan and Marijn Janssen who gave valuable comments for improvement of the thesis draft during my WIP 2 seminar. And last but not least I must mention the assessment committee that consists of three great scholars; Helle Zinner Henrikssen (CBS), John Leslie King (University of Michigan) and Matti Rossi (Aalto University). I am honored to have them read and approve of this dissertation. I also want to thank all the anonymous reviewers and editors of published articles, who contributed valuable ideas for improvement.

I want to thank my husband, Magnús Böttvar Eyþórsson, who left his job and his son in Iceland and encouraged me to follow my life-long dream. He has been willing to listen to endless hours of monologue about my research interests, give feedback to my presentations and read my academic papers. My children have always inspired me. I left my lovely first-born Álfheiður María in Iceland at 16 years of age but throughout she has encouraged me and told me how proud she is of PhD mom. She has grown into a lovely young woman and I am so proud of her. My second daughter, Ásta Björk, who came with her mother to Denmark at the fragile age of 13, just because she knew it would break my heart to leave both of them behind. She keeps inspiring me with her intelligence, her views and her opinions. Þórey Margrét my youngest is my little ray of sunshine. Always smiling, always happy (despite her grumpy mother), always curious and ready to learn. She reminds me that we must explore the world and never settle for how things are. We can always do better. My stepson Eyþór, who has experience beyond his years from travelling around the world, and who keeps opening my eyes to so many things.

Finally, friendship is necessary for everyone. I want to thank my group of friends, here in Denmark and in Iceland. Sometimes it is important to remember that life has more to offer than sitting in front of a computer ☺

To all of you, THANK YOU! If not for all these wonderful people, I doubt I would have managed to finish this journey successfully.



## English Abstract

The impact of the digital revolution on our societies can be compared to the ripples caused by a stone thrown in water: spreading outwards and affecting a larger and larger part of our lives with every year that passes. One of the many effects of this revolution is the emergence of an already unprecedented amount of digital data that is accumulating exponentially. Moreover, a central affordance of digitization is the ability to distribute, share and collaborate, and we have thus seen an “open theme” gaining currency in recent years.

These trends are reflected in the explosion of Open Data Initiatives (ODIs) around the world. However, while hundreds of national and local governments have established open data portals, there is a general feeling that these ODIs have not yet lived up to their true potential. This feeling is not without good reason; the recent Open Data Barometer report highlights that strong evidence on the impacts of open government data is almost universally lacking (Davies, 2013). This lack of evidence is disconcerting for government organizations that have already expended money on opening data, and might even result in the termination of some ODIs. This lack of evidence also raises some relevant questions regarding the nature of value generation in the context of free data and sharing of information over networks. Do we have the right methods, the right intellectual tools, to understand and reflect the value that is generated in such ecosystems?

This PhD study addresses the question of *How is value generated from open data?* through a mixed methods, macro-level approach. For the qualitative analysis, I have conducted two longitudinal case studies in two different contexts. The first is the case of the Basic Data Program (BDP), which is a Danish ODI. For this case, I studied the supply-side of open data publication, from the creation of open data strategy towards the dissemination and use of data. The second case is a demand-side study on the energy tech company Opower. Opower has been an open data user for many years and have used open data to create and disseminate personalized information on energy use. This information has already contributed to a measurable world-wide reduction in CO<sub>2</sub> emissions as well as monetary savings. Furthermore, to complement the insights from these two cases I analyzed quantitative data from 76 countries over the years 2012 and 2013. I have used these diverse sources of data to uncover the most important relationships or *mechanisms*, that can explain how open data are used to generate sustainable value.

I conceptualize liquid open data as a multi-dimensional construct, consisting of seven different dimensions. Moreover, I propose that when data become open across more of these dimensions, the opportunity for new use, and subsequent value generation, will increase. Use of data also depends on other factors, at both the societal and the organizational level. Citizens and companies must have the ability to generate value from data. The most relevant societal level enabling factors are having access to low-cost, high-speed networks and a skilled workforce. The soft infrastructure is also important for supporting sustainable value generation through open data, especially the existence of robust regulatory data and privacy protection frameworks and governmental leadership. At the organizational level, I recognize the importance of absorptive capacity. Absorptive capacity defines organizations ability to recognize the value of external data and information, assimilate them, and apply them to commercial ends.

Value generation from data can happen through the markets via mechanisms like efficiency and innovation. Alternatively, value generation can happen through a class of value generating mechanisms I call *information sharing mechanisms*. I propose this as a new archetype of value generating mechanisms that is becoming more relevant for organizations operating in increasingly networked societies. Moreover, I propose that organizations that are effectively utilizing Multi-Sided Platforms (MSPs) are in fact capitalizing on the synergies between the information sharing mechanism and market mechanisms for superior value generation. However, I also propose that we currently lack the right tools to make much of the resulting value explicit, which has resulted in marginal interest in open data from the private side and under-investment from the public sector side.

Thus, I find it is extremely important for stimulating cross-boundary generation of sustainable value that we strive to understand how value is generated through open digital resources. Otherwise, we might miss an unprecedented opportunity for a positive paradigm change. We must untangle and clarify new concepts, emphasize the most important constructs and underlying relationships and create a holistic map of their relation to each other. I thus propose that we need a new mid-range theory that can explain how value is generated in an open data ecosystem. I hope that this PhD study has made a contribution towards creating such a theory. Moreover, to advance further research in this emerging but relevant field of study, I propose a research agenda by the end of this study.

## Dansk Abstrakt

Virksomheden af den digitale revolution på vore samfund kan sammenlignes med ringe forårsaget af en sten der kastes i vand: de spreder udad og påvirker en større og større del af vores liv med hvert år der går. En af de mange virkninger af denne revolution er fremkomsten af en allerede hidtil uset mængde digitale data, der akkumuleres eksponentielt. Desuden er evnen til at distribuere, dele og samarbejde en central affordance af digitalisering. Vi har således set en stærkt øget interesse i "åben temæt" i de seneste år.

Disse tendenser afspejles i en eksplosion i antallet af åbne data initiativer rundt omkring i verden. Men mens hundredvis af nationale og lokale regeringer har etableret åbne data portaler, så observerer vi, at disse initiativer endnu ikke har levet op til deres sande potentiale. Den seneste Open Data Barometer rapport fremhæver, at stærke beviser på virkningerne af åbne offentlige data mangler i næsten universelt grad (Davies, 2013). Denne mangel på beviser er foruroligende for offentlige organisationer, der allerede anvender betydelige beløb på åbne data, og det kan måske endda resultere i suspension af nogle åbne data initiativer. Denne mangel på beviser rejser også nogle relevante spørgsmål om hvordan værdiskabelse foregår i forbindelse med fri data og udveksling af oplysninger via netværk. Har vi de rigtige metoder og de rigtige intellektuelle værktøjer, som vi behøver for at forstå og reflektere den værdi, der er genereret i sådanne økosystemer?

Denne ph.d.-afhandling omhandler spørgsmålet om *Hvordan skaber åbne data værdi?* Spørgsmålet belyses gennem brug af mixed-methods på makroniveau. Til kvalitativ analyse, har jeg gennemført to casestudier i to forskellige sammenhænge. Den første handler om Grunddataprogrammet, som er det danske åbne data initiativ, der allerede i 2013 frikøbte relevante data som geografisk data og CVR data. I dette case studerede jeg udbudssiden af åbne data, lige fra oprettelsen af åbne data strategi over formidling til anvendelse af data. Den anden case er en undersøgelse af energi teknologi virksomheden Opower. Opower bruger data til at skabe gratis personlige oplysninger om energiforbrug, som de stiller til rådighed og udbreder via forskellige digitale kanaler. Denne information har allerede bidraget til en målbar verdensomspændende reduktion i CO<sub>2</sub>-udledningen samt monetære besparelser. For at supplere indsigter fra disse to cases, har jeg analyseret kvantitative data fra 76 lande for årene 2012 og 2013. Jeg har brugt disse forskellige datakilder til at afsløre de vigtigste relationer eller mekanismer, der kan forklare, hvordan åbne data anvendes til skabe bæredygtig værdi.

Jeg konceptualiserer hvad der kaldes 'liquid open data'. Dette er en multi-dimensional konstruktion, der består af syv forskellige dimensioner. Desuden foreslår jeg, at når data bliver åbent på tværs af flere af disse dimensioner, så stiger muligheden for ny anvendelse, og efterfølgende generation af bæredygtig værdi. Anvendelse af data afhænger også af andre elementer både på det samfundsmæssige og organisatoriske niveau. Borgere og virksomheder skal have evnen til at generere værdi fra data. De mest relevante samfundsmæssige faktorer er at man skal have adgang til billigt højhastighedsnet og en kvalificeret arbejdsstyrke. "Soft" infrastruktur er også særdeles vigtigt for at understøtte en bæredygtig generation af værdi gennem åbne data, især eksistensen af solid lovgivningsmæssig beskyttelse af data og privatlivets fred, samt styrke i ledelsen fra regeringen. På det organisatoriske niveau anerkender jeg vigtigheden af organisatorisk absorptionskapacitet. Absorptionskapacitet definerer organisationens evne til at anerkende værdien af eksterne data og oplysninger, assimilere dem, og at anvende dem til kommercielle formål.

Værdiskabelse fra data kan ske via markedsmekanismer som effektivitet og innovation. Alternativt kan generation af værdi ske gennem en klasse af mekanismer, som jeg kalder informationsdelingsmekanismer. Jeg foreslår disse mekanismer som en ny arketype af mekanismer, som skaber værdi, og jeg argumenterer for, at den type bliver stadig mere relevant for organisationer, der opererer i vores netværksbaserede samfund. Desuden foreslår jeg, at organisationer, der effektivt udnytter Multi-Sided Platforms, faktisk udnytter synergien mellem informationsdelingsmekanismerne og markedsmekanismerne for at kunne skabe overlegen generation af værdi. Men jeg konkluderer, at vi mangler de rigtige værktøjer til at gøre en del af den resulterende værdi synlige, hvilket har resulteret i en marginal interesse for åbne data fra den private side og en potentiel underinvestering fra det offentlige side.

Således foreslår jeg, at det er ekstremt vigtigt for at få stimuleret generation af bæredygtig værdi på tværs af grænser, at man forstår, hvordan værdi skabes gennem åbne digitale ressourcer. Ellers kan vi gå glip af en enestående mulighed for et positivt paradigmeskift. Vi bliver nødt til at udrede og afklare nye koncepter, at fremhæve de vigtigste konstruktioner og underliggende relationer og at skabe et sammenhæng omkring disse begrebers relation i forhold til hinanden. Jeg foreslår således, at vi har brug for en ny mid-range teori, der kan forklare, hvordan værdi genereres i åbne data økosystemet. Jeg håber at denne Ph.d. afhandling har bidraget til at skabe sådan en teori, og jeg foreslår desuden en forskningsdagsorden for fortsat forskning på dette fagområde.

# Table of Contents

Preface .....	i
English Abstract .....	v
Dansk Abstrakt .....	vii
Table of Contents .....	ix
List of Tables (Summary paper) .....	xiii
List of Figures (Summary paper) .....	xiv
1. Introduction .....	1
1.1 Problem statement .....	2
1.2 Research Context and Research Questions .....	4
1.3 Research Design .....	6
1.4 Contributions and Future Research .....	9
2. State of the Art .....	13
2.1 Open Data Strategies, Enablers and Barriers .....	16
2.2 Supply Side: Open Data Dissemination .....	17
2.3 Demand Side: Using Open Data .....	20
2.4 Value Generation in a Networked Society .....	22
3. Methodology .....	24
3.1 Research Philosophy .....	24
3.2 Research Design .....	27
3.3 Data Collection .....	31
4. Assumptions and Meta-theory .....	33
4.1 Basic Assumptions .....	34
4.2 Social Mechanisms .....	36
4.3 Coleman’s Framework – The Micro-Macro Conundrum .....	37
5. Towards a Theory of Value Generation through Open Data .....	39
5.1 Seven Dimensions of Liquid Open Data .....	42
5.2 A Framework of Four Mechanisms .....	48
5.3 Sustainable Value .....	54
5.4 Research Model .....	56
5.5 Hypotheses .....	58
5.5.1 Liquid open data .....	58
5.5.2 Robust regulatory data and privacy protection frameworks .....	59
5.5.3 Digital leadership of government .....	60

5.5.4	Cost of high-speed networks.....	61
5.5.5	Ease of reaching a skilled workforce.....	62
5.5.6	Shared digital content dissemination.....	62
5.5.7	New digital product and services.....	63
5.5.8	Accountability of private sector organizations.....	64
5.6	Research Model Estimation.....	66
5.7	The Role of Open Data Intermediaries.....	72
5.8	Explaining the Open Data Value Paradox.....	76
6.	Conclusion.....	77
6.1	Revisiting the Research Questions.....	77
6.2	Contributions.....	82
6.3	Final Contribution: Research Agenda.....	85
6.4	Research Limitations and Validity of Results.....	88
6.5	Final Reflections.....	91
7.	References.....	92
	Appendix A: Summary of the Included Papers.....	103
	Paper I: The Value of Open Government Data: A Strategic Analysis Framework.....	110
I.1	Introduction.....	112
I.2	Open data and government.....	114
I.3	Strategic analysis framework for OGD initiatives.....	115
I.4	Discussion.....	123
I.5	Conclusion.....	125
I	References.....	126
	PAPER II: The Generative Mechanisms of Open Government Data.....	131
II.1	Introduction.....	132
II.2	The unique features of OGD.....	134
II.3	The multifaceted nature of value.....	135
II.4	Research Strategy.....	136
II.5	Findings and Analysis.....	138
II.6	Discussion and Conclusion.....	147
II	References.....	148
	PAPER III: Generating Value from Open Government Data.....	151
III.1	Introduction.....	153
III.2	Theoretical background and model development.....	154

III.3	Measurements and data collection .....	167
III.4	Analysis and findings .....	171
III.5	Discussion .....	177
III.6	Conclusion .....	180
III	References .....	182
PAPER IV:	Data-driven Innovation through Open Government Data .....	187
IV.1	Introduction .....	189
IV.2	Critical Realism and Open Government Data .....	190
IV.3	Value Generating Mechanisms .....	196
IV.4	Research Design .....	202
IV.5	Event Analysis .....	204
IV.6	Structural and Contextual Analysis .....	209
IV.7	Retroduced Mechanisms .....	212
IV.8	Empirical Corroboration .....	219
IV.9	Discussion and Conclusion .....	221
IV	References .....	223
PAPER V:	Generating Sustainable Value from Open Data in a Sharing Society .....	229
V.1	Introduction .....	230
V.2	Open data empowers the sharing society .....	232
V.3	Behavioral assumptions for value generating <i>mechanisms</i> .....	234
V.4	A framework of four value generating mechanisms .....	236
V.5	Generating sustainable value from data in the sharing society .....	238
V.6	Discussion: The transformative power of data .....	243
V.7	Conclusions .....	249
V	References .....	251
PAPER VI:	Managing Complexity across Multiple Dimensions of Open Government Data .....	255
VI.1	Introduction .....	257
VI.2	Motivation and Frameworks for Analysis .....	258
VI.3	Research Setting and Approach .....	265
VI.4	Case Context and Description .....	270
VI.5	A Process Model of Open Data Supply and Value Generation .....	283
VI.6	Analysis and Discussion .....	286
VI.7	Contributions, Limitations and Conclusions .....	291
VI Appendix A:	Overview over interviews .....	294

VI	References .....	296
PAPER VII:	Driving Sustainable Value: A Conceptual Model of Open Data as a Resource.....	300
VII.1	Introduction.....	302
VII.2	Dimensions of Open Data as a Strategic Resource.....	304
VIII.3	Value Generation in a Networked Economy .....	307
VII.4	The Social Mechanisms of Value Generation .....	311
VII.5	Research Framework and Hypotheses Development .....	313
VII.6	Research Design .....	322
VII.7	Data Analysis and Results .....	326
VII.8	Discussion.....	336
VII.9	Conclusion .....	339
VII	References .....	340
VII Appendix A:	Operationalization of variables.....	352
VII Appendix B:	Descriptive statistics .....	358
VII Appendix C:	Cross loadings .....	359
VII Appendix D:	Differences in parameters between PLSc and PLS .....	360
PAPER VIII:	Innovation in the Open Data Ecosystem.....	361
VIII.1	Introduction.....	363
VIII.2	The Sustainable Value of Open Data.....	365
VIII.3	The Role of Intermediaries in Open Data Ecosystems .....	368
VIII.4	The Economics of Two Sided Markets.....	370
VIII.5	Investing in Open Data MSPs: Insights from Real Options Theory.....	378
VIII.6	Growing sustainable value from open data.....	381
VIII.7	Discussion.....	385
VIII.8	Limitations, Implications and Conclusion .....	387
VIII	References .....	390



**List of Abbreviations**

BDP	Basic Data Program
CR	Critical Realism
DS	Data Steward
GDP	Gross Domestic Product
IS	Information Systems
MSP	Multi Sided Platforms
ODI	Open Data Initiative
OGD	Open Government Data
PLS	Partial Least Squares

**List of Tables (Summary paper)**

<i>Table 1: Stratified Ontology of Critical Realism</i>	<i>Page 26</i>
<i>Table 2: The Seven Dimensions of Liquid Open Data</i>	<i>Page 45</i>
<i>Table 3: Effect Sizes Measured with Cohen’s <math>f^2</math></i>	<i>Page 69</i>
<i>Table 4: Open Data Research Agenda</i>	<i>Page 88</i>
<i>Table 5: An Overview over Included Papers; Publication Outlets, Main Contributions and Co-Authors</i>	<i>Page 103</i>

## List of Figures (Summary paper)

<i>Figure 1: Research Overview</i>	<i>Page 6</i>
<i>Figure 2: The Chicken-and-Egg nature of the Open Data Value Paradox</i>	<i>Page 10</i>
<i>Figure 3: The Open Data Value Generation Lifecycle: Four Events and Four Classes of Unobservable Mechanisms</i>	<i>Page 12</i>
<i>Figure 4: Four Main Areas in Current Literature</i>	<i>Page 15</i>
<i>Figure 5: Meta-theory based on Coleman's Framework</i>	<i>Page 38</i>
<i>Figure 6: Retroductive Research Process</i>	<i>Page 40</i>
<i>Figure 7: Development of Explanatory Variables over course of Study</i>	<i>Page 42</i>
<i>Figure 8: Evaluation of Datasets based on the Seven Dimensions of Liquid Open Data</i>	<i>Page 46</i>
<i>Figure 9: A Framework of Four Value Generating Mechanisms</i>	<i>Page 54</i>
<i>Figure 10: Research model</i>	<i>Page 57</i>
<i>Figure 11: Results from PLS Estimation</i>	<i>Page 67</i>
<i>Figure 12: Moderating Relationships</i>	<i>Page 68</i>
<i>Figure 13: Liquid Open Data, Shared Digital Content, New Digital Products and Services and Sustainable Value</i>	<i>Page 70</i>
<i>Figure 14: Extent of Shared Digital Content as a function of Liquid Open Data</i>	<i>Page 71</i>
<i>Figure 15: Extent of New Digital Products and Services as a function of Liquid Open Data</i>	<i>Page 72</i>
<i>Figure 16: MSP type Intermediaries Create Synergies between Content Sharing and Commercial Products and Services</i>	<i>Page 75</i>
<i>Figure 17: The Open Data Value Generation Lifecycle Revisited</i>	<i>Page 86</i>
<i>Figure 18: Paper Overview</i>	<i>Page 102</i>

# 1. Introduction

The amount of data in the world is increasing rapidly. In addition to all the data that have been manually entered in various information systems since the beginning of digitization, we now discern a rapid increase in data that are generated through means such as IoT devices, smart phones and social media. This trend signifies that data are created in higher volumes than before, they are generated at a faster pace than previously and they come from a much larger variety of sources (McAfee & Brynjolfsson, 2012). These new type of data are commonly termed big data and are often described by means of four dimensions - Volume, Velocity, Variety and Veracity (the 4 V's). Big data, being a very recent concept, have generated great interest in both practice and research communities. A growing body of literature on this subject reflects the relevance of the concept. Big data can belong to, or be utilized by, either the public or private sectors. An analysis of use cases where companies are using and transforming big data demonstrates a growing diversity and complexity of data use and data ownership. While previously it was easy to conceptualize a relatively stable progress from data generation/collection to data use and creation of value, this reality no longer holds true.

I propose that the concept of openness of data constitutes an important dynamic towards making the process from data collection to generation of value more complex and intricate, while at the same time increasing the opportunities for value generation through data immensely.

As early as 1942, Robert Merton emphasized the importance of all results of research being freely accessible to all. In order to allow knowledge to move forward, each researcher must contribute to a common pool of knowledge and should give up intellectual property rights (Merton, 1942). The concept of open access gained traction in the scientific community and later extended to the public sector, especially in the U.S., where to cite a far-reaching example, all data from the U.S. Global Positioning Satellite (GPS) system were made freely accessible for civil use in the eighties. The foundation of the World Wide Web was another significant step towards open access and open standards, presenting the world with an opportunity to publish websites that anyone could freely connect to and access. We can safely say that the World Wide Web has fundamentally revolutionized how society operates, on both a social as well as a technical level. Citizens and organizations have gained access through the Web to an infrastructure that allows them to share information freely. Openness has now

become not only ideologically attractive, but also a technically and commercially viable strategy.

For my PhD thesis, I have elected to explore how *openness* is relevant to value generation from use of data, with particular reference to what is described as Open Government Data (OGD). It is important to categorically state at this point that open government data is not an equivalent to, but a subcategory or subset of open data, which may equally originate in the commercial, academic or third sectors (Heimstädt et al., 2014). However, as governments currently provide a great majority of open data and as these concepts are mostly used interchangeably in the current discourse; I will merely use the term open data for the remainder of this dissertation. Moreover, I make extensive use of the concept of mechanism, which several disciplines of science have adopted to explain how a phenomenon comes about (Hedström & Ylikoski, 2010). In the context of this PhD study on the value of open data, we can define a mechanism as frequently occurring and easily recognizable causal patterns (Elster, 2007). My main emphasis has accordingly been to find high-level, generalizable patterns that are used to explain the transformation from open data to sustainable value.

## **1.1 Problem statement**

Following the terrorist attacks of 9/11 in 2001, the heightened emphasis on national security in the U.S. led to a perceived lack of transparency in government (Roberts, 2006; Peled, 2011). Barack Obama addressed this concern in his presidential campaign, promising an unprecedented level of openness in government. Right from the outset, Obama's open government initiative was very technology oriented (Yu & Robinson, 2012). After his election as President of the United States in 2009, this focus area developed into a full-fledged Open Data Initiative (ODI) (Peled, 2011). The initiative included the appointment of a state Chief Information Officer (CIO) and Chief Technology Officer (CTO) and the advent of a strong push towards open data through the "Data.gov" open data portal. More recently, the "Open Government Partnership (OGP)" was launched in 2011, to make good on a pledge made by President Obama to the United Nations General Assembly in September 2010. The primary goal of the OGP was to foster the development of more open governments around the world, in order to combat corruption and increase accountability (Harrison et al., 2011). These initiatives are representative of a twofold agenda. The first point in the agenda represents a push towards increasing technological innovation through use of data, oftentimes for the "greater good", while the latter point in this agenda

highlights the need for more transparency in government, and more participation and collaboration with citizens.

Following Obama's precedence, most countries have now launched open data programs, but with varying underlying motives. While the perceived lack of transparency has spurred many of the ODIs to envision an open and transparent government, the potential of data-driven innovation that might rekindle their stagnating economies has inspired leaders of the European Union (EU) (Janssen, 2011; Zuiderwijk et al. 2012). In yet other initiatives, the focus is predominantly on how open data might improve governmental efficiency through sharing of data across public organizations, building on World Wide Web-like ideas of interoperability and open standards. There is comparatively little research available on ODIs at the municipal or city levels. Nonetheless, emerging research indicates a focus on the participatory aspects of open data, wherein local governments aim to stimulate civic engagement in their local contexts (Kassen, 2014; Lassinantti et al. 2014). With so many diverse views on the value proposition of open data, referring to numerous old and new theoretical concepts across multiple disciplines, the currently accessible discourse comes across as fragmented, lacking in common foundations.

While this conceptual vagueness is natural given the early stages of the open data phenomenon, the various interpretations may lead to misunderstanding and frustration, (Peled, 2011; Yu & Robinson, 2012) and a lack of clarity about the Why's, How's and Who's of open data (Jetzek et al., 2014b). Moreover, while there appears to exist an overall belief in the potential benefits of open data, governments are continually struggling with practical issues related to financing, data quality, conceptual, technical and organizational interoperability, lack of motivation and incentives, and an overall shortage of skills and resources (Conradie & Choenni, 2014; Janssen et al., 2012; Martin et al., 2014; Zuiderwijk & Janssen 2014a). In spite of the motivation and drive, that has identified many of the early open government data initiatives, the first signs of disillusionment are appearing. The experiences of hundreds of initiatives worldwide have uncovered a high level of complexity with yet little or no evidence of value generation (Davies, 2013, Huijboom & den Broek, 2011, Zuiderwijk & Janssen, 2014b, Zuiderwijk et al., 2014a). There are no simple solutions, which are capable of transforming open data to sustainable value. Numerous trials and checks are necessary to test the feasibility of open government data, representing multiple challenges related to informational, functional and structural complexity. However, there is undeniably great value potential in open data - value that would be revealed if these challenges are solved.

If these important initiatives are to be sustained, their leaders must be able to justify the spending of public money for the objective of opening government data. For such justification, we need to understand how open data can generate value and how we can stimulate this value generation. In spite of the complexity involved in this process, or perhaps because of it, we must disentangle, clarify and simplify concepts for improved understanding. We need to draw out the most important constructs and underlying mechanisms and create a holistic map of their relation to each other. Or, in other words, we need a theory.

## **1.2 Research Context and Research Questions**

This PhD research was motivated by an initiative in Denmark, the Basic Data Program (BDP). The BDP originally focused on methods by which the public sector could better utilize certain basic datasets for improved efficiency, but later evolved into an Open Data Initiative (ODI). The program is unique amongst other ODIs in view of the fact that it has maintained a strong focus on data quality, data standardization and interoperability within the public sector. Moreover, KMD, a private company with an interest in following these developments, hired the author of this dissertation as an industrial PhD student. My role was to observe, document, and explain the trends towards dissemination and use of open data in the public sector and to present my results as a potential input to their data strategy.

In 2012, the publication of the report Good basic data for everyone – a source of growth and efficiency marked the beginning of the BDP (Agency for Digitization, 2012). This publication was a practical outcome of certain objectives of the Danish e-Government Strategy 2011-2015 (Agency for Digitization, 2011). The Agency for Digitization had classified a number of core societal level reference or master data, used widely for different purposes by public and private sectors alike, as basic data. Basic data include, but are not limited to: data from the person register, business register and real property register, as well as tax data, address data, place names and geographic data. In the beginning, the BDP's main goal was to create an infrastructure that would enable more efficient use of the basic data across administrations and sectors (Horst et al., 2014). However, for various reasons explored in Paper VI, the BDP evolved into a full-fledged ODI during the program definition phase. As a first step towards opening data, geographic data (including maps) and data from the business register were made available free-of-charge as of January 2013. However, this particular event was only a beginning of a complicated process involving many challenges of both social and technical nature.

The private sector IT Company KMD that supports this PhD study is responsible for implementing the BDP's open data platform. For KMD, as well as for the members of the BDP, an important aspect of this study is to contribute to their understanding of how value is generated through open data and how this value can be captured and evaluated. Open data have the features of a public good, as they are open to all and can be easily accessed, reproduced or reconfigured and shared over networks. Public goods have two defining characteristics: they are non-excludable - one individual's use of the data will not exclude the use of another; and non-rivalrous - one individual's use will not reduce the amount available to another. As a result, it is extremely difficult to trace and evaluate the impact of open data use. This implies a necessity for a new approach and a new understanding of the way in which we perceive value and value generation in our increasingly digital, open and networked economies.

During the course of this PhD, a number of macro-trends in society have influenced my research. To name but a few: the increasing use of technology that has resulted in the creation of a gigantic volume of data (big data); the general trend towards peer-to-peer resource sharing (sharing economy); ongoing changes in business models and the advent of Multi Sided Platforms (MSPs); and finally, the realization by numerous economists that the overarching focus on society-wide economic value generation has resulted in undesired developments like economic inequality and an excessive emphasis on financial assets vs. other social and environmental elements.

The overarching research question is:

*How is value generated from open data?*

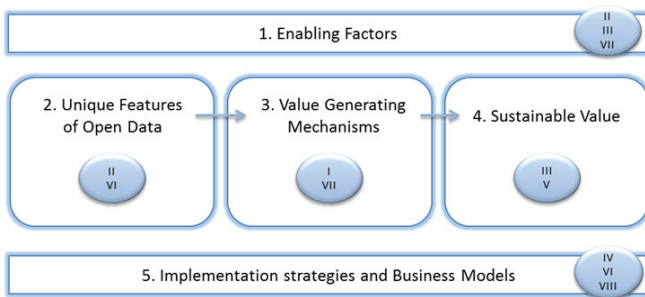
While the primary theme of this research study is to understand how we generate value from open data, the study also addresses five sub-questions in more detail.

The five sub-questions are:

1. What are the main enabling factors for value generation through open data ?
2. What are the unique features of open data?
3. What are the value generating mechanisms of open data ?
4. How can we identify, conceptualize and measure the value that is generated from open data?
5. What are the key implementation strategies and business models that can promote long term generation of value from open data?

### 1.3 Research Design

This PhD study follows a paper-based publication approach. The initial papers of this research are conference papers, and the chief reason for choosing this platform was to acquire initial feedback on the topic and the approach. The following papers are lengthier and more comprehensive and intended for journal publication. The process has been iterative and marked by a gradual systematic progression, with a different combination of sub-questions addressed through different methods, ultimately leading to incremental discovery and overall progress towards the end goal, i.e. to answer the overarching research question. Figure 1 summarizes the main contributions of each paper to the different sub-questions.



*Figure 1: Research Overview*

Each of the papers addresses one to three of the five sub-questions specifically, while continuously reflecting on the overall theme of explaining how value is generated from open data. The sub-questions are represented by the boxes in figure 1. The five circles with roman numbers indicate which papers include the most relevant contribution towards each question. Appendix A provides a summary of the individual papers.

The research is positioned at the societal level for three key reasons: 1) The Open Data Initiative (ODI) viz. the Basic Data Program of Denmark, which was the foundation of this study is a central government program and as such, is intended to deliver value for the whole of society. 2) The public good features of open data make the process of exploring value generation from open data without looking at the wider societal consequences very challenging. 3) Due to the novelty of the open data phenomenon, it is the author's perception that an overall framework showing high-level constructs and their relationships can provide a common conceptual basis from which an in-depth study of individual elements can take departure.



I selected a phenomenon based research approach because open data is an emerging phenomenon. This approach is focused towards capturing and reporting on new or recent phenomena of interest (von Krogh et al., 2012). A phenomenon based approach is recommended when no currently available theory presents sufficient scope to account for the phenomenon or for the relevant cause and effect relationships associated with that phenomenon. Therefore, the target of phenomenon-based research is to capture, describe, document, and conceptualize a phenomenon, in order to facilitate more comprehensive theoretical work and development of research designs (von Krogh et al., 2012). My focus has accordingly been on conceptualizing and explaining constructs and relationships. I believe that this groundwork is vital for both businesses and government organizations to be able to document the potential benefits of open data. In two of the papers that were written as part of this research, we moved further to actually measure and validate relationships, but with the ultimate goal of improving our understanding of the phenomenon through empirical, as well as theoretical work.

For meta-theory, I have used an explanatory meta-theoretical framework based on Coleman's (1990) framework (sometimes called Coleman's boat) that takes into account both the macro and micro perspectives required to fully explain societal level phenomena (Elster, 2007). Coleman's framework can be used to explain how micro-level action is linked to macro-level structures (and vice versa). To state this briefly, there are certain societal-level conditions that will influence individuals' actions, which in turn, will collectively form new societal level structures. Explanatory research in general seeks explanations of observed phenomena, problems, or behaviors and seeks answers to why and how types of questions. Furthermore, explanatory research attempts to identify causal factors and outcomes of the target phenomenon (Bhattacharjee, 2012). As I was conducting research on an emerging phenomenon, I employed the empirical data not only to test causal relationships, but also to identify such relationships through triangulation between qualitative and quantitative data. An iterative approach between empirical research and theoretical modeling assisted me to capture, describe and conceptualize the most important constructs relevant to the means by which open data generates value and thereby, model the most relevant relationships.

This PhD study is an industrial PhD project, and thus makes specific demands towards practical contributions. The study loosely follows the iterative process recommended for Engaged Scholarship (van de Ven, 2007). Engaged Scholarship recommends that practitioners engage in the following four stages of research: 1) research design, when

experts are required to share their insight into interesting problems and to provide easy access to information; 2) theory building, where knowledge experts in the relevant disciplines and functions should be involved and invited to participate; 3) problem formulation, when those that experience and know the problem should be engaged; and finally, 4) problem solving, when the intended audience should be engaged to interpret meanings and their usage. I have attempted to follow this approach by engaging continuously with the participants of the BDP, both from the government and private company sides. I interacted with them during the study design, through observation, participation and interviews over the course of the study, and by requesting feedback on proposed models and ideas. This iterative process has proved to be very rewarding, providing a contribution to both research and practice, by intertwining knowledge dissemination and learning, and thereby, building directly on the two components of theory and practice.

If choice of research philosophy were based on a dichotomy of positivism on the one hand and interpretivism on the other hand, I would probably lean towards positivism. The reason has probably something to do with my background in economics. My personal views are thus not fully aligned with the interpretivism paradigm on how to generate (generalizable) knowledge claims. However, I was never fully satisfied with the strong focus on mathematical generalizations in economics in the past and after working in the ICT industry for around twelve years, I was very aware of the inherent complexities of Information Systems (IS) related phenomena. Moreover, I have become more pragmatic after years of industry work. All of these considerations, as well as the industrial context of my PhD project, directed me towards examining an alternative background philosophy.

Relatively early in the process, my main supervisor introduced me to Critical Realism (CR), a philosophical approach associated with Roy Bhaskar (1975, 1978). CR is often viewed as a middle approach between positivism and interpretivism, thus introducing a more nuanced version of realist ontology (Zachariadis et al., 2013). CR focuses on providing causal explanations in the form of generative structures or mechanisms. A small body of research by critical realists also proposes the use of the logic of inference called retrodution, which can be used to uncover these unobservable underlying structures or mechanisms (Baskar, 1975, Danermark et al., 2002). Retrodution allows researchers to move between the knowledge of empirical phenomena to the creation of explanations or hypothesizing, and is capable in theory, of providing some indications on the existence of unobservable entities (Zachariadis et al. 2013). It has also been argued that a retroductive approach to research embraces a wide variety of methods

(Downward & Mearman, 2006, Venkatesh et al., 2013, Wynn & Williams, 2012, Zachariadis et al., 2013). I decided to use a mixed method approach to satisfy my requirement for empirical generalizations complemented with an in-depth understanding and explanation of the open data phenomenon. Thus, I became convinced that critical realism was the right philosophy for my project. I shall discuss the research philosophy and methods at length in Chapter 3.

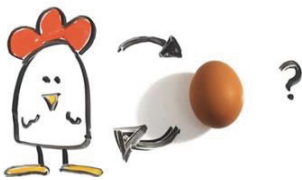
## **1.4 Contributions and Future Research**

Clear constructs are simply robust categories that extract phenomena to create precise distinctions that are comprehensible to a community of researchers (Suddaby, 2010). The first theoretical contribution of this PhD project is the definition of liquid open data, explained in more depth in section 5.1. In many of the papers published as a part of this PhD study, I have reiterated my view that to date, the open data phenomenon has not been adequately explained. To contribute to conceptual clarity in the field of open data value, I have identified the main features of open data, both in the sense of economic features but also by proposing the liquid open data construct containing seven clearly identified dimensions. The second contribution of this PhD project is a definition and conceptualization of sustainable value, which is explained more carefully in section 5.3. The definition of sustainable value represents a shift from the previous focus on dominant economic value. Sustainable value as a concept offers an emphasis on proactive, concerted efforts of businesses, government institutions and the overall community, to address social challenges in innovative and holistic ways that generate social, environmental and economic value for all stakeholders and future generations (van Osch & Avital, 2010).

Any theory must not only provide construct clarity but also identify the relationships among constructs (Suddaby, 2010). Critical realists argue that general underlying but unobservable “generative mechanisms” can explain the occurrence of phenomena. As a third contribution, I have created a two-by-two framework with a taxonomy of the most relevant value generating mechanisms that have been identified within the context of using open data to generate value. The framework is discussed in-depth in section 5.2. The framework highlights two principal mechanisms that facilitate how value is generated through open data: the Information sharing mechanism and the Market mechanism. It also emphasizes that for each of those mechanisms, value generation can happen either through exploitation of current resources or through exploration, focused on driving change. As a fourth theoretical contribution I have identified and proposed a conceptual model that illustrates the nomological network of

causal relationships. The model illustrates the relationships that occur between empirical observations of certain events produced by the underlying mechanisms. In addition to the constructs of liquid open data and sustainable value, the model suggests two types of empirical manifestations of the information sharing and market mechanisms in the context of open data use. Four additional societal level structures are proposed as enabling factors that can influence these underlying mechanisms. Finally, I propose that to indicate the importance of cross-sector collaboration for addressing societal challenges, we can model the moderating influence of private sector accountability on sustainable value generation.

The fifth contribution of this PhD study is the Open Data Value Paradox. Several existing ODI efforts have fallen short of prior expectations regarding value generation, and there is a universal lack of documented evidence of impact from open data (Davies, 2013). This is disconcerting for government organizations that have already expended money on opening data, and might result in the termination of some ODIs. What are the reasons behind this lack of evidence? I suggest that many of these initiatives may not provide the key dimensions of liquid open data that I propose are important for value generation to happen, indicating an underinvestment problem. Furthermore, the contexts in which these ODIs operate may not be productive for value generation. Most importantly, however, ODIs have generally not been evaluated from a wider macro-economic point of view, due to insufficient understanding of how value generation happens and a consequent dearth of appropriate evaluation methods. At present, we do not have any standard methods to evaluate the impact of programs and initiatives that depend on information sharing to the same extent as ODIs.



*Figure 2: The Chicken-and-Egg nature of the Open Data Value Paradox.*

Source: <http://www.brainpickings.org/2013/02/01/which-came-first-the-chicken-or-the-egg/>

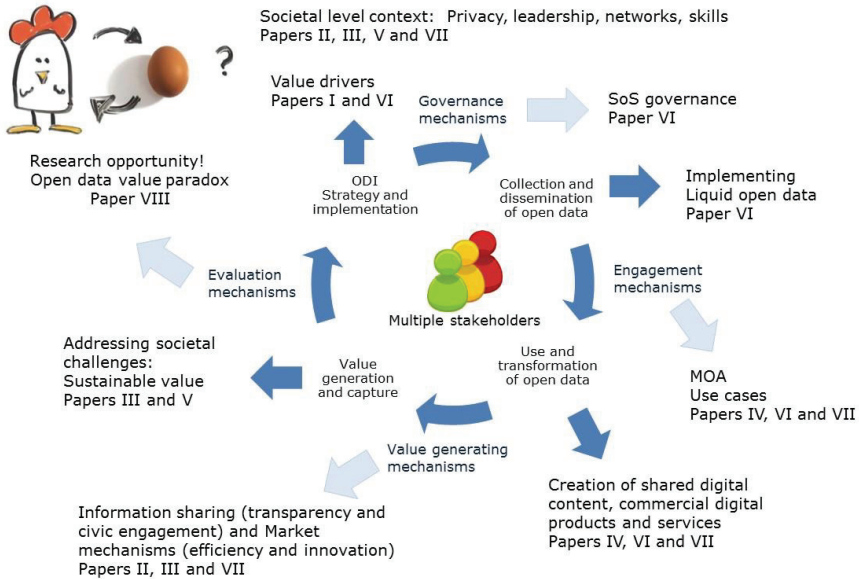
This open data value paradox states that in order to generate value from open data we require more investment for making open data more useful. However, public and private investors are not willing to expend additional capital on open data unless they are able to perceive evidence of value. This therefore becomes a chicken-and-egg type of paradox. The open data value paradox is both theoretically interesting and

practically relevant. It is interesting to private sector businesses, as it applies to a very central problem they are currently facing. This problem is reflected in the fact that consumers are accustomed to having access to free information services delivered over mobile devices, the World Wide Web and even through wearable technology. The companies that specialize in information services must however, eventually produce income and profits if they are to survive. The same applies for government institutions providing open data. They must be able to understand, articulate, document and evaluate the inherent worth of the open data they produce. In many cases, these data are transformed to free information, which is furthermore used to generate intangible value. This type of value creation cannot be traced through company or national accounts.

The final contribution of this research is a result of the method of redtroduction. Based on a triangulation between qualitative and quantitative data I have developed four events and four classes of generative mechanisms in the Open Data Value Lifecycle, as shown in figure 3. The lifecycle is an extension of a process model presented in Paper VI and shows how we move between the decision to open data, through implementation of open data infrastructures and dissemination of data towards use of the data and the eventual generation and capture of value, that will furthermore influence strategy.

The four main classes of mechanisms I have identified are: 1) Governance mechanisms that elucidate how open data strategies and policies are shaped and explain how ODIs are governed. 3) Engagement mechanisms explain how and why users engage (or not) with the openly disseminated data. 4) Value generating mechanisms are a consequence of use of open data and explain how the use of open data contributes to the generation of value. 5) Evaluation mechanisms explain how the value that is generated through use of open data is perceived and accounted for. The evaluation of open data will furthermore influence strategy and decision-making.

I propose this cycle uncovers multiple gaps in the literature covering the relationship between open data and value. Furthermore, I propose that this PhD study has contributed in a meaningful way to some of these gaps, but in other areas I have not moved much farther than recognizing the need for more research. Consequently, I propose a research agenda in section 6.3, which I consider a contribution in itself in such an emerging field.



*Figure 3: The Open Data Value Generation Lifecycle: Four Events and Four Classes of Unobservable Mechanisms*

In the same manner in which I have adopted the insights from practitioners to help create relevant theory, I have correspondingly used the theoretical lenses that I have counted as contributions of this PhD project to generate several practical contributions.

Firstly, I have exercised Engaged Scholarship by involving the Danish BDP in all stages of this study. The participants of the BDP and KMD have in turn appropriated value from the research, benefitting from the outsider, helicopter view provided by the author as I had access to a great deal of information on the program without being heavily involved in the implementation itself. This academic-practice exchange of knowledge and sentiments has been fruitful to all partners, and has enabled us to acquire a more holistic view of the phenomenon of open data.

Secondly, the BDP case study (Paper VI) has offered a unique insight into the real-world technical and governance related obstacles encountered by an ODI over a period of almost four years. The BDP case study offers an insight into how an investment in open data infrastructure can contribute to public sector efficiency, and illustrates the challenges inherent in governing such a complex initiative. I have attempted to transform the experiences of the BDP group into four practical principles, which can

be utilized by other ODIs that operate in similar contexts. These principles are discussed in more detailed in Paper VI and in section 5.1 of this cover paper.

Thirdly, the Opower case study (Paper IV) offers another perspective on value generation. Unlike the other papers in this PhD research, this article is positioned at an organizational level. The Opower case highlights how and why various dimensions of data (such as usability, discoverability and accessibility) are important to private sector users and explains how open data can be utilized by the private sector to generate societal level value. The case supported an emerging realization: most of open data are used in combination with proprietary data. Open data are therefore not the only, or even the main, resource for most open data users. This is not to say that open data are not important to these users, as they gain access to external data they could not have produced themselves. However, this finding indicates that the impact from use of open data is hardly separable from the impact from using other resources, and thus not clearly visible even to the open data users themselves.

As a practical contribution, I suggest that private companies who: a) access and link open and proprietary datasources; b) use big data tools and analytics for the development of free information as well as commercial data-driven products and services, and; c) utilize the dynamics and interactions enabled by MSPs, are in a superior position to address a variety of societal challenges and simultaneously generate economic and social value, for themselves as for all of society.

## **2. State of the Art**

As open data are a shared or common resource, they hold great potential for a number of stakeholders, including public sector agencies, private businesses, the academia, citizens and civic organizations. However, this potential has so-far not been very clearly articulated due to the novelty of the open data phenomenon. In 2012 when this work started, a limited number of published scholarly papers were available on the subject. However, a small body of European literature on what was typically termed “Public Sector Information” was readily obtainable. This corpus concentrated mostly on the potential of open access to certain categories of commercially relevant government data, such as geographic data. The main driver for the PSI discussion in Europe was the fact that datasets like GPS, maps and weather data had been open for many years in the U.S., leading to vast and vibrant markets and countless innovations. Beyond these publications, it was difficult to locate scholarly material about the more recent concepts of Open Data, Open Government or Open Government Data (OGD).

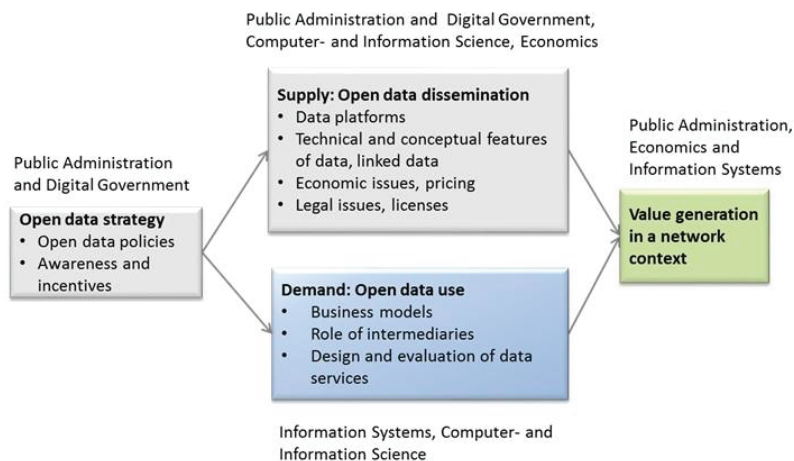
In the initial literature search, I used the keywords “Public Sector Information” and “Open Government Data” and “Value” and different combinations thereof. There were no articles in core Information System (IS) journals (Basket of eight) or core IS conference proceedings. Nonetheless, I identified seventy-nine different articles, mostly reports or conference proceedings from other disciplines on the above subjects. The majority of identified articles were published in digital governance or public administration publications or in computer and information science publications. Furthermore, I identified a number of reports where authors had attempted to measure the economic value of particular datasets, and most of these pertained to geographic data. While this PhD study is by design essentially grounded within the IS and e-Government streams of literature, I have drawn from research in different disciplines and from both scholarly and popular media, as recommended by the phenomenon based research approach (von Krogh et al., 2012).

The open data phenomenon has references in different streams of research in different disciplines, as indicated in figure 4. After an initial period of working with the above-mentioned sources, I created the first classification scheme in an attempt to develop a conceptual model. I classified the articles based upon whether their focus was on any of the following topics: 1) open government data related policy making and initiation of ODIs; 2) data platforms, technical and conceptual features of open data or linked data, or economic and legal features of open data; 3) data engagement and use, including business models, role of intermediaries and design of data services; 4) value generation through use of open data, see figure 4.

Since 2012, the volume of literature on open (government) data has increased phenomenally. Over the three years that I have worked on this dissertation, I have collected articles on open data from various sources. Nevertheless, in order to ensure that I was up-to-date with the most recent articles relevant to the study, I initiated a fresh search in Web of Science (in April 2015) by using the following key words: TOPIC: (“open government data” OR “open data”) and “value”). This search yielded seventy-two results, some of which were from fields far removed from IS. Upon scrutinizing the list, I selected the articles that seemed relevant to my topic. Still today, most articles referring to open data are published in the computer or information science streams of literature, and in a vast majority of those cases, the main topic is linked data and the semantic web. Linked data are relevant to the usability and functionality of data for value generation; however, an in-depth technical discussion is not the core topic of this dissertation. Upon a thorough review of the abstracts from



this search, a mere nine of those papers were added to my current collection of papers on open data.



*Figure 4: Four Main Areas in Current Literature*

I had identified two special issues on open data in the Journal of Theoretical and Applied Electronic Commerce Research; the first issue focused on innovation through open data and included seven articles (thereof one article by the author of this dissertation); and the second issue focused on transparency and open data policies, including six articles. By using these sources to look backwards and forwards as recommended by Webster & Watson (2002), I managed to extend the pool of papers further. The final sample added seventy-seven articles to the original seventy-nine; the majority of added papers (forty-two) was published in 2014. Upon a brief review of the abstracts, I classified these articles according to the model in figure 4 (above). While my sample is not an all-inclusive overview of the open data literature, it provides a satisfactory indication of the direction in which current research is mostly focused, besides identifying gaps in the literature. To summarize, most of the papers identified focus on the supply side of open data, including open data policies and technical aspects of data dissemination. Additionally, there is an emerging body of literature where demand side topics are discussed, for instance the role of intermediaries, emerging business models and the role and impact of innovation contests. There was, however, a huge gap in the scientific literature on the subject of explaining how open data generates value. Thus, I needed to seek aspiration from literature on value generation in different contexts. Consequently, I synthesized the literature on open data

with literature on the changing perception of value generation within the networked society.

## **2.1 Open Data Strategies, Enablers and Barriers**

Governments around the world are increasingly defining and implementing open data strategies and policies. Some of the most common goals of these policies are to increase public sector transparency (Bertot et al, 2010; 2012; Peled, 2011), civic engagement (Bertot et al., 2012; Janssen, 2011), government efficiency (Tinati et al., 2102; Ubaldi, 2013) and/or private sector innovation (Kundra, 2012; Zuiderwijk et al. 2014a). As these open data strategies are relatively new, evidence of the expected impact is still limited (Davies, 2013; Huijboom & van den Broek, 2011). It has been suggested that governments can manage and control impediments to the open data supply process by developing efficient open data policies (Zuiderwijk et al. 2012). Nevertheless, many important questions pertaining to this issue continue to be debated. For example, what is an appropriate open data strategy for governments? Another pertinent question concerns why some governments are achieving success in opening up their data, while others struggle (Huijboom & van den Broek, 2011).

Open data policies are created at many levels of government, ranging from individual cities to international organizations like the European Union (EU). An example of a national level open data policy is the U.S. Open Government Directive and subsequent Executive Order, which makes open and machine readable the new default for all government information (White House, 2009; 2013). The EU Public Sector Information (PSI) Directive is an example of an international open data policy. The PSI Directive presents a common legislative framework, which regulates in what way public sector bodies should make data open and available for re-use (European Commission, 2003). Building on former policies, the EU has now presented an Open Data Strategy for Europe, on making the best use of government-held information through well-defined and precise rules (European Commission, 2011a). An important amendment of the EU strategy of 2011, as compared with previous EU directives and guidelines, is that the EU recommended a general rule that all documents, which are made accessible by public sector bodies, can be re-used for any purpose, commercial or non-commercial, unless protected by third party copyright (Zuiderwijk et al., 2012). Moreover, the strategy states that public bodies should not be permitted to charge more than marginal costs elicited by an individual request for data (ibid).

Since open data policies are a recent phenomenon, applicable systematic research that identifies different policies and relates those differences to policy impacts has not

emerged yet (Huijboom & van den Broek, 2011). A framework for comparison of open data policies is a recent contribution (Zuiderwijk & Janssen, 2014c). The framework includes factors of environment and context, policy content, performance indicators and public values. The authors use the framework to compare seven different open data policies in the Netherlands and suggest that current policies are inward looking, and that more collaboration with external stakeholders could improve these strategies. It is furthermore recommended that ODIs increase focus on stimulating the use of open data and pay more attention to the impact of the policy (Zuiderwijk & Janssen, 2014c).

Several papers have investigated the socio-technical impediments of implementing an open data policy (for instance, Barry & Barrister, 2014; Conradie & Choenni, 2014; Esterman, 2014; Meijer & Thaens, 2009; Janssen et al., 2012; Zuiderwijk et al., 2012). A pilot survey among heritage institutions in Switzerland suggests that while many of those institutions had yet to adopt open data policies, several among them were already considering and examining such policies (Esterman, 2014). The institutions' views on open data, however, were characterized by skepticism of the potential impact and fear of the consequences of making data freely available, including fear of losing control (Esterman, 2014). Janssen et al. (2012) mention numerous barriers to open data adoption, including lack of awareness and incentives from both the supply and demand side, incomplete or non-existing open data policies, a risk-averse culture, immature legislative infrastructure, low data quality and shortage of supporting technical infrastructure. Additional reservations or obstacles include fear of false conclusions when data are used for analysis, adverse financial effects on government institutions and lack of resources and incentives (Barry & Bannister, 2014; Conradie & Choenni, 2014). Due to these complexities, governments are inclined to find it a challenge to open up their data for public scrutiny, and must subsequently develop appropriate processes and procedures to execute this action (Zuiderwijk et al., 2012).

## **2.2 Supply Side: Open Data Dissemination**

In most cases, the dissemination of open data does not constitute an integral part of the data collection or creation process, and it is usually viewed as a separate activity, which is not integrated into the data custodian's daily procedures and routines (Zuiderwijk & Janssen, 2013; van Veenstra & van den Broek, 2013). Although opening government data has the potential to generate countless advantages, the process of opening data is a complex and ill-understood activity (Zuiderwijk et al., 2012). Insufficient attention has been devoted to the dilemma of publishing data and thereafter, sustaining and renewing the data disseminated through open data portals

(Zuiderwijk et al., 2014a). The challenges confronting the supply of open data are broadly classified into the following areas: technical features or data platforms; conceptual features, commonly revolving around linked data; economic issues, largely related to pricing of data and funding of data collection; and legal issues that center on licenses, privacy and the legislative role of public institutions.

Insufficient research exists to analyze the individual steps that are required to publish open data after development of policy. Information science literature offers insights into specific areas such as data quality and use of metadata and computer science literature offers a sizable corpus of papers on linked data principles, models and platforms. Nonetheless, these papers do not present a holistic review of the technical issues and other relevant factors to be taken into consideration by organizations, which aim to publish open data. Moreover, open data governance is not well represented in the literature, although some similarities are discernible in the literature on data quality. It has been proposed that ODIs require a central governance organization that is willing to provide various support activities (Hofman & Rajagopal, 2014), but this is not widely supported.

Barry & Bannister (2014) conclude that organizations have to deal with various socio-technical issues regarding data cleansing, understanding the data structures and combining different sources of data. Other currently identified barriers to supplying open data are data formats, which must be machine readable, standardized and easily accessible, and a lack of data portals, which influences discoverability of data (Barry & Bannister, 2014; Janssen et al., 2012; Zuiderwijk et al., 2012). By building upon the current literature on technical barriers to open data dissemination, we can address these barriers by formulating functional requirements for open data platforms (Hofman & Rajagopal, 2014). A platform should provide sufficient functionality for data users who are required to access the data, as well as for data providers who are required to publish their data sets.

The lack of availability of technically re-usable formats and lack of semantic interoperability of open data continues to be a cause of concern (Petychakis et al., 2014). Data sets are often published in a format defined by the system from which the data originate. This poses a challenge for users to reuse the data, since they may lack the context, the original structure and semantics. Openly publishing data models, including syntax (structure), semantics (understanding) and context (metadata) is an imperative cornerstone towards making data more usable (Bountouri et al., 2009; Petychakis et al., 2014). It is recognized that the semantics of open data need to be clearly and concisely specified, for instance by ontologies (Hofman & Rajagopal,

2014). Computer science literature includes semantic models, which are represented as ontologies to address schema heterogeneity (see for instance Alani et al, 2007; Garcia et al., 2015; Höchtl & Reichstädter, 2011). However, interoperability of data does not only deal with semantics, but also with other pragmatics of sharing data that originate from heterogeneous systems (Hofman & Rajagopal, 2014). Results of a recent study confirm that to date, there is no uniform policy on the subject of provision of public sector information across data sources in the countries of the EU, and the current portals seldom support a semantic search (Petychakis et al., 2014). These results indicate the importance of further in-depth investigation into the relationship between the conceptual foundations of open data and their ongoing use.

Among research topics concerning dissemination of open data are the economic issues related to funding such initiatives. While collecting, processing and storing digital information may incur high fixed costs, the online dissemination of open data has very low marginal costs (Pollock, 2008). Prior research thus supports the notion that supplying open government data free of charge is generally justifiable on grounds of economic efficiency. This scenario is relevant particularly when there are no apparent obligations and risks related to nondisclosure (Cook, 2010). Furthermore, there seems to be a relationship between the public sector funding model and the economic results displayed by firms re-using open data. Based upon a study of 14,000 firms in the architectural and engineering business in fifteen countries during the period from 2000-2007, it was established that firms operating in countries where public sector agencies provided fundamental geographical information either freely or at maximum marginal costs had grown significantly. Their growth rate was, on average, about fifteen percent higher per annum, as compared to firms in countries where public sector geographic information was priced based upon a cost-recovery principle (Koski, 2011). Nevertheless, many public sector organizations have funding models that prevent them from freely disseminating open data (Barry & Bannister, 2014). Moreover, public sector officials do not comprehend the reasoning behind supplying valuable data free of charge and thereby, enabling commercial firms to profit from that data, while the public sector experiences an acute lack of resources (ibid).

A large proportion of the data made available as part of ODIs, are currently not published under an open license. A recent survey amongst Swiss heritage organizations revealed that most of them (83%) possessed no experience with alternative licensing models, such as Creative Commons licenses (Esterman, 2014). Heterogeneous licenses pose an added challenge, in view of the fact that users of open data may have trouble understanding and distinguishing between different licenses, especially when

combining different data and using that data commercially (Martin et al., 2013). Furthermore, additional legislative barriers complicate the act of opening up government data. For instance, the stakeholders in the Danish BDP needed to modify nearly ten different laws, in order to make geographic data and data from the business register free and open. Several public sector organizations are required by law to sell data to fund their operations (see discussion in Paper VI). Finally, while open government data are by definition not personal data, the data risk becoming person sensitive when combined with other publicly available data or if the data are de-anonymized (Kulk & Loenen, 2012; Meijer et al., 2014). Public sector organizations are apprehensive about matters like the threat of privacy violations by virtue of releasing data and being legally liable if open data are misused. Moreover, accidentally publishing data that could reveal privacy or security related information could damage the reputation of an organization (Zuiderwijk et al., 2014b).

### **2.3 Demand Side: Using Open Data**

The underlying motivation behind the establishing of many ODIs is that external innovators will benefit from the data and employ the data to develop novel digital products and services (Hjalmarsson et al., 2015a; Kuk & Davies, 2011). However, contrary to expectations, evidence of use of open data in innovations is still lacking (Davies, 2013; Foulonneau et al., 2014). Previously acknowledged barriers to use of open data are bad data quality, data being inaccessible or not usable (for example not published in machine-readable formats), and data which are not coherent and discoverable. Other identified obstacles include license restrictions, a lack of dialogue between data providers and re-users, lack of relevant skills, immature business models and risks related to the sustainability of data delivery (Davies, 2013; Janssen et al. 2012, Jetzek, 2014a; Martin et al., 2013; Zuiderwijk and Janssen 2014b). A knowledge gap has been identified, particularly for explaining methods by which to establish a sustainable open data ecosystem populated by interdependent stakeholders who supply and consume open data and digital services produced for this market (Lindman et al., 2013).

Ecosystem is a metaphor often used to denote an interdependent social system of actors, organizations, material infrastructures, and symbolic resources that can be created in technology-enabled, information-intensive social systems (Harrison et al., 2012). Hjalmarsson et al. (2015b) classify the key stakeholders in an open data ecosystem based upon the following four categories: 1) data providers that make data available beyond their organizational borders; 2) open data brokers or intermediaries;

3) third party developers (entrepreneurs); and 4) consumers or end users who use the results from the market to resolve challenges in everyday situations. All members of an ecosystem contribute to the generation of value from open data, although at present their actions are not well-coordinated (Mayer-Schönberger & Zappia, 2011). Research indicates that a digital ecosystem functions optimally if there is fruitful interaction between cooperating and competing actors (Heimstädt et al., 2014). A widely applied method to increase such interaction is to host innovation contests. These contests are in practice typically defined as hackathons. These events provide a meeting platform for data suppliers (most frequently from the public sector) and developers, and may lead to collaboration on a specific dataset or cooperation to confront a specific social challenge. Nevertheless, beyond the impact on catalyzing product and service development through open data, we know little about how innovation contests contribute towards the establishment of sustainable open data ecosystems (Juell-Skielse et al. 2014).

From the literature review, I infer the following three conclusions:

- a) There is a growing body of literature on barriers and enablers, generally relevant to open data dissemination and less focused on open data use. While the current classifications of enablers and barriers of open data dissemination are beneficial, they exhibit three main shortcomings. Firstly, the enablers and barriers originate from a variety of structures. While some of them are features of the data themselves, others are dimensions that reflect the ability of external users to access and work with the data, and yet others are factors that exhibit the relevance of the context in which data are collected and disseminated. Secondly, enablers and barriers are commonly classified together over multiple levels of analysis. For instance, legal data protection frameworks are placed at the societal level, while open data licenses can differ between individual datasets. Funding issues vary between individual ODIs and even datasets. Capabilities such as IT skills and technical infrastructures vary between organizations but also between countries. Finally, we recognize a lack of longitudinal research on individual initiatives, which would offer the ability to understand the relationships between these enablers and barriers and dynamic issues like policymaking, governance processes and implementation procedures.
- b) Rather few of the articles in available literature focus on the use of open data and open data engagement mechanisms. Nonetheless, we can confidently state that this body of literature is rapidly growing, especially in the context of innovation contests and open data services. There is an urgent necessity to understand the

phenomenon of open data from the demand side, since publishing open data has limited benefits if there are no users for that data.

- c) Current research offers practically no discussion on the topic of open data value, except for a minimal number of reports that attempt to identify the macro-level economic value of specific datasets. There is a huge gap in the literature when it comes to understanding how open data generate value, the type of value that is generated, and the best means by which this value can be exploited and identified. A few papers briefly mention the concept of public value, but usually on a very abstract level, without discussing the relevant mechanisms or how public value can be measured. This lack of research motivated me to review the value literature in various disciplines, generally documented in a different context, yet relevant enough to impart inspiration and provide theoretical foundations for the study.

## **2.4 Value Generation in a Networked Society**

The two categories of value frequently discussed in different streams of literature are economic value, which is defined as the worth of a good or service as determined by the market; and social value, which is created when resources, inputs, processes or policies are combined to generate improvements in the lives of individuals or society as a whole. The primary function of the structures of the industrial economy's markets was to facilitate the creation and documentation of economic value. In the industrial economy, value is traditionally measured through the price of a good, as most exchanges make use of the markets. The free market mechanism leads to a state of equilibrium, wherein the price is a reflection of the worth of each good to the "average buyer." However, with the advent of the World Wide Web and the adoption of open access, open source software, open data and similar ideologies, it has been suggested that the market mechanisms are only capable of explaining a part of the modern day value exchanges (Benkler, 2006).

Another feature of the industrial economy is a perceived dichotomy of the market and the state (Porter & Kramer, 2011). The free markets operate under the principle of each player maximizing his own economic value (profits), which should eventually lead to the most efficient distribution of goods and services (based on assumptions of rational behavior and full information). The state, however, is generally responsible for creating a structure around alleged market failures, when the potential to generate economic value on the market does not lead to an optimal situation for society. Yet, this perceived divergence between the state and the markets is being increasingly contested due to the inadequate ability of current social and economic structures to



address emerging social challenges. The concept of shared value has been introduced as an attempt to amalgamate the concepts of social and economic value (Porter & Kramer, 2011). Shared value is based on the belief that societal needs, and not purely economic needs, define markets. It also recognizes that social harms, such as wasted energy, health problems and/or the need for remedial training to compensate for inadequacies in education, create internal costs for companies (Porter & Kramer, 2011). The underlying notion behind the concept of shared value is essentially about expanding the overall pool of economic and social value, and simultaneously increasing the competitiveness of businesses and society.

The concept of public value has been studied for over twenty years, but has recently received renewed attention in the public administration and digital government literature, specifically in the context of open data. The central proposition in Moore's (1995) public value framework is that public resources should be used to increase value, not only in an economic sense, but also in broader terms of what citizens and communities value. A recent reframing of Moore's definition of public value includes a discussion of "What adds value to the public sphere" (Benington, 2011, p. 31), where the public sphere includes the state, the market and civil society. The dimensions of public value include: 1) economic value: adding value to the public realm through the generation of economic activity; 2) social and cultural value: adding value to the public realm by contributing to social cohesion and well-being; 3) political value: adding value to the public realm by stimulating and supporting democratic dialogue and active public participation, as well as citizen engagement; and 4) ecological value: adding value to the public realm by actively promoting sustainable development and reducing pollution, waste, and global warming (Benington, 2011).

Welfare economists apply the concept of social welfare to describe a construct that reaches further than the commonly used measure of Gross Domestic Product (GDP) as an indicator of value. As GDP primarily measures market production, it ignores many determinants of social value such as the environment, freedom, health and education (Fleurbaey, 2009; Michaelson et al., 2009). For example, traffic jams may raise GDP as a result of increased volume of gasoline bought on the markets, but traffic jams obviously do not improve the quality of citizens' lives. However, companies that develop and sell products that reduce traffic jams by providing commuters with real-time traffic information, not only contribute directly to GDP, but also simultaneously improve wellbeing through reduced stress, decreased pollution and by saving the valuable time of commuters. Hence, they manage to increase social welfare through their commercially viable products. While there is still no generally accepted method

of value accounting that can replace current accounting practices, we already see an advent of emerging societal-level indices, intended to extend the GDP measures and bring attention to other important aspects of society. To name a few, there are OECD Better Life index, the United Nations Human Development index and Social Progressive Imperative's Social progressive index.

The need to assess intangible value has been also been emphasized recently within the IS value research stream (Kohli & Grover, 2008). In today's complex organizational settings, both businesses and consumers are perceived as the arbitrators of value creation. Therefore, the true benefits of IT to these stakeholders may not be accurately reported if pure financial metrics, or even ex-ante market value are overemphasized (Kohli and Grover, 2008). In view of this, the relationship between the generation and appropriation of value for multiple stakeholders in the context of networks and alliances is increasingly being viewed as a multifaceted construct (Sarker et al., 2012). Hence, it is only through a multidimensional view that we are able to acquire a true picture of the value generated within network relationships (Gil-Saura et al., 2009). Co-creation of value occurs when different companies join together to create new value, most often in situations where the collaborating organizations are not capable of creating similar value on their own (Grover & Kohli, 2012). Co-creation "represents one of the most important streams in the IT value research area" and will gain greater importance as firms expand collaborative relationships with other firms (Grover & Kohli, 2012, p. 231). However, despite the recent focus on the creation of value in collaborative settings, little is known about the underlying mechanisms (Sarker et al., 2012).

### **3. Methodology**

This PhD study was planned as an explanatory research, which seeks explanations of observed phenomena, problems, or behaviors and seeks answers to why and how types of questions. Furthermore, explanatory research attempts to identify causal factors and outcomes of the target phenomenon (Bhattacharjee, 2012). The research philosophy is that of critical realism, which falls between positivism and subjectivism in the post-positivistic tradition, as discussed in section 3.1.

#### **3.1 Research Philosophy**

The term epistemology originates from the term epistēmē, which is the Greek word for knowledge. Epistemology is the philosophy of how we come to generate knowledge. It

is therefore not necessarily a description of methods, but rather the underlying beliefs. Positivism is a specific philosophy of science, often linked to the physical sciences, which use a scientific method to prove or disprove claims of knowledge. Positivism commonly evaluates results with the help of statistical methods and aims for generalizations and quantifiable observations. On the other hand, interpretive philosophy or interpretivism adopts the position that our knowledge of reality is a social construction by human actors (Walsham, 1995). Interpretivism incorporates no objective data, since it works on the assumption that the researcher uses his or her preconceptions to guide the process of inquiry (ibid). Followers of interpretivism frequently favor use of qualitative data, yet the research philosophy by its very nature, does not predict what type of data is used.

Critical Realism (CR) is a growing intellectual movement in various social science disciplines, including the IS discipline (Tsang, 2014; Wynn & Williams, 2012; Williams & Karahanna, 2013; Zachariadis et al., 2013). Two basic philosophical theses comprise the foundation of CR. Firstly, CR expounds the existence of a reality independent of human perception and cognition. Secondly, this reality exhibits its own inherent order (Tsang, 2006). CR predicts that there is a difference between scientific concepts, laws, and theories or what is referred to as transitive objects of knowledge on the one hand, and the intransitive structures and mechanisms of the world, which our theories refer to, on the other hand (Bhaskar, 1978). Intransitivity implies that these structures and mechanisms exist independently of our knowledge of them (Tsang, 2014). CR acknowledges the role of the subjective knowledge of social actors in a given situation, along with the existence of independent structures that constrain or empower human actors pursuing certain actions in a particular setting (Wynn & Williams, 2012). Researchers who apply methodological approaches consistent with the CR paradigm are thus in a more suitable position to provide detailed causal explanations of a given set of phenomena, than those researchers who adhere to positivism. At the same time, these researchers are able to propose general relationships within a given context, which would not fit with the interpretivist paradigm.

Ontology is the philosophical study of what constitutes reality and how we can understand existence. Critical realism assumes a stratified ontology divided into three domains: the real, the actual, and the empirical (Bhaskar 1975, 1978, 1989). This important illustration highlights the fact that even though there is one reality, it does not necessarily follow that researchers have immediate access to it or that they are able to observe and realize every aspect of that reality (Zachariadis et al., 2013). Moreover,

the philosophy of critical realism emphasizes the explanatory role of generative mechanisms (Bhaskar, 1998). Generative mechanisms are used to describe and explain causal relationships through the specifying of how central events or outcomes are produced and reproduced by the structures, actions and contextual conditions in a particular setting (Williams & Karahanna, 2013). Generative mechanisms exist in the real, but can trigger events that happen in the domain of the actual (Bygstad & Munkvold, 2011). These events may (or may not) be observed in the empirical domain. The stratified ontology of CR is illustrated in table 1.

**Table 1: Stratified Ontology of Critical Realism**

	<b>Domain of Real</b>	<b>Domain of Actual</b>	<b>Domain of Empirical</b>
<b>Mechanisms</b>	Causal relationships or tendencies that explain how events happen		
<b>Events</b>		Events result from mechanisms which are triggered (or not) depending on the context	
<b>Manifestations</b>			If events are observed or measured they exist in the empirical domain

*Source: Adapted from Bhaskar, 1975, 2008 and Wynn & Williams, 2012*

CR is important for IS research for three main reasons: 1) CR enables us to take a fundamentally realist stance, while accepting the major critiques of naïve realism. 2) CR endorses both natural and social science, and thus, comprises both hard and soft (and critical) approaches. 3) CR is in conformity with the reality of IS as an applied discipline (Mingers, 2004). CR posits that a special mode of inference must be adopted, in order to explain events “by postulating (and identifying) the mechanisms which are capable of producing them” (Sayer, 1992, p. 107). This logic of inference is entitled retrodution (Bhaskar, 1998; Zachariadis et al., 2013). Retrodution allows researchers to move between the knowledge of empirical phenomena as expressed through measured or observed instances of events to the creation of explanations (or hypothesizing), in ways that hold “ontological depth” and potentially offers some

indications on the existence of unobservable mechanisms (Downward & Mearman, 2006, Zachariadis et al., 2013).

Retroduction is a form of abductive reasoning. In brief, abductive reasoning is a form of logical inference that goes from an observation to a hypothesis that accounts for the observation, ideally seeking to find the simplest and most likely explanation. Abduction involves analyzing data that fall outside of an initial theoretical framework or premise, while retroduction is primarily a method of conceptualizing and theorizing causal relationships. Retroduction refers to the application of previously identified mechanisms and/or identification of new mechanisms to explain the causal events that can generate an observed outcome in a specific context (Wynn & Williams, 2012). Used in conjunction, these forms of inference can lead to the formation of a new conceptual framework or theory (Danermark et al., 2002).

It is generally argued that the retroductive approach to research embraces a wide variety of methods (Downward & Mearman, 2006; Venkatesh et al., 2013; Wynn and Williams 2012; Zachariadis et al., 2013). This approach integrates qualitative and quantitative approaches, with the aim of identifying and hypothesizing about the generative mechanisms that cause the events we experience. CR advocates a marginally different interpretation from the standard data analysis validation criteria of qualitative and quantitative research (Zachariadis et al., 2013). In CR-based quantitative research, internal validity goes beyond confirming observed correlations and interprets them as manifestations of the particular generative mechanism in the context of the field. Findings from correlational statistics can therefore prove beneficial, by virtue of providing information about the relationships of events observed in the empirical domain, while causal assumptions must be hypothesized or tested through further means such as experimental design.

## **3.2 Research Design**

Researchers opine that the preference for either qualitative or quantitative methods may be an obstacle to scientific progress (von Krogh et al., 2012; Venkatesh et al., 2013; Zachariadis et al., 2013). The research questions under consideration and the nature of the phenomenon, as well as the degree of closure that is assumed, ultimately determine the suitability of methods (Downward & Mearman, 2006). For example, a study can address more than one research question, with all the questions focusing on the same phenomenon, which will underscore “different aspects of abstraction and uncover alternative characteristics of the same layered reality” (Zachariadis et al., 2013, 866). Methods such as case studies help establish the context specific

understanding further, by exploring the meaning and mechanisms of particular processes, while econometric methods can be used to explore their generality in the sense that similar demi-regularities might be detected (Downward & Mearman, 2002). Thus, combined use of qualitative and quantitative data may serve to generate unique insight into a complex social phenomenon in an unfolding reality (Bhattacharjee, 2012; Von Krogh et al., 2012).

Mixed methods research has been termed the third methodological movement or paradigm (Venkatesh et al., 2013). Researchers who apply a mixed methods research approach analyze both qualitative and quantitative data and therefore, are capably positioned to use their substantial body of observations to formulate a unified theory consisting of valid concepts and theoretical mechanism, or meta-inferences (Venkatesh et al., 2013). Meta-inferences are defined as “theoretical statements, narratives, or a story inferred from an integration of findings from the quantitative and qualitative strands of mixed methods research.” (Venkatesh et al., 2013, 38).

A concurrent mixed methods design approach refers to quantitative and qualitative data that are collected and analyzed in parallel and consequently, the data are used to create a more holistic view of the phenomenon (Creswell, 2003). As open data are an emerging phenomenon, I undertook my analysis within a constantly evolving reality. Moreover, I wanted to examine the phenomenon from up close and from afar, as is recommended for doing Engaged Scholarship (Van de Ven, 2007). I adopted a complementary approach for the use of mixed methods, using different methods to gain complementary views of the phenomenon (Venkatesh et al., 2013; Zachariadis et al., 2013). I favored this type of design, in view of the fact that the overall goal of my inquiry was firstly to understand and document the phenomenon of open data, and then to explain how open data generates value. Moreover, my design evolved into a developmental approach, as I applied the findings from my qualitative studies to guide the development of constructs and search for quantitative data, which I would subsequently use to test the hypothesized relationships (Venkatesh et al., 2013).

I selected a longitudinal case study as the main qualitative research method, because this method conforms to the fundamental doctrines of the CR philosophy (Tsang, 2014). A case study is defined as “an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (Yin, 2009, 18). In general, the case study is a preferred method when: 1) the investigator asks “how” or “why” questions; 2) the investigator has limited control over events; 3) the focus is on a contemporary phenomenon within a real-life context (Yin, 2009). The

inquiry should rely on multiple sources of evidence, and the data (for example, interviews, observations, documents and archival records) should converge in a triangulating fashion. Strong triangulation of data sources is vital towards establishing the necessary reliability and validity of a research study (Yin, 2009). The case study also benefits from the prior development of theoretical propositions to guide data collection and analysis (Yin, 2009). The strength of the case study research method is its ability to discover a wide variety of social, cultural, and political factors, which are not determined beforehand to be potentially related to a phenomenon of interest that may not be well-known in advance (Bhattacharjee, 2012).

For this PhD study, I conducted two longitudinal case studies. The first case study pertained to the global energy tech vendor Opower, and was carried out with the goal of understanding how value is generated through data-driven innovation. This case study primarily examines the demand or user side of open data. The second case study investigates and analyzes a particular ODI, the Danish Basic Data Program (BDP), and reports on the tensions and governance challenges the program encounters. This inquiry was focused on the supply side of open data. I thus used the qualitative strand for creating a detailed insight into the open data phenomenon based on two real-life cases. I explored the phenomenon through observation and interviews and made use of these data to select and conceptualize relevant constructs and to understand comprehensively the hypothesized relationships. The limitations of a case study analysis, however, are that this form of inquiry demonstrates a tendency to be heavily contextualized and nuanced (Bhattacharjee, 2012). Moreover, interpretation of findings may depend on the observational and integrative abilities of the researcher. The lack of control over events may lead to difficulty in establishing causality, and findings from a single case site may not be readily generalized to other case sites (ibid).

In order to address these limitations I also conducted two empirical quantitative studies (Paper III and Paper VII). I present the conceptual model that is used in those studies in Chapter 5. The model has foundations in state-of-the-art research, but is extended through triangulation of results from qualitative and quantitative data analysis. I encountered numerous barriers when confronted with the task of designing a study that would statistically estimate whether or not openness of data is a relevant enabling factor for societies aiming to stimulate value generation from data. The first barrier relates to the severe complications that arise if one wants to trace the value that governments, companies and individuals gain from using these data. It is understandably difficult to comprehend how societies would have fared without access to these data and experimentation was not an option in this context. Moreover, due to

the very nature of open access, it is a demanding task to identify all the users and usage of multiple sources of data. Consequently, I utilized a correlational approach as my remaining option. I conducted a macro level study comparing countries that are at various distinct stages in the process of opening data. The primary advantage of performing cross-sectional analysis is that it does not rely on the researcher being able to identify the advent of open data in all countries. In many countries, data have been open for years, much prior to the commencement of the mainstream open data initiatives. Instead, it allows the researcher to determine the status of selected variables in one particular year with the aim of detecting whether countries at more advanced stages of opening data will systematically demonstrate superior performance, conceptualized as sustainable value.

The next barrier was to analyze and deduce a method of reflecting the level of sustainable value present in different countries. Numerous macro-level analyses have previously attempted to estimate the overall economic value of open data (see for instance de Vries et al., 2011, Houghton, 2011, McKinsey, 2013a; Vickery, 2011). However, none of these studies have attempted to document the intangible or social dimensions of value, recognized to offer even more benefits than the economic value (McKinsey, 2013a). While ODIs generally highlight the economic potential of increased use of data, in fact, even more attention has been paid to concepts such as transparency and collaboration, with the underlying motive of increasing social value. McKinsey, a premier multinational consulting firm, has highlighted the implications of various informational benefits from open data to consumers, for example the ability to decide which school best satisfies our educational requirements, or the most convenient mode of transport suited to our needs (McKinsey, 2013a). The above examples illustrate the need for a model that can connect the availability of open data with the level of available information and give evidence to whether the existence of such information indicates higher levels of societal level phenomena, such as the overall level of education.

By virtue of their contextual nature, the basic structure of mechanisms is frequently described as a context-mechanism-outcome pattern. Hence, only by employing a structural model can we adequately reflect this structure, consequently calling for use of statistical methods that can estimate structural models. A structural model contains formulas representing the relation of every dependent variable to its independent variables, whereas the reduced form exhibits only the net or overall relation between the dependent variable and the ultimate independent variable (Henseler et al., 2009; Tsang, 2006). The ability to operationalize theoretical but unobservable phenomena



has long been viewed as the core strength of the structural equation modeling method. Thus, I elected to conceptualize sustainable value as a latent variable, applying economic, social and environmental indicators as recommended in the earlier works of Stiglitz et al. (2009). Additionally, the model needed to reflect the importance of the context in which value generation through open data takes place, including elements of the soft infrastructure (King & Uhler, 2014).

The partial least squares (PLS) analysis method was chosen for several reasons. Firstly, PLS is a recommended structural equation modeling method for secondary data analysis (Henseler & Sarstedt, 2013). Secondly, the theory development efforts needed to clarify what are the most essential antecedents of the target constructs. PLS is recommended when the goal of a study is to build rather than test theory, which was the case in this inquiry due to the emergent state of the open data phenomenon (Hair et al., 2011; Ringle et al., 2012; Sarstedt et al., 2014). Thirdly, I used a formative variable to model the robustness of regulatory data and privacy protection framework, and PLS is recommended for models that make use of formative variables. Moreover, the model included both mediating and moderating relationships, indicating high model complexity (Sarstedt et al., 2014). Finally, the data used was a relatively small set of cross-country data, including data that are not normally distributed. Since PLS is based on a series of ordinary least squares or linear least squares (OLS) regressions, it makes minimum demands regarding sample size, and generally achieves high levels of statistical power (Hair et al., 2011). Nevertheless, as is necessary, I did consider the overall characteristics of the data and model in order to ensure that the sample size was sufficient to achieve adequate statistical power (Marcoulides & Saunders, 2006).

The empirical papers (paper III, IV, VI and VII) offer a more detailed discussion of the methodology applied in each of them.

### **3.3 Data Collection**

I collected the qualitative data for the Danish BDP case study between November 2011 and March 2015 through my interactions with various stakeholders directly or indirectly involved in the BDP. Interviews were the main method for primary data collection, but additionally I participated in various events such as the initial meetings for different project phases and project conferences. Moreover, as I worked for the company responsible for implementing the technical data infrastructure, I had the opportunity to observe the BDP from a vendor perspective. Over the project period, I conducted twenty-eight formal interviews with twenty-two interviewees, ranging from half an hour to one and a half hours. In a majority of cases, the interviews occurred in

the respondents' offices and all the interviews were conducted face-to-face. Most of the interviewees communicated in Danish, but for a few interviews, we used English as the medium of communication (mostly non-Danish speaking interviewees). In order to sufficiently document the progress of the BDP over the case study period, specified respondents were interviewed multiple times. The key informant was the program leader, who had extensive dialogues with the author, with reference to the program and its major challenges and opportunities. An overview of the dates, the number of interviewees participating in each interview, the number of interviews with each interviewee, and their roles in the organization, appears in table 6 in Appendix B.

After a few initial exploratory interviews, all the interviews had a semi-structured format. Key questions prepared beforehand addressed the interviewees' perceptions of the BDP, their motivations and their opinion regarding main tensions and challenges. I also asked about what they understood to be the primary value proposition of the program. Some of the interviewees had worked with open data in Denmark for many years. With these interviewees, I investigated further with additional questions that focused on the program's "historical" aspects, in order to improve understanding of the social, economic and political contexts in which this program operated. Furthermore, I examined selected project documents (I was unable to gain access to all documents, due to confidentiality restrictions necessary for the technical infrastructure tender process). Correspondingly, I analyzed publications from the organizations that participated in the BDP to acquire insights on their perceptions of the program's contribution to value generation, and reviewed newspaper articles that discussed the subject of value of open data, or more specifically basic data, in Denmark. Finally, I gained access to a report containing an external review of the BDP, which included information about the main governance tensions. Additional material that supported my analysis comprised of policy documents, tender documents and meeting notes, in conjunction with scientific research papers on open data, interoperability and management of complex systems. These different qualitative sources were triangulated to increase the credibility and dependability of inferences (Teddlie & Tashakkori, 2009; Yin, 2009).

The collection of data for the Opower case study took place over several months in 2013. I conducted one group interview and afterwards engaged in correspondence over email for follow-up questions. Most of the qualitative data used in this study was secondary data, collected over the web, although I gained access to some primary documents we got directly from Opower. I gave Opower an opportunity to comment on the storyline and conclusions before publication to correct any factual errors and to

improve descriptive validity and confirmability of the conclusions (Venkatesh et al., 2013; Zachariadis et al., 2013).

The quantitative part of the study demanded specificity and precision in conceptualization of the constructs, and this proved useful in order to overcome the common problem of deficient construct clarity (Suddaby, 2010). As discussed in section 3.2, I opted to perform secondary data analysis for testing the hypothesized relationships between the liquid open data construct, supplementary enabling factors, the mediating variables and sustainable value. Using secondary data provided an effective means of research in a situation where primary data collection was not feasible due to the very nature of the sample (cross-country comparison was necessary, since the research is positioned at the societal level). I aimed to observe the manifestation of events I had proposed would be instantiated, if the hypothesized mechanisms were activated as data became open and liquid. An important limitation of this method is that I could merely look for these manifestations within the pool of pre-existing measures. Fortunately, I could choose from a large volume of openly available data from respectable data providers, which proved to be highly relevant to our theoretical constructs (an illustration of the value of having open access to data that are easily accessible and downloadable in machine readable formats). One benefit of using secondary data from diverse sources is that they circumvent problems such as common-method/single source bias. The final sample used in Paper VII was a convenience sample of the seventy-six countries that were included in the Open Data barometer (the pilot study in Paper III included sixty-one countries), published by the Open Data Institute (Davies, 2013). Table 7 in Appendix B displays individual item measurements.

#### **4. Assumptions and Meta-theory**

All theories comprise a set of assumptions from which empirical generalizations are derived (Merton, 1949). In the social sciences, a satisfactory explanation must ultimately be anchored in hypotheses or assumptions about individual behavior (Elster, 2007). Constructing a set of explicit, behavioral assumptions enables us to underline the point that all macro-level, societal phenomena are inherently derived from human beliefs and actions. Based upon these assumptions, we thereupon apply the method of retroduction to hypothesize about the mechanisms that are capable of explaining the process by which initial conditions work together to produce human actions, which collectively result in a (potentially) observed outcomes at the societal level. Since the core behavioral assumptions of a theory often form the foundation of its mechanistic

explanations, it is crucial to define these assumptions explicitly, during the early stages of empirical research (Tsang, 2006).

## 4.1 Basic Assumptions

The classical economic model of strategic interaction assumes that people are: a) largely concerned about bettering their own material situation; and b) in order to achieve this, they determine an optimal strategy through perfectly rational judgments. However, new results from behavioral economics discredit these assumptions, since new evidence discloses that people are, in fact, concerned with unobservable outcomes such as reputation, fairness or the well-being of others, as well as the ways in which their own immediate actions might affect their and others future well-being (Shogren, 2012). A comprehensive review of empirical studies of behavioral assumptions pointed to hundreds of studies that discredit the validity of the assumption of self-interest (Rabin, 1998). Individual preferences may undoubtedly be at odds with public welfare. This is characterized as a social choice problem, which may be alleviated via constructive social choice (Sen, 1999). Constructive social choice stipulates that individuals do not simply act on their planned personal and known preferences, but interact with each other to rearrange and shape those preferences (York et al., 2013). Information dissemination through social networks may, as an example, act as a medium for constructive social choice.

The rationality of individuals is limited by the information they possess, the cognitive limitations of their minds, and the finite amount of time they have to make a decision (Shogren, 2012). People typically do not apply sufficient cognitive effort to calculate an optimal strategy, but rather resort to heuristics, which may be influenced by context (Simon, 1957). Bounded rationality is “behavior that is intendedly rational but only limitedly so” (Simon, 1957, xxiv). Limited quality and quantity of information available to firms and individuals introduces a “boundary” to rational behavior, as it reduces an individual’s ability to assess present and future environmental states (Hitt et al., 2011). We must therefore assume that people face a multi-dimensional value function consisting of complex relationships and even contradictions between individual and social well-being. Moreover, we as individuals have limited ability to choose rationally between different options, especially when lacking information.

Proceeding from the above, we should clarify the principal distinction between data and information. The same meaning seems to be frequently attached to these two important concepts in the open data/big data discourse, in some measure due to the novelty of these phenomena. Nevertheless, the data-information-knowledge

relationship is already well established, and constitutes the foundation of the IS discipline (Kettinger & Li, 2010). Data are recorded (captured and stored) symbols and signal readings, whereas information is a message that contains relevant meaning, implication, or input for decision and/or action (Liew, 2007). Knowledge, on the other hand, is reflected in (1) cognition or recognition (know-what); (2) capacity to act (know-how); and (3) understanding (know-why) that is contained within the mind or in the brain (Liew, 2007). If we are to understand the relationship between raw data and value, we must have a basic understanding of the underlying value chain, or perhaps more accurately, value network. It has been suggested that the ultimate purpose of knowledge is for value creation (Liew, 2007).

Kettinger and Li (2010) note that information represents a status of conditional readiness for action, which is generated from the interaction between the states measured in data and their relationship with future states predicted in knowledge (Kettinger & Li, 2010). “Information is the meaning produced from data based on a knowledge framework that is associated with the selection of the state of conditional readiness for goal-directed activities.” (Kettinger & Li, 2010, 415). Thus, new information forms the basis for action through its potential influence on people’s decisions, and consequently, their behavior. An important role of technology is therefore to augment people’s capabilities in dealing with multiple sources of data, in order to generate information as a basis for better operations or decisions (Kettinger & Li, 2010). Following this interpretation, with the aim of facilitating value generation from data, the data should be transformed to information that serves as a catalyst for enhanced decision-making, applicable behavioral change and the ability to act swiftly, more accurately and with optimal use of resources.

As a foundation for the proposed theory, I advance two basic assumptions.

The first assumption is that individuals in general aspire to go beyond increasing their own material wealth in their efforts at value generation. In other words, most individuals possess the necessary intrinsic motivation for sustainable value generation. Accordingly, I assume that individuals will strive to generate sustainable value, including the economic, social and environmental dimensions, for all stakeholders and future generations (van Osch & Avital, 2010). However, while intrinsic motivation is a necessary condition for sustainable level value generation at the individual level, it is not sufficient. A class of behavioral theories, based on the Motivation- Opportunity- Ability (MOA) framework, further explored in Paper V, indicates how extrinsic motivation combined with the opportunity and ability to perform certain tasks influences the behavior of individuals (Blumberg & Pringle, 1982; Reinholt et al.

2011). Consistent with this framework, motivation is defined as goal-directed arousal (MacInnis et al. 1992; Rothschild, 1999). In the context of this study, motivation implies that individuals are provided with an incentive to allocate resources to generate sustainable value from data; opportunity refers to the environmental or contextual factors that enable action; and ability represents the power or capacity to act.

The second assumption I advance is based on the notion of bounded rationality. People will be able to make better decisions if provided with information that is relevant to the situation they face. Access to relevant information therefore pushes the boundaries of our ability to choose rationally and contribute to the generation of sustainable value, which is what we wish to do. Recent research on online shopping behavior has supported the notion that increased transparency of information can significantly influence user behavior, consequently contributing to customers being more satisfied with choices they make (Xu et al., 2014). While having access to too much information can have a paralyzing effect, the right amount of information sufficiently contextualized can really help us make better decisions. Why is this assumption relevant for this study? The underlying motive is my proposal that a large part of value generation from use of open data essentially comes about through the creation and sharing of information over networks. Access to this information subtly influences individual behavior, utilizing the forces of constructive social choice to influence collective behavior, and eventually creating value through means such as improved decisions making.

## **4.2 Social Mechanisms**

Let us assume that we have observed a systematic relationship between two types of events, I and O. The way in which the two sets are linked to one another is expressed through the mechanism M:  $I \rightarrow M \rightarrow O$ . This high-level model represents what can be described as a structural equation, in which the mechanism mediates the effect of I on O. A mechanism can be explained by opening up the so-called black box and thereby making "...explicit the causal cogs and wheels through which effects are brought about." (Hedström & Ylikoski, 2010, 54). The mechanisms approach thus develops a causal reconstruction of a phenomenon by identifying the overarching processes through which an observed outcome was generated (Bunge, 2004; Hedström & Swedberg, 1998; Machamer et al., 2000; Mayntz, 2004). A majority of the accepted mechanism definitions across various disciplines share several general concepts (Hedström & Ylikoski, 2010). First, a mechanism is identified by the type of effect or phenomenon it produces. A mechanism is always a mechanism for something (Darden,

2006). In this study, the mechanisms I aim to uncover are for value generation. Second, a mechanism is an irreducibly causal notion. It refers to the entities of a causal process, which produce the effect of interest. Third, the mechanism has a structure. When a mechanism-based explanation opens the so-called black box, it discloses this structure.

Mechanisms are a key construct in many disciplines of science and are defined as frequently occurring and easily recognizable causal patterns (Elster, 2007). “In the natural sciences, no event or process is regarded as having been satisfactorily understood unless its actual or possible mechanism has been unveiled” (Bunge, 1999, p. 63). Mechanisms are not like the deterministic laws of physics, in which certain inputs positively lead to certain outputs (Elster, 2007; Hedström & Swedberg, 1998). Instead, mechanisms allow us to address the probabilistic nature of social life. In social theory IS research, collective entities and processes unfolding in a social context are understood to typify and represent the generation of aggregate effects from the actions of individuals (Avgerou, 2013). Therefore, a methodological strategy to cope with theoretical complexity is the acceptance of mechanisms in terms of collective actors (ibid). Avgerou (2013) identifies two main spheres wherein social mechanism based explanations augment social theory IS research: First, social mechanisms make explicit the causal paths that produce outcomes of IS phenomena, and thus, lead to a more constructive explanatory theory. Second, being empirically driven, the tracing of social mechanisms is likely to produce new insights beyond those implied by the theories that frame existing research, thus contributing to a more complete, multi-causal explanation (Avgerou, 2013).

### **4.3 Coleman’s Framework – The Micro-Macro Conundrum**

Melville (2010) identified three distinct classes of sustainability phenomena: 1) how cognitive states about sustainability (beliefs, opportunities, motives etc.) emerge; 2) actions of organizations and individuals on the topic of sustainability practices and processes; and 3) environmental and financial performance outcomes. These classes encompass the act of generating sustainable value from data, ranging from factors that influence individual cognitive states to the individual actions that collectively contribute to the generation of value. Viewed as a whole, the three classes of phenomena comprise micro and macro constructs (Melville, 2010). Coleman’s theoretical foundation underscores the mediating role of individuals in linking macro-level structural variables to social level impacts as shown in Figure 5.

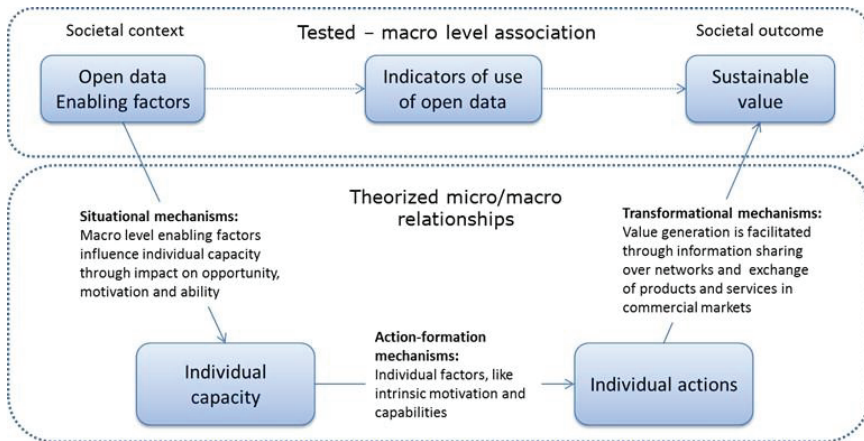


Figure 5: Meta-theory based on Coleman's Framework, adapted from Hedström and Ylikoski (2010)

Following Coleman (1986, 1990), we deduce that the goal of the social system is to maximize utility or what we have defined as sustainable value. We use Coleman's framework as a meta-theory to explain the micro to macro level relationship between use of open data and the generation of sustainable value. Coleman (1990) proposes that a theory which can generate macro-level empirical generalizations as specific propositions may be thought of as "a theory of individual action, together with a theory of how these actions combine, under specific rules, to produce systemic behavior" (Coleman, 1990, p. 20).

Three types of relations are included: 1) macro-level variables that explain the societal context which affects individual capacity through situational macro-micro mechanisms (by influencing the motivation, opportunity and ability of individuals); 2) individual capacity affects individual action through action-formation micro-micro mechanisms; and 3) collective individual action affects macro-level constructs such as sustainable value through transformational micro-macro mechanisms (Hedström & Swedberg, 1998). In Paper V, I propose that the opportunity for value generation through use of data is created through the act of making data open. By constructing a robust regulatory data protection and privacy framework that produce trust and stability in information markets, we gain the means to influence the extrinsic motivation of stakeholders to use data. Furthermore, governments can motivate individuals if they demonstrating strong IT leadership, for instance by organizing events like hackathons, by promoting use of data or by being forerunners in utilization of digital solutions.



Finally, individual ability and expertise can be influenced by the supply of relevant skills and tools in the open data ecosystem.

## **5. Towards a Theory of Value Generation through Open Data**

The stratified ontology of Critical Realism (CR) posits that the underlying mechanism of the domain of the real generate events in the domain of the actual. While the causal powers themselves are unobservable, it is nonetheless possible to observe the generated events through manifestations or measurements. Different events, E1, E2 etc. can thus leave behind empirical traces or measures. The domain of the empirical reflects these observed events. Although these measurements may be applied in traditional empirical methods, cognizance of the fact that these manifestations are caused by underlying mechanisms that operate within a specific context, is essential (Zachariadis et al., 2013). Therefore, causal laws should be ontologically distinct from patterns of events (Tsoukas, 1989). As stated by CR, generalizations are valid when we are confident that similar or other events that arise (or may arise) in other contexts are caused by the same generative mechanisms that led to the actual events in our research domain. To ensure external validity, distinction must be drawn between the contextual or external elements that partially influence the effects we identify, and the internal (necessary) properties of the phenomenon we study (Danermark et al., 2002; Smith, 2006; Zachariadis et al., 2013).

Retroductive methodology focuses on research not as a discrete event but as a creative process with different phases that involve different types of activities (Bhaskar 1978; Zachariadis et al., 2013). Its main objective is to link the structures and causal powers of the objects under study, to the events we attempt to explain through the notion of causal mechanisms (Wynn & Williams, 2012). From a CR perspective, even though structures are comprised by situated interactions which are better understood qualitatively, it is also useful to employ quantitative measures to identify certain characteristics of a phenomenon. In that respect, statistical descriptions are regarded as helpful (Zachariadis et al., 2013). In the spirit of CR, the quantitative empirical work in this study was intended to gain more understanding of a complex phenomenon, rather than to test and confirm an already developed theory. During the quantitative analysis, I had to face multiple challenges related to construct definitions, measurements and relationships; challenges that have both brought me increased understanding of the open data phenomenon and provided a fertile ground for future research. The conceptual model that is presented as a conclusion of this PhD study should be considered as a stepping stone towards a more complete understanding of the

relationship between the open data phenomenon and subsequent generation of sustainable value, and as a contribution to the nascent but growing scientific literature on open data.

The overall approach in this study is similar to that described in Zachariadis et al. and illustrated in figure 6. First, we defined the topic of interest and purpose of study. This resulted in a research proposal where the research questions were identified. The research proposal was presented at IRIS 2012, the Scandinavian Chapter of AIS yearly conference where I got good feedback (this paper is not included in the study). Afterwards, I started the literature review and early, exploratory interviews with the aim of advancing theory development.

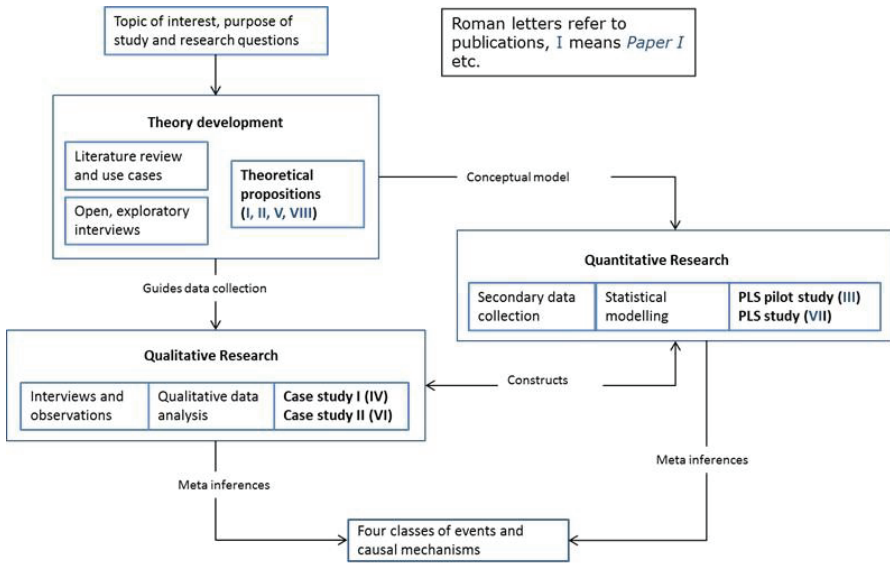


Figure 6: Retroductive Research Process, adapted from Zachariadis et al. (2013)

The first published conference paper (Paper I) introduced an early version of the framework discussed in more detail in section 5.2. The second conference paper was presented at ECIS 2013, introducing an early version of the conceptual model discussed in section 5.4 (Paper II). I continued to collect qualitative data, using these two frameworks to guide the data collection, as recommended by Yin (2009). Originally, I had planned to do a survey, but after Paper II had uncovered the public good features of the phenomenon, I opted to look for secondary, societal level data to

test the conceptual model. The pilot study (Paper III) was published at ICIS 2013 and has gotten good response, although the approach used is not traditional for IS research. After this study, I felt I needed more information from the user side to complement my thoughts about the relevant dimensions of open data and the most important societal level enabling factors. Due to delays in the Basic Data Program, I searched globally for a successful open data user, someone who could demonstrate documented societal level impacts. I found the company Opower, who fortunately agreed to cooperate on the development of a case study (Paper IV).

Consequentially, reflecting on the strong motivation to generate sustainable value that guided the founders of Opower, another conference paper was presented at IFIP WG 8.6 conference. This article (Paper V) was the first to introduce Coleman's framework. A class of behavioral theories called MOA (Motivation-Opportunity-Ability) was used as a theoretical lens for proposing how open data and other societal level enabling factors could influence individual action and stimulate the four previously identified generative mechanisms. In this paper, I used four use cases to demonstrate four different ways in which individuals and companies have used open data to address difficult societal challenges, consequently generating what I conceptualized as sustainable value.

Using the MOA-framework as a guideline, alongside new findings from data and literature, I changed the explanatory variables originally presented in Paper II, as illustrated in figure 7. Simultaneously, the case study of the BDP (Paper VI) was gradually taking shape. Paper VI presents a final version of the framework of four mechanisms and the final version of the seven dimensions of liquid open data, both of which I use as theoretical lenses for guiding data analysis. The main contribution of Paper VI is a process model that reflects the four main events and causal mechanisms of an ODI initiative. The final PLS study (Paper VII) uses the conceptualization of liquid open data from Paper VI and consequently attempts to measure liquid open data with secondary data over seventy-six countries. This study also suggests that the two overarching mechanisms from the framework of four mechanisms should leave some identifiable manifestations or empirical traces. Thus, I searched for empirical data that would indicate the dissemination of shared digital content and use of IT for developing new digital product and services (discussed further in section 5.4).

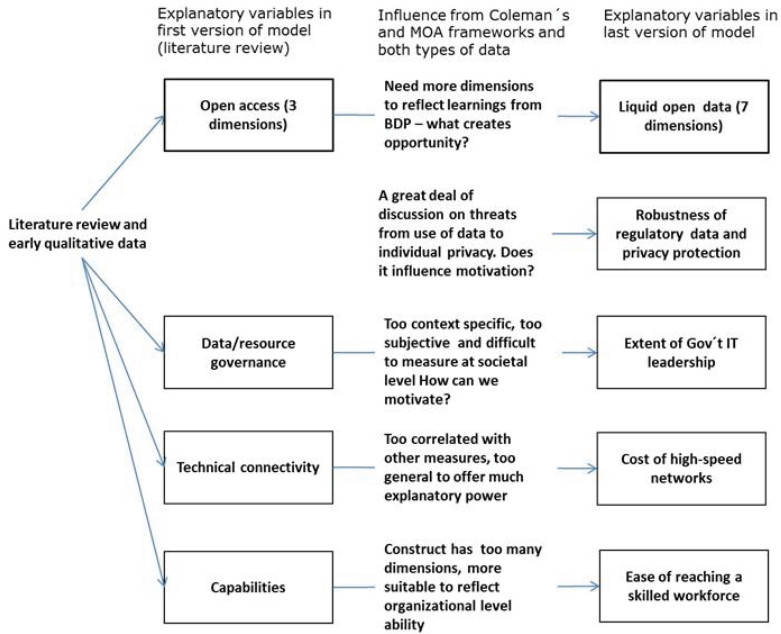


Figure 7: Development of Explanatory Variables over course of Study

Finally, I reflected back on the use cases I had collected on individual companies and users, which had already managed to transform open data into sustainable value. Paper VIII discusses these reflections in an exploratory study, suggesting how intermediaries have facilitated this transformation by capitalizing on the interplay between the market and the information sharing mechanisms. Paper VIII also discusses how investment in open data in general, and in intermediary platforms more specifically, depends on how we evaluate the network based interactions that generate value through non-market based mechanisms such as information sharing.

## 5.1 Seven Dimensions of Liquid Open Data

Constructs are the foundation of theory. Theory can be viewed as a “system of constructs and variables in which the constructs are related to each other by propositions and the variables are related to each other by hypotheses.” (Bacharach, 1998, 498). Just as constructs are the building blocks of strong theory, clear and accurate terms are the fundament of strong constructs (Suddaby, 2010). Accordingly,

in order to offer conceptual clarity we must articulate the constructs of our proposed theory.

Providing open data (supplier side view) has been proposed as being a matter of availability, accessibility, format and license (Davies, 2010). From the demand point of view, openness is proposed to combine unrestricted availability of data with accessibility and technical interoperability (Tammisto & Lindman, 2012). In practice oriented literature, the term open data is interpreted in a variety of fashions, as evidenced from the many different working definitions found online. The Open Knowledge Foundation defines open data as “data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike” (OKF, 2015). However, this definition lacks references to the technical dimensions of open data. Alternatively, Berners Lee’s five stars of linked data specify a number of technical dimensions. However, the five stars are not really an open data definition, but rather a maturity model that focuses on how to gradually transform data into linked data. Linked data is a method of publishing structured data so they can be interlinked and discovered through semantic queries.

An overview over multiple definitions of open data is presented in Verhulst et al. (2014). This overview shows that the currently used definitions usually highlight 2-4 dimensions, and no two definitions are based on exactly the same dimensions. Thus, I turned to my data see if they could reveal the most relevant dimensions. From my qualitative data from the Danish BDP, I could infer that members of the program had very different opinions on the most important attributes of data as a valuable resource. Some of the basic data were already free and program participants working with these particular datasets focused mostly on improving data quality and coherence of data. Other BDP members were working on data modeling and yet others were responsible for developing a common license. A group of members and data users advocated the case for making data available free-of-charge. Programmers I interviewed were focused on the technical challenges involved in developing APIs and web services. Other technical people were designing and building the technical open data infrastructure. I concluded that many of the challenges faced by program participants were directly related to the multi-dimensionality of the open data phenomenon.

For the quantitative studies, I had been using empirical data from The Open Data Barometer, which defines “truly open” data as data that are available online, in bulk, and under an explicit open license (Davies, 2013). In spite of the fact that considerable efforts have been exercised to make diverse government data available to the public, less than one in ten public datasets reviewed in seventy-seven countries in 2013 could

be classified as truly open according to the Open Data Barometer definition (Davies, 2013; Höchtl et al., 2014; Zuiderwijk & Janssen, 2014). Thus, the question remained, which dimensions are important for stimulating engagement with and use of open data? In order to corroborate my data against previous scientific research I reviewed a number of academic articles that have suggested the main barriers and enablers encountered by several ODIs. The barriers that could be traced to attributes related to the data themselves can be broadly classified as follows (Conradie & Choenni, 2014; de Vries et al., 2011; Halonen, 2012; Janssen et al., 2012; Martin et al., 2014; Meyer-Schönberger & Zappia, 2011):

- 1) Availability issues (data only available to specific groups or not available at all).
- 2) Economic issues (too high prices limit use).
- 3) Legal issues (not standardized licenses, unavailable licenses, laws that limit dissemination or potential use of data).
- 4) Usability issues (data are not published under open standards or in machine-readable formats, data quality lacking).
- 5) Discoverability issues (links to data buried in website hierarchies, no use of metadata to facilitate search, data not linked to other data, no central repository or portal).
- 6) Accessibility issues (lack of download possibilities, no bulk download nor web services or open API's, not secure access, not standardized or sustainable access opportunities).
- 7) Interoperability issues (data buried in silos, not published with compatible identifiers, no data models that explain syntax and semantics, or data models not openly published).

Thus, through triangulation of qualitative and quantitative data, as well as a synthesis of the dimensions found in the literature, I finally managed to create and define a construct I call Liquid open data. The extra word liquid is intended to indicate the importance of the technical dimensions for making open data more useful to external re-users. Truly liquid open data are defined as *data that are available online, free-of-charge and under an open access license, published in machine-readable formats, easily discoverable, accessible and conceptually coherent. Liquid open data can be re-used without discrimination or limitation, linked to other data and streamed across systems.*

Table 2 demonstrates a framework consisting of seven dimensions of liquid open data.

**Table 2: The Seven Dimensions of Liquid Open Data**

<b>Dimension</b>	<b>Description</b>
<b>Strategic: Availability</b>	Availability reflects the strategic importance of open data, ranging from all government data that are not subject to privacy or national security limitations being open to all by default, to government data in general not being available outside of organizational boundaries.
<b>Economic: Affordability</b>	Affordability is an economic dimension and refers to the pricing of data, ranging from data being completely free-of-charge to data being extremely expensive.
<b>Legal: Reusability</b>	Reusability is a legal dimension and depends on the type of license used for government data intended for reuse. The license can range from a type of creative commons licenses that allows anyone to use the data for whatever purposes they like, to very strict licenses that allow use for a single purpose only.
<b>Technical: Usability</b>	Usability is a technical dimension and refers to the clarity and ease with which we can interact with the data. If data are not usable, it is difficult to use them for purposes other than originally intended. Data need to be of high quality and presented using standardized machine-readable data formats.
<b>Technical: Discoverability</b>	Discoverability is also a technical dimension refers to whether potential users can easily discover the data and find information about the data. Highlights use of metadata.
<b>Technical: Accessibility</b>	Accessibility refers to whether data are easily, consistently and securely accessible and downloadable or streamable. Highlights use of open standards.
<b>Technical/ conceptual: Interoperability</b>	Interoperability of data refers to data being conceptually open. Interoperable data are published in a way that enables the data to be used outside of the context within they were collected. Interoperability of data depends both on structure (syntax) and meaning (semantics) and calls for use of unique identifiers for linkability. If data are interoperable, they are also “liquid” in the sense that they can stream across systems and easily be linked to other data.

Openness as defined in the context of liquid open data is not a binary feature, where data are either open or closed in the sense of them being available or not. Rather, we

can view openness as a continuous, multidimensional variable, ranging from data being closed over all dimensions, to data being truly liquid and open. Accordingly, data can have different degrees of openness, distributed across these dimensions, as explained in Table 2 and in more detail in Paper VI. All of the above dimensions will influence how users can and will engage with the data. Within each dimension, ODIs should balance the costs and the benefits related to supporting the desired value generating mechanisms (these are discussed in section 5.2). As a practical result and extension of this study, I am using this framework is to design a methodology for use in ODIs that are interested in stimulating the desired impact from open data with the limited resources at hand. Figure 8 illustrates as an example evaluation of three important datasets in Denmark over the seven dimensions. The evaluation is based on my own observations as well as on data from the Open Data Barometer in 2013 and 2014 and from the Open Data Index (Open Knowledge Foundation).

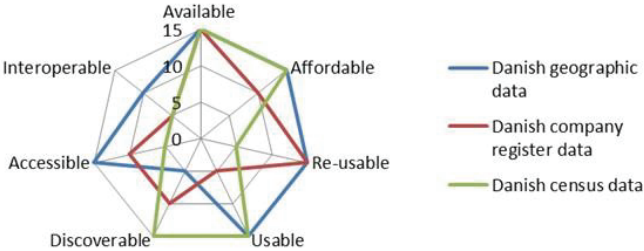


Figure 8: Evaluation of Datasets based on the Seven Dimensions of Liquid Open Data

An important aspect of achieving liquid open data is the implementation of a technical open data infrastructure, which can be crucial for the last four dimensions of liquid open data. Implementation and governance of such an infrastructure was amongst the topics of discussion in Paper VI. As can be seen in figure 8, open data originate from a variety of organizations, which I like to call data custodians. Accordingly, successful dissemination of liquid open data can rarely be achieved by an individual public organization in isolation. Collaboration between multiple stakeholders across the public and the private sectors is an important success factor for an ODI (Ubaldi, 2013). The choice of governance mechanisms can be a deciding factor for the success of such collaborations (Provan & Kenis, 2008). ODI governance must address potentially competing motives and ensure that individual data custodians have sufficient autonomy, without introducing too much governance related overhead, as resources are



generally scarce. Paper VI addresses the four following ODI governance tensions in more detail.

#### *Tension 1: Simplicity vs. Comprehensiveness*

Every ODI has to balance their ambitions for open data with the level of funding they receive. While disseminating high quality data that are liquid and open across all seven dimensions of liquid open data is a tempting idea, it is very difficult to achieve in reality. The approach chosen by the BDP was to focus on a limited number of key datasets and to develop comprehensive data modeling principles for these data. These data modeling principles are based on EU Inspire standards whenever possible and are general enough to be reused by other public sector organizations in their efforts to publish liquid open data. Interoperability of data across different domains can thus be improved, without including too many sets of data in the first round of data modeling and data publishing.

#### *Tension 2: Autonomy vs. Control*

Open government data do not originate from a single organization and therefore can be difficult to publish in a coherent manner. However, while each data custodian is collecting data for their own (regulatory) purposes, these data can undeniably be of much use to other organizations. The data custodians must have enough autonomy to fulfill their individual roles, while contributing towards a common goal of liquid open data, which means using a set of common standards. In order to achieve a balance between those two competing demands, the BDP chose a governance approach called System-of-Systems governance (see Paper VI for more detail). A Systems-of-Systems can be defined as a collaborative set of systems where the components are independent dedicated systems that are separately acquired and integrated to form a single system, yet maintain a continuous operational existence independent of the collaborative system (Rechtin & Maier, 2000). This style of governance seems to be well suited for a constellation of loosely coupled participants, although demanding a high level of network governance skills (Provan & Kenis, 2008).

#### *Tension 3: Exploration vs. Exploitation*

Promoting a common goal for publishing liquid open data is highly important. However, it is not likely that all public and private organizations share a common view on the most important value drivers of open data. While some may emphasize the importance of private sector innovation, others may be more interested in efficiency gains within the public sector. Consequently, these stakeholders may subscribe to different, perhaps competing implementation methods. The BDP case study indicates

that these tensions might actually be resolved by maintaining a focus on all the seven dimensions of liquid open data, which was possible in this case due to a limited number of datasets included in the program. The infrastructural features of open data have resulted in some unexpected synergies across value generating mechanisms, reflecting the serendipitous value generation opportunities offered by open data. Infrastructural resources are considered as shared means to many ends, which satisfy the following three criteria: 1) they are non-rivalrous, 2) social demand is driven primarily by downstream productive activities, 3) the resource can be used as an input for a wide range of purposes (general purpose criteria) (Frischmann, 2012).

#### *Tension 4: Short term gains vs. long-term investment*

While it is natural for any ODI to focus on low-hanging fruits, it is important not to do so at the expense of future, currently unknown applications of the data. Thus, it remains a challenge to find a balance between being openly publishing a large number of datasets and ensuring that continuing publication of these data is actually sustainable. The two case studies conducted as a part of this study revealed that users hesitate to engage with open data unless they are convinced of the sustainability of the data resource. The chosen strategy of the BDP was to finance their ODI upfront, by effectively transferring funds from organizations that are expected to benefit from the liquid open data to the data custodians, which are responsible for remodeling data and building the technical infrastructure. This ensured enough funding to reorganize data-collection and data modeling efforts across central and local government, which is a long-term investment. Simultaneously, individual data stewards had the means to publish their open data through new data services, making the data available for interested users, although not yet in a fully coherent manner.

These governance mechanisms are important for understanding how data can become open and liquid. Therefore, they fall beyond the scope of the conceptual model I present as the overall biggest contribution of this dissertation. The conceptual model focuses solely on the relationship between data that are already liquid and open to some degree or another, the mediating variable that are intended to reflect manifestations of underlying generative mechanisms, and the resulting impacts, conceptualized as sustainable value.

## **5.2 A Framework of Four Mechanisms**

It is imperative to emphasize the importance of creating a basic understanding of the main mechanisms that explain how value is generated in today's networked digital society, irrespective of whether the value is reflected through monetary measurement

(added income or lesser costs), or through other more subjective indicators (a more meaningful life for instance). In order to gain this understanding, we must move beyond positing a simple statement of desired outcomes, which in turn, would imply knowledge of the underlying mechanisms. Alternatively, we should endeavor to open the black box of mechanisms that would explain how value is generated. Detecting and consequently recognizing these mechanisms would greatly improve our ability to understand the complex interactions that happen when value is generated through the use of open data.

Paper I discussed open data strategies and identified four leading categories of value drivers, which explicate the demands for governments to open up their data. Firstly, open data are considered vital for creating transparency and accountability (Meijer et al., 2014). A second driver is participatory governance: Open data are used as a tool to enable citizens to participate in decision-making processes in an informed and structured manner (Bartenberger and Grubmüller, 2014). Thirdly, many analysts view open data as a catalyst for creating new data-driven applications and services, in addition to new business models (Lindman et al., 2014; Zuiderwijk et al., 2014a). Finally, open data correspondingly offer an important internal value for the public sector itself. Public sector organizations may obtain access to data held by other public authorities, or else they may design potentially improved and novel techniques to use their own data. This enables them to significantly improve the efficiency of public services, and enhance their internal understanding of their fundamental tasks and objectives (Halonen, 2012).

After writing Paper I, I proceeded to discover if these value drivers could be conceptualized as mechanisms. Using the method of retroduction, I initially examined previously identified mechanisms that would assist in explaining value generation (Wynn and Williams, 2012).

*The one mechanism that economists relate most of their analysis to – their master mechanism, so to speak – is the market. (Hedström and Swedberg, 1998, p. 3).*

A natural first choice was to consider market mechanisms. Market-type mechanisms are a broad concept. The Organization for Economic Co-operation and Development (OECD) has adopted a very comprehensive definition of market mechanisms as “encompassing all arrangements where at least one significant characteristic of markets is present” (OECD, 2005, 131). The types of mechanisms that are classified as market mechanisms utilize the forces of demand and supply to determine prices and quantities of goods and services offered for sale in a free market. We may infer that the

transaction generates value for both parties, as they are at liberty to enter the transaction and, would do so only if they perceive value from the exchange. The value generated for the buyer is to fulfill his or her specific needs, while for the seller, the value generated is the monetary value received to actualize the transaction. Governments in western societies are increasingly making use of the market mechanisms or market-like practices, the chief motive being the necessity for governments to attain increased value for money expended in their operations (Blöndal, 2005; Henriksen, 2006).

*Mechanisms form a hierarchy (Hedström & Ylikoski, 2010, p. 52).*

While a mechanism at one level presupposes or takes for granted the existence of certain entities with characteristic properties and activities, it is acknowledged that there are lower-level mechanisms, which are capable of explaining them comprehensively (Hedström & Ylikoski, 2010). I propose that markets do indeed play a major role in facilitating value generation through use of open data. This occurs primarily through two classes of lower level (micro to macro) mechanisms: efficiency mechanisms and innovation mechanisms, as identified both in open data literature and in the qualitative data collected for this study.

The significance of the efficient use of public resources for economic growth and stability, along with general welfare and security, has been brought to the forefront by numerous developments over the past decades (Afonso et al., 2010). Public sector organizations are capable of achieving efficiency by cutting processing costs, making strategic connections between and among government agencies, and creating empowerment (European Commission, 2006). The aim of efficiency is to improve resource allocation, in order to minimize waste and maximize the outcome value, given a fixed pool of resources. The markets are important for creating efficiency because their primary goal is to facilitate efficient use of resources. While public sector efficiency is widely cited as one of the potential benefits of open data, the literature review conducted in this study confirms a lack of empirical and theoretical research that extensively examines the core relationships between open data and public sector efficiency. Nevertheless, Paper VI suggests that openly sharing data reduces administration and transaction costs through lesser demands for manual data entry, reduction in IT infrastructure investment needs, and by providing faster and easier access to information. Moreover, liquid open data enable automation of cross-organizational processes and increased interoperability between organizations. All these factors work towards improvement in public sector efficiency, in the sense of getting more value for money.

In the particular case of the Danish BDP, I could soon conclude that the Ministry of Finance was mostly motivated by the potential of using open data to improve efficiency and reduce costs through less use of resources (refer to Paper VI for details). Moreover, it was apparent that different public sector organizations subscribed to different value drivers. Alternatively, the Ministry of Business and Growth emphasized the potential impact from private stakeholders who would use the public data to design and market new products and services, spawning new companies and contributing to job creation. Drawing from Schumpeter's economic theory, innovation is the source of value creation, resulting in novel combinations of resources, new production methods, as well as new products and services, which, in turn, lead to the transformation of markets and industries and thus contribute to increasing value. Numerous studies have confirmed the relationship between macro-level business innovations and economic value (commonly conceptualized as economic growth). The social impacts of innovations have, however, been much less discussed and analyzed, with the possible exception of Simon Kuznets (1974), who divided the economic and non-economic consequences of technological innovations.

The economic properties of open data suggest that data is an infrastructural resource, which in theory, is used by an unlimited number of users and for an unlimited number of purposes as an input to produce goods and services (OECD, 2014). Recent technological developments have provided firms with the ability to collect, manage and use different types of data in multiple ways to innovate, and subsequently create value (Koski, 2013). Overall, empirical studies suggest a positive impact from innovative use of data of approximately 5% to 10% on productivity growth, depending on a number of enabling and complementary factors (OECD, 2014). New digital products driven by the increasing use for and use of data include new technologies which help to manipulate large volumes of unstructured data (for example, Hadoop), as well as multitude of new data analytics and visualization tools (for example, Qlik and Tableau). Moreover, driven by increasing availability of data, we are now able to witness the rise of platforms that use government data to provide services, which among other things help people locate the fastest route, save energy or buy their dream house (refer to narratives and further analysis in Paper IV, Paper V and Paper VII).

I have categorized the two drivers of efficiency and innovation as lower level market mechanisms due to their strong economic focus and use of the markets to allocate resources. However, this was evidently not the case for the two other value drivers, transparency and civic engagement.

*Sunlight is said to be the best of disinfectants; electric light the most efficient policeman. (Brandeis, 1914).*

Transparency does not depend on commercial markets, although it can be beneficial to markets as transparency is expected to reduce information asymmetry. Most definitions of transparency recognize the extent to which an entity reveals relevant information about its own decision processes, procedures, functioning and performance. Unfortunately, it is common to see the concepts of open data, open government and transparency used to convey similar ideas in the open data literature. Opening access to chosen public documents does not necessarily contribute to a transparent government (Gurstein, 2011; Relly et al., 2009; Yu & Robinson, 2012). A government can provide open data on politically neutral topics, even as it remains deeply opaque and unaccountable (Yu & Robinson, 2012). In this PhD study I tend to view transparency as a tool against information asymmetry, moral hazard and corruption, as it creates a more equal access to important information and helps shed light on otherwise hidden activities. The transparency mechanism only generates value if it reduces information asymmetry through information creation and sharing over networks. While the power of market mechanisms can be measured via currency, we must measure the power of transparency through the “informativeness” that is created, not by counting numbers of published datasets.

Open data is also widely believed to contribute to increased public participation and collaboration. Public participation allows citizens to contribute their ideas and expertise allowing governments to create policies that reflect constituent driven information from all reaches of society. By this definition, participation provides a wider breadth of citizens with a voice in government (Linders & Wilson, 2011). Information is considered vital to democratic participation (Jaeger, 2007). Information matters “in the processes by which citizen preferences are formed and aggregated, in the behaviors of citizens and elites, in formal procedures of representation, in acts of governmental decision making, in the administration of laws and regulations, and in the mechanisms of accountability that freshen democracy and sustain its legitimacy.” (Bimber, 2003, p. 11). Currently, citizen participation in public administration decision-making is entering a new phase, as many government agencies have utilized information-based applications to communicate with constituents in order to provide online services, often termed e-participation (Kim & Lee, 2012). The idea behind using these applications is to lower the barriers for those willing-but-unable, and to make participation more attractive to those able-but-unwilling (Axelsson et al., 2010).

However, the exact benefits of the so-called e-participation are still not fully understood (Andersen et al., 2007).

In this study, I prefer to utilize the terminology civic engagement to move beyond the traditional meaning of public participation in government decision making and address both the participatory and collaborative impacts of open data. The civic engagement mechanism is defined as a collective source of positive change that occurs when multiple stakeholders start sharing information across boundaries.

*Civic engagement means working to make a difference in the civic life of our communities and developing the combination of knowledge, skills, values and motivation to make that difference. It means promoting the quality of life in a community, through both political and non-political processes. (Ehrlich, 2000, vi).*

Neither the mechanism that explains how value is generated from open data via transparency, nor the mechanism that explains how value is generated via civic engagement, make explicit use of the markets. In fact the shift towards an economy centered on information, and the move to a networked Internet-based environment have caused significant attenuation of the limitations that market-based production places on the pursuit of value (Benkler, 2006). In general, I propose that the previously dominant role of the market mechanisms is slowly giving way to new types of value generation mechanisms, which have not been previously identified. Private sector companies are slowly shifting towards a use of mechanisms that generate value through network effects. These effects are most commonly realized when dissemination of free information attracts relevant stakeholders to digital platforms where market mechanisms can subsequently be enacted (see analysis in Paper VIII). This new type or classification of mechanisms is termed the information sharing mechanisms.

The information sharing mechanisms create mutual gain for a network of parties through the simultaneous creation, dissemination and appropriation of information by many stakeholders. Unlike the bilateral market mechanisms, these types of mechanisms are many-to-many, indicating a value network rather than a value chain. In such a network, the size of the networks and the interactions themselves are primary sources of value (Bowman, 2014; Viscusi et al. 2014). When an increasing amount of information is available free of charge online, and networks are used to share this free content, we have no currency-like 'token' that indicates how the users of the network value this information. Accordingly, it is very difficult to quantify the resulting value. However we can assume that the same principle applies to the information sharing

mechanisms as the market mechanisms; the information content will only be created and disseminated, and the ‘transaction’ initiated, if the information creator perceives that doing so generates some kind of value. Correspondingly, the content will only be used if the user perceives a similar value.

Figure 9 illustrates a framework of four archetypes of value generating mechanisms.

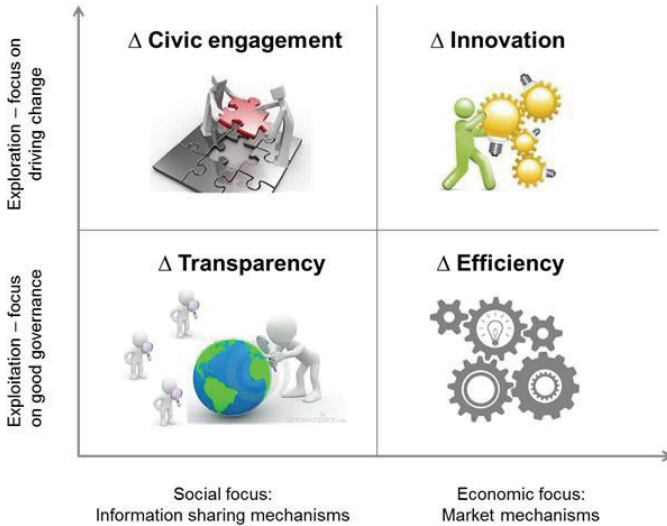


Figure 9: A Framework of Four Value Generating Mechanisms

The framework in figure 9 highlights two principal types of mechanisms that facilitate how value is generated through open data: the information sharing mechanism type and the market mechanism type. Additionally, the figure emphasizes that for each of these mechanisms, value generation can happen either through exploitation of current resources or through exploration, where the focus is on driving change.

### 5.3 Sustainable Value

The 2014-2015 Global Competitiveness Report highlights the fact that policymakers, businesses, and citizens are increasingly aware that economic growth must strike a balance between the creation of opportunities and benefits for all segments of the population, and a long-term, comprehensive focus on a sustainable environment (Schwab, 2014). An examination and discussion of the social and environmental dimensions of an economy is considered to be fundamental in any growth or



development agenda. This PhD study supports the notion that the world is at an inflection point where technological advances and boundary crossing societal challenges have come together to create a paradigm shift in our collective thinking. This notion was very evident at the 2014 and 2015 Annual Meeting of the World Economic Forum in Davos, where both technological shifts and social challenges were repeatedly mentioned. The societal challenges are numerous and urgent, besides being social, environmental and economic in nature. They range from economic inequality, unemployment and poor social conditions to chronic diseases and climate change. I propose that given the complexity and cross-boundary nature of these challenges, a new approach is necessary. In particular, we need an approach where social and technological progress co-evolves in order to generate value (OECD, 2011).

Despite the previously mentioned shifts, the valuation structures of our economies have not changed much in recent years and are still predominantly based on market based economic value. We use monetary measures to compare the progress of societies (growth), to make decisions on which companies to invest in (profits), and so on and so forth. This is not at all surprising, as we have an entire system based on these measures, a system which has served us very well. However, due to the technological and social shifts our societies are undergoing, the predominant monetary based measurements are becoming increasingly problematic. For instance, despite the fact that creation and use of digital information is growing exponentially, the share of the information sector (software, publishing, motion picture and sound recording, broadcasting, telecom, and information and data processing services) has not grown significantly in the national accounts since around the 1960s, and is still around 4% of GDP (Brynjolfsson and Saunders, 2009). As the national accounts present the sum of value-added in market-based transactions, dissemination and use of free data and information services are not accurately accounted for in the GDP measure. While GDP is an accurate metric of value in industries like steel or automobiles, it will miss most of the value in information goods (Brynjolfsson and Saunders, 2009; Brynjolfsson and Oh, 2012).

*What we measure affects what we do; and if our measurements are flawed, decisions may be distorted (Stiglitz et al., 2009, 7).*

To give an example: No financial transaction takes place when a citizen uses free online information created from open government data to improve their decisions regarding energy use, which will consequently impact CO2 emissions. Thus, according to common accounting methods, such a process generates no value. The conceptual distinction between value creation and value capture has been growing in importance

with the recent trends towards openness, technical connectivity and collaborative ventures. Value creation materializes when the utility of members in society increases, after accounting for the resources used in that activity. Value capture materializes when an actor is able to capture a portion of the value created by an activity (Bowman & Ambrosini, 2000). Traditional markets facilitate equilibrium between the two, bringing together a buyer and a seller in a transaction where currency is used to indicate how valuable the products or services are to the buyer. Alternatively, in information sharing mechanisms, multiple stakeholders collaboratively create value for society without there being any record of currency changing hands. Moreover, external stakeholders will benefit from this value generation without paying, as the resulting value includes not only economic, but also social and environmental dimensions.

Building upon these notions, I have formulated two definitions: 1) Sustainable value is defined as a multidimensional construct, *reflecting the simultaneous generation of sustainable economic, social and environmental value in society by multiple collaborating actors*. 2) The open data ecosystem is defined as a *distributed multi-stakeholder network built around the sharing of data across boundaries in order to create information, products and services that contribute to sustainable value generation*.

## **5.4 Research Model**

The overarching expectation is that better use of the troves of data that are currently being created and shared will offer considerable benefits for society, organizations, and individuals in the coming years (Beardsley et al, 2014; Kundra, 2012; European Commission, 2011b). There exist many good examples that make a strong case for increased use of open data and information sharing. For instance, while the genomics revolution accelerated genome mapping through gene-sequencing machines, the real revolution began when researchers started to share data (King & Uhler, 2014). Such payoffs have already spread to other fields, for instance through exploitation of open geographic information systems and open social network data (ibid). However, the open data phenomenon is still at an early stage, which demands an exploratory approach to create a new explanatory theory. The emerging new patterns of networked community governance, driven by openness and sharing, have thus far lacked the support of a distinctive economic and social theory (Benington et al., 2011).

To exploit the opportunities offered by open data, we must move beyond technical challenges, to the challenges of "soft infrastructure", for example institutional factors, governance, and cultural inertia, which tend to impede payoffs from the rapid

evolution of techniques and methods (King & Uhler, 2014). From the open data literature, various reports on the subject and a collection of relevant use cases and case studies, I have identified a number of societal level contextual factors that I propose are important for stimulating value generation through open data. Among the macro-level enabling factors that have been said to be most important for increasing use of data are; robust high-speed Internet networks; a workforce with relevant skills and education; and consumer trust regarding the protection of personal data and privacy (Bilbao-Osorio et al., 2014). Moreover, the public sector plays an important role in facilitating value generation from data. Apart from producing, disseminating and using data themselves, the public sector must, through strong digital leadership, promote and foster data-driven innovation and growth throughout (Andrade et al., 2014).

The research model is presented in Figure 10.

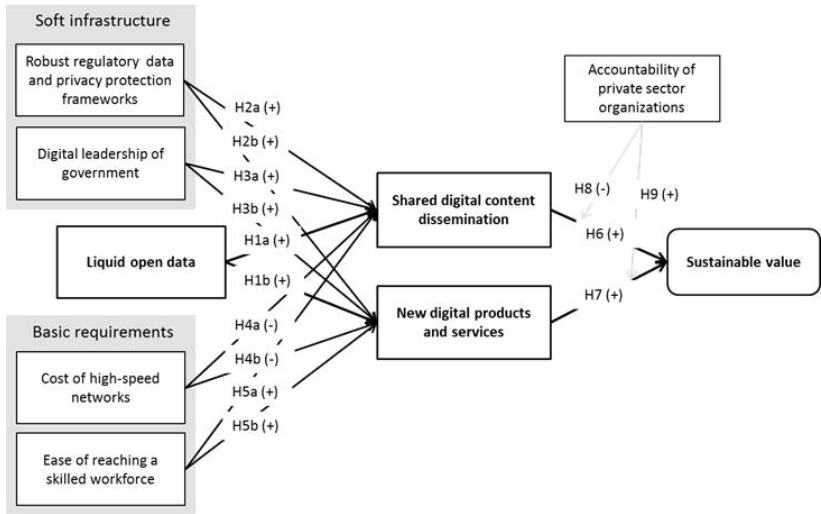


Figure 10: Research model

The model portrays a nomological network centered on what can be perceived as manifestations of the interaction between open data engagement and the two archetypes of value generating mechanisms discussed in section 5.2. The two constructs are modeled as mediating variables, mediating the impact of liquid open data on sustainable value. I propose that when multiple stakeholders start engaging with liquid open data, and using them to generate a variety of information, we can find empirical traces of such actions by looking for examples of dissemination of shared

digital content. I also propose that when users engage with liquid open data they will, to some extent, use these data for creating new and innovative digital products and services. We can find empirical traces or manifestations of such events in available secondary data.

By opening government data, public authorities can stimulate value generation, which happens through the creation of shared digital content and commercial digital products and services. These events reflect the underlying but unobservable information sharing and market mechanisms. The extent of new content, products and services is enabled or prohibited by the context in which value generation happens, which impacts not only the opportunity but also the motivation and ability to generate value from data.

Individual relationships and hypotheses are discussed in section 5.5.

## **5.5 Hypotheses**

### **Liquid open data**

I have previously discussed how publishing open data creates an opportunity for value generation through new use of data, and defined the construct liquid open data. The real estate information platform company Zillow can now be used as a real-life example to explain how this happens. Zillow is a notable example of technology startup flourishing on the availability of open data (Capgemini, 2013).

Zillow firmly believes that free and open access to data of all kinds – from real estate information to demographic, education and crime statistics – can only help empower the people that use it. That’s why we couldn’t be happier that the nation founded on the principle of “We, the people” has officially adopted this stance, too.

Zillow’s mission is to empower consumers with information and tools to make smart decisions about homes, real estate and mortgages. The company has established a successful business by creating a living database of homes across the United States. The database is built from a range of linked sources such as tax data; listing of available homes for sale or rent; mortgage information; geographic data and data on local land value and housing prices. Accordingly, open government data are one of Zillow’s most valuable resources. The company offers homebuyers free information (shared content) on available properties through their website. Zillow has also built a variety of digital products and services on top of this database. One of Zillow’s main offerings is their so-called “Zestimates”, an advanced statistical predictive tool providing up-to-date information on home values and rental prices (Capgemini, 2013).

As Zillow uses open data to create both shared content and commercial products and services, both information sharing and market mechanisms are activated in order to generate sustainable value from liquid open data.

*H1a: Liquid open data positively influences stakeholder's opportunity to disseminate shared digital content*

*H1b: Liquid open data positively influences stakeholder's opportunity to develop and market new digital products and services*

## Robust regulatory data and privacy protection frameworks

Data is the new currency of the digital economy, where instead of market transactions in which partners exchange currency, people get access to free information services, often in exchange for their own data (Bean & Koeppel, 2013). We can define personal data as any data that can be attributed to an identifiable person either directly or indirectly (Beardsley et al., 2014). By definition, open government data do not include personal data. Anonymization methods are in many cases used to remove any information that could be traced back to individuals and put them at risk. Even so, many worry that open data could be traced to individuals, for instance through a combination of datasets. The protection of personal data has long been viewed as a fundamental personal right. It is important that citizens feel safe about the protection of their personal data and privacy. Governments and regulators will thus need to frame data protection policies that safeguard the privacy of citizens (Beardsley et al., 2014). However, policymakers must also identify the appropriate balance between protecting the privacy of individuals' data and allowing for innovation in service delivery and product development (Pepper and Garrity, 2014).

The free flow of any fiat currency depends on trust. It is necessary to ensure that the same norms, principles and values that underpin physical markets are employed online, instilling confidence that publication and use of liquid open data will not result in personal data being unethically or unlawfully used. A 2011 survey by the European Union revealed that 92% of Europeans were concerned about mobile apps collecting their data without consent. 89% of those surveyed said they wanted to know when the data on their smartphone was being shared with a third party, asking for the option to give or refuse permission (European Commission, 2011a). EU citizens were also concerned about the risks linked to cyber-security. 74% of respondents agreed that the risk of becoming a victim of cybercrime had gone up (ibid). The key security issues for use of data include the reliable prevention of hacking and access by unauthorized and unwanted users to large databases and data flows. In order to ensure a healthy

ecosystem where users, consumers and businesses feel safe in engaging in data-driven activities, protection against cybercrime becomes essential. The trust in legislative ability to protect people's data is likely to affect positive sentiment towards the dissemination and use of liquid open data and thus influence the motivation for engaging with open data for value generation.

I propose that a robust regulatory data protection framework will positively influence the continuing trend towards creation and dissemination of shared digital content and the creation of digital products and services, both of which rely to an increasing degree on use of personal data.

*H2a: Robust of regulatory data and privacy protection frameworks positively influence stakeholder's motivation to disseminate shared digital content*

*H2b: Robust regulatory data and privacy protection frameworks positively influence stakeholder's motivation to develop and market new digital products and services*

## Digital leadership of government

The relationships between where value is added and where value is captured, and by whom, are becoming increasingly complex in the context of shared resources used by a network of stakeholders (Bowman & Ambrosini, 2000; Bowman, 2014). Networks have been shown to support value co-creation due their flexibility and adaptability (Morgan et al., 2013). However, networks present major challenges in terms of governance (Benington et al. 2011; Morgan et al., 2013). For example, it can be challenging to sustain clear strategic direction and appease multiple stakeholders, with different mandates (see discussion in Paper VI). In the context of open data, current research is revealing that ODIs, where public bodies have strong ties to the eventual data-users, have been most successful (van Veenstra and van den Broek, 2014a), calling for a more structured network governance (Eckartz et al., 2014). Moreover, innovative activities intended to generate sustainable value should be supported by an innovation ecosystem, which the public sector should establish in order to increase use of their data (Bason, 2010). In general, governments should foster an environment supportive of innovation and an ICT-friendly government policy (Bilbao-Osorio et al., 2014). Moreover, governments should have a clear vision for the internal use of data and support cross-government initiatives (Ubaldi, 2013). Strong government leadership is thus proposed as an important factor for motivating users to engage with open data.

*H3a: Digital leadership of governments positively influences stakeholder's motivation to disseminate shared digital content*

*H3b: Digital leadership of government positively influences stakeholder's motivation to develop and market new digital products and services*

### Cost of high-speed networks

The generative ability of the mechanisms that facilitate value generation through use of data depends on certain capabilities. A capability can be defined as a measure of the ability of an entity to achieve an objective. Here, these capabilities reflect the ability of individuals and organizations to create and disseminate shared digital content or develop new commercial digital products and services.

An important foundation for the ability to create and disseminate digital content, products and services is having access to high-speed network infrastructure. Improved information infrastructure has vastly increased our ability to store, retrieve, sort, filter and distribute information, thereby greatly enhancing the value of the underlying information itself (Shapiro & Varian, 1999). Due to increasing digitization and exponential growth of data, the need to provide high-speed broadband to all segments of the population has gained importance in recent years (Bilbao-Osorio et al., 2014). Unfortunately, universal connectivity and global access still remain out of reach for many citizens of the world (Avital et al., 2007). Moreover, bandwidth constraints will increasingly become an obstacle in transmission over existing networks, due to the growing amount of supply and demand for data (Pepper & Garrity, 2014). I use a measure of the relative cost of internet bandwidth as a proxy for the general affordability and availability of network infrastructure (see table 7, appendix B for detailed measures). The higher the relative cost of using high-speed internet, the less general affordability. I propose high costs will negatively influence the ability of individuals and organizations to create and disseminate shared digital content or develop new digital products and services.

*H4a: Cost of high-speed networks negatively influences stakeholder's ability to share digital content*

*H4b: Cost of high-speed networks negatively influences stakeholder's ability to develop and market new digital products and services.*

## Ease of reaching a skilled workforce

Another important construct that affects the ability of individuals and organizations to create and share digital content or develop new commercial digital products and services is the availability of skilled workers. The constant upgrading of workers' skills has never been more important for continuing competitiveness of our society (Bilbao-Osorio et al., 2014). If the education system does not equip students with the necessary expertise to stimulate innovation, both in and outside of organizational contexts, the capacity for innovation will suffer, both in regards to general civic engagement and product and service development. In fact, the demand for people with the deep analytical skills relevant for big data analytics might outstrip current projections of supply by 50 to 60 percent (Brown et al., 2011).

I propose that societies which can easily access a workforce offering relevant technical skills, also will present more ability to generate value through creation and sharing of digital content or the development of new digital products and services.

*H5a: Ease of reaching a skilled workforce positively influences stakeholder's ability to create and disseminate shared digital content*

*H5b: Ease of reaching a skilled workforce positively influences stakeholder's ability to develop and market new digital products and services*

## Shared digital content dissemination

Timely access and availability of previously disparate data sources can positively contribute to new knowledge, insight and awareness, ultimately creating a society equipped to make information-driven decisions. Studies show that companies who adopt data-driven decision making are on average 5-6% more productive than their intuition-driven counterparts (Brynjolfsson et al., 2011). The energy-tech company Opower is an example of the generative power of information dissemination. Opower is a double-bottom-line company, motivated by a corporate mission to drive global energy savings. Opower merges and analyzes various types of data, including open government data, to generate reports for individual energy users outlining personalized insights on their energy use. The reports are freely available to consumers through multiple channels. The energy reports include a comparison of the individual's energy consumption with that of their "neighbors" – that is a calculated average of users with a similar profile, based on for example demographic information, house types and location based weather (see discussion in Paper IV). Moreover, to motivate low-energy users to maintain their limited consumption, they are rewarded for their good behavior



with a symbolic “smiley token” (Allcott, 2011). In May 2015, the Opower home energy reports had collectively enabled a global savings of over 7.7 terawatt hours of electricity, 12 billion pounds of CO<sub>2</sub> emissions and a savings of approximately \$1 billion on individual energy bills. This example of people’s engagement with contextualized information demonstrates how such engagement can lead to a collective generation of sustainable value.

In the not so distant past, dissemination of information happened mostly on paper and made use of the market mechanisms via channels like subscription newspapers and bookstores. At present, however, this dissemination occurs primarily through the creation of freely shared digital content and is driven by information sharing mechanisms that connect immensely diverse people on a global scale. The information sharing mechanisms are enabled by the interactive, collective intelligence of today’s mobile and web-based information technologies, such as mobile apps and social media. The key components of these mechanisms are not based on financial remuneration, but on user involvement, accreditation, and tools that promote collaboration between individuals (Benkler, 2006). Value generation and appropriation through information sharing mechanisms in the context of open data emerge as a constant interaction between governments, citizens and companies, where continuous exchange of data and information is the key facilitator. For the purpose of this dissertation, it is assumed that people in general wish to make choices that benefit society, whilst safeguarding their own wellbeing. Furthermore, it is assumed that providing people with more information gives them the means to make choices that are more reflective of said wishes. Therefore, I propose that the creation and dissemination of shared digital content containing valuable information will positively influence the generation of sustainable value.

*H6: Shared digital content positively influences the generation of sustainable value*

## New digital product and services

When previously unavailable data-sources are released to the public, new stakeholders are able to employ them, possibly generating innovation and an opportunity for new digital product and service creation. For example, the availability of open GPS data in the US in the 1980s stimulated the creation of many new businesses specializing in GPS and mapping services. Since then, so-called geo-services have become increasingly widespread, and are estimated to contribute annually between \$150 and \$270 billion in monetary value to global markets (Oxera, 2013). Additionally, open

data from the US National Weather Service supports a private weather industry worth over \$1.5 billion per year (Kundra, 2012).

As an example of how product and service innovation through open data contributes to sustainable value, we can refer to INRIX, which is a leading provider of traffic services. INRIX's mission is to globally reduce congestion caused by traffic and in order to do so they have utilized crowdsourcing as a data collection method (see Papers V and VIII). Their traffic intelligence platform combines and analyzes data from public and private sources, including government road sensors, official accident and incident reports and data on real-time traffic speed, sourced from a large community of local drivers that have either downloaded an application on their smartphones, or provided information through their in-car GPS device. INRIX has developed a variety of products and services based on these data. They sell both enriched data and data-driven products and services to stakeholders such as automobile manufacturers and public traffic engineers.

While business models of many companies are transforming as they increasingly make use of information sharing mechanisms, most innovative companies still depend on the market mechanism to create revenue. These market mechanisms are driven by the forces of demand and supply, and can be used to identify business opportunities and promote new digital products, services or processes that offer the generation of sustainable value. In the specific case of data-driven innovation products that route traffic and reduce congestion, direct and indirect impacts on sustainable value can happen through: a) a reduction of unproductive time spent by commuters; b) a greener and more sustainable environment; and c) reduction in stress and pollution related health-issues. Naturally, other innovations offer different contributions to sustainable value. In general, I propose that the creation of new digital products and services will influence the generation of sustainable value in a positive way.

*H7: New digital products and services positively influence sustainable value.*

## Accountability of private sector organizations

We are currently facing numerous and urgent economic, social, environmental societal challenges. Given the complexity and cross-boundary nature of these challenges, a new approach is necessary (Jetzek et al., 2014b; OECD, 2011). The silos of government departments are in many cases poorly equipped to tackle challenges that span sectors and nation states. Civil society often lacks the capital, skills and resources to take promising ideas to scale. In section 1.4, I discussed various concepts of value and how the dichotomy between the state and the markets was being increasingly contested

(Porter and Kramer, 2011). Following Porter and Kramer (2011), I suggest that there is a role for private companies in solving these challenges and producing sustainable value.

The concepts of market failures and negative externalities have for a long while shaped company strategies that have had a tendency to exclude social and environmental considerations from their economic thinking. Thus, firms have traditionally resisted regulatory standards as being contrary to their interests (Porter and Kramer, 2011). However, continued advances in digital technologies have disrupted the status quo and altered the way many people think, live and work, rearranging value pools (McKinsey, 2013b). Governments seeking to provide social protection to their citizens, while strengthening national economic competitiveness are becoming more interested in cross-sector co-creation of value. The evolving sense of corporate social responsibility is portrayed in the number of companies that participate in collective efforts to develop international standards that go well beyond what is required by international law (Haufler, 2013). Increasingly, companies are adopting policies that address urgent social issues (ibid). However, if companies are to play a bigger role in terms of social responsibility, they must consistently perform in the interests of shareholders and wider society. To do this companies will need to make visible the precise ways in which they are contributing to sustainable value generation, allowing shareholders, government officials and the greater population to hold them accountable.

In previous sections, I have outlined a number of use cases where private companies have utilized open data to create both shared content and commercial products and services that contribute to sustainable value for society. The motives of these companies are altruistic, while simultaneously remaining economically grounded. Thus, these businesses have reconnected company success with social progress. Accordingly, I propose a moderation perspective, implying that the relationship between content dissemination and sustainable value on the one hand, and new digital products and services and sustainable value on the other hand depends on the level of private sector accountability. However, the moderating effect of accountability of private sector organizations differs between the two proposed mechanisms.

Firstly, I propose that while dissemination of shared content is positively related to sustainable value, this effect will be more prominent in countries exhibiting less accountability of private institutions. When citizens in these countries uncover various sources of information, for instance regarding worker's rights or environmental protection, they are able to put social pressure on the corrupt or unaccountable companies. This pressure, I argue, will have an impact on the behavior of companies

and positively influence the generation of sustainable value. The impact of shared content dissemination on sustainable value is, using a similar logic, suggested to be less pronounced in countries where companies already adhere to various standards and regulations, as those companies are already held accountable by shareholders and other stakeholders. To rephrase more clearly: the positive impact of increased content sharing (via transparency and civic engagement) on sustainable value is stronger in countries with less accountable private institutions.

*H8: The positive effect of shared digital content on sustainable value decreases with increased accountability of private sector institutions*

Secondly, I argue that while the generation of commercial digital products and services is positively related to sustainable value, this effect will be more prominent in countries with more accountable private sectors. When companies in general adhere to standards and regulations that hold them accountable for their actions, they are more likely to conform to social responsibility. Thus, the positive impact of new digital products and services (via innovation and efficiency) on sustainable value is stronger in countries with more accountable private institutions.

*H9: The positive effect of commercial digital products and services on sustainable value increases with increased accountability of private sector institutions*

## **5.6 Research Model Estimation**

I used Partial-Least-Squares (PLS) method for assessing both the measurement model and structural model. For the measurement model, I used secondary data for seventy-six countries in 2013. Construct definitions and item measures are illustrated in table 6, Appendix B. Paper VII offers a more thorough discussion, including measurement model validity indicators. Figure 11 illustrates the results of the PLS estimation.

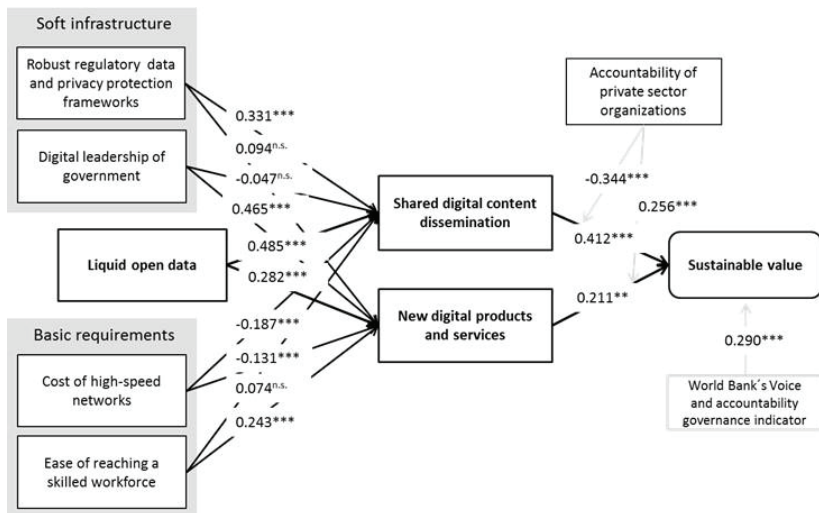


Figure 11: Results from PLS Estimation

The overarching conclusion from the estimated model is that both digital content sharing and new digital products and services contribute significantly to the generation of sustainable value. Surprisingly however, the path through digital content sharing is considerably more relevant. The case of climate data is a demonstrative example of how such value generation might happen. Government-released open climate data is at present pioneering countless discoveries within the environmental sustainability field. Today's climate data partnerships are creating unique ventures that cross boundaries between business, government and academia. For instance, IBM offers free super-computing hours on its World Community Grid for researchers who are studying climate change, utilizing and generating mass quantities of open data. One of the projects making use of the grid, Harvard University's Clean Energy Project, has screened and publicly catalogued 2.3m compounds in its search for new materials that could potentially double the efficiency of current carbon-based solar cells. The findings of the Clean Energy Project's Database are open to the public to help further the discovery of new materials. While the value of such initiatives is not easily measured using traditional methods, it seems plausible that these synergistic network-relationships have contributed to sustainable value.

Adopting the configurational perspective (Henfridsson and Bygstad, 2013), I conclude the following: shared digital content dissemination is enabled by liquid open data, a robust regulatory data protection framework and cost of high-speed networks, as those

are the only significant relationships. Similarly, new digital products and services are enabled by liquid open data, digital leadership of government, cost of high-speed networks and ease of reaching a skilled workforce. A more complete discussion on the individual relationships is presented in Paper VII. There it is proposed that the non-significance of the relationship between a robust regulatory data protection framework and new digital products and services could indicate a two-sided relationship, through increased trust on the one hand (positive) and perceived burden of government regulation on the other hand (negative).

Liquid open data contributes to both shared digital content and new digital products and services. The parameter that indicates the indirect effect of liquid open data on sustainable value is 0.312 and highly significant, as well as fully mediated via the shared digital content and new digital product and services variables.

The hypothesis that private sector accountability influences the relationship between the mediating variables and sustainable value is supported. The relationship between shared digital content and sustainable value is stronger in countries with less accountable private organizations. The relationship between new digital products and services and sustainable value is stronger in countries with more accountable private organizations, see figure 12.

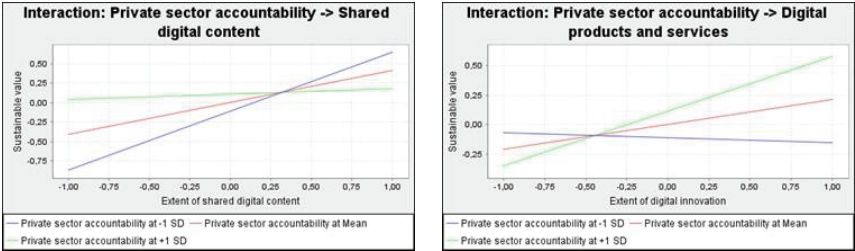


Figure 12: Moderating Relationships

I checked the effect of each of the mediating and moderating variables on sustainable value by comparing the R2 of the sustainable value construct with, and without, the variable in question, using Cohen’s f2 measure, as is shown in table 3.

**Table 3: Effect Sizes Measured with Cohen's  $f^2$**

<b>Impact on Impact from</b>	<b>Sustainable value</b>	<b>Shared digital content</b>	<b>Digital products and services</b>
<b>Shared digital content</b>	0.254 (medium/high)		
<b>New digital products and services</b>	0.075 (weak/medium)		
<b>Interaction: Accountab. / Content</b>	0.264 (medium/high)		
<b>Interaction: Accountab. / Products</b>	0.174 (medium)		
<b>Liquid open data</b>		0.434 (high)	0.216 (medium)

I can conclude that both mediating variables are important for explaining the variance of the sustainable value construct, but the effect of shared content is much stronger than the effect of digital products and services. Moreover, both interaction effects are important, but the effect of interaction between private sector accountability and shared digital content dissemination is stronger than the effect of interaction between private sector accountability and digital product and service innovation. Additionally, I checked the effect of the liquid open data construct on the mediating variables. Liquid open data is highly important to the configuration of variables that explain shared digital content dissemination and moderately important for the configuration of variables that explain new digital product and services.

Figure 13 shows the relationships between standardized results of the latent variables from the estimated model.

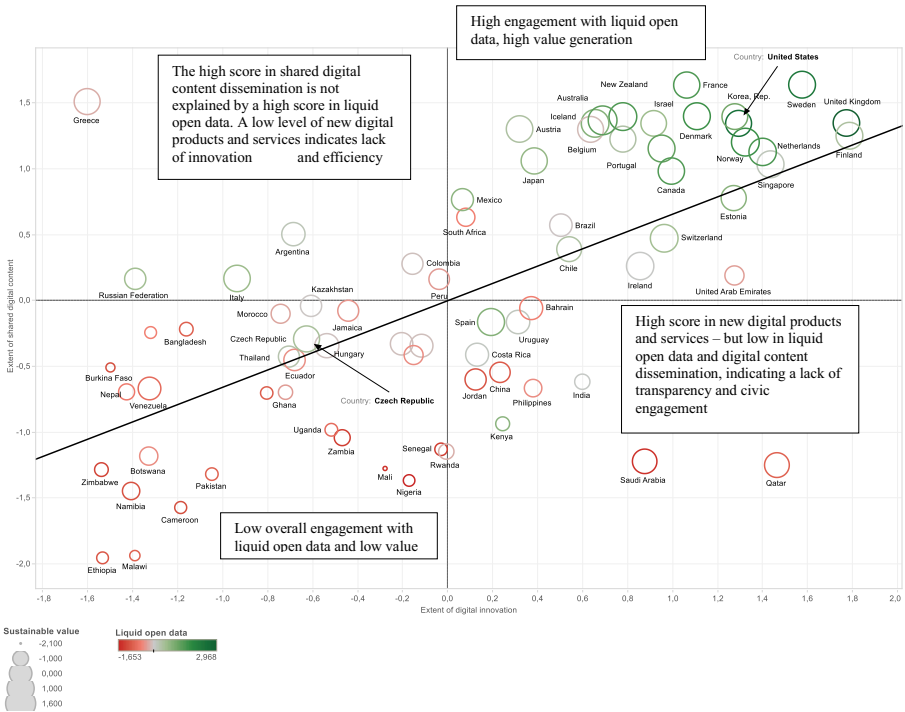


Figure 13: Liquid Open Data, Shared Digital Content, New Digital Products and Services and Sustainable Value

The x-axis in figure 13 shows the measured extent of new digital products and services. The y-axis shows the measured extent of shared digital content across the seventy-six countries included in the study. The color of each circle shows the extent of liquid open data in the country (red is low, green is high) and the size of each circle reflects the degree of sustainable value (bigger circle indicates higher levels of sustainable value). Figure 13 indicates that countries that score disproportionately high in content sharing also tend to generate more sustainable value. The outliers, which score very low content sharing mechanisms but high in new products and services, do worse in terms of sustainable value than other countries who have similar scores in the latter. The same goes for the outliers who score very low in terms of new products and services. Figure 13 indicates that there could be synergies between the two underlying mechanisms.



Figures 14 and 15 show the relationship between the liquid open data latent variable on the one x-axis and the latent variables that measure shared digital content dissemination and digital product and service innovation on the y-axis, respectively.

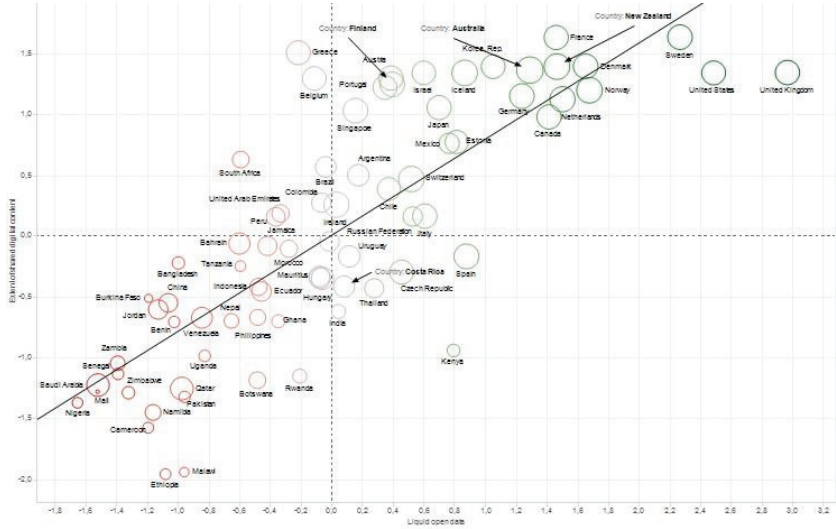


Figure 14: Extent of Shared Digital Content as a Function of Liquid Open Data (Color indicates Sustainable Value)

In both cases the figure illustrates a strong correlation between the two latent variables, but again, there are outliers that indicate that there could indeed be a somewhat synergetic relationship between the two underlying value generating mechanisms. For instance, United Kingdom comes across as an outlier if we only look at the relationship between liquid open data and shared digital content. However, the UK is directly on the trend-line in the case of new digital products and services where they are also the clear leader. Other countries like Qatar and Saudi-Arabia are scoring disproportionately high in digital innovation but are laggards when it comes to liquid open data and content sharing. Overall, they rank number 29 and 30 out of 76 countries in sustainable value as measured by this model. Hence, it is tempting to conclude that they could see positive results if they could activate both mechanisms through a more open approach to data and information sharing.



Figure 15: Extent of New Digital Products and Services as a Function of Liquid Open Data

## 5.7 The Role of Open Data Intermediaries

The emerging relationships I have discussed at length in previous sections are the topic of Paper VIII, where the role of intermediaries in the open data ecosystem was explored. The aim with this paper was to increase understanding of how users engage with liquid open data, and to explore the previously mentioned potential synergies between the market and information sharing mechanisms. I reviewed a body of literature that focuses on the theory of two sided markets, in order to improve my understanding of the role of intermediaries in reconciling the quest for market-driven economic profits, and the valuable network externalities that arise from information sharing. The business models that have been developed in these types of markets, which are commonly referred to as Multi Sided Platforms or MSPs in IS literature, are generally based on two or more sides of affiliated customers that interact via digital platforms.

The theory of two-sided markets has emerged over the past decade as one of the most active areas of research in information systems, economics, and strategy. It has also drawn considerable interest from practitioners. Two-sided markets are defined as “markets in which one or several platforms enable interactions between end-users, and

try to get the two (or multiple) sides “on board” by appropriately charging each side” (Rochet and Tirole 2006, 2). The importance of this approach for the analysis of how open data are used to generate sustainable value is that, unlike classical economic theory, the two sided market approach explicitly recognizes the value of network externalities (Katz & Shapiro 1985, 1986). The theory of two-sided markets builds on the notion that there are non-internalized externalities among end-users: “The starting point for the theory of two-sided markets by contrast is that an end-user does not internalize the welfare impact of his use of the platform on other end-users” (Rochet & Tirole 2006, 3). Two-sided markets, by playing an intermediary role, will facilitate an interaction that would not occur without them and therefore create value for both sides through direct and indirect interactions and network effects.

Intermediaries help overcome five limitations of direct transactions: search costs, lack of privacy, incomplete information, contracting risk and pricing (Resnick et al. 1995; Janssen & Zuiderwijk 2014). Accordingly, intermediaries essentially have four roles: a) information aggregation, 2) providing trust, 3) facilitating, and 4) matching (Bailey & Bakos, 1997). The intermediary can be an agent of any kind, a government organization, an individual, or a private company. The recent democratization of content, as well as the subsequent sharing, remixing, redistribution, and re-syndication of content in newer and more useful forms, has caused dramatic power shifts in the intermediary market (Bharadwaj et al., 2013). These trends have challenged traditional value chain thinking, while exploring innovative ways to generate value through information sharing. For instance, internet based peer-to-peer business models, often aggregated under the term Sharing Economy, have greatly disrupted the traditional intermediary model. Taxi companies, for example, are now competing against new peer-to-peer ridesharing services, which connect drivers and their personal vehicles with riders seeking transit through intermediary platforms (Cannon & Summers, 2014). These new types of intermediaries are creating an important layer that matches demand and supply for services, utilizing economies of scale and digital technologies as well as the business models of two-sided markets.

What then, is the role of intermediaries in making a digital infrastructural resource such as open data, easily reusable and reproducible? Research shows that in its current state, open data may not be as reusable as previously assumed. A recent review of the open datasets provided in Berlin, Germany showed that approximately 90% of the data provided were published in a non-machine readable PDF format (Martin et al., 2013). The Open Data Barometer reveals that of the 821 datasets surveyed in 2013, less than 7% of the datasets were published both in bulk machine-readable forms, and under

open licenses. Only 1.2% of open data were published as linked data (Davies 2013). Moreover, there are multiple ODIs in most countries, where different local governments, and state government each have their own policies and standards for open data. During these early days of open data, ODIs are still heterogeneous in nature, licenses differ between initiatives and open data standards are still underdeveloped and underused. Reuse of these data suffers as a result of these issues, as well as from limited network activity in the open data ecosystem (Mayer-Schönberger & Zappia, 2011, Martin et al., 2013). For most individuals and smaller developers, these issues have actually created a substantial barrier to entry, as the efforts involved in acquiring, manipulating, and analyzing these disparate data are simply too extensive, in comparison to an uncertain and potentially non-economic payoff.

As stated in section 5.1, liquid open data is not a binary construct, but rather a complex constellation of many dimensions, ranging from licenses, to prices, to usability and technical accessibility. Making data available in their current form is, by far, not the only barrier to overcome when it comes to enabling value generation through use of open data (Conradie & Choenni 2014; Janssen et al., 2012; Martin et al., 2014; van Veenstra and van den Broek 2013; Zuiderwijk & Janssen 2014a; 2014b; Zuiderwijk et al., 2012; 2014). Another important factor is that datasets are becoming increasingly larger, introducing a new barrier for those that cannot easily download or move the data. Moreover, raw data themselves are often of little or no use to end-users who have neither the capabilities nor time to manipulate and process them (Janssen & Zuiderwijk, 2014). Accordingly, publishing open data that are ready for use is a costly and time-demanding task. As public sectors are frequently constrained by budgetary demands, these costs constrain government's aspirations for open data, despite the potential for value generation. However, these same issues can also introduce an opportunity for private parties that can create a business from removing these barriers to use of open data.

I propose that open data intermediaries will likely play a crucial role in the emerging open data ecosystem by facilitating data and information access for smaller organizations that may not have the capacity and capabilities to store, integrate, and analyze large and heterogeneous datasets. Intermediaries might also contribute directly to value generation by augmenting and amplifying the circulation of open data and by sanitizing and curating data coming from both public and private sources. By making data easier to access, manipulate, and use, intermediaries could drive information creation and product, service, or process innovation based on these data. The model

shown in figure 16 is an extension of the research model in figure 10, including MSPs acting as open data intermediaries.

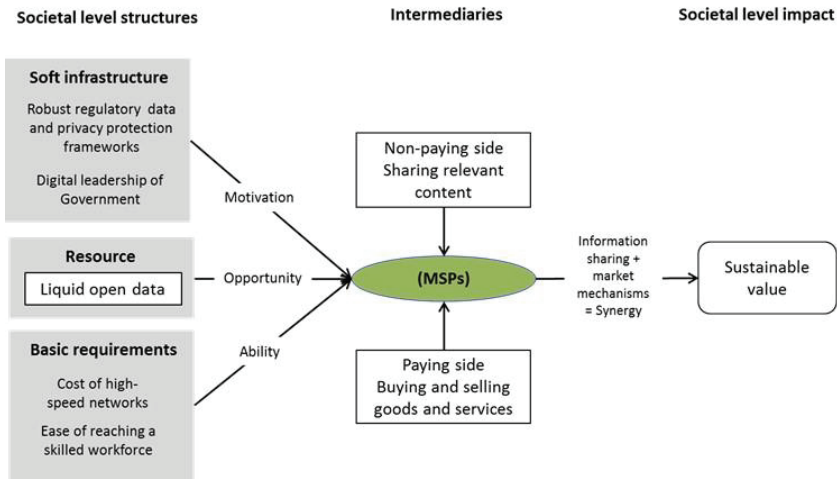


Figure 16: MSP type Intermediaries create Synergies between Content Sharing and Commercial Products and Services

Having easy, one-stop access to data services offers a value proposition for companies striving to create a competitive advantage in an increasingly data driven world (Lindman et al., 2014). However, a large share of data-driven services is provided for “free”, oftentimes in exchange for access to personal data (OECD, 2014). Data intermediaries need to adapt to market conditions where users are accustomed to having free access to data, information, and information services. To enable the ongoing generation of valuable but free information, the data intermediaries could implement two-sided market business models. This would allow them to generate economic profit by capitalizing on the positive network externalities that arise from the interaction of multiple stakeholders that use the platform to gain access to the services provided by these intermediaries and their affiliates. Three examples of the use of these business models in platforms that make use of open data are discussed in Paper VIII.

I have previously discussed the mechanisms that explain how sustainable value can be created through use of open data. I now add a further proposition, positing that MSPs are facilitating such value generation through the creation of synergies between information sharing and market mechanisms. MSPs enable companies to move beyond simple rent-seeking and integrate the quest for economic profits with the generation of

society wide sustainable value. These platforms act as a medium between stakeholders that are collectively addressing complicated societal challenges, previously the responsibility of governments alone. Governments have already started to realize the power of these models, which thrive on sharing and interaction, and are even implementing their own MSPs where public sector, businesses and citizens can meet and interact to create sustainable value (Janssen & Estevez, 2013).

## **5.8 Explaining the Open Data Value Paradox**

In spite of the promising role of MSPs as open data intermediaries, there are still various challenges present. One barrier that has been identified in prior work on MSPs is the chicken-and-egg problem (Caillaud & Jullien, 2003; Hagiu, 2015). MSPs need to attract a sufficient number of participants on one side of the platform in order to attract the other side, which in the case of open data is the side that is paying for products and services. In the case of government provision of open data, this translates to governments attracting enough users to justify the investments required for making data open, and finding a way to make the resulting value more explicit to decision makers. In this scenario, governments do not get any direct payments but will benefit indirectly from sustainable value generation. When the users start engaging with the data, value will be generated, although what kind of value, and how much is uncertain. However, the users will generally not engage unless they have a relatively clear perception of their own future value proposition. This is what I term the open data value paradox. The open data value paradox describes a situation where entrepreneurs do not use the data because the data are not liquid and open enough, and there is too much uncertainty over factors like the sustainability of ODIs. Paradoxically, the data custodians are not willing to invest in the people and technology necessary to make the data more usable and sustainable until they observe concrete evidence of value generated from their use.

I have proposed that the action of making data liquid and open is equivalent to creating an opportunity for all stakeholders that care to use the data. To rephrase, governments are presenting their citizens with an option to utilize these data. The economics of real options can help us conceive the perceived future value of such an option by building on the same ideas that underpin the financial options markets. Having an option provides the option owner with valuable flexibility. If we conceptualize having access to sustainable liquid open data as an option, companies know that they can use these data for creating new information or data-driven products and services at any suitable point in time. Conceptualizing open data as an option might help resolve the open data

value paradox. By using option valuation methods we gain a tool that enables us to systematically evaluate the uncertain payoff derived from making data liquid and open, under different scenarios. Moreover, unlike most other valuation methods, option valuation methods promote unpredictability and variability as a positive factor rather than a negative one. Entrepreneurs gain valuable flexibility from not having to incur large investments in data. Rather, they can experiment with the data, ultimately choosing those datasets that turn out to be most suitable. Moreover, if governments can evaluate these currently unknown future gains, they will be able to justify their investments in liquid open data.

## 6. Conclusion

In this final chapter I will revisit the research questions, outlining the answers I propose as a result of this study. I will then discuss the theoretical and practical contributions and introduce an agenda for future research. Additionally I will review the limitations, and validity associated with my study. Lastly, I discuss my personal reflection, and give closing remarks of my journey through this dissertation.

### 6.1 Revisiting the Research Questions

In his recent TED talk, Kenneth Cukier provides a succinct, yet accurate answer to the overarching research question of this study, How is value generated from open data?: “You have more information. You can do things you couldn’t do before” In an attempt to provide more theoretical depth to this answer, I have suggested that two overarching value generating mechanisms can explain how value is generated from open data, the information sharing mechanism and the market mechanism. New value is generated through new information that contributes to increased transparency, and reduces the adverse effects of information asymmetry. Open data can also drive civic engagement through members of society engaging with and enriching available information, using it to positively impact their decision making capabilities, and to align their behavior to practices that contribute to sustainable value. Alternatively, value is generated through the markets when data are used to support process improvements, resulting in increased efficiency and cost reduction, or as a resource in new products or services, resulting in innovations that can transform markets, creating new companies and job opportunities.

I will now examine the five sub-questions in greater detail. The first sub-question posed was: *What are the main enabling factors for value generation through open*

*data?* After synthesizing between literature and qualitative and quantitative data, I have identified four measurable, contextual factors I propose are an important part of a societal level infrastructure that is conducive to value generation through liquid open data. These factors are: 1) robust regulatory data and privacy protection frameworks; 2) digital leadership of government; 3) cost of high-speed networks; and 4) ease of reaching a skilled workforce. Moving down one level of abstraction, other factors will be important for individual organizations. The qualitative data collected for Paper IV revealed one particularly important enabling factor, namely absorptive capacity. Absorptive capacity is a function of organizational and technical capabilities, which allow organizations to reach outside of their boundary for data, information, or knowledge and to productively utilize these external resources for value generation. It remains a research opportunity to extend the conceptual model presented in this dissertation to include organizational level variables and consequently test the boundary-crossing relationships, using observations at both the societal and organizational level.

The second sub-question asked was: *What are the unique features of open data?* The answer is in fact two-fold. The first concerns the theoretical economic features, which include open data being (in theory) non-rivalrous and non-excludable. These features have a direct impact on how value is generated from open data, as they influence how market participants view open data as a resource. While the openness factor offers a clear value proposition at the societal level, the value proposition is more complicated for individual companies, which generally desire unique resources (Barney, 1991; Wade and Hulland, 2004). Lindman and Kuk (2015) suggest that open data should be viewed as a common pool resource, subject to subtractability, as the feature of subtractability acts to encourage commercial use and investment. OECD (2014) utilizes the terminology infrastructural resource to highlight that digital data can be easily reproduced and reused by many stakeholders for many different purposes. Unfortunately, we still do not have a widely accepted economic concept that can fully describe the unique features of open data.

The second answer lies in the more detailed conceptualization of liquid open data, where I have synthesized different definitions of open data. Diverging from those definitions, I have increased the emphasis on the features, or attributes, that reflect the technical and governance related complexity of making data liquid and open. I would also like to draw attention to the fact that open data can mean many things to many people. Each of the seven dimensions of liquid open data has different relevance for different use cases. I propose, however, that governments should strive to make data as



liquid and open over all of the seven dimensions, as is economically feasible. Doing so will create a bigger opportunity or, alternatively, a more valuable option, potentially resulting in synergistic value generation.

The third sub-question posed was: *What are the value generating mechanisms of open data?* To which I identified four distinctive value generating mechanisms. As these mechanisms form a hierarchy, I have further categorized two as market mechanisms: efficiency and innovation mechanisms. These two are already widely recognized as value generating mechanisms. However, as a contribution of this study, I propose two additional mechanisms that do not utilize the markets to generate value. As the overarching mechanism guiding these was not previously recognized, I chose to conceptualize it as an archetypical value generating mechanism and named it the information sharing mechanism. The information sharing mechanism explains how value is generated through positive network externalities that are created when multiple stakeholders share valuable information, effectively reducing information asymmetry and encouraging civic engagement.

The fourth sub-question raised was: *How can we identify, conceptualize and measure the value that is generated from open data?* This question raises a number of complexities and the answer is manifold. Primarily I intended to highlight the immense value that is both generated and captured from open data, yet never enters formal accounting ledgers. For example, access to superb education, clean air and an efficient and supportive healthcare system are amongst the things valued highly by many, yet this value is not easily quantifiable. Moreover, when I started to work with the secondary data, I saw that there was a high correlation between wealth and the other chosen value indicators such as health, education and sustainable environment. It is, however, not a perfect correlation, and the resulting sustainable value construct rates countries rather differently than the GDP measure. For instance, Qatar ranks number 3 out of 76 countries in GDP dollars per capita but number 29 in sustainable value. Norway, however, ranks number 1 in GDP dollars per capita and number 2 in sustainable value, due to their more balanced approach to information sharing and market-driven innovation. United States rank number 8 in GDP dollars per capita but 23 in sustainable value. Thus, the model does confirm that the countries that value openness and information sharing tend to rank highly in sustainable value as well.

Regarding measurement, it is extremely difficult to trace how much value is generated from use of open data or how much these data are worth. There are many different methods that can be, and are currently used, each of them with their own strengths and limitations. Firstly, we can use previously created nationwide indicators to measure

sustainable value. There are a number of recently developed indexes that offer a nuanced and multi-dimensional view of societal progress to draw on. For instance OECD Better Life index, United Nations Human Development index and Social Progressive Imperative's Social Progressive index. The difficult part, however, is to adequately link the act of providing society with open data to the generation of sustainable value. One possible approach is to use correlational, cross sectional comparisons as was done in this dissertation. This method certainly offers a valuable indication, but remains silent in terms of direct causality. Therefore, I have opted to use the CR-based method of retroduction to hypothesize about the possible underlying causal mechanisms that could explain the relationships illustrated in the model.

One method that has been used in previous studies is the cost benefit analysis. This type of analysis evaluates the costs and benefits of opening a single dataset, or type of data (Houghton, 2011). Use of the cost-benefit analysis method has revealed that the socio-economic benefit of making commercially attractive data, such as geographic data, available is typically higher than the costs of making said data available. Cost benefit analysis is however a difficult exercise in the case of open data as the data are used by multiple unknown stakeholders for a variety of purposes. Informed by this research, it is my conclusion that the most comprehensive way to obtain an estimation of the use of data for various purposes is through a bottom-up approach, as is employed for calculating the GDP. One possible way to understand use of open data in the private sector is through self-reporting, a method used by the GovLab's Open Data 500 index. As a second approach, the Danish BDP has adopted a reporting template where individual organizations use a predefined method to report on both tangible and intangible benefits they observe from the use of data. As a third approach, the Danish Geodata Agency used a survey methodology to create a baseline estimate of use of geographic data, before they were made free-of-charge in January 2013.

However, the cost-benefit analysis approach is by definition an ex-post method. From my identification of the open data value paradox, I conclude that we need to evaluate ex-ante the potential benefits in relation to various scenarios consisting of different configurations of enabling factors or barriers. To accomplish this I have suggested the use of option value methods (further discussion in Paper VIII). Option value thinking can benefit from evaluating opportunity costs: how would the geo-service industry have developed if the USA had not made GPS data available? Alternatively, how would our medical knowledge have progressed if genomics data had not been shared? I propose that this way of thinking will draw our focus towards the importance of

creating the right environment for individuals, businesses and governments that would like to utilize open data, providing them with an option for value creation.

The final sub-question raised was: *What are the key implementation strategies and business models that can promote long-term generation of value from open data?*

This is in fact two questions in one. I will first address the part that considers key implementation strategies.

I propose that for most ODIs, the seven dimensions of liquid open data can be used to benchmark the desired outcome. I expect that more standardized, more interoperable and more readily reusable data are preferred everywhere. However, as ODIs operate in very different contexts, appropriate strategies for implementing liquid open data can vary considerably. In some countries, the need to digitize data might introduce a first implementation barrier, while in other countries the challenge revolves around making already available data open for external use. These two cases introduce very different implementation barriers. For instance, Estonia, one of the frontrunners in E-Government, has in fact benefitted from having next to no pre-existing systems, thus being able to develop them effectively from the start. This in contrast to a country like Denmark who struggles with outdated legacy systems, legislation and a soft infrastructure that is socially entrenched and difficult to change. Denmark, on the other hand, benefits from high capacity and professionalism in data collection and widespread use of central sources of data across government. In the specific case of Denmark, explored in Paper VI, the main governance challenge was finding a balance between the safeguarding the autonomy of existing data custodians and ensuring an overall coordination between program participants, including use of common standards. Accordingly, I suggest that the approach the BDP adopted could be suitable for other ODIs with a strong focus on government efficiency through open data and interoperability, operating within a system that already supports cross-government collaboration.

The second part of the question concerns business models. The answer to this part remains in observation, however the business models behind two-sided markets or MSPs do provide a hopeful avenue of exploration for creating sustainable value, as discussed in Paper VIII. These business models have solved the complications associated with capitalizing on the intangible value created through information sharing mechanisms. The value inherent in the provided content attracts the non-paying stakeholders to the platform. Moreover, this group of users incentivizes the other group of stakeholders, the other side so to speak, to use the platform. The latter group, or the other “side” of the platform, utilizes market mechanisms create value for

money or new income, and thus contributes to the financial sustainability of the platform.

## **6.2 Contributions**

I have proposed that the open data construct has thus far lacked clarity. Clear constructs are simply robust categories that distill phenomena into sharp distinctions that are comprehensible to a community of researchers (Suddaby, 2010). The first theoretical contribution of this PhD project is the definition of liquid open data, which is explained in more detail in section 5.1. In order to increase user engagement, ODIs should consider questions like: 1) what datasets are in demand and likely to be used now or in the future? (Availability); 2) how important is it for users to have free access to the data? (Affordability); 3) what guarantees do the users need to start working with the data? (Reusability); 4) what formats are easiest for the different users to work with and what level of accuracy, timeliness and completeness is necessary? (Usability); 5) how easy is it for users to find the data and information about the data (Discoverability); 6) how easy is it for users to connect to the data or download for further use? (Accessibility); 7) how easy is it for users to understand the data out of context and how easy is it for users to link different sets of data together? (Interoperability).

The second contribution of this PhD project is a definition, conceptualization and measurement of the sustainable value construct. I have suggested a method to measure the sustainable value that is generated through content sharing and the creation of new digital products and services. In Paper VII I have modeled sustainable value as a complex multidimensional indicator using structural equation modeling and PLS path analysis (Trincherá & Russolillo, 2010). Using secondary data from reliable sources to estimate a structural model, I suggest the resulting latent variable scores can be used to indicate the relative level of sustainable value across countries. This method has been suggested for creating multidimensional indicators, as it relieves us from having to make subjective judgments about the proportion of each indicator, as is often the case when building composite indexes (Costanza et al., 2009; Trincherá & Russolillo, 2010).

Any theory must not only provide construct clarity but also identify the relationships among constructs (Suddaby, 2010). Critical realists argue that unobserved generative mechanisms can explain how phenomena come about. While Coleman's framework highlights that the relationship between societal context and societal outcomes must happen through action of individuals, we can use prior research regarding certain

individual level tendencies to generalize at the macro level. I propose this is an important step towards facilitating a discussion about the societal level impact of open data. Naturally, theorizing at the macro level means we must sacrifice the richness and diversity of individual action, but instead we gain a high-level framework from which detailed individual or organizational level analysis can depart. As a third contribution, I have suggested two general and overarching value generating mechanisms, the information sharing mechanism and the market mechanism. Moving down one level of abstraction, I suggest two types of information sharing mechanisms: transparency mechanisms and civic engagement mechanisms as well as two types of market mechanisms: efficiency mechanisms and innovation mechanisms. Throughout the papers that form the backbone of this study my co-authors and I have introduced a number of in-depth use cases and case studies intended to illustrate these causal relationships, and to bring about a more nuanced understanding of how use of open data contributes to the generation of sustainable value through these two different mechanisms.

Moreover, the information sharing mechanism has thus far not been explicitly discussed or conceptualized as a value generating mechanism in the context of open data. Therein lies my biggest contribution. I have contributed to a vibrant academic discussion on how we can better understand and articulate the value generation process that happens through information sharing in network relationships with the aim of generating a midrange theory. Midrange theory starts with an empirical phenomenon and abstracts from it to create general statements that can be verified by empirical data (Gregor, 2006; Merton, 1949; 1968). While the proposed mechanisms are general, the context and boundary of the theory is clearly the phenomenon of open data. According to Gregor's (2006) classification the proposed theory of open data value generation can be classified as level IV theory for explanation and prediction, including causal explanations (mechanisms), testable relationships and the ability to predict as PLS-structural equation modeling is well-suited for predictions. My hope is that in developing a midrange theory, I have made a theoretical contribution to the emerging field of open data.

The generation of sustainable value from open data does not happen in a vacuum, and there are many contextual factors at differing levels of analysis that play a significant role. These factors can help us explain why, or why not, the suggested value generating mechanisms are activated in different countries. As a fourth theoretical contribution, I have identified a preliminary conceptualization of four societal level enabling factors

that together form a societal level infrastructure I propose can stimulate value generation through open data.

The fifth contribution of this PhD research is the identification of the open data value paradox. The open data value paradox is both theoretically interesting and practically relevant. It is interesting to private sector businesses as it applies to a very central problem they are currently facing. This problem is a manifestation of consumer adaptation to free information access, delivered over mobile devices, the World Wide Web, and most recently though wearable technology. Moreover, it is relevant for governments, as they must make decisions on how to spend the limited funds they possess. In many Western societies, governments are using evaluation methods based on the new public management paradigm, which is rooted in a logic based or market like practice (Henriksen, 2006). This paradigm makes it difficult for many ODIs to survive unless new ways to evaluate E-Government initiatives are suggested. The emerging New Public Value paradigm (the concept of Public Value was discussed briefly in section 2.4) might be more sympathetic towards initiatives such as ODIs but still faces considerable challenges of governance and evaluation (Stoker, 2006). I posit that it is imperative that businesses and governments find a way to understand, articulate, document and evaluate the inherent worth of open data.

Just as I have used the insights from practitioners to help create relevant theory, I have used the theoretical lenses that I have counted as contributions of this PhD project for a number of practical contributions.

Firstly, by continuously engaging the BDP participants and exposing them to different conceptualizations and models, as well as discussing the practical relevance of the findings (which has admittedly revealed that some of them are more readily utilized than others), I have managed to infuse and test some of the theory on a real-world ODI. This academic-practice exchange of knowledge has been fruitful to both partners, as suggested by the Engaged Scholarship paradigm (Van de Ven, 2007). Moreover, it has helped all involved gain a holistic view of the phenomenon of open data. The Basic Data case study (Paper VI) concentrates on the supplier side of open government data. This case study has offered a unique insight into the real technical and governance related struggles of an ODI over almost four years. It also offers an insight into how open data can contribute to public sector efficiency and how the impacts of an open data program can be evaluated. I have attempted to transform the experiences of the BDP group into a number of practical guidelines that can be used by other similar initiatives and summarized the findings in a process model that shows the open data value lifecycle from the supplier point of view.

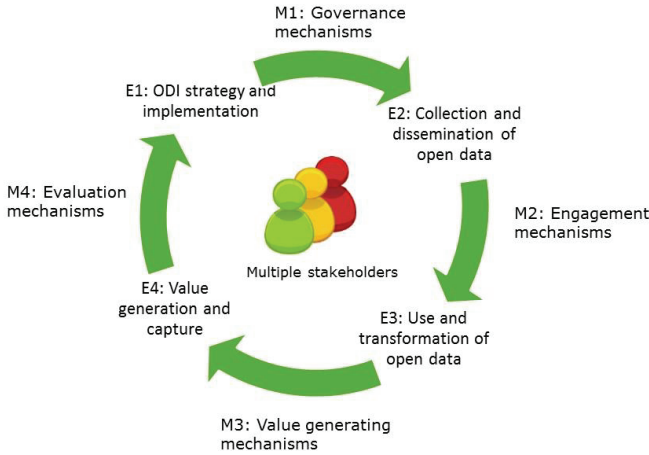
Secondly, the Opower case study (Paper IV) is focused on the demand side of open data. Unlike the other papers in this PhD research, it is positioned at the organizational level. The Opower case gave some interesting insights into the dimensions of data that are of interest to private sector users, such as data usability, discoverability and accessibility. Moreover, the study highlights that open data are for the most part combined with proprietary data when utilized by private business users. Open data are thus an important resource for multiple stakeholders, but represent in many cases only a fraction of the overall resources used. Therefore, I consider the proposed similarities between open data and infrastructural resources as being both powerful and accurate. Drawing on this approach, open data compare in certain aspects to public roads; they provide the path to destination, yet are seldom viewed as the largest contributor to the journey. Accepting this similarity might subtly affect the private sector's view on the importance of open data.

Thirdly, the conceptualization of liquid open data and sustainable value combined with the four value generating mechanisms can be of practical relevance. As previously mentioned, I am designing a methodology for evaluating individual datasets, and ODIs in general, using these constructs. Moreover, I am continuing to develop the evaluation methodology based on real option value. These efforts will be part of a future research agenda, paying specific attention to the design related implications of my findings.

### **6.3 Final Contribution: Research Agenda**

The final contribution is a result of the method of retrodution. Based on a triangulation between qualitative and quantitative data I have determined four classes of relevant mechanisms in the open data value lifecycle, illustrated in figure 17 (recreated from figure 3 for clarity of representation). Furthermore, I have identified four distinct "events" that happen in the domain of the actual when these mechanisms are activated.

- 1) Event1: ODI strategy and implementation of open data infrastructure.
- 2) Event 2: Active dissemination of open data.
- 3) Event 3: Use and transformation of open data.
- 4) Event 4: Value generation through open data and subsequent value capture by multiple stakeholders.



*Figure 17: The Open Data Value Generation Lifecycle Revisited*

The events are connected through four distinctive classes of underlying and unobservable generative mechanisms: 1) M1: Governance mechanisms, explaining how the ODIs are governed, progressing from open data strategy towards active dissemination of open data. 2) M2: Engagement mechanisms that explain how users engage with open data and choose which of them to use. 3) M3: Value generating mechanisms that explain how value is generated from the open data. 4) M4: Evaluation mechanisms that explain how governments and other decision makers evaluate the impacts of the ODI and how this evaluation impacts future decision making.

The first event is described to some degree in Paper I and then in more detail in Paper VI, where the efforts of the BDP to refine and execute the open data strategy are examined. While a great deal of the open data research corpus focuses on the barriers and enablers of implementing open data policies, I have focused more on the implementation process itself, and consequently the governance mechanisms. In the case of the BDP, the governance approach that was eventually adopted by the program leaders was what I have identified as System-of-Systems (SoS) governance. SoS governance is a type of network governance that utilizes several governance principles. The first item on the proposed research agenda is to examine the types of governance mechanisms used in ODIs in more depth and to determine the appropriateness of governance mechanisms depending on the relevant context.



The second event is also described in Paper VI. However, as the implementation of the technical infrastructure intended for open data dissemination was not complete when the PhD research period ended, the full impact of the BDP has not yet emerged. Accordingly, I did not have the opportunity to thoroughly observe user engagement. However, I intend the liquid open data construct as an evaluation of “output” of an ODI. Regarding the mechanisms that explain how users engage with the disseminated data, I have briefly addressed those through the Opower case (Paper IV) and through the MOA theoretical frameworks (Paper V). Open data engagement, however, is still an unexplored research theme. The second item of the proposed research agenda is to advance more in-depth examination of how users engage with data, represented by different dimensions of liquidity and openness.

The third event is the use and transformation of the data, which I have described by using various use cases (Paper IV, Paper V, Paper VII and Paper VIII). The third item of the proposed research agenda is to collect more use cases and case studies to create a better understanding of the different ways in which open data are being used for value generation. User stories function as good examples and can furthermore be used to motivate both data users and data publishers. I, however, find the value generating mechanisms that explain how use of data contributes to sustainable value to be more theoretically interesting as they act as a tool for theoretical demi-generalizations. While I propose I have already contributed significantly to this particular area, it remains an interesting research opportunity to traverse yet another level of abstraction and explore these mechanisms in more detail within a more limited context.

The fourth event reflects the value generation and capture. For this event, my focus was primarily on creating a concept or construct that can be used to identify both the economic and the social or intangible value generated through use of open data. I argue that such a construct needs to move beyond the all-encompassing monetary valuation methods of our current times. The evaluation mechanism will reflect how governments, and other potential investors in the open data ecosystem, will look at the ex post and ex ante value generated through open data. My hope is that additional research in this specific arena will bring greater clarity to the relationships that explain the nature of value that is generated from shared resources such as open data. Furthermore, I suggest that more focus on this area will contribute to the improvement of evaluation methods. In Paper VIII, I propose that if evaluation methods were to reflect the intangible social and environmental value to a higher degree, interest in ODI related investment would also increase. The final item on the research agenda is to

contribute to a thorough understanding of how evaluation methods influence the willingness to invest in open data, both in the public sector and the private sector.

**Table 4: Open Data Research Agenda**

Category	Research question
Governance mechanisms	<i>What types of governance mechanisms are ODIs using and how do these types relate to the success/failure of the ODI?</i>
Engagement mechanisms	<i>How are users engaging with open data?</i>
Value generating mechanisms	<i>How is open data used for value generation?</i>
Evaluation mechanisms	<i>How do evaluation methods influence the willingness to invest in open data?</i>

## 6.4 Research Limitations and Validity of Results

This section aims to both outline the limitations of this research, while in parallel establishing the validity of the results. The use of mixed methods was chosen to overcome the limitations of individual methods. I claim that use of mixed methods contributes positively to the quality of analysis, as the researcher can draw on the individual strength of each method. However, when utilizing multiple methods, researchers face limitations regarding the depth with which they can explore each method. In the case of this study, I focused on a robust quantitative data analysis, while using the qualitative data in an exploratory manner, for increasing understanding. During this PhD study I conducted two longitudinal single case studies (Paper IV and Paper VI). Much of my personal inspiration and insight was derived from the interaction with the people participating in the Danish BDP, as well as from the Opower case. In both cases, I was witnessing events that were unfolding during my window of observation, which made it very difficult to decide appropriate cutoff points. Admittedly, while I tried to convey historical information, the latter cutoff points were based on convenience, related to the individual publications.

For the qualitative data analysis I used Yin’s (2009) case study approach, with some inspiration from the Harvard Case Study method. Although the latter method has traditionally been applied to teaching cases, I found value in the Harvard method’s use of a chronological timeline. Instead of the grounded theory approach using coding to uncover themes in the data, I created “stories” that connected events in a process and

described traits and abilities within a structural framework. Such stories are instrumental in explaining complex results (Yin, 2009). Afterwards, I applied a coding scheme from the theoretical lenses to highlight events in the storyline. A challenge of this approach is that the researcher must use creative insight and careful interpretation to make sense of, and explain observed findings (explanation building), as well as identify events and mechanisms that have not been observed. Reliability can be difficult to establish in such a study. In an attempt to improve conformability, I used secondary sources of data to corroborate my primary interview and observation data. For the Basic Data Program case study, I had access to a number of program documents as well as to an external review of the program. For the Opower case, I used a Harvard Teaching Case and a published economics paper that reviewed Opower's methodology for evaluating energy efficiency gains. Additionally, I made use of online material and internal documents.

To create credibility is always an issue in case study research and is certainly a limitation of the qualitative part of this research. To improve credibility I had research participants review my narrative and correct factual errors. Of course, all interpretations and conclusions remain my own (and my co-author's in the Opower case). To introduce traceability, I created an Outlook meeting request, which the interviewee approved. While most interviews were recorded, technical difficulties prevented this in two cases, and descriptive notes were relied on. Both case studies are single cases and thus it is difficult to generalize outside of the specific context in which the events happened. In order to make the context more explicit and introduce a level of transferability (Venkatesh et al., 2013), I used secondary data to create a thick context description so that the readers could themselves evaluate how the context influenced the events.

There are also a number of limitations related to the quantitative work.

Firstly, the sample size used in both PLS studies (Paper III and Paper VII) was small as there is limited secondary data to be found that are consistent across multiple countries and are fit for the purpose of reflecting different degrees of openness of government data. I was essentially limited by the number of observations in the Open Data Barometer. However, I have reasonable evidence to believe that the model contains enough power to draw preliminary conclusions from results.

Secondly, while the majority of the secondary data providers are reliable and have a long history of collecting and disseminating data, the Open Data Barometer is relatively new as it was constructed for the first time in 2013, collecting data for 2012, which was used in the pilot PLS study. However, I participated in the data collection

myself and as such can verify that scientific methods were used in the process. The data in the Open Data Barometer are based on an expert survey, using similar methodology as the Web Index and the World Economic Forum's global technology reports and global competitiveness reports.

Thirdly, comparing impacts from open data between countries, where many ODIs are still in their infancy, might be premature. That being said, there are types of open government data that have been available for many years although worldwide interest in open data is relatively recent. For instance, Obama's open government initiative was born six years ago, and even at the time, important types of data such as geographic data and weather data were already available free of charge. In Denmark, address and property data were made available free of charge in 2003.

The use of mixed methods introduces its own limitations. While I believe my conceptualizing efforts benefitted from this approach, there are limits concerning triangulation of different types of data and research methods. For instance, some attributes of open data are more difficult to measure than others. It is relatively easy to estimate whether data are published under open licenses, or if they are disseminated in machine-readable formats. However, it is much more difficult to evaluate qualitative attributes like data quality as they depend on the nature of use, the needs of the users and other less generalizable attributes. The quality of basic data was certainly one of the main areas of emphasis in the BDP. However, in the econometric modeling, I could not find appropriate societal level measures that reflected data quality. While quality aspects or dimensions do appear in the final liquid open data construct, I could not emphasize the importance of data quality in the quantitative work to the same degree as was done in the case studies. Accordingly, not all of the insights from the qualitative studies could be carried forward to the quantitative studies.

There are also limitations and biases related to the researcher herself. I needed to familiarize myself with PLS and a case study research methods when I started working on the PhD study, and as such much time was spent learning the two approaches. I am relatively confident of the robustness of the qualitative analysis due to fact that the quality indicators in quantitative studies are well documented and easily comparable between studies, and I had relevant statistical experience from previous education and work. However, when reflecting on this PhD study, I conclude that it is difficult to learn the skills required for qualitative data analysis from a book or a single PhD course. I believe these skills are best acquired through experience over time, and conclude that this is an area with growth opportunities for myself as a researcher. While it is hard to evaluate one's own biases, I suspect I have leaned towards

positivism in spite of my focus on the CR-based mechanism approach. I also must make explicit that I truly believe open data offer the opportunity to generate value, which does make me biased in my evaluation (which is incidentally not a very positivistic trait).

## **6.5 Final Reflections**

This PhD study has been a rewarding journey, and to my surprise, I did not lose my enthusiasm for the topic over these last three years. I still believe that use of data can have a real impact on the world, and that openness is a very important factor in this context. I furthermore suspect that the current level of value that is being directly or indirectly generated through use of open data is only a small fraction of the value we could generate. I know, however, that this potential value generation is by no means inevitable. Interest in open data might fade away, and ODIs and our societies in general need to overcome many barriers if they are to be able to introduce new societal structures. One of the more difficult barriers is the way we are accustomed to think about value and value generation. I sincerely hope this PhD dissertation and the included papers will influence the current way of thinking. Furthermore, I hope that future research will identify how we can change current structures, rendering them more supportive of openness and sharing and the collaborative generation of sustainable value. If my work can motivate people to look for new ways of generating value my aspirations for this study will have been reached. I look forward to a future where open data are used in cross-boundary collaborative initiatives, simultaneously using the forces of the markets and the information sharing networks to generate sustainable value for all of us and our future generations.

## 7. References

- 1 Afonso, A., Schuknecht, L. and Tanzi, V. (2010). Public sector efficiency: evidence for new EU member states and emerging markets. *Applied Economics*, 42(17), 2147-2164
- 2 Agency for Digitization (2012). Good basic data for everyone – a driver for growth and efficiency. Available <http://www.digst.dk/Home/ServiceMenu/English/Digitisation/Basic%20Data>
- 3 Agency for Digitization (2011). eGovernment strategy 2011-2015: The digital path to future welfare. Available <http://www.digst.dk/ServiceMenu/English/Policy-and-Strategy/eGOV-strategy>
- 4 Alani, H.; Dupplaw, D.; Sheridan, J.; O'Hara, K.; Darlington J.; Shadbolt, N.; Tullo, C. (2007). The Semantic Web, *Lecture Notes in Computer Science*, 4825(2007), 708-721.
- 5 Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95( 10-11), 1082–1095.
- 6 Andersen, K.M., Henriksen, H.Z., Secher, C. & Medaglia, R. (2007). Costs of eparticipation: the management challenges. *Transforming Government: People, Process and Policy*, 1(1), 29 – 43.
- 7 Andrade, P.L., Hemerly, J., Recalde, G. and Ryan, P. (2014). From Big Data to Big Social and Economic Opportunities: Which Policies Will Lead to Leveraging Data-Driven Innovation's Potential? In Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014): The Global information technology report 2014: Rewards & Risks of Big Data. World Economic Forum, 81-86.
- 8 Avgerou, C. (2013). Explaining Trust in IT-Mediated Elections: A Case Study of E-Voting in Brazil. *Journal of the Association for Information Systems*, 14(8), 399-419.
- 9 Axelsson, K., Melin, U. & Lindgren, I. (2010). Exploring the Importance of Citizen Participation and Involvement in E-government Projects - Practice, Incentives and Organization. *Transforming Government: People, Process and Policy*, 4(4), 299-321.
- 10 Avital, M., Lyytinen, K., King, J.L., Gordon, M.D., Granger-Happ, E., Mason, R.O. & Watson, R.T. (2007). Leveraging Information Technology to Support Agents of World Benefit. *Communications of the Association for Information Systems*, 19 (25), 567-588.
- 11 Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Journal*, 14, 496–515.
- 12 Bailey, J. P. & Bakos, Y. (1997). An exploratory study of the emerging role of electronic intermediaries. *International Journal of Electronic Commerce* (1997), 7-20.
- 13 Barber, B.R. (1984). *Democracy Participatory Politics for a New Age*. Berkeley University of California Press
- 14 Barney, J. (1991). Firm resources & sustained competitive advantage. *Journal of management*, 17(1), 99-120
- 15 Barry, E. & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, 19(1), 129-152.
- 16 Bartenberger, M., & Grubmüller, V. (2014). The Enabling Effects of Open Government Data on Collaborative Governance in Smart City Contexts. Available: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2474974](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474974)
- 17 Bason, C. (2010). *Leading Public Sector Innovation: Co-creating for a better society*. Bristol: Policy Press.
- 18 Bean, R & Koeppel, H. (2012). Big Data Analytics: The Currency of the 21st Century Enterprise. [Online]. Available at <http://www.information-management.com/news/big-data-analytics->

- the-currency-of-the-21st-centuryenterprise-10023139-1.html
- 19 Beardsley, S. Enríquez, L. Grijpink, F. Sandoval, S., Spittaels, S. & Strandell-Jansson, M. (2014). Building Trust: The Role of Regulation in Unlocking the Value of Big Data. In Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014): The Global information technology report 2014: Rewards & Risks of Big Data. World Economic Forum, 73-81.
  - 20 Benington, J. (2011). From Private Choice to Public Value? In Benington, J. & Moore, M. (eds.) *Public Value: Theory and Practice*. Houndmills, Basingstoke: Palgrave MacMillan.
  - 21 Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven CT: Yale University Press.
  - 22 Bertot, J.C., Jaeger, P.T. & Grimes, J.M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27(3), 264-271.
  - 23 Bertot, J.C., Jaeger, P.T. & Grimes, J.M. (2012). Promoting transparency and accountability through ICTs, social media, and collaborative e-government, *Transforming Government: People, Process and Policy*, 6(1), 78-91.
  - 24 Bharadwaj, A., El Sawy, O.A., Pavlou, P.A. & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. *MIS Quarterly* 37(2), 471-482.
  - 25 Bhaskar, R. (1975). *A Realist Theory of Science*. Leeds, UK: Leeds Books.
  - 26 Bhaskar, R. (1978). *The Possibility of Naturalism*. Sussex, UK: Harvester Press.
  - 27 Bhaskar, R. (1989). *Reclaiming reality*. London: Verso.
  - 28 Bhaskar, R. (1998). Philosophy and Scientific Realism. In Archer, M. S., Bhaskar, R., Collier, A., Lawson, T. & Norrie, A. (Eds.). *Critical Realism: Essential Readings* (pp. ix-xxiv). London: Routledge.
  - 29 Bhaskar, R. (2008). *A Realist Theory of Science*. London: Routledge.
  - 30 Bhattacharjee, A. (2012). *Social Science Research: Principles, Methods, and Practices*, (2. ed.). Tampa, FL: Global Text Project.
  - 31 Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2013). The global information technology report 2013: Growth & Jobs in a Hyperconnected World. World Economic Forum.
  - 32 Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014). The global information technology report 2014: Rewards & Risks of Big Data. World Economic Forum.
  - 33 Bimber, B. (2003). *Information and American democracy: Technology in the evolution of political power*. New York: Cambridge University Press.
  - 34 Blumberg, M., & Pringle, C. D. (1982). The missing opportunity in organizational research: Some implications for a theory of work performance. *Academy of Management Review* 7(4), 560-569.
  - 35 Blöndal, J. R. (2005). Market-type mechanisms and the provision of public services. *OECD Journal on Budgeting*, 5(1), 79-106.
  - 36 Bountouri, L.P., Christos Soulikias, V. & Stratis, M. (2009). Metadata interoperability in public sector information. *Journal of Information Science*, 35(2), 204-231.
  - 37 Bowman, C. & Ambrosini, V. (2000). Value Creation Versus Value Capture: Towards a Coherent Definition of Value in Strategy. *British Journal of Management*, 11(1), 1-15.
  - 38 Bowman, C. (2015). The role of technology in the creation and capture of value. *Technology Analysis & Strategic Management*, 27(2), 237-248.
  - 39 Brandeis, L.D. (1914). Other people's money. [Online]. Available: <http://www.law.louisville.edu/library/collections/brandeis/node/196>
  - 40 Brynjolfsson, E., & Saunders, A. (2009). *Wired for Innovation*. MIT Press.

- 41 Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in Numbers: How does data-driven decision-making affect firm performance? Paper presented at the 32nd International Conference on Information Systems, Shanghai, China.
- 42 Brynjolfsson, E., & Oh, J. (2012). The attention economy: measuring the value of free digital services on the Internet. Paper presented at the 33rd International Conference on Information Systems, Orlando, USA.
- 43 Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. *McKinsey Quarterly*, 4, 24-35.
- 44 Bunge, M.A. (1999). *The Sociology-Philosophy Connection*. New Jersey: Transaction Publishers.
- 45 Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the social sciences*, 34(2), 182-210.
- 46 Bygstad, B. & Munkvold, B. E. (2011). In search of mechanisms. Conducting a critical realist data analysis. Paper presented at the Thirty Second International Conference on Information Systems (ICIS), Shanghai, 2011.
- 47 Caillaud B. & Jullien, B. (2003). Chicken and Egg: Competition Among Intermediation Service Providers. *RAND Journal of Economics* 34(2), 309-328.
- 48 Cannon, S. and Summers, L.H. (2014). How Uber and the Sharing Economy Can Win Over Regulators. *Harvard Business Review*. [Online]. Available: <https://hbr.org/2014/10/how-uber-and-the-sharing-economy-can-win-over-regulators/>
- 49 Capgemini Consulting (2013). The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data. [Online]. Available: [https://www.capgemini-consulting.com/resource-file-access/resource/pdf/opendata\\_pov\\_6feb.pdf](https://www.capgemini-consulting.com/resource-file-access/resource/pdf/opendata_pov_6feb.pdf)
- 50 Coleman, J.C. (1986). Social Theory, Social Research and a Theory on Action. *American Journal of Sociology*, 91(6), 1309-1335.
- 51 Coleman, J.C. (1990). *Foundations of Social Theory*. Boston: Harvard University Press.
- 52 Conradie, P. and Choenni, S. (2014). On the barriers for local governments releasing open data. *Government Information Quarterly*, 31(1), S10-S17.
- 53 Costanza, R., Hart, M., Talberth, J., & Posner, S. (2009). Beyond GDP: The need for new measures of progress. [Online]. Available: [http://pdxscholar.library.pdx.edu/iss\\_pub/11/](http://pdxscholar.library.pdx.edu/iss_pub/11/)
- 54 Cook, J. S. (2010). Economic issues in funding and supplying public sector information. In Fitzgerald, B. (ed.). *Access to public sector information: law, technology and policy: Volume 1*, Sydney: Sydney University Press.
- 55 Creswell, J. W. (2003). *Qualitative, quantitative and mixed method approaches* (2. Ed). Thousand Oaks, California: Sage Publications Inc.
- 56 Danermark, B., Ekström, M., Jakobsen, L. & Karlsson, J.C. (2002). *Explaining Society: Critical Realism in the Social Sciences*. London: Routledge.
- 57 Darden L. (2006). *Reasoning in Biological Discoveries. Essays on Mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge, UK: Cambridge Univ. Press
- 58 Davies, T. (2013). Open Data Barometer: 2013 Global Report. Available: <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>
- 59 de Vries, M.; Kapff, L.; Achiaga, M. N.; Wauters, P.; Osimo, D.; Foley, P.; . . . & Whitehouse, D. (2011). Pricing of public sector information study (POPSIS), Brussels: European Commission, Information Society and Media Directorate General.
- 60 Downward, P. & Mearman, A. (2002). Critical Realism and Econometrics: Constructive Dialogue with Post-Keynesian Economics. *Metroeconomica*, 53(4), 391-415.



- 61 Downward, P. & Mearman, A. (2006). Retrodution as mixed-methods triangulation in economic research: reorienting economics into social science. *Cambridge Journal of Economics*, 31(1), 77-99.
- 62 Eckartz, S. M., Hofman, W. J., & Van Veenstra, A. F. (2014). A Decision Model for Data Sharing. Paper presented at the 13<sup>th</sup> IFIP WG 8.5 Conference, Ireland. In Janssen, M., Scholl, H.J., Wimmer, M.A. & Bannister, F.E. (eds.) *Electronic Government*, 253-264, Springer Berlin Heidelberg.
- 63 Ehrlich, T. (Ed.). (2000). *Civic responsibility and higher education*. Greenwood Publishing Group.
- 64 Elster, J. (2007). *Explaining Social Behavior*. Cambridge: Cambridge University Press,
- 65 Estermann, B. (2014). Diffusion of open data and crowdsourcing among heritage institutions: results of a pilot survey in Switzerland. *Journal of theoretical and applied electronic commerce research*, 9(3), 15-31.
- 66 European Commission. 2003. Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the Re-Use of Public Sector Information [Online]. Available: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF>
- 67 European Commission. (2006). eGovernment Economics Project (eGEP). Measurement Framework Final. DG Information Society and Media, European Commission
- 68 European Commission (2011a). Digital Agenda: Commission's Open Data Strategy, Questions & answers [Online]. Available [http://europa.eu/rapid/press-release\\_MEMO-11-891\\_en.htm?locale=en](http://europa.eu/rapid/press-release_MEMO-11-891_en.htm?locale=en)
- 69 European Commission. (2011b). Open data: An engine for innovation, growth and transparent governance. European Commission Information Society. [Online]. Available: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF>
- 70 European Commission (2011c). Special Eurobarometer 359: Attitudes on Data Protection and Electronic Identity in the European Union. Brussels: TNS Opinion & Social. [Online]. Available at [http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_359\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf)
- 71 Fleurbacy, Marc. (2009). Beyond GDP: The Quest for a Measure of Social Welfare. *Journal of Economic Literature*, 47(4), 1029-75.
- 72 Foulonneau M, Turki S, Vidou G & Martin S. (2014). Open data in Service design. *Electronic Journal of e-Government*, 12(2), 99-107.
- 73 Frischmann, B. M. (2012). *Infrastructure: The Social Value of Shared Resources*. Oxford: Oxford University Press.
- 74 García, R., Paulheim, H., & Di Maio, P. (2015). Special Issue on Semantic Web Interfaces. *Semantic Web*, 6(3).
- 75 Gil-Saura, I., Frasquet-Deltoro, M. & Cervera-Taulet, A. (2009). The Value of B2B Relationships. *Industrial Management & Data Systems*, 109(5), 593-609.
- 76 Gregor, S. (2006). The nature of theory in information systems. *MIS quarterly*, 30(3), 611-642.
- 77 Grover, V. & Kohli, R. (2012). Cocreating IT Value: New Capabilities and Metrics for Multifirm Environments. *MIS Quarterly*, 36(1), 225-232.
- 78 Gurstein, M. (2010). Open Data: Empowering the Empowered or Effective Data Use for Everyone? *First Monday*, 16(2), February 7.
- 79 Hagiú, A., & Wright, J. (2015). Multi-sided platforms. *International Journal of Industrial Organization (forthcoming)*.
- 80 Harrison, T.M.; Guerrero, S.; Burke, G.B.; Cook, M.E.; Cresswell, A.M.; Helbig, N.; Hrdinová, J. & Pardo, T.A.. (2011). Open government and e-government: democratic challenges from a public value perspective. In Bertot, J.C. Editor, Nahon, K. Editor,

- Chun, S.A. Editor, Luna-Reyes, L.F. Editor & Atluri, V. Editor. *Digital Government Innovation in Challenging Times*, Paper Presented at the 12th Annual International Conference on Digital Government Research, College Park, MD, 245-25.
- 81 Hair, J.F., Ringle, C.F., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, 19(2), 139–151.
- 82 Hair, J.F., Ringle, C.F., & Sarstedt, M. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Management Science* (40), 414-433
- 83 Halonen, A. (2012). Being Open About Data. Available: <http://finnish-institute.org.uk/images/stories/pdf2012/being%20open%20about%20data.pdf>
- 84 Harrison, T.M, Pardo, T.A. & Cook, M. (2012). Creating Open Government Ecosystems: A Research and Development Agenda. *The Future Internet*, 4(4), 900-928.
- 85 Haufler, V. (2001). *A public role for the private sector: Industry self-regulation in a global economy*. Carnegie Endowment.
- 86 Hedström, P. & Swedberg, R. (1998). *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press, 1-31.
- 87 Hedström P., & Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Annual Review of Sociology*, 36, 49-67.
- 88 Heimstädt, M., Saunderson, F., & Heath, T. (2014). Conceptualizing Open Data ecosystems: A time-line analysis of Open Data development in the UK. Discussion Paper (No. 2014/12). School of Business & Economics.
- 89 Henfridsson, O. & Bygstad, B. (2013). The Generative Mechanisms of Digital Infrastructure Evolution. *MIS Quarterly*, 37(3), 907-931.
- 90 Henriksen, H. Z. (2006). Fad or investment in the future: an analysis of the demand of e-services in Danish municipalities. *Electronic Journal of e-Government*, 4(2), 19-26.
- 91 Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in international marketing*, 20, 277-319.
- 92 Henseler, J. & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path model-ing. *Computational Statistics*, 28(2), 565–580.
- 93 Hitt, M. A., Ireland, R. D., Sirmon, D. G., and Trahms, C. A. (2011). Strategic entrepreneurship: Creating value for individuals, organizations, and society. *The Academy of Management Perspectives*, 25(2), 57-75.
- 94 Hjalmarsson, A. Juell-Skielse, G., Ayele, W., Rudmark, D. & Johannesson, P. (2015a). From Contest to Market Entry: A Longitudinal Survey of Innovation Barriers Constraining Open Data Services Development. Paper presented at the 23rd European Conference on Information Systems (ECIS) 2015, Münster, Germany.
- 95 Hjalmarsson, A., Johansson, N. & Rudmark, D. (2015b). Mind the Gap: Exploring Stakeholders' Value with Open Data Assessment. Paper presented at the 48th Hawaii International Conference on System Sciences, Hawaii, USA, 5-8.
- 96 Hofman, W., & Rajagopal, M. (2014). A technical framework for data sharing. *Journal of theoretical and applied electronic commerce research*, 9(3), 45-58.
- 97 Horst, N. L., Bjerre, S., Lind, M., & Hvingel, L. (2014). Grunddataprogrammet dansk infrastrukturmodel for offentlige data. *Geoforum Perspektiv*, 13(24), 36-44.
- 98 Houghton, J. (2011). Costs and Benefits of Data Provision: Report to the Australian National Data Service, Centre for Strategic Economic Studies, Victoria University.
- 99 Huijboom, N. & Van den Broek, T. (2011). Open data: an international comparison of strategies, *European Journal of e Practice*, 12(1), 4-16.
- 100 Höchtel, J., & Reichstädter, P. (2011). Linked open data-a means for public sector information management. In Andersen,

- K.M., Francesconi, E., Grönlund, A. & van Engers, T.M. (eds). *Electronic Government and the Information Systems Perspective*. Paper presented at the Second International Conference, EGOVIS 2011, Toulouse, France, 330-343. Springer Berlin Heidelberg.
- 101 Höchtel, J., Davies, T., Janssen, M. & Schieferdecker, I. (2014). Open Data: Growing Up and Getting Specific. *JeDEM: eJournal of eDemocracy and Open Government*, 6 (1), i-iii.
- 102 Jaeger, P. T. (2007). Information policy, information access, and democratic participation: The national and international implications of the Bush administration's information politics. *Government Information Quarterly*, 24(4), 840-859.
- 103 Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258-268.
- 104 Janssen, M. & Estevez, E. (2013). Lean government and platform-based governance - Doing more with less. *Government Information Quarterly* 30(1), S1-S8.
- 105 Janssen, M., & Zuiderwijk, A. (2014). Infomediary business models for connecting open data providers and users. *Social Science Computer Review*, 32(5), 694-711.
- 106 Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28(4), 446-456.
- 107 Jetzek, T. Avital, M & Bjørn-Andersen, N. (2012). The Value of Open Government Data: A Strategic Analysis Framework. Paper presented in the 2012 Pre-ICIS e-Government Workshop, Orlando, USA 12/2012.
- 108 Jetzek, T., Avital, M. and Bjørn-Andersen, N. (2014a). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2), 100-120.
- 109 Jetzek, T., Avital, M. & Bjørn-Andersen, N. (2014b). Generating sustainable value from open data in a sharing society. In Bergvall-Kåreborn, B & Nielsen, P.A. (eds.) *Creating Value for All Through IT*, edited by Birgitta and Peter Axel, 243-260. Berlin Heidelberg: Springer.
- 110 Juell-Skielse, G., Hjalmarsson, A., Johannesson, P., & Rudmark, D. (2014). Is the Public Motivated to Engage in Open Data Innovation? *Electronic Government: Lecture Notes in Computer Science Volume*, (8653), 277-288
- 111 Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, 30(4), 508-513.
- 112 Katz, M. L. & Shapiro, C. Network externalities, competition, and compatibility. *The American economic review* (1985), 424-440.
- 113 Katz, M.L. & Shapiro, C. Technology adoption in the presence of network externalities. *The Journal of Political Economy* (1986), 822-841.
- 114 Kettinger, W. J., & Li, Y. (2010). The infological equation extended: towards conceptual clarity in the relationship between data, information and knowledge. *European Journal of Information Systems*, 19(4), 409-421.
- 115 Kim, S. & Lee, J. (2012). E-Participation, Transparency, and Trust in Local Government. *Public Administration Review*, 72(6), 819-828.
- 116 King, J. L., & Uhlir, P. F. (2014). Soft infrastructure challenges to scientific knowledge discovery. *Communications of the ACM*, 57(9), 35-37.
- 117 Kohli, R. & Grover, V. (2008). Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times. *Journal of the Association for Information Systems*, 9(1), 23-39.
- 118 Koski, H. (2011). Does marginal cost pricing of public sector information spur firm growth? Discussion Papers (1260), Helsinki:

- ETLA – The Research Institute of the Finnish Economy
- 119 Koski, H. (2013). ICT Outsourcing, User-Driven and Open Innovation Strategies in the Generation of New Data-Based Solutions. ETLA Working Papers No. 7. Available: <http://pub.etla.fi/ETLA-Working-Papers-7.pdf>
- 120 Kuk, G. & Davies, T. (2011). The Roles of Agency and Artifacts in Assembling Open Data Complementarities. In the proceedings of the 19th International Conference of Information Systems (ICIS), Shanghai, China
- 121 Kulk, S. & Van Loenen, B. (2012). Brave new open data world. *International Journal of Spatial Data Infrastructures Research*, 7(2012),196-206.
- 122 Kundra, V. (2012). Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect. Harvard: Joan Shorenstein Center on the Press, Politics and Public Policy. Available: [http://www.hks.harvard.edu/presspol/publications/papers/discussion\\_papers/d70\\_kundra.pdf](http://www.hks.harvard.edu/presspol/publications/papers/discussion_papers/d70_kundra.pdf)
- 123 Kuznets, S. (1974). *Population, Capital and Growth: Selected Essays*. London: Heinemann Educational.
- 124 Lassinantti, J., Bergvall-Kåreborn, B., & Ståhlbröst, A. (2014). Shaping local open data initiatives: politics and implications. *Journal of theoretical and applied electronic commerce research*, 9(2), 17-33.
- 125 Liew, A. (2007). Understanding data, information, knowledge and their inter-relationships. *Journal of Knowledge Management Practice*, 8(2), 1-16.
- 126 Linders, D. & Wilson, S.C. (2011). What is Open Government? One Year after the Directive, Paper presented at the 12th Annual International Conference on Digital Government Research: Digital Government Innovation in Challenging Times, Bertot, J.C.; Nahon, K.; Chun, S.A.; Luna-Reyes, L.F. & Atluri, V. (eds.), College Park, MD, pp. 262-271.
- 127 Lindman, J., Rossi, M. & Tuunainen, V.K. (2013). Open Data Services: Research Agenda. Paper presented at the 47th Hawaii International Conference on System Sciences (HICSS), 1239-1246.
- 128 Lindman, J., Kinnari, T. & Rossi, M. (2014). Industrial open data: Case studies of early open data entrepreneurs. Paper presented at the 47th Hawaii International Conference on System Science (HICSS), 739-748.
- 129 Lindman, J., & Kuk, G. (2015). From Open Access to Open Data Markets: Increasing the Subtractability of Open Data. Paper presented at the 48th Hawaii International Conference on System Sciences (HICSS), 1306-1313.
- 130 Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science*, 67(1), 1-25.
- 131 MacInnis, D.J., Moorman, C. and Jaworski, B.J. (1991). Enhancing Consumers' Motivation, Ability, and Opportunity to Process Brand Information from Ads: Conceptual Framework and Managerial Implications. *Journal of Marketing*, 55(1991), 32-53.
- 132 Marcoulides, G. A. & Saunders, C. (2006). PLS: A Silver Bullet? *MIS Quarterly*, 30(2), iii-ix.
- 133 Martin, S., Foulonneau, M., Turki, S. & Ihadjadene, M. (2014). Risk Analysis to Overcome Barriers to Open Data. *Electronic Journal of e-Government*, 11(2), 348 -359.
- 134 Mayer-Schönberger & Zappia, Z. (2011). Participation and Power: Intermediaries of Open Data. Conference draft prepared for the Berlin Symposium. Available: <http://tinyurl.com/bo3pyl8>
- 135 Mayntz, R. (2004). Mechanisms in the analysis of social macro-phenomena. *Philosophy of the social sciences*, 34(2), 237-259.
- 136 McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data. The management revolution. *Harvard Business Review*, 90(10), 61-67.

- 137 McKinsey (2013a). Open data: Unlocking innovation and performance with liquid information, McKinsey Global Institute, McKinsey Center for Government and McKinsey Business Technology Office.
- 138 McKinsey (2013b). Disruptive technologies: Advances that will transform life, business, and the global economy. McKinsey Global Institute.
- 139 Meijer, A. and Thaens, M. (2009). Public Information Strategies: Making Government Information Available To Citizens. *Information Polity*, 14(1), 31–45.
- 140 Meijer, A. J., Curtin, D., & Hillebrandt, M. (2012). Open government: connecting vision and voice. *International Review of Administrative Sciences*, 78(1), 10-29.
- 141 Meijer, R., Conradie, P., & Choenni, S. (2014). Reconciling contradictions of open data regarding transparency, privacy, security and trust. *Journal of theoretical and applied electronic commerce research*, 9(3), 32-44.
- 142 Melville, N. (2010). Information Systems Innovation for Environmental Sustainability, *MIS Quarterly*, 34(1), 1-21.
- 143 Merton, R. K. (1942). Note on Science and Democracy. *American Journal of Legal & Political Sociology*, 1, 115.
- 144 Merton, R. K. (1949). *Social Theory and Social Structure*. New York: Free Press.
- 145 Merton, R. K. (1968). *Social theory and social structure (enlarged edition)*. New York: Free Press.
- 146 Michaelson, J., Abdallah, S., Steuer, N., Thompson, S. and Marks, N. (2009). *National Accounts of Well-Being: Bringing Real Wealth onto the Balance Sheet*. London: New Economics Foundation.
- 147 Mingers, J. (2004). Realizing Information Systems: Critical Realism as an Underpinning Philosophy for Information Systems. *Information and Organization*, 14(2), 87-103.
- 148 Moore, M.H. 1995. Creating public value: strategic management in government. Harvard University Press.
- 149 Morgan, L., Feller, J., & Finnegan, P. (2013). Exploring value networks: Theorising the creation & capture of value with open source software. *European journal of information systems*, 22(5), 569-588.
- 150 OECD. (2005). Modernising government: The way forward. [Online]. Available: <http://www.oecd.org/gov/modernisinggovernmenttheforward.htm>
- 151 OECD (2011). Fostering Innovation to Address Social Challenges. Workshop Proceedings. [Online]. Available: <http://www.oecd.org/sti/inno/47861327.pdf>
- 152 OECD. (2014). Data-driven Innovation for Growth and Well-being. Interim Synthesis Report.[Online]. Available: <http://www.oecd.org/sti/inno/data-driven-innovation-interim-synthesis.pdf>
- 153 Open Knowledge Foundation (OKF) Available at <http://www.opendefinition.org/> and <http://opengovernmentdata.org>
- 154 Oxera. (2013, January). What is the economic impact of geo-services. [Online]. Available: [http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services\\_1.pdf](http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services_1.pdf)
- 155 Peled, A. (2011). When transparency and collaboration collide: The USA open data program. *Journal of the American society for information science and technology*, 62(11), 2085-2094.
- 156 Pepper, R and Garrity, J. (2014). The Internet of Everything: How the Network Unleashes the Benefits of Big data. In Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014): The Global information technology report 2014: Rewards & Risks of Big Data. World Economic Forum, 35-42
- 157 Petychakis, M., Vasileiou, O., Georgis, C., Mouzakitis, S., & Psarras, J. (2014). A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective. *Journal of theoretical and applied electronic commerce research*, 9(2), 34-47.

- 158 Pollock, R. (2008). *Economics of PSI*. Cambridge: University of Cambridge. Available at [http://rufuspollock.org/economics/papers/economics\\_of\\_psi.pdf](http://rufuspollock.org/economics/papers/economics_of_psi.pdf)
- 159 Porter, M. E. & Kramer, M.R. (2011). Creating Shared Value. *Harvard Business Review*, 89(1), 62-77.
- 160 Provan, K. G., & Kenis, P. N. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory*, 18(2), 229-252.
- 161 Rabin, M. (1998). Psychology & Economics. *Journal of Economic Literature*, 3(1), 11-46.
- 162 Rechtin, E., & Maier, M. W. (2009). *The art of systems architecting (3rd edition)*. CRC Press.
- 163 Reinholt, M. Pedersen, T. & Foss, N.J. (2011). Why a Central Network Position isn't Enough: The Role of Motivation and Ability for Knowledge Sharing in Employee Networks. *Academy of Management Journal*, 54(6), 1277-1297.
- 164 Relly, J.E. and Sabharwal, M. (2009). Perceptions of Transparency of Government Policymaking: A cross-national study. *Government Information Quarterly*, (26), 148-157.
- 165 Ringle, C.M., Sarstedt, M. & Straub, D.W. (2012). A Critical Look at the Use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, 36(1), iii-xiv.
- 166 Rochet, J.C. & Tirole, J. 2006. Two-sided markets: A progress report. *The RAND Journal of Economics* 37(3), 645-667.
- 167 Rothschild, M.L. (1999). Carrots, Sticks, and Promises: A Conceptual Framework for the Management of Public Health and Social Issue Behaviors. *Journal of Marketing*, 63(4), 24-37.
- 168 Sarker, S.; Sarker, S.; Sahaym, A. & Bjørn-Andersen, N. (2012). Exploring Value Cocreation in ERP Vendor–Partner Relationships. *MIS Quarterly*, 36(1), 317-338.
- 169 Sarstedt, M., Ringle, C. M., Henseler, J., & Hair, J. F. (2014). On the emancipation of PLS-SEM: A commentary on Rigdon (2012). *Long range planning*, 47(3), 154-160.
- 170 Sayer, A. (1992). *Method in Social Science: A Realist Approach* (2nd ed.), London: Routledge.
- 171 Schumpeter, J.A. (1934). *The Theory of Economic Development*, Cambridge: Harvard University Press.
- 172 Sen, A. (1999). *Development as Freedom*, New York: Anchor Books.
- 173 Shapiro, C. & Varian, H.R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press, Boston.
- 174 Shogren, J. (2012). *Behavioural Economics and Environmental Incentives*. OECD Environment Working Papers 49, OECD Publishing
- 175 Simon, H. (1957). *A Behavioral Model of Rational Choice. Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. Wiley, New York
- 176 Smith, M. (2006). Overcoming Theory–Practice Inconsistencies: Critical Realism and Information Systems Research, *Information and Organization*, 16(3), 191-211
- 177 Stiglitz, J.E., Sen, A. & Fitoussi, J. P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress, OECD.
- 178 Stoker, G. (2006). Public value management a new narrative for networked governance?. *The American review of public administration*, 36(1), 41-57.
- 179 Suddaby, R. (2010). Editor's comments: Construct clarity in theories of management and organization. *Academy of Management Review*, 35(3), 346-357.
- 180 Tammisto, Y. & Lindman, J. (2011). Definition of open data services in software business. Paper presented at the third International Conference on Software

- Business (ICSOB 2012), Cambridge, MA, USA. In Cusumano, M.A., Iyer, B., & Venkatraman, N. (eds.), *Software Business Lecture Notes in Business Information Processing*, 114(2012), 297–303,
- 181 Teddlie, C. & Tashakkori, A. (2009). *Foundations of Mixed Methods Research*, Thousand Oaks, CA: Sage Publications
- 182 Tinati, R., Carr, L., Halford, S., & Pope, C. (2012). Exploring the Impact of Adopting Open Data in the UK Government. *Digital Futures 2012*. Available: <http://eprints.soton.ac.uk/id/eprint/344808>
- 183 Trinchera, L. & Russolillo, G. (2010). On the use of Structural Equation Models and PLS Path modeling to build composite indicators. Working paper no. 30, Università Degli Studi Di Macerata
- 184 Tsang, E. W. K. (2006). Behavioral assumptions and theory development: The case of transaction cost economics. *Strategic Management Journal*, 27(11), 999-1011.
- 185 Tsang, E.W.K. (2014). Case studies and generalization in information systems research: A critical realist perspective. *Strategic Management Journal*, 23(2), 174-186.
- 186 Tsoukas, H. (1989). The Validity of Idiographic Research Explanations, *Academy of Management Review*, 14(4), 551-561.
- 187 Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. OECD Working Papers on Public Governance, No. 22, OECD Publishing.
- 188 Van de Ven, A. H. (2007). *Engaged Scholarship - A Guide for Organizational and Social Research*. Oxford:Oxford University Press.
- 189 van Osch, W. & Avital, M. (2010). Generative Collectives. Paper presented at the International Conference on Information Systems (ICIS), Saint Louis, Missouri.
- 190 van Veenstra, A.F. & van den Broek, T.A. (2013). Opening Moves – Drivers, Enablers and Barriers of Open Data in a Semi-public Organization. Paper presented at the 12th IFIP WG 8.5 International Conference, EGOV 2013, Koblenz, Germany, 50-61.
- 191 van Veenstra, A. F., & van den Broek, T. (2014). A Community-driven Open Data Lifecycle Model Based on Literature & Practice. In Boughzala, I., Janssen, M. & Assar, S. (eds.). *Case Studies in e-Government 2.0*, 183-198. Springer International Publishing
- 192 Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, 37(1), 21-54.
- 193 Verhulst, S., Noveck, B. S., Caplan, R., Brown, K. & Paz, C. (2014). The Open Data Era in Health and Social Care. The Governance Lab. [online]. Available: <http://images.thegovlab.org/wordpress/wp-content/uploads/2014/06/nhs-full-report.pdf>
- 194 Vickery, G. (2011). Review of Recent Studies on PSI Re-Use and Related Market Development, Paris: European Commission.
- 195 Viscusi, G., Castelli, M. & Batini, C. (2014). Assessing Social Value in Open Data Initiatives: A Framework. *Future Internet*, 6(3), 498-517.
- 196 von Krogh, G.; Rossi-Lamastra, C. & Hefliger, S. (2012b). Phenomenon-based Research in Management and Organisation Science: When is it Rigorous and Does it Matter. *Long Range Planning*, 45(4), 277-298
- 197 Wade, M. & Hulland, J. (2004). Review: the resource-based view and information systems research: review, extension, and suggestions for future research. *MIS Quarterly*, 28(1), 107-142.
- 198 Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of information systems*, 4(2), 74-81.
- 199 Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), xxiii-xxiii.

- 200 White House. (2009, December). [Online]. Open Government Directive.
- 201 Available: <https://www.whitehouse.gov/open/document/s/open-government-directive>
- 202 White House. (2013, May) Executive order - making open and machine readable the new default for government information. [Online]. Available:
- 203 <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>
- 204 Williams, C.K. and Karahanna, E. (2013). Causal explanation in the coordinating process: A critical realist case study of federated IT governance structures, *MIS Quarterly*, 37(3), 933-964.
- 205 Wynn Jr. & Williams, C.K. (2012). Principles for Conducting critical realist case study research in information systems. *MIS Quarterly*, 36(3), 787-810.
- 206 Xu, J., Benbasat, I. & Cenfetelli, R. T. (2014). The nature and consequences of trade-off transparency in the context of recommendation agents. *MIS Quarterly*, 38(2), 379-406.
- 207 Yin, R.K. (2009). *Case Study Research: Design and Methods* (4th ed.). Los Angeles: Sage.
- 208 York, J.G., Sarasvathy, S.D. & Wicks, A.C. (2013). An Entrepreneurial Perspective on Value Creation in Public-Private Ventures. *Academy of Management Review*, 38(2), 307-315.
- 209 Yu, H. & Robinson, D.G. (2012, February). The new ambiguity of 'open government'. 59 *UCLA L. Rev. Disc.* 178. Available: <http://ssrn.com/abstract=2012489>
- 210 Zachariadis, M., Scott, S. & Barrett, M. (2013). Methodological Implications of Critical Realism for Mixed-Methods Research. *MIS Quarterly*, 37(3), 855-880.
- 211 Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R. & Alibaks, R.S. (2012). Socio-technical Impediments of Open Data. *Electronic Journal of e-Government*, 10(2), 156 – 172.
- 212 Zuiderwijk, A. & Janssen, M. (2013). A coordination theory perspective to improve the use of open data in policy-making. *Electronic Government, Lecture Notes in Computer Science (8074)*, 38-49.
- 213 Zuiderwijk, A., & Janssen, M. (2014a). The negative effects of open government data-investigating the dark side of open data. Paper presented at the *15th Annual International Conference on Digital Government Research*, Aguascalientes, Mexico. 147-152. ACM.
- 214 Zuiderwijk, A. & Janssen, M. (2014b). Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View. In Gascó-Hernández, M. (ed). *Open Government, Opportunities and Challenges for Public Governance*. New York: Springer.
- 215 Zuiderwijk, A., & Janssen, M. (2014c). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17-29.
- 216 Zuiderwijk, A., Helbig, N. Gil-Garcia, J. R. & Janssen, M. (2014a). Special Issue on Innovation through Open Data: Guest Editors' Introduction. *Journal of theoretical and applied electronic commerce*, 9(2), 1-13.
- 217 Zuiderwijk, A., Janssen, M., Choenni, S. & Meijer, R. (2014b). Design principles for improving the process of publishing open data. *Transforming Government: People, Process and Policy*, 8(2), 185-204.



## Appendix A: Summary of the Included Papers

This PhD study is based on the following articles, as shown in figure 18:

Methods	Focus	Open data supply: Policy and dissemination	Open data use: Business models and evaluation	Open data value generation
Theoretical (Models and Frameworks with use cases)		The Value of Open Government Data: A Strategic Analysis Framework	Innovation in the Open Data Ecosystem: Exploring the role of real options thinking and multi-sided platforms	The Value Generating Mechanisms of Open Government Data First version of conceptual model Paper II – ECIS
		First version of framework Paper I – pre-ICIS eGov SIG workshop	Theoretical paper on the role of intermediaries Paper VIII - Book chapter (forthcoming)	Generating Sustainable Value from Open Data in a Sharing Society Micro foundations, assumptions Paper V – IFIP
Empirical (Case study or quantitative data analysis)		Managing Complexity across Multiple Dimensions of Open Government Data: The Case of the Danish Basic Data Program	Data-Driven Innovation through Open Government Data	Generating Value from Open Government Data PLS analysis of 63 countries in 2012 Paper III – ICIS
		Case study analyzing strategy to dissemination Paper VI – Government information Quarterly (under review)	Case study analyzing how data are used for driving change Paper IV - JTAER	Driving Sustainable Value: A Conceptual Model of Open Data as a Resource PLS analysis of 76 countries in 2013 Paper VII - #NA Journal

Figure 18: Paper Overview

Table 5 gives a more comprehensive description of each paper and contribution towards overall study

**Table 5: An Overview over Included Papers; Publication Outlets, Main Contributions and Co-Authors**

#	Full name and publication	Brief description	Contribution towards overall study	Co-authors
I	The Value of Open Government Data: A Strategic Analysis Framework Paper presented in the 2012 Pre-ICIS e-Government Workshop, 2012, Orlando, U.S.	A theoretical framework of four value drivers of open data, based on preliminary qualitative data, use cases and literature review.	Paper I discusses the importance of open data <i>strategy and goal setting</i> . The paper also identifies the four most prominent value drivers that are frequently cited as goals of ODIs. The paper also discussed what type of value (social and economic) is expected from each of them. Finally, the paper proposes that <i>evaluation</i> of ODIs is highly important, as it influences strategy and affects the financial sustainability of ODIs.	Michel Avital Niels Bjørn-Andersen
II	The Generative Mechanisms of Open Government Data In Proceedings of the 21st European Conference on Information Systems (ECIS), 2013, Utrecht, The Netherlands	A conceptual model illustrating the nomological network consisting of open data, societal level enabling factors, four value generating mechanisms and value. Based on preliminary qualitative data, use cases and	Paper II discusses unique (infrastructural) features of open data and how they influence value generation. The article utilizes narratives from use cases to give examples of <i>how use of open data contributes to generate social and economic value</i> through four distinct <i>value generating mechanisms</i> . Paper II also discusses the <i>contextual factors</i> that	Michel Avital Niels Bjørn-Andersen

#	Full name and publication	Brief description	Contribution towards overall study	Co-authors
III	Generating Value from Open Government Data In Proceedings of the 34th International Conference on Information Systems (ICIS), 2013, Milan, Italy	literature review. The first (pilot) PLS study, which extends and tests the conceptual model presented in paper II with secondary country-level data for 2012.	are important to value generation. Paper III extends Paper II by empirically testing the hypotheses proposed in the conceptual model. Due to the challenge of measuring different theoretical constructs related to open data value generation, this study called for more detailed descriptions and definitions of constructs, which contributed back to the qualitative studies. The paper introduced a first version of a value construct and a first version of the open data construct.	Michel Avital Niels Bjørn-Andersen
IV	Data-driven Innovation through Open Government Data Journal of Theoretical and Applied Electronic Commerce Research 9(2), pp. 100-120 (2014)	A case study of the company Opower, focusing in particular on use of data for innovation purposes. This study is multilevel in nature, and discusses the micro-macro mechanisms and the informal interactions between public and	Paper IV focuses on the demand side, particularly on which aspects or dimensions of <i>open data</i> are most important for private sector users ( <i>engagement</i> ). The case discusses use of open data in the specific case of Opower. The study analyses the interaction between levels; the importance of certain societal and organizational level enabling factors for a private company that uses	Michel Avital Niels Bjørn-Andersen

#	Full name and publication	Brief description	Contribution towards overall study	Co-authors
V	<p>Generating Sustainable Value from Open Data in a Sharing Society Presented at IFIP WG 8.6 in Aalborg, 2014</p> <p>In Creating Value for All Through IT, edited by Birgitta Bergvall-Kårebom and Peter Axel Nielsen, pp. 243–260. Berlin Heidelberg: Springer</p>	<p>private stakeholders.</p> <p>A theory development paper that delves deeper into the previously mentioned micro-macro level relationships. The overarching focus is on how private and public sectors can collaboratively address societal challenges by using open data as a “tool”.</p>	<p>open data for creating societal level value. Moreover, the article touches upon how recognition of this value (<i>evaluation</i>) will influence government open data <i>strategy and goal setting</i>.</p> <p>Paper V discusses how societal challenges defy current structures and boundaries. The article proposes that both public and private stakeholders must act collectively in order to address these challenges. It is suggested that open data can facilitate boundary crossing without demanding rigorous cross-boundary governance. Paper V also discusses the importance of creating the right <i>societal level context</i> that will provide individuals with the motivation, opportunity and ability to generate sustainable value from open data. This paper is the first to conceptualize <i>Sustainable value</i>.</p>	<p>Michel Avital Niels Bjørn-Andersen</p>

#	Full name and publication	Brief description	Contribution towards overall study	Co-authors
VI	<p>Managing Complexity across Multiple Dimensions of Open Government Data</p> <p>Currently under review in Government Information Quarterly. (corresponding editor Marijn Janssen)</p>	<p>A case study of the Danish Basic Data Program (BDP). The relationships between open data value drivers/evaluation, the ODI strategy process and the actual ODI planning and deployment are explored in-depth in this study.</p>	<p>Paper VI focuses on the supply-side of open data. Discusses relevant <i>Governance mechanisms</i>, which emerged from the qualitative interviews and observations as being one of the most important factors for successful dissemination of liquid open data. The article identifies four tensions relevant to this specific ODI. I furthermore suggest how ODIs can actually attend to competing demands simultaneously, and consequently create synergies from these tensions. This paper is the first to conceptualize <i>Liquid open data</i>, which is later used as a construct in the quantitative study.</p>	Single authored

#	Full name and publication	Brief description	Contribution towards overall study	Co-authors
VII	<p>Driving Sustainable Value: A Conceptual Model of Open Data as a Resource</p> <p>This is an unpublished and un-edited manuscript which we intend for a journal.</p>	<p>The second PLS study where we used secondary country-level data from 2013 to test the final version of conceptual model.</p>	<p>Paper VII introduces the final conceptual model and empirically tests the proposed relationships between liquid open data, four societal level contextual factors, manifestations of the event of using open data for value generation and sustainable value. Paper VII is the first to define the <i>Information sharing mechanism</i> as an alternative to the <i>Market mechanisms</i>. We furthermore suggest that we can measure these separately by looking for shared digital content on the one hand and the extent to which society provides innovative digital products and services on the other hand. Suggests how we can measure <i>Liquid open data</i> and <i>Sustainable value</i>.</p>	<p>Michel Avital Niels Bjørn-Andersen</p>

#	Full name and publication	Brief description	Contribution towards overall study	Co-authors
VIII	<p>Innovation in the Open Data Ecosystem: Exploring the role of real options thinking and multi-sided platforms for sustainable value generation through open data</p> <p>In Analytics, Innovation and Excellence Driven Enterprise Sustainability, edited by Rick Edgeman, Elias G. Carayannis and Stavros Sindakis, Palgrave Studies in Democracy, Innovation and Entrepreneurship for Growth. (Forthcoming in 2015).</p>	<p>A book chapter exploring the role of two sided markets to stimulate synergies between information sharing and market type mechanisms. Suggest use of real options methods for evaluation purposes, to stimulate interest and investment in ODIs.</p>	<p>Paper VIII is an exploratory theory development paper that discusses the importance of intermediaries for facilitating conversion of open data to shared digital content and commercial digital products and services. Discusses how Multi Sided Platforms (MSPs) can create synergies between the two types of value generating mechanisms identified in paper VII for superior sustainable value generation. The paper proposes that we currently experience an <i>Open Data Value Paradox</i>. To boost investment in open data, we need to understand the most important features that influence generation of value and how we can measure this value (<i>evaluation mechanisms</i>). Otherwise, ODIs might never be able to justify continuing investment in open data platforms. I suggest the use of real options valuation methods to accommodate the infrastructural features of open data.</p>	Single authored

# **Paper I: The Value of Open Government Data: A Strategic Analysis Framework**

Paper presented in the 2012 Pre-ICIS e-Government Workshop, 2012, Orlando, U.S.



# **The Value of Open Government Data: A Strategic Analysis Framework**

*Research-in-Progress*

## **Abstract**

Government data has been accumulated for centuries in protected repositories and registries as public record and a matter of civil order. Recently, the Open Government Data (OGD) movement has emerged as a group that focuses on facilitating open access to government data. Proponents of OGD initiatives argue that it can strengthen democracy and improve government through increased participation, collaboration and transparency. OGD advocates are also motivated by its potential contribution to greater productivity and economic growth through increased government efficiency and the creation of new businesses and services. However, as most OGD initiatives are relatively recent, the key questions regarding the value propositions and return on investment of these initiatives remain unanswered. In this theory development paper, we offer a strategic options framework that offers criteria for generating and prioritizing OGD initiatives. The framework can guide structured analysis of the economic and social impacts of OGD with an emphasis on its value propositions at both the public and private sectors. Building on a literature review and fieldwork-based anecdotal evidence, we expect OGD initiatives to be able to generate value through either increased transparency, efficiency of government activities, citizen participation or entrepreneurship, depending on different enablers and barriers, mechanisms and actors. We conclude that OGD initiatives have a potential to generate a substitutive return on investment.

**Keywords:** Open government data, Public sector information, open data value assessment, Open access.

## I.1 Introduction

Data is the fuel of the information age revolution. Nowadays, we generate and store more data than any other time in history. Computing and networking capabilities combined with openness enhance the potential impact of the accumulated data and offer society an opportunity to drive massive social, political and economic change (Kundra, 2012). The public sector and especially the various government branches are one of the main sources of data. Government data has traditionally been accumulated in protected repositories and registries as public record and a matter of civil order. Consequently, over centuries, archived government data has had limited access via proprietary interfaces and often cumbersome fee-based procedures. In response to the Budapest Open Access Initiative (2002) and the Open Access movement (see Suber, 2007), governments have started to look into the prospects of providing open access to their data repositories. The Open Government Data (OGD) movement has emerged as a distinct group that focuses on facilitating open access to government data in consideration with its unique characteristics, political significance, and legal stature.

Clearly, OGD has far-reaching effects that go beyond direct economic gain. OGD implies that the public sector relinquishes its role as information gatekeeper in lieu of a new role as information publisher. Thereby, OGD involves a realignment of the power dynamics between the public-private sectors (Davies, 2010). Proponents of OGD hope that such shifts will readjust the power balance between government and citizenry and subsequently strengthen democracy and improve government work through increased participation, collaboration and transparency. OGD advocates are also motivated by its potential for innovative entrepreneurs, who can use the open data to propel economic growth as well as to address social challenges (Gigler et al., 2011). Moreover, advocates of OGD argue that it enables greater government efficiency through an information infrastructure that allows for better data re-use within the public sectors and inter-agency coordination.

The impact and ramifications of providing open access to government data, let alone its value propositions are, however, still debated in professional and academic circles. Huijboom and Van den Broek (2011) argue that the precise economic impact of open-data policies remains largely unclear and that calculations differ substantially. According to Uhlir (2009), there are relatively little empirical data available on the effects of the various policy approaches used when opening up data. This state of affairs leaves policy makers and information managers without the facts they need to assess and improve these policies (Uhlir, 2009). Halonen (2012) finds it likely that the

release of government data has the potential to lead to significantly enhanced efficiency in the long run, but he points out that in the UK certain obstacles have emerged, namely the operability of current IT systems, the lack of context in data, data literacy and negative attitudes among public officials (Halonen, 2012). Other possible reasons for less than expected gains are lack of technological knowledge in the public sector and a lack in recognition of the value of data (Halonen, 2012). There is also a need for a common understanding of the concepts and terms used in the OGD discourse. Questions like what is OGD, why should data be open, what is the value of OGD, how can it be evaluated and captured and what are the real-life enablers and inhibitors that governments face, need to be answered.

The relatively short experience with OGD initiatives, scarcity of case studies and evidence-based research on the topic, let alone the complexity of the underlying issues, have led us to develop a framework that addresses the following core questions:

- How to address evaluating the effect of open government data on government work?
- How to address evaluating the effect of open government data on the general public welfare?

The following theory development paper aims to provide a strategic options framework that offers criteria for generating and prioritizing OGD initiatives. The framework is meant to aid the design of the OGD value proposition and how it is positioned and delivered, as well as to enable a structured analysis of the potential wide-range effects of OGD. Building on the OGD literature and fieldwork-based anecdotal illustrations, we identify the main mechanisms that drive value from opening government data. Based on the evidence from current literature, we conclude that it is very likely that OGD initiatives generate economic value, i.e. that the benefits are higher than the costs of such initiatives. According to Emerson et al. (2001), social value is created when resources, inputs, processes or policies are combined to generate improvements in the lives of individuals or society as a whole. This could mean cleaner air, increased life expectancy, more equality and inclusiveness or more trust and knowledge. Social value is much harder to measure than economic (financial) value and takes longer to become evident. However, current evidence supports that OGD has the potential to induce social benefits derived from government due process and increased citizenship behavior.

Next, we review the context and nature of OGD, afterwards we describe the proposed framework and finally we conclude with a discussion of the implications to research and practice.

## **I.2 Open data and government**

Open Government Data (OGD) is one specific, but very important, part of open data. A definition of government data can be found on the website [opengovernmentdata.org](http://opengovernmentdata.org): “By government data we mean data and information produced or commissioned by government or government controlled entities” (OKF, 2012). But what is the meaning of Open in OGD? According to [OpenDefinition.org](http://OpenDefinition.org), to be considered open, data must be accessible online, published without technical restrictions to re-use, and provided under a license that allows the data to be re-used without limitation, including across different "fields of endeavor" (e.g. commercial and non-commercial alike) (OKF, 2012).

One strand of the OGD discussion focuses on the economic aspects of use and re-use of government data. The general economic idea of openness is to stimulate economic activity by increasing the use and re-use of this valuable resource, which has historically been subject to multiple barriers to use (Houghton, 2011; Kundra, 2012; Vickery, 2011). Being free for use is of course not the same as being free of use (Clark, 2006), however it is common in the economic discussion to interpret the concept of openness as meaning that the data is/should be free of charge. This results from research showing that in the cases where governments charge nothing (or at most the marginal cost of dissemination of electronic data which is near zero), re-use, and consequently economic activity, is markedly increased (de Vries et al., 2011; Koski, 2011, Pollock, 2008). The use of the concept and idea of openness in the economics domain is in this sense different from the use of the same concept in the open source software discourse. The Free Open Source Software (FOSS) movement has for instance clearly differentiated between the concepts of open and free.

In the last few years, OGD has been associated more prominently with the more social aspects of the Free Open Source Software (FOSS) culture of sharing and collaboration. Perhaps more specifically, this trend has arrived with the rise of the Open Government ideology, where public participation, collaboration, interagency and cross-sector partnerships have become part of the discussion (Linders and Wilson, 2011). The idea of Open Government draws in part on the philosophy and methods of the open source movement and means government where citizens not only have access to information, documents, and proceedings, but can also become participants in a meaningful way

(Harrison et al., 2011). It is important to notice, that in both the economic and social aspects of the discussion, opening government data is seen as a means to an end, not an end in itself. The importance of openness comes from the extent that it helps us achieve other goals, not because it is valuable in and of itself.

A number of economic studies on OGD have tried to estimate market size, whether by estimating the total turnover of the information industry (PIRA, 2000), the “use value” or added-value from use of the information (Dekkers et al., 2006) or exchange value in the meaning that the information is worth what buyers are willing to pay for it (DECA, 2010). Vickery (2011) made a survey on existing findings on the economic impact of OGD and estimated that the overall direct and indirect economic gains could be around EUR 140 billion throughout the EU. In general, the consensus of multiple studies on the value of OGD is that the economic value seems to substantially outweigh the costs of collection and dissemination (Uhlir, 2009; Houghton, 2011; PIRA, 2000; Weiss, 2001). But when the Open Government/FOSS movement kind of collaboratively created social value is considered as well, the estimation of value becomes more complicated. The public value perspective, introduced by Moore (1995), describes public value as the product of governmentally-produced benefits, part of which is derived from the direct usefulness of such benefits and part that is derived from the fairness and equitability of their production and distribution. A simplification of this perspective, often seen in the OGD discourse, is that value generated by government actions can either be of an economic nature (more economic activity, increased productivity) or social nature (better democracy, less corruption, happier and healthier citizens).

### **I.3 Strategic analysis framework for OGD initiatives**

After going through the literature on OGD we saw that the value generating mechanisms set forth by OGD initiatives resemble those set forth by a value network. In a value network, value is co-created or co-produced (Morgan et al., 2010). Creating value cannot be done unilaterally based on the efforts of a single organization, nor can it be done without keeping in mind the different and divergent interests of all collaborating partners (Vanhaverbeke, 2008). Verna Allee defines value networks as any web of relationships that generates both tangible and intangible value through complex dynamic exchanges between two or more individuals, groups or organizations (Allee, 2008).

In the case of OGD, different initiatives have the ability to create value of both social and economic nature for both the private and public sector. However, when we looked

at these mechanisms from the practical implementation oriented point of view, we could see important differences. Using the two-by-two matrix to show the main strategic options for government, we simplified the value generating mechanisms into value driven primarily by the actions of the public sector on one hand and the private sector on the other. The resulting value can be both social and economic, but some value propositions are more geared towards the social types of value while other deliver proportionally more economic value. This classification resulted in four drivers of value, illustrated in Figure 1.

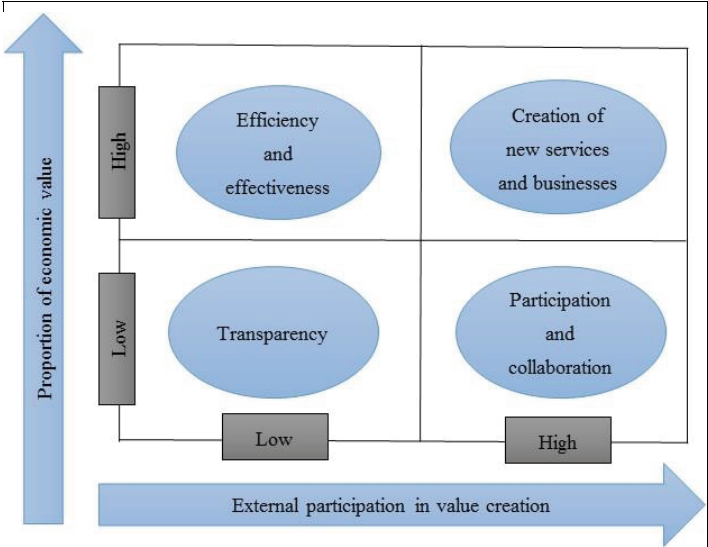
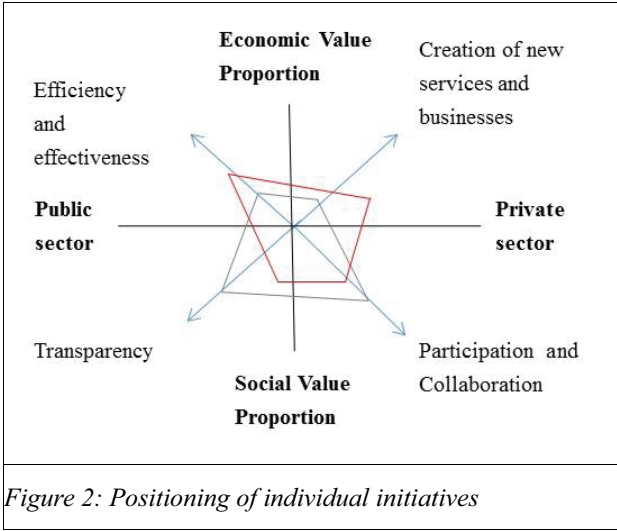


Figure 1: Strategic options for OGD initiatives

The matrix explores two key dimensions: Sector, where the involvement of the private sector is bigger in the right hand column and Type of Value, where the proportion of economic value generated is bigger in the top row. In each of the quadrants we have a different value proposition, namely Transparency, Participation and Collaboration, Public sector efficiency and effectiveness and Creation of new businesses and services. Different philosophies or ideologies driving OGD initiatives can be illustrated in the rows of the framework. Open Government ideology types of initiatives are more focused on the social types of value (bottom row) while other initiatives (EU) focus more on the ability of open data to increase efficiency and drive economic growth (top row).

Each of these value drivers is enabled by OGD but when examined more thoroughly there are different implementation considerations to each of them. Different levels and types of investment in processes and technology are needed as well as different data sets, licenses and even business models. In each case there is a direct and indirect cost, that can incur only (or mostly) in the public sector (left hand column) or be shared with the private sector (right hand column). Therefore, we conclude that the return on investment is dependent on how well these value generating mechanisms are understood. OGD initiatives can have the ambition to implement more than one value driver and it can be argued that the most interesting synergies occur on the margins. There can also be spillovers between the squares, for instance greater transparency can lead to increased effectiveness. Figure 2 shows how initiatives can be designed using the framework.



Bigger diamond means more total value created but also bigger investment and more need for collaboration. The balance between public and private involvement is shown in the horizontal axis and the balance between economic and social value created on the vertical axis.

In the next section, we discuss each of the four value propositions and provide arguments for why and how each of them is considered to be able to generate value from opening up government data.

## *Transparency*

“Sunlight is said to be the best of disinfectants; electric light the most efficient policeman.” (Brandeis, 1914, pp.xx).

Our first identified value proposition is *Transparency*. Transparency represents an action by the *public sector* that drives *social value*. The relationships between information, transparency, and democracy are fundamental and basic (Harrison et al., 2011). But why and how does transparency drive value? Transparency provides citizens and other stakeholders with a window into what government is doing. Open data enables better government through transparency of government activities and processes that encourage due process and fairness. In economic terms, increased transparency means less information asymmetry. Asymmetry of information can lead to adverse selection and moral hazard (Cook, 2010) resulting in corruption, defined as the misuse of public power for private benefits. The Open Budget Index found in 2008 that 80% of the world's governments fail to provide adequate information for the public to hold them accountable for managing their money: Nearly 50 percent of 85 countries provided such minimal information that they were able to hide unpopular, wasteful, and corrupt spending (Fioretti, 2011). Transparency is also valuable for the public sector itself as transparency can create trust in public operations. In breaking down information silos between agencies, government officials can also consume information from other parts of the bureaucracy to benefit their work (Gigler et al, 2011).

However, it is not obvious that any OGD initiative automatically leads to increased transparency, at least not one that is valuable for everyone. Here we have to consider the context of the project in question. Is the data current, is it of high quality, is it secure, and is it available and accessible for all? Data collection, management, access, and dissemination practices all have strong effects on the extent to which datasets are valid, sufficient, or appropriate for policy analysis or any other use (Dawes and Pardo, 2006). Data literacy and skills of individual groups of citizens and their access to technology should also be considered. Benjamin et al. (2007) studied the Bhoomi program in Bangalore and found out that the digitization of land records led to increased corruption, much more bribes and substantially increased time taken for land transactions. And it eventually enabled very large players in the land markets to capture vast quantities of land (Benjamin et al., 2007). Malensky et al. (2011) point out that transparency can cause perverse effects in systems where agents (politicians) understand the relationship between behavior and outcome better than their principals



(the voters). In general, in order to achieve social value through transparency, equal access to information, equal opportunities for use and the context and quality of data become the prerequisites.

### ***Participation and Collaboration***

The US Open Government Directive, issued on December 8, 2009 foregrounded the principles of transparency, participation, and collaboration as the cornerstone of an open government (see for instance Harrison et al., 2011; Linders and Wilson, 2011; Noveck, 2009). Our second value proposition is Participation and Collaboration. This driver represents a set of actions by the private sector that drives social value and is representative of the latter two principles of open government. Participation, according to Lee and Kwak (2011), refers to public engagement in relatively simple interactive communications such as blogging and social networking and relies primarily on expressive social media to connect people and help share their ideas. Noveck (2009) argues that collaboration is “a form of democratic participation” that differs in important ways from traditional participative and deliberative practices, which often take place in circumstances disconnected from decision making. This driver includes both types of participation and describes the ability of citizens to help governments with difficult decisions and even workload.

Open data and use of information technology enables increased citizen participation and collaboration, leading to improved citizenship and collaborative behavior through crowdsourcing activities. In this case, OGD not only transforms how services are delivered, but opens the opportunity for citizens to control those services. A good example of citizen collaboration is the crowdsourcing activities that have been immensely helpful in natural disaster incidents, such as hurricane Katrina and the earthquake in Haiti. While government agencies and formal organizations failed to respond quickly, open collaboration among the public demonstrated it as a viable and effective mechanism to respond to those daunting challenges (Lee and Kwak, 2011). Another example is the Web and SMS-accessible platform called the Public Participation Information System (LAPOR), launched by the Government of Indonesia in 2011. The new unit lets citizens monitor and verify the delivery of government services in real time. It also uses this information to improve the way it allocates public resources in areas ranging from education and health to energy and defense (McKinsey, 2012).

Participation and collaboration must, however, be meaningful and directed toward goals that are carefully defined and acknowledged by ample government feedback. Further, the citizen input generated must be represented in outcomes that are visible to stakeholders in the decisions and the value produced (Harrison et al., 2012). These kinds of changes are not easily made; they call for considerable change of processes and even mindset within the public sector. They also demand investment in technology and moreover, deliberative design of collaboration platforms, including both the community and society dimensions (de Cindio, 2012).

### ***Public sector efficiency and effectiveness***

The third value proposition is Public sector efficiency and effectiveness. This proposition represents an action by the public sector to create economic value. OGD is, in this context, strongly related to digital or e-government activities where the goal is to modernize and streamline government with the help of information technologies. By opening government data, efficiency can be increased through consolidation of overlapping repositories, improved information infrastructure, inter-agency coordination and better financial controls. One example of such an initiative is the Danish “Better Access to Public Data” free-of-charge access to address data agreement from 2002. The aim of the agreement was to improve public and private services and to promote public safety (ambulance, police and other emergency services) by using the official addresses as a common reference which could promote interoperability in different IT systems. In 2010 a study on the benefits of the agreement concluded that the direct financial benefits in the period 2005-2009 amounted to around EUR 62 million. Until 2009 the total costs of the agreement were around EUR 2 million (DECA, 2010). The Danish government is now running a similar, but broader based, initiative where over the next four years all basic government data will be improved in quality and context and collection and dissemination of the data will be coordinated within the public sector. A common infrastructure will be established for stable and efficient distribution of government data, with the aim to make the administration of the basic data registers easier and more efficient (Digitaliseringsstyrelsen, 2012). At the same time the data will be opened, so that it will be free and available for the private, as well as the public, sector. The estimate of the project leaders is that when the project has been fully implemented (from 2020) the annual savings to the public sector will be around EUR 35 mio (Digitaliseringsstyrelsen, 2012).

A special effort is required in order to ensure that opening data leads to increased efficiency. Schematic heterogeneity and lack of consistency complicate access and integration of the data. Adoption of standards for the documentation, organization and dissemination of information is an important part of government systems for keeping and managing data (Bountouri et al., 2010). The key information architecture principles include treating data as an asset through a value, cost and risk lens and thereby ensuring timeliness, quality and accuracy of the data. Finally, the security of information must be considered, a holistic approach to data governance begins with an understanding of the information life cycle—the collection, updating, processing, and eventual deletion of personal information—and the adoption of a technology framework that enables governments to set controls which safeguard individuals' privacy (Lampri, 2012).

### ***Creation of new businesses and services***

The last value proposition identified is the creation of new businesses and services. This proposition represents a set of actions by the private sector that generates economic value. Generally this means that organizations outside of the public sector use OGD to create new services (private sector innovation) ultimately leading to economic growth. The 2009 Digital Britain Report described data as 'an innovation currency' and 'the lifeblood of the knowledge economy (Department for Culture, Media and Sport and Department for Business, Innovation and Skill, 2009). Open data is an essential raw material for a wide range of new information products and services that build on new possibilities to analyze and visualize data from different sources (European Commission, 2011).

And the opportunity is there as large part of this innovation currency has already been produced, collected and paid for by governments. Since 2003, the Spanish Oficina del Catastro (the Spanish Cadastre/Land Registry) has put increasing amounts of geographical data online and, from 2010, has facilitated electronic land registry certification. From June 2004, free access to cadastral maps for non-commercial users was provided and in April 2011 free access was also extended to commercial re-users, and a new model allowed mass downloads. Since obtaining free access in 2011, the number of private companies downloading data increased 15 fold; alphanumeric data download volume per week increased 20 fold; total digital map downloads increased by a factor of 80

and downloads increased by 100 fold (De Vries, 2012, Koski, 2011). And Open Data can create business opportunities even when not all potential customers or beneficiaries have internet access. Question Box, a mobile phone-based tool developed with support from the Grameen Foundation, allows Ugandans to call or message operators who have access to a database full of information on health, agriculture and education (Fioretti, 2011).

The networked value creation is demonstrated clearly in many OGD based innovations. Collaboration between students, the creative industry, local government, and inhabitants was used to stimulate idea exchange and foster innovation in an OGD initiative in Rotterdam (Conradie et al., 2012). The conclusion after this project was that such an approach, where crucial partners collaborate together, can create a sustainable infrastructure to co-create public services and fosters further innovation based on OGD. Another good example of possible OGD based innovations is the TWC LOGD Portal, an open source infrastructure supporting government data conversion, publishing, enhancement and access. A team of graduate and undergraduate students have used this infrastructure to create over 40 different mashups and visualizations (Ding et al., 2011). These mashups are diverse, some demonstrating the integration of data from multiple sources or deploying data via web and mobile interfaces, others showing how open data can support interactive analysis for specific domains including health, policy and financial data and yet others showing the design of data access and semantic data integration tools.

Making data available and making it re-usable are, however, two very different things (Alani et al., 2007). A big part of the economic value generation possibilities depend on the ability to mash up different sets of data to gain new insights and knowledge, for instance by linking sensor data, government data and company data. One enabler of such activities is the linked data and Semantic Web technology. A lot of promising work is being done on showing how these technologies can solve the need for integrated and interconnected datasets (Böhm et al.; 2012, Ding et al., 2011; Hausenblas, 2011; Höchtl and Reichstädter, 2011; Zuiderwijk et al., 2012). Platforms designed to make use of and work with big, connected datasets for use in various applications are also an enabling technology. Grid and service-oriented high performance systems can be used as an effective cyber infrastructure for implementing and deploying geographically-distributed services and applications (Talia and Trunfio, 2010). The value of OGD can also be discovered through statistical, visual or semantic

models, designed to deliver new knowledge. Parallel to the increased access and coherency to government data, we are witnessing a revolution in the technologies for analyzing, exploiting and processing data. However, the results of a data mining process depend strongly on the quality of the data it processes (Paulheim and Fümkrantz, 2012), which again strengthens the argument for solid data governance.

## **I.4 Discussion**

Government data is a rich source of valuable information and currently an untapped resource as the data is frequently locked up within the public sector, even only usable at the institution that creates or collects the data. It can be seen from the examples above that OGD is capable of generating both economic and social value. The value generated can range from being of a social nature like increased trust in government activities to financial benefits like those resulting from increased economic growth. We identified four distinct, but complementary, value generating mechanisms. Each of them has its own enablers and inhibitors as well as technical and organizational requirements. The value generation is initiated by different sectors in the value network, but these value generating mechanisms are complex and there can be substantial synergies and spillovers between the drivers. In all cases the value can be captured by both the public and the private sector and the relationship between who generates and who captures the value is not a simple one. We conclude that there is need for a new approach to explain the generation, capture and measurement of OGD value, based on complex co-creation of social and economic value in value networks, rather than the value of ownership and exchange. The understanding of the complex value generating mechanisms as well as the possible inhibitors and enablers is imperative to the ability to maximize value captured from OGD.

One possible inhibitor to OGD value generation is a general lack of awareness of the value of government data, both by public officials and the private sector in general. And the opportunities to capture value from OGD might also differ within the private sector, even if the data is made available and free, as some citizens don't have the required skills to capitalize on the opportunities brought forth by OGD (Gurnstein, 2010; Rath, 2012). However, while the digital divide is a fact, this should not be seen as a reason not to publish government data, just as illiteracy should not be conveyed a reason not to publish books. Rather, governments should consider these possible adverse effects and optimally implement some measures to counter-affect them. One way of doing so is to encourage crowdsourcing activities to help increase computer and

data literacy and to educate and increase the awareness of public officials and the general public.

The quality of government data is another possible inhibitor to value generation, even to the extent that data is not available in a digital format (Hogge, 2011). It is also important to give the correct context to the data. Government data is in many cases collected or created for specific purposes, creating substantial risks for validity, relevance, and trust if taken out of that context. In spite of that, information is often seen as a given, used uncritically, and trusted without examination (Dawes, 2012). Dawes (2012) suggests that we are more likely to achieve the promised benefits of OGD if we look at government data as one of four linked phenomena - policy, management, technology, and data - embedded in social, organizational, and institutional contexts that have substantial influences on data quality, availability, and usability. Many of the tools and techniques that are being developed today can make a huge difference for the collection, management and dissemination of government data. New technologies also offer new ways of capturing value from OGD, via the semantic web, mobile apps, network platforms and big data connectors and analytics. Use of information technology can be the most important enabler to enhance the value of OGD initiatives.

Two of the most discussed enablers for OGD success are open licenses and low prices. Still, the distribution of costs and value between the public and private sectors is the cause of some controversy. The benefit from lower prices is increased use and re-use of data, leading to increased economic activity (Pollock, 2008; Koski, 2011). But observers like Bates (2012) worry that the OGD model is essentially becoming little more than a corporate subsidy. The question remains: "Is OGD being made available for private gain at public expense?" (Grupe, 1995). Even if many OGD initiatives should be able to create positive financial return on investment to the public sector, this is not by definition true of every OGD initiative, especially those focused on creating social value. Furthermore, administrations that first see the extra money generated by OGD are almost never the same who created them in the first place (Fioretti, 2011). And finally, improved quality and governance of government data, as well as the use of new technologies, can significantly boost OGD generated value; however, these improvements call for public sector investment and therefore increased costs. We suggest that new business models based on cost sharing, as well as co-creation of value between the public and private sector, could help solve this paradox.

## **I.5 Conclusion**

The discussion above brings us to our two questions:

- How to address evaluating the effect of open government data on government work?
- How to address evaluating the effect of open government data on the general public welfare?

Many empirical research papers have shown evidence towards the potential benefit of opening government data. To name one example, Houghton (2011) estimates the benefit/cost ratio from opening up Australian spatial data as 13/1. Of course the value generated will be different between datasets, but datasets that do not offer much economic value might have a lot of potential for generating social value. However, in order to safeguard the ability of government to collect high-quality, high-value data, new business models where costs as well as benefits are shared between the public and private sectors should be explored. From the perspective of the private sector the value captured from OGD could be of an economic nature like direct cost savings (cheaper services), indirect cost savings (saving time through better services) or increased opportunity to generate revenue (new businesses). The value captured can also be of a social nature like increased trust, more equal and fair society and increased life expectancy. From the public value perspective (Moore, 1995), the value of OGD is also derived from the fairness and equitability of the economic benefits. Eventually, more market activity and increased market efficiency should benefit everyone to a point, at least in democratic societies. And increased transparency, participation and collaboration through citizenship and due procedure could be the mechanisms needed to ensure that opportunities and value are distributed more evenly.

From the evidence we have collected we conclude that:

- OGD initiatives are likely to create both economic and social value and for certain datasets the direct financial benefits will substantially outweigh the costs;
- OGD initiatives need investment in data management and technology, but the type and level of investment, as well as the implementation approach, depends on what value propositions are of importance;
- Consequently, considerable effort should be spent on agreement towards what kind of value or value propositions are of interest and in what way this value should be captured and by whom;

For these deliberations to be successful, the potential inhibitors (and enablers) in the current OGD initiative environment need to be identified and new business models should be explored.

We conclude that there is a need for further research to identify more clearly the enablers and inhibitors that governments operating in different social and economic contexts have to consider when opening up government data. There is also a need for a more structured set of goals and measurements for governments to use when estimating the captured value of OGD initiatives. From a theoretical perspective, the notion of openness requires a reconsideration of the processes that usually generate value creation and capture, from a value chain perspective to a value network perspective. Open Government and Open Data have the ability to facilitate networks of collaboration and co-creation that produce real economic and social impact, but more research on these value generating mechanisms is needed in order to guide future initiatives.

## I References

- Alani, H.; Dupplaw, D.; Sheridan, J.; O'Hara, K.; Darlington, J.; Shadbolt, N. and Tullo, C. 2007. Unlocking the potential of public sector information with Semantic Web technology. In: The 6th International Semantic Web Conference (ISWC), 11-15 Nov 2007, Busan, Korea.
- Allee, V. 2008. "Value Network Analysis and Value Conversion of Tangible and Intangible Assets," *Journal of Intellectual Capital* (9:1), pp. 5 – 24.
- Bates, J. 2012. "This is what Modern Deregulation Looks Like" [1]: Co-Optation and Contestation in the Shaping of the UK's Open Government Data Initiative," *Journal of Community Informatics* (8:2), 2012.
- Bellamy, C. and Taylor, J. 1998." *Governing in the information age,*" Philadelphia: Open University Press.
- Benjamin, S.; Bhuvanewari, R. and Rajan, P. 2007. "Bhoomi: 'E-Governance', Or, An Anti-Politics Machine Necessary to Globalize Bangalore?" CASUM-m working paper (Internet, <http://casumm.files.wordpress.com/2008/09/bhoomi-e-governance.pdf>).
- Bountouri, L., Papatheodorou, C. and Gergatsoulis, M. 2010. "Modelling the Public Sector Information through CIDOC Conceptual Reference Model. On the Move to Meaningful Internet Systems," *OTM 2010 Workshops Lecture Notes in Computer Science* (6428/2010), pp. 404-413.
- Böhm, C.; Freitag, M.; Heise, A.; Lehmann, C.; Mascher, A.; Naumann, F.; Ercegovac, V.; Mauricio; Haase, P. and Schmidt, M. 2012. "GovWILD: integrating open government data for transparency," In *Proceeding WWW '12 Companion Proceedings of the 21st international conference companion on World Wide Web*, New York: ACM, pp. 321-324.
- Budapest Open Access Initiative. 2002. Available at <http://www.soros.org/openaccess/read.shtml>
- Brandeis, L.D. 1914. *Other people's money*. Available at <http://www.law.louisville.edu/library/collections/brandeis/node/196>



- Clark, M. 2006. "Public sector information: Renaissance, Reformation and the Age of Reason," *Business Information Review* (23:2), pp. 125-134.
- Conradie, P.; Mulder, I. and Choenni, S., 2012. "Rotterdam Open Data: Exploring the release of public sector information through co-creation," In *ICE 2012: International Conference On Engineering, Technology And Innovation*, 18-20 June 2012, pp. 187-196.
- Cook, J. S. 2010. "Economic issues in funding and supplying public sector information," In *Access to public sector information : law, technology and policy: Volume 1*, Fitzgerald. B. (eds.), Sydney: Sydney University Press.
- Davies, T. 2010. "Open data, democracy and public sector reform: A look at open government data use from data.gov.uk," "Unpublished Master's thesis, University of Oxford. Available at <http://practicalparticipation.co.uk/odi/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf>
- Dawes, S. S. and Pardo, T.A. 2006. "Maximizing knowledge for program evaluation: critical issues and practical challenges of ICT strategies," In *Proceedings of the 5th International Conference, EGOF*. Springer: Lecture Notes in Computer Science.
- Dawes, S.S. 2012. "A Realistic Look at Open Data," Center for Technology in Government, University at Albany/SUNY Available at [http://www.w3.org/2012/06/pmod/pmod2012\\_submission\\_38.pdf](http://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf).
- de Cindio, F. 2012. "Guidelines for Designing Deliberative Digital Habitats: Learning from e-Participation for Open Data Initiatives," *Journal of Community Informatics* (8:2).
- de Vries, M. 2012. "Re-use of public sector information," Report for Danish Ministry for Housing, Urban and Rural Affairs.
- de Vries, M.; Kapff, L.; Achiaga, M. N.; Wauters, P.; Osimo, D.; Foley, P.; . . . and Whitehouse, D. 2011. "Pricing of public sector information study (POPSIS)," Brussels: European Commission, Information Society and Media Directorate General.
- DECA. 2010. "The value of the Danish address data: Social benefits from the 2002 agreement on procuring address data etc. free of charge," Available at [http://www.adresse-info.dk/Portals/2/Benefit/Value\\_Assessment\\_Danish\\_Address\\_Data\\_UK\\_2010-07-07b.pdf](http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf) [Accessed July 2012]
- Dekkers, M; Polman, F.; te Velde, R. te, and de Vries, M. 2006. "Measuring European Public Sector Information Resources, Final Report of Study on Exploitation of public sector information – benchmarking of EU framework conditions," Brussels: HELM and Zenc.
- Department for Culture, Media and Sport and Department for Business, Innovation and skill. 2009. "Digital Britain: Final Report," Richmond: Office of Public Sector Information.
- Digitaliseringsstyrelsen. 2012. "Basic Data," Available at <http://www.digst.dk/Home/ServiceMenu/English/Digitisation/Basic%20Data> [Accessed October 5<sup>th</sup>]
- Ding, L.; Lebo, T.; Erickson, J.S.; DiFranzo, D.; Williams, G.T.; Li, X.; Michaelis, J.; Graves, A.; Zheng, J.G.; Shangguan, S.; Flores, J.; McGuinness, D.L.; Hendlar, J.A. 2011. "TWCLOGD: A portal for linked open government data ecosystems," *Web Semantics: Science, Services and Agents on the World Wide Web* (9:3), pp. 325-333.

- Emerson, J.; Wachowicz, J. and Chun, S. 2001. "Social Return on Investment (SROI): Exploring Aspects of Value Creation," Available at <http://hbswk.hbs.edu/archive/1957.html> [Accessed October 2012]
- European Commission. 2011. "Open Data: An Engine for Innovation, Growth and Transparent Governance," European Commission Information Society. Available at [http://ec.europa.eu/information\\_society/policy/psi/docs/pdfs/directive\\_proposal/2012/open\\_data.pdf](http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/open_data.pdf) [Accessed July 2012]
- Fioretti, M. 2011. "Open Data: Emerging trends, issues and best practices," Available at [http://www.lem.sssup.it/WPLem/odos/odos\\_2.html](http://www.lem.sssup.it/WPLem/odos/odos_2.html) [Accessed July 2012]
- Gigler, S; Custer, S. and Rahmetulla, H. 2011. "Realizing the Vision of Open Government Data- Opportunities, Challenges and Pitfalls," World Bank Report. Available at <http://www.scribd.com/doc/75642397/Realizing-the-Vision-of-Open-Government-Data-Long-Version-Opportunities-Challenges-and-Pitfalls>.
- Grupe, F. 1995. "Commercializing public information: A critical issue for governmental IS professionals," *Information & Management* (28:4), pp. 229-241.
- Gurstein, M. 2010. "Open Data: Empowering the Empowered or Effective Data Use for Everyone?," *First Monday* (16:2), February 7.
- Halonen, A. 2012. *Being Open About Data*. Available at <http://finnish-institute.org.uk/images/stories/pdf2012/being%20open%20about%20data.pdf> [Accessed July 2012]
- Harrison, T.M.; Guerrero, S.; Burke, G.B.; Cook, M.E.; Cresswell, A.M.; Helbig, N.; Hrdinová, J. and Pardo, T.A. 2011. "Open government and e-government: democratic challenges from a public value perspective," In *The Proceedings of the 12th Annual International Conference on Digital Government Research: Digital Government Innovation in Challenging Times*, Bertot, J.C.; Nahon, K.; Chun, S.A.; Luna-Reyes, L.F. and Atluri, V. (eds.), College Park, MD, pp.245-253.
- Hausenblas, M. 2011. "Utilising linked open data in applications," In *'11 Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, Article No. 7, New York: ACM.
- Hogge, B. 2011. "Open Data Study: New Technologies," London: Transparency and Accountability Initiative.
- Houghton, J. 2011. "Costs and Benefits of Data Provision: Report to the Australian National Data Service," Centre for Strategic Economic Studies, Victoria University.
- Huijboom, N., and Van den Broek, T. 2011. "Open data: an international comparison of strategies," *European Journal of e Practice*.
- Höchtel, J. and Reichstädter, P. 2011. "Linked Open Data - A Means for Public Sector Information Management," *Lecture Notes in Computer Science* (6866/2011), pp. 330-343.
- Koski, H. 2011. "Does marginal cost pricing of public sector information spur firm growth?" Discussion Papers (1260), Helsinki: ETLA – The Research Institute of the Finnish Economy.
- Kundra, V. 2012. "Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect," Jan 2012, Harvard: Joan Shorenstein Center on the Press, Politics and Public Policy. Available at

[http://www.hks.harvard.edu/presspol/publications/papers/discussion\\_papers/d70\\_kundra.pdf](http://www.hks.harvard.edu/presspol/publications/papers/discussion_papers/d70_kundra.pdf)  
[Accessed September 2012]

- Lampri, M. 2012. "E-Government and Data Protection in the Public Sector," In *Virtual Communities, Social Networks and Collaboration, Annals of Information Systems*, Athina A Lazakidou (ed.) pp. 135-147.
- Lee, G. and Kwak, J.H. 2011. "Open Government Implementation Model: A Stage Model for Achieving Increased Public Engagement," In *The Proceedings of the 12th Annual International Conference on Digital Government Research: Digital Government Innovation in Challenging Times*, Bertot, J.C.; Nahon, K.; Chun, S.A.; Luna-Reyes, L.F. and Atluri, V. (eds.), College Park, MD, pp. 254-261.
- Linders, D.; Wilson, S.C. 2011. "What is Open Government? One Year after the Directive," In *The Proceedings of the 12th Annual International Conference on Digital Government Research: Digital Government Innovation in Challenging Times*, Bertot, J.C.; Nahon, K.; Chun, S.A.; Luna-Reyes, L.F. and Atluri, V. (eds.), College Park, MD, pp. 262-271.
- Malensky, E.; Schuler, P. and Tran, A. 2011. "The Adverse Effects of Sunshine: Evidence from a Field Experiment on Legislative Transparency in an Authoritarian Assembly," Prepared for presentation at Princeton Political Economy Conference (Polecon), May 14, 2011.
- McKinsey and Company. 2012. "Innovation in government: Indonesia and Colombia," McKinsey Quarterly: Public sector practice.
- Moore, M.H. 1995. *Creating public value: strategic management in government*, Harvard University Press.
- Morgan, L; Feller, J. and Finnegan, P. 2010. "Value Creation and Capture with Open Source Software: A Theoretical Model for Understanding the Role of Value Networks," In *Proceedings of the 18th European Conference on Information Systems*, Alexander, P. M.; Turpin, M. and van Deventer, J. P. (eds.), Pretoria, South Africa.
- Noveck, B.S. 2009. "Wiki Government: How Technology Can Make Government Better, Democracy Stronger and Citizens More Powerful." Washington: Brookings Institution Press.
- Open Knowledge Foundation (OKF) Available at <http://www.opendefinition.org/> and <http://opengovernmentdata.org/> [Accessed September 15, 2012].
- Paulheim, H. and Fümkrantz, J. 2012. "Unsupervised generation of data mining features from linked open data," In *The Proceedings of the 12th Annual International Conference on Digital Government Research Conference: Digital Government Innovation in Challenging Times*, Bertot, J.C.; Nahon, K.; Chun, S.A.; Luna-Reyes, L.F. and Atluri, V. (eds.), College Park, MD, pp. 20-25.
- PIRA. 2000. "Commercial exploitation of Europe's public sector information," Final Report, Pira International, European Commission, Directorate General for the Information Society.
- Pollock, R. 2008. "Economics of PSI," Cambridge: University of Cambridge. Available at [http://rufuspollock.org/economics/papers/economics\\_of\\_psi.pdf](http://rufuspollock.org/economics/papers/economics_of_psi.pdf) [Accessed August 2012]
- Rath, S. 2012. "Data templates for district Economic Planning," *Journal of Community Informatics* (8:2).
- Suber, P. 2007. "Open Access Overview," Available at <http://www.earlham.edu/~peters/fos/overview.htm>

- Talia, D. and Trunfio, P. 2010. "How distributed data mining tasks can thrive as knowledge services," *Communications of the ACM* (53:7), pp. 132-137.
- Uhlir, P. F. 2009. "The Socioeconomic Effects of Public Sector Information on Digital Networks: Toward a Better Understanding of Different Access and Reuse Policies: Workshop Summary," US National Committee CODATA, in cooperation with OECD ([http://www.nap.edu/catalog.php?record\\_id=12687](http://www.nap.edu/catalog.php?record_id=12687))
- Vanhaverbeke, W. 2008. "The Interorganizational Context of Open Innovation," In *Open Innovation: Researching a New Paradigm*, H. Chesbrough, W. Vanhaverbeke and J. West (eds.), Oxford: Oxford University Press, pp. 258-281.
- Vickery, G. 2011. "Review of Recent Studies on PSI Re-Use and Related Market Development," Paris: European Commission.
- Weiss, P. 2002. "Borders in Cyberspace: Conflicting Public Sector Information Policies and Their Economic Impacts - Summary Report," Available at [http://www.nws.noaa.gov/sp/Borders\\_report.pdf](http://www.nws.noaa.gov/sp/Borders_report.pdf)
- Zuiderwijk, A.; Janssen, M.; van den Braak, S. and Charalabidis, Y. 2012. "Linking open data: challenges and solutions," In *Proceedings of the 13th Annual International Conference on Digital Government Research*, Bertot, J.; Luna-Reyes, LF. and Mellouli, S. (eds.), College Park, MD, pp. 304-305.

## **PAPER II: The Generative Mechanisms of Open Government Data**

In Proceedings of the 21st European Conference on Information Systems (ECIS),  
2013, Utrecht, The Netherlands

# THE GENERATIVE MECHANISMS OF OPEN GOVERNMENT DATA

Jetzek, Thorhildur, Copenhagen Business School, Frederiksberg, Denmark,  
[tj.itm@cbs.dk](mailto:tj.itm@cbs.dk)

Avital, Michel, Copenhagen Business School, Frederiksberg, Denmark,  
[michel@avital.net](mailto:michel@avital.net)

Bjorn-Andersen, Niels, Copenhagen Business School, Frederiksberg, Denmark,  
[nba.itm@cbs.dk](mailto:nba.itm@cbs.dk)

## Abstract

The exponentially growing availability of data, global connectivity, and trends toward increased openness and sharing are turning into a powerful force that is changing the global economy and society. Governments around the world have become active participants in this evolution by opening up their data for access and re-use by public and private agents alike. The recent phenomenon of Open Government Data (OGD) has spread around the world, driven by the proposition that opening government data has the ability to generate both economic and social value. However, a review of the academic research and the popular press reveals only sporadic attention given to various aspects with no overarching framework that explains how OGD generates value. We apply a critical realist approach to uncover the generative mechanisms that serve to explain this relationship. First, we present a strategic framework with four archetypical generative mechanisms. The framework outlines the different pathways to value generation and highlights the current tension between the private/public and economic/social domains. Second, we offer a conceptual model that provides a systematic way of articulating and examining further the generation of value from OGD.

**Keywords:** *Open Government Data, Open Access, Big Data, Generative Mechanisms, Value Creation.*

## II.1 Introduction

Today, we generate and store more data than at any other time in history. Computing and networking capabilities combined with openness enhance the potential impact of the accumulated data and offer society an opportunity to drive massive social, political

and economic change (Kundra, 2012). Thus far, there is an apparent tension between the social value driven networks of data, information and knowledge sharing, on the one hand, and the economic value driven products and services markets of ownership and exchange, on the other. Traditionally, these have been seen as two different ends, with businesses on one end and governments and not-for-profit organizations on the other. The presumed trade-offs between economic efficiency and social progress have almost become institutionalized (Porter and Kramer, 2011). However, in the last few years, the boundaries between the two are blurring, following the increased connectivity, openness and sharing. These increasingly complex interdependencies are forcing us to re-think the concepts of economic and social value in a world moving towards digital interoperability, information sharing, co-creation and collaborative networks (Avital et al, 2007; Hess and Ostrom, 2006).

Having *Open Government Data* (OGD) implies that the public sector relinquishes its role as information gatekeeper in lieu of a new role as information provider, leading to a realignment of the power dynamics between the public and private sectors (Davies, 2010). Proponents of OGD hope that such shifts will strengthen democracy and improve the impacts of government work through increased transparency, participation and collaboration. Moreover, advocates of OGD argue that it enables greater government efficiency through an information infrastructure that allows for better data re-use within the public sectors and inter-agency coordination. Open data advocates are also motivated by its potential for innovative entrepreneurs who can use OGD to generate new products and services. However, due to the early stage of OGD initiatives, their impact and ramifications are still debated in professional and academic circles. While economic and social value has been expected to emerge from the dissemination of OGD, evidence is still scarce. The economic and social impact of open-data policies remains largely unclear, and there are relatively little empirical data available on the effects of the various policy approaches, thus leaving policy makers without the facts they need to assess and improve these policies (Huijboom and Van den Broek, 2011; Zuiderwijk and Janssen, 2012).

This paper addresses the question: How do open government data stimulate value generation? To explore and illustrate how value can be generated from the use of OGD, we turned to the critical realist concept of a micro-to-macro mechanism as a causal structure that contingently generates observable outcomes. We use these mechanisms to portray the instrumental pathways between OGD, social factors and technological factors that have the potential to generate value. We concentrate on the overall economic and societal impacts from OGD and examine OGD initiatives as the

unit of analysis. We propose that there are four archetypical generative mechanisms that explain how OGD can stimulate value generation, contingent on three enabling conditions. We furthermore propose that synergies between different generative mechanisms can lead to enhanced value generation and more widely dispersed value appropriation. Finally, we propose that by contrasting different mechanisms and different perspectives, we can contribute to knowledge on how ambidextrous strategies and cross-sector collaboration within value networks can benefit OGD value generation and appropriation.

## **II.2 The unique features of OGD**

The amount of data accumulating in our increasingly digital world is breathtaking. In the past two years alone, increased Internet access and data generation from surveillance cameras and smart meters have doubled the amount of information available in the digital universe to its current rate of 2.8 ZB, a number that is expected to double every year. The massive increase in the amount of data, combined with openness and technologies that allow global distribution, has dramatically changed the nature of data from a closed proprietary resource to a common shared resource. Additionally, the continuously developing technical and semantic ability to merge disparate datasets, combined with sophisticated data analysis, has the potential to increase vastly the overall value of data. It is estimated that better use of data resources in the US health care could generate USD 300 billion annually and in Europe's public sector it could generate EUR 250 billion annually (McKinsey, 2011).

*Open data* can be defined as data that are freely accessible online, available without technical restrictions to re-use, and provided under open access license that allows the data to be re-used without limitation, including across different 'fields of endeavor' (e.g., commercial and non-commercial alike) (OKF, 2012). The concept of Open Government Data refers to government data, defined as "data and information produced or commissioned by government or government controlled entities" (OKF, 2012), that are opened up for use and re-use by public and private agents alike. In the currently used terminology, OGD does not include data that are subject to valid privacy, security or privilege limitations, as governed by other statutes. Government data sets are an interesting subset of open data because such subsets have already been collected for specific use, have been paid for by taxpayers, are relevant and offer value beyond what is captured from the originally intended use. When opened up, government data become a common, shared resource (i.e., public good) that is



provided by the government. Building on Nilsen (2010) and Shapiro and Varian (1999), the typical features of OGD as a resource are as follows:

It is non-rivalrous: One person's use of the data does not reduce availability to others.

It is not excludable: One person cannot exclude another person from using the data.

It has high fixed costs: Costs of collecting, processing and storing the data are generally high.

It has (almost) zero marginal cost, i.e., it is inexpensive to reproduce the data after they have been collected.

It offers valuable information and has high potential for re-use.

While the generative mechanisms discussed in section 5 can be applied to all types of open data as a resource, the outcomes generated from the use of open data provided by government differ from the outcomes generated by use of open data provided by the private sector. In this paper, we focus specifically on the value generated by the use of open data that are produced and provided by the government.

### **II.3 The multifaceted nature of value**

The most advanced economies in the world today have made two parallel shifts that, paradoxically, make possible a significant attenuation of the limitations that market-based production places on the pursuit of value that is central to liberal societies. The first is the move to an economy centered on information and the second is the move to a networked Internet-based environment. These changes have increased the role of non-market and non-proprietary production, both by individuals alone and by cooperative efforts in a wide range of loosely or tightly woven collaborations (Benkler, 2006). We can define value as objectively established or as perceived worth for somebody. In order to highlight the apparent tensions between the monetary and market based value creation and the non-proprietary and collaboration based value creation, we base our discussion on two types of value: *economic value*, defined as the worth of a good or service as determined by the market, and *social value*, which is created when resources, inputs, processes or policies are combined to generate improvements in the lives of individuals or society as a whole (Emerson et al., 2001).

The conceptual distinction between *value generation* (creation) and *value appropriation* (capture) has been growing in importance with the trend towards openness, technical connectivity and collaborative ventures. Value generation materializes when the utility of society's members increases after accounting for the

resources used in that activity. Value appropriation materializes when an actor is able to capture a portion of the value created by an activity (Bowman and Ambrosini, 2000). Due to the unique features of OGD, the value that is generated from the use of data is not necessarily exclusively available for appropriation by the owner of the resource, nor even by the value creator. The relationship between the generation and appropriation of the value of multiple stakeholders in the context of alliances is increasingly being viewed to be multifaceted in nature (Sarker et al., 2012). Gil-Saura et al. (2009, p. 595) suggest that value is a subjective, multidimensional construct; accordingly, it is only through a multidimensional view that we get a true picture of the value generated within alliance relationships. However, despite the recent focus on the creation of value in collaborative settings, little is known about the underlying mechanisms (Sarker et al., 2012). The fundamental transformation towards organizations collectively creating value has raised important new issues that cannot easily be addressed by currently used frameworks in IT value research (Grover and Kohli, 2012).

## **II.4 Research Strategy**

### **Research philosophy and method**

The critical realist notion of *generative mechanism* has served as a springboard for our study. The basic assumption of critical realism is the existence of a real world independent of our knowledge of it. The objects and structures of the real give rise to causal powers, called generative mechanisms, which cause the events that we may observe (Bhaskar, 1998). The aim of critical realism is not to investigate the level of events, but rather to uncover and describe the mechanisms producing these events. The mechanisms are associated with the real nature of the underlying objects, and it is thereby assumed that they are relatively stable structures created by an interplay of objects. Critical realists emphasize that causality is contingent - in the sense that the outcome of a mechanism is contextual. In that regard, the observed outcome emerges from the intersection of appropriate pre-conditions, and the mechanisms then act transfactually (Henfridsson and Bygstad, 2013). Furthermore, the critical realist approach identifies an intermediary level of analysis, between pure description and story-telling, on the one hand, and grand theorizing and universal social laws, on the other (Hedstrom and Swedberg 1998).

Due to the emergent nature of open data and their value, the research approach is phenomenon-based with the aim to capture, describe, document and conceptualize the

underlying phenomenon, as well as to identify the mechanisms that explain how value is generated through the application of OGD. The investigation of a new phenomenon calls for a relatively unrestricted gathering not only of primary data, for example, interviews, but also of secondary data, for example, news reports or online sources (von Krogh et al., 2012b).

Our study was conducted in five consecutive phases: 1) identification of definitions, concepts and keywords, 2) review of the literature, 3) collection of primary data, 4) creation of a framework with the main types of generative mechanisms and 5) specification of a conceptual model. The literature review aims to: 1) reveal in what ways, and to what degree, the literature has addressed our research question, 2) organize and classify the literature according to the topics covered and 3) identify gaps in the current literature to justify the creation of a new theoretical framework.

## Data collection

First, we conducted a literature search based on the main concepts relevant to our topic: ‘government,’ ‘open,’ ‘data’ and ‘value’ and various combinations thereof. In order to be comprehensive, we searched for other similar terms that are used in the extant literature, e.g., ‘public sector information.’ We also collected diverse literature on value generation. We searched for the terms ‘value generation/creation,’ ‘value appropriation/capture,’ ‘value networks’ and ‘value co-creation.’ We searched in scientific databases, but in consideration of the nescient phenomenon, we extended the search to the blogosphere and popular press. Furthermore, we did not constrict the search to certain disciplines because we aimed to bring together previously disparate streams of work to help shed light on the phenomenon.

Second, we collected primary data from an open government data initiative in Denmark: *The Danish Basic Data Program* (Digitaliseringsstyrelsen, 2012). The obtained data were derived from project documents, nine semi-structured interviews and notes from participation in meetings and conferences related to the project. A short interview guide was used in order to keep the discussion of interviews focused on the topic of value, particularly on what kind of value could be obtained from OGD initiatives and what might be the main enablers and inhibitors for value generation. The nine interviewees included: four project managers from different sub-projects within the main Basic Data Program; two product owners from a large supplier for public sector IT-solutions and a big user of government data in Denmark (KMD); one consultant and open data evangelist from the Netherlands; one project manager from the municipalities’ common IT/project organization (KOMBIT), who was also a

participant in the program; one representative from the ministry that provides property data and one representative from the organization that collects and provides geographical data. All interviews by one author lasted one to two hours and were performed face to face.

## II.5 Findings and Analysis

### Generative Mechanisms for OGD value

Mechanisms can be portrayed as small pieces of theory that specify how a given input will reliably create a specific output (Hedstrom and Swedberg, 1998). Mechanisms do not merely describe what happened but also how it happened, thereby allowing us to see beyond the surface-level description of a phenomenon. Mechanisms may be classified on three levels: contextual mechanisms (macro–micro level), action-formation mechanisms (micro-micro) and transformational mechanisms (micro–macro level). The latter type explains how different components interact in order to produce an outcome at the macro level (Hedstrom and Swedberg 1998). In this paper, we focus on the transformational mechanisms as socio-technical mechanisms, i.e., generative mechanisms that are triggered by the interaction of social and technological constructs. Our aim was to explain *how* use of OGD can generate value and *what* contextual elements may lead to a desired outcome.

In order to identify the main generative mechanisms, we conducted a wide search that focused on various operational definitions of open government data. We found that two distinct ideologies drive most open government data initiatives: the ‘Re-use of data’ perspective and the ‘Open Government’ perspective. We thus reviewed the respective tensions and contributions of these two unique streams. The literature on re-use of OGD is mostly focused on the *economic value* of government data, often in connection to the European PSI-directive (Janssen, 2011). The literature in the context of Open Government is mostly derived from Obama’s 2009 Open Government Directive, and, in a higher grade, is directed towards government policy that is centered around how use of OGD can contribute to the generation of *social (or public) value* in collaborative settings (Linders and Wilson, 2012). The emergent open government movement is said to offer the possibility to reconcile the divergent paths of e-democracy and e-government by creating shared understanding, using new sources of expertise and building civic capacity (Harrison et al., 2011). However, the OGD discourse is increasingly citing both social and economic reasons for opening data, and the

principles of supplying data for open government and re-use are converging (Janssen, 2011).

Through the lens of economic and IS-based theories on value generation, we were able to identify four distinct generative mechanisms that can explain how OGD enable generation of value. Two come from the Open Government literature: *transparency of government* and *citizen participation/collaboration* (Cordis and Warren, 2012; Harrison et al., 2011; Linders and Wilson, 2012) and two from the Re-use literature: *efficiency* and *innovation* (Gigler et al, 2011; Halonen, 2012; Jansen, 2011). Commonly discussed barriers to value generation in both streams are: 1) closed or inaccessible datasets, 2) lack of comprehensive data policies, 3) lack of validity, completeness and exhaustiveness of datasets, 4) insufficient metadata, 5) lack of consistency in cross-border access regimes, 6) lack of motivation within public sector, 7) lack of technical and semantic interoperability between governmental systems and datasets and 8) too fragmented and disparate open data community (Davies, 2010; Dawes, 2012; Halonen, 2012; Janssen, 2011; Lee and Kwak, 2011; Mayer-Schönberger and Zappia, 2011; Zuiderwijk et al., 2012).

Finally, we looked for barriers to value appropriation and identified the following: 1) lack of technical ability to extract value from data, 4) the digital divide and 5) power differences between data users and unequal access opportunities (Bertot et al., 2010; Halonen, 2012). In order to overcome these barriers, we propose that governments should focus on three enabling factors: open access, data governance and technical connectivity that apply to all four mechanisms, as subsequently described in section 5.3.

## Strategic framework

We use a two-by-two matrix to represent the strategic framework of the generative mechanisms that explain how OGD can generate value (Figure 1). The framework spans the boundaries between the public and private sector, as well as the different types of strategic focus of OGD initiatives (Harrison et al., 2011; Janssen, 2012). The horizontal dimension focuses on the *sector* that generates value through OGD initiatives, spanning between public sector-based initiatives (e.g., efficiency and transparency of public services) and private sector-based initiatives (e.g., innovation and e-participation.) The vertical dimension focuses on *value*, spanning between OGD initiatives that are focused on the generation of social value (e.g., strategies focused on

the softer measures of transparency, participation and collaboration and directed towards citizens) and economic value (e.g., re-use of data strategies focused on the monetary benefits that are expected to arise from increased efficiency and creation of new services and businesses). We now discuss each of the four types of generative mechanisms in greater detail.

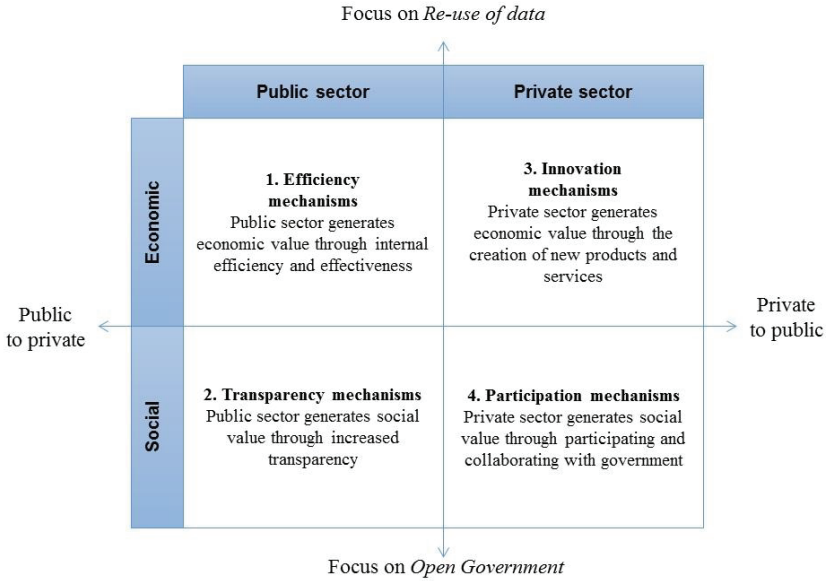


Figure 1: Strategic framework of four archetypical generative mechanisms

### Efficiency mechanisms

The efficiency type of generative mechanisms enables value generation by better utilizing current resources. The general economic theory that describes how this mechanism works is *Transaction Cost Economics*, where value is generated by reducing transaction costs in operations. In the case of OGD, such transaction costs might be incurred by keying in the same data many times, saving the same data in multiple repositories or by charging for the data. The creation of more effective methods of collection, management, distribution and use of data can create direct and indirect cost-savings. In these instances, the strategy that drives value generation is motivated by the vision of a more efficient government. Moreover, these types of initiatives also have the capability of generating value for the private sector through

more effective public services. As an example, the transaction costs incurred by selling and delivering geographic data to users in Australia before data were made freely available online in 2002 were estimated to be between 17 - 33% of the revenues. Assuming that transaction and access costs of users mirror those of the agencies, the net private cost savings may have been around \$1.7 million annually (de Vries, 2012).

The Danish authorities have recently initiated an OGD project (the Basic Data Program), where the aim is to generate economic value through more efficient collection, dissemination and use of government data. The number of basic data registers will be reduced from five to three and master data will be synchronized via a common data model. A common platform (where both public and private users can get access to the same, high-quality data) is being implemented. As a result of these changes, the possibility for automated business processes across authorities is greatly increased. Furthermore, as data will be freely available online, transaction costs related to user support and billing should also be diminished. The total yearly savings for the public sector are projected to be around €35 million (Digitaliseringsstyrelsen, 2012).

The focus of the Danish authorities is on industry-wide or market-wide collective savings due to the large initial costs incurred by making these big changes to the data model and data distribution channels. Moreover, the positive external effect from this project is that integrated government data of better quality will also benefit private industries, such as real estate dealers, insurance companies, the financial sector and the telecom industry, which previously had to spend significant resources on creating usable information from heterogeneous data-sources. The cost-savings for the private industry are estimated to be around €65 million per annum when the program is fully implemented.

### **Innovation mechanisms**

This type of generative mechanisms generates value through *transformational effects*, where data are supplied as a service or leveraged in applications in ways that are new and innovative. Innovation is the source of value creation in *Schumpeter's economic theory*, bringing about novel combinations of resources, new production methods, as well as new products and services, which, in turn, lead to the transformation of markets and industries, thus increasing value. An example of the positive effect of providing government data to the private sector can be found in the Netherlands, where openness and technical availability of meteorological data with an emphasis on data governance has led to the creation of a competitive and innovative private weather market. Impacts include 400% increase in turnover for private sector re-users, 250% increase in high-

end users, a rise in the use activity of re-users of 300% and an increase of over €35 million on corporate tax returns (de Vries, 2012).

One specific example of the innovative combination of map data with data on drug prescriptions can be found at <http://www.prescribinganalytics.com>. This website was collaboratively created by a group of NHS doctors, academics and a big-data analytics start-up company, Mastodon C. Their analytics show how prescriptions of statins, drugs used to lower cholesterol, differ between different municipalities in England. The entrepreneurs used open prescriptions data made available by the NHS in the UK, and combined them with geographic data. They used modern data analytics to produce a visualization map showing the different proportions between expensive (branded) and inexpensive (generic) statin prescriptions in different counties. Wherever the proportion of branded items were high, it represented a potential to make big savings by switching to a generic form of the same drug. According to their analysis, if two thirds of the proprietary drugs had been substituted with generic forms of the same drugs in the year to June 2012, public healthcare in the UK could have saved £200 million pounds. In this case, the innovative use of OGD, enabled by open access to reliable government data and use of technology, has generated economic value that can be appropriated by entrepreneurs as well as the UK government, subsequently improving their healthcare services.

### **Transparency mechanisms**

This type of mechanism enables value generation by information effects. The general economic theory that explains how value is generated is based on the concept of *Information Asymmetry*. Information Asymmetry describes situations where one party has more or better information than the other while participating in transactions, negotiations or communications. Information asymmetry can cause all sorts of sub-optimal results and behaviors, such as *Moral Hazard*, where the more informed make decisions on their own benefits, with the cost falling on others. In the case of government, the consequences of misuse of public power for private benefits can be particularly dire for society in general. While empirical studies have given conflicting evidence on the relationship between transparency and corruption, the results of a recent study show that corruption conviction rates almost doubled when Freedom of Information Act (FOI) laws were strengthened in various states in the US (Cordis and Warren, 2012).

The promise of openness is to provide a source of pressure that counteracts the tendency of technology enactment to reproduce existing rules, routines, norms and



power relations, despite the new and innovative capabilities introduced by these technologies. However, this promise can only be fulfilled if open government changes the nature of relationships between stakeholders and governments, thereby producing innovative forms of organizing that enable groups to link across organizational boundaries and functions (Harrison et al., 2011). One such transparency agenda for tackling poverty in the global economy was presented by the British Prime Minister, David Cameron, in the G8 meeting at the World Economic Forum in Davos in January, 2012. The plan is to tackle: illicit financial flows, the hidden company ownership that makes such flows possible, land grabs, and the secrecy by which big oil, gas and mining corporations are doing business. The claim is that citizens in developing countries are regularly robbed of the benefits of their countries' mineral wealth through poorly negotiated or corrupt backroom deals. In this case, open access to government data on company ownership, natural resources and tax information - combined with technical connectivity and governance - could enable greater cross-border transparency which are the mechanisms that could uncover corrupt practices, subsequently generating social value that could be appropriated by governments and citizens alike.

### **Participation (and collaboration) mechanisms**

These mechanisms generate value through the *positive effects of scale*, where openness and sharing enable value generation drawing from a larger pool of resources. In the case of OGD, the generative mechanisms of *participation* lead to improved citizenship and collaborative behavior through crowdsourcing activities. A similar theoretical argument is used in the literature on *Open Innovation* (Chesbrough et al., 2006) where the principal idea is that an open approach to sharing knowledge across boundaries expands the firm's innovative potential, as the firm is able to tap into a much larger pool of ideas and find such ideas faster. But what drives individuals and organizations to share their resources without direct monetary reimbursement? The answer might lie in the notion of social value. A substantial amount of academic work has theorized about, and empirically examined, the motivations of those contributing to the development of Open Source Software, where it is argued that individual motivation should not be looked at in isolation, but in interplay with institutions, goods and the social practice: "...people's pursuit of visible carrots is at times interrupted by the larger quest for the invisible gold at the end of the rainbow" (von Krogh et al., 2012a, p. 671).

Participation in the context of OGD focuses on engaging the public to inform government solutions and decision-making. This can take two discrete forms: 1) collecting opinions (citizen engagement) and 2) collecting ideas and solutions or

crowdsourcing (Linders and Wilson, 2012). An example of the former mechanism can be found in Iceland, where the public sector turned to the private sector to create and vote on a draft for a new constitution by using open data, the enabling social media technologies and open data governance. A good example of citizen collaboration is the crowdsourcing activities that have been immensely helpful in natural disaster incidents, such as hurricane Katrina and the earthquake in Haiti (Lee and Kwak, 2011). Just a few hours after the earthquake hit Haiti in January 2010 the OpenStreetMap (OSM) Community began tracing roads from imagery that was previously available from Yahoo. Within 48 hours high resolution imagery taken post-earthquake became available and in the first month over 600 people added information to the OSM. OSM communities have continued to work with NGO's and the Government of Haiti to further develop the OSM data. The collaboration between public and private stakeholders around data creation and collection is enabled by access to open data, the OSM technical platform, as well as the OSM community's access to and knowledge about geo-data. This collaboration is generating social value, appropriated by the Haitian government as well as the citizens of Haiti.

## Conceptual model

An interesting finding that emerged from the study is the understanding that opening government data is not in itself sufficient for value generation. A number of barriers have to be overcome in order to enable the mechanisms that allow for value generation. Accordingly, we propose that the key *enablers* for OGD value generation are as follows: open access to data, data governance procedures and technical connectivity. Furthermore, we propose that the synergies created by the interacting mechanisms and cross sector collaboration enhance the generation of value, thus allowing both sectors to appropriate the generated value. We offer a conceptual model that depicts the relationships between the three enabling factors of ODG, the four generative mechanisms, and the resulted social and economic value (Figure 2).

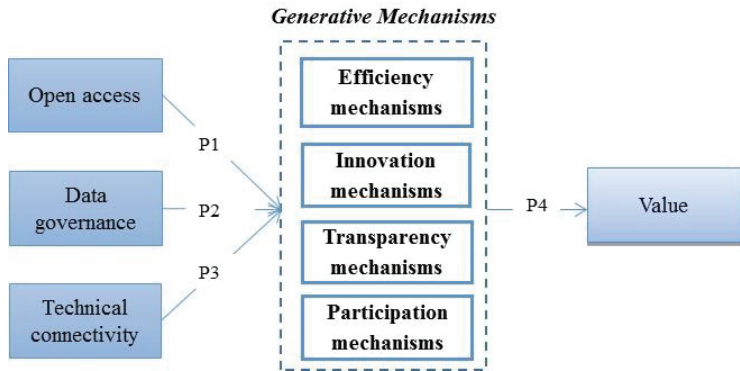


Figure 2: Conceptual model of OGD value generation

### Open access

Open access to government data is a composite construct that represents the use of *open access licenses*, the *availability* of OGD and the *accessibility* of OGD. The current literature, as well as anecdotal evidence, supports the proposition that opening government data enables the generation of social and economic value. Thereby,

**Proposition 1:** *Open access to government data has a positive effect on the Generative Mechanisms.*

### Data governance

Data governance is a composite construct that describes the actions and policies needed in order to provide the efficient dissemination of data of good quality and usefulness, as well as the sustainable and equitable dissemination of these data. Unknown, inconsistent or unsatisfactory *quality* of OGD leads to substantial risks for validity and relevance. *Relevance* of the data can be increased if organizations carefully consider which datasets support the strategy of the initiative (Lee and Kwak, 2011). It is important to give the correct context to the data, as government data are in many cases collected or created for specific purposes, and thus could be misleading if taken out of that context (Dawes, 2012). Accordingly, use could be stimulated if more information about the way open data are collected and processed were to be provided by including metadata (Zuiderwijk and Janssen, 2012). Citizen's access to the Internet and their ability to utilize the provided information are important for ensuring *equitable* dissemination (Bertot et al., 2010). Finally, governance must also ensure the *sustainability* of the resource, and therefore includes the creation of sustainable

business models that enable the government to guarantee the continuing collection of data (Hess and Ostrom, 2006). We propose that OGD is a common (public) resource and argue that resource governance is an enabler of all mechanisms that generate value from OGD. Thereby,

**Proposition 2:** *Data governance has a positive effect on the Generative Mechanisms.*

### **Technical connectivity**

Technical connectivity is a composite construct that describes the availability and usability of a technical infrastructure that allows users to access and combine the data. The technological backbone of any OGD initiative is an infrastructure that facilitates data exchange between government agencies and the public (including telecommunications infrastructure, connections between front-end web interfaces and back-end information management systems, system interoperability between agencies or government levels, and adequate availability of hardware and software within government bureaucracies) (Gigler et al., 2011). Moreover, governments have to consider not only the technical infrastructure as a tool to ensure availability and accessibility of data, but also the need for users to be able to understand and use the data as well as the technologies through which data are disseminated (Bertot et al., 2010). Schematic heterogeneity and lack of consistency can decrease usability and complicate access and integration of the data. Due to the decoupling of data from its original creation context, it is the semantic interoperability, identity resolution and ontologies that are central methodologies to ensure consistency and meaningful results and allow third parties to connect different data-sources (Alani et al., 2007). All of the identified mechanisms depend on the dissemination of data via technical platforms. Furthermore, the ability to access these platforms and to make sense of the data for different purposes is also supported by technology. Thereby,

**Proposition 3:** *Technical connectivity has a positive effect on the Generative Mechanisms.*

### **Generative mechanisms**

Finally, the generative mechanisms—efficiency, transparency, innovation, and citizen participation—also form a composite. We suggest that when the different generative mechanisms interact within an open system, economic and social value is generated. This interaction can be encouraged by collaboration between sectors within value networks. Value networks are important to facilitate the sharing of not only data, but

also information, know-how and other resources. In this way, value enhancement can happen, where value is extended to network participants within the value network. While value networks around open data have still not emerged to the same extent as that in the world of Free/Libre Open Source Software (Mayer-Schönberger and Zappia, 2011), communities such as the OSM and the proposed co-operation between governments proposed at the G8 meeting at the World Economic Forum in Davos give promise that we might be on the verge of a new era where governments and private sector collectively generate and appropriate value from OGD. Sarker et al. (2012) term this phenomenon, Synergistic Integration, where value is co-created through amalgamation. Thereby,

**Proposition 4:** *The Generative Mechanisms have a positive effect on value.*

## II.6 Discussion and Conclusion

The aim of this research was to identify the generative mechanisms that enable the extraction of value from OGD. To this end, we synthesized and conceptualized previous work on OGD and value generation and appropriation, and reviewed the identified relationships via the lens of current theories of IT-based value. To return to our research question: *How do open government data stimulate value generation?*, we have proposed four different archetypical generative mechanisms, each of which represents a certain type of cause-and effect relationship between OGD and value. However, we further propose that the highest level of value should be expected where there is synergistic integration between these mechanisms and where both public and private sectors are active in generating social and economic value from OGD. Unfortunately, in the many open data initiatives that we have been able to identify in the last few years, synergies do not seem to have emerged to any great extent. We suspect that the key reason for this deficiency is the prevalent tendency to implement OGD initiatives in silos, with unclear goals and a lack of both appropriate governance and proper enablement of the use of the data.

The contributions of this study to knowledge are as follows: a) the illustration of how generative mechanisms can be used to explain the relationship between OGD and value, b) the strategic framework that depicts different pathways to value generation and c) the conceptual model that illustrates the relationship between OGD, the generative mechanisms, and value. Moreover, the contribution to practice is first and foremost the identification of the factors that can enable (or hinder) the generation of value with OGD. We believe that the insights offered in this study can be applied, at

least in part, not only to OGD initiatives, but also to open data initiatives in private organizations. Open *private* data are also shared resources, and we suggest that the generative mechanisms in these instances are the same: efficiency (consolidation of databases, cross-company automated business processes), open innovation, transparency in company operations, and participation/collaboration as illustrated for example in the open design and open source software communities. The generalizability and validity of the study are inherently limited. As a theory development study, we offer a set of propositions but no empirical validation. Further confirmatory research is required to examine the mechanisms at multiple levels of analyses and various contexts.

We propose that openness is in itself an important enabler to the creation of value from data, as openness enables both the generation and appropriation of value, not only by the organization that owns the data, but also by external stakeholders. However, while openness might be seen to be the necessary condition in this context, it is insufficient on its own. Governments should increase the usability and re-usability of their data by focusing on data governance, where the aim is to ensure the quality and sustainability of the resource and to minimize risk for external users. This requires the use of technical platforms built on open standards that can increase efficiency and allow users the opportunity to link disparate data sources. Finally, governments should facilitate the emergence of value networks around open data in order to support the recursive mechanisms of synergistically co-created value. We suggest that economic markets of self-interested participants aiming to maximize their own value in value networks of sharing and co-creation enabled by openness and technical connectivity have huge potential for all participants. If the synergies between public/private and social/economic domains were to be exploited, we might have the potential for a quantum leap in increased productivity and social progress.

## II References

- Avital, M., Lyytinen, K., Iacono, S., Kraemer, K.L., Sambamurthy, V., and Sawyer, S. (2007) "Data Rich and Data Poor Scholarship: Where Does IS Research Stand?" Proceedings of the International Conference on Information Systems (ICIS), Montreal, Canada.
- Bhaskar, R.A. (1998). General Introduction, in Critical Realism: Essential Readings, M. S. Archer, R. Bhaskar, A. Collier, T. Lawson, and A. Norrie (eds.), London: Routledge, ix-xxiv
- Benkler, Y. (2006). The Wealth of Networks: How Social Production Transforms Markets and Freedom, New Haven CT: Yale University Press.

- Bertot, J.C., Jaeger, P.T. and Grimes, J.M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, 27 (2010), 264-271.
- Bowman, C. and Ambrosini, V. (2000). Value Creation Versus Value Capture: Towards a Coherent Definition of Value in Strategy. *British Journal of Management*, 11(1), 1-15.
- Chesbrough, H., Vanhaverbeke W. and West, J. (eds.), (2006). *Open Innovation: Researching a New Paradigm*, Oxford: Oxford University Press.
- Cordis, A.S. and Warren, P. L. (2012). Sunshine as Disinfectant: The Effect of State Freedom of Information Act Laws on Public Corruption. Available at <http://ssrn.com/abstract=1922859>
- Davies, T. (2010). Open data, democracy and public sector reform: A look at OGD use from data.gov.uk. Available at <http://tinyurl.com/7joks46>
- Dawes, S.S. (2012). A Realistic Look at Open Data. Center for Technology in Government, University at Albany/SUNY Available at [http://www.w3.org/2012/06/pmod/pmod2012\\_submission\\_38.pdf](http://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf)
- de Vries, M. (2012). Re-use of public sector information. Report for Danish Ministry for Housing, Urban and Rural Affairs.
- Digitaliseringsstyrelsen. (2012). Basic Data. Available at <http://tinyurl.com/dxn2jhh>
- Emerson, J.; Wachowicz, J. and Chun, S. (2001). Social Return on Investment (SROI): Exploring Aspects of Value Creation. Available at <http://hbswk.hbs.edu/archive/1957.html>
- Gigler, S; Custer, S. and Rahmetulla, H. (2011). Realizing the Vision of OGD- Opportunities, Challenges and Pitfalls. World Bank Report.
- Gil-Saura, I., Frasquet-Deltoro, M., and Cervera-Taulet, A. (2009). The Value of B2B Relationships. *Industrial Management & Data Systems*, 109(5), 593-609.
- Grover, V. and Kohli, R. (2012). Cocreating IT Value: New Capabilities and Metrics for Multifirm Environments. *MIS Quarterly*, 36 (1), 225-232.
- Halonen, A. (2012). Being Open About Data. Available at <http://tinyurl.com/c8mz5vt>
- Harrison, T.M.; Guerrero, S; Burke, G.B; Cook, M; Cresswell, A; Helbig, N.; Hrdinová, J; Pardo, T. (2011). Open Government and E-Government: Democratic Challenges from a Public Value Perspective. In Proceedings of the 12th Annual International Conference on Digital Government Research (dg.o 2011), 245-253.
- Hedstrom, P. and Swedberg, R. (1998). *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge: Cambridge University Press, 1-31.
- Henfridsson, O. and Bygstad, B. (2013). The Generative Mechanisms of Digital Infrastructure Evolution, *MIS Quarterly*, forthcoming.
- Hess C. and Ostrom, E. (2006). *Understanding Knowledge as a Commons – From Theory to Practice*. Cambridge, Massachusetts: MIT Press.
- Hujiboom, N., and Van den Broek, T. (2011). Open data: an international comparison of strategies. *European Journal of e Practice*.
- Jansen, K. (2012). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28 (2011), 446-456.

- Kundra, V. (2012). Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect. Jan 2012, Joan Shorenstein Center on the Press, Politics and Public Policy, Harvard.
- Lee, G. and Kwak, J.H. (2011). Open Government Implementation Model: A Stage Model for Achieving Increased Public Engagement. In The Proceedings of the 12th Annual International Conference on Digital Government Research (dg.o.2011), 254-261.
- Linders, D.; Wilson, S.C. (2011). What is Open Government? One Year after the Directive. In The Proceedings of the 12th Conference on Digital Government Research (dg.o.2011), 262-271.
- Mayer-Schönberger and Zappia, (2011). Participation and Power: Intermediaries of Open Data. Conference draft prepared for the Berlin Symposium. Available at: <http://tinyurl.com/bo3pyl8>
- McKinsey & Company. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
- Nilsen, K. (2010), Economic theory as it applies to Public Sector Information. Annual Review of Information Science and Technology, 44(1), 419-489.
- Open Knowledge Foundation (OKF). (2012a). Open Data Handbook version 1.0.0 (revised November 14, 2012). Available at <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>
- Sarker, S.; Sarker, S.; Sahaym, A. and Bjørn-Andersen, N. (2012). Exploring Value Cocreation in ERP Vendor-Partner Relationships. MIS Quarterly, 36(1), 317-338.
- Shapiro, C. and Varian, H.R. (1999) Information Rules: A Strategic Guide to the Network Economy. Harvard Business School Press, Boston.
- Suber, P. (2007). Open Access Overview. Available at <http://tinyurl.com/lb3r3>
- von Krogh, G.; Haefliger, S.; Spaeth, S.; Wallin, M.W. (2012a). Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development. MIS Quarterly, 36(2), 649-676.
- von Krogh, G.; Rossi-Lamastra, C.; Hefliger, S. (2012b). Phenomenon-based Research in Management and Organisation Science: When is it Rigorous and Does it Matter. Long Range Planning, 45(4), 277-298.
- Zuiderwijk, A. and Janssen, M. (2012). A comparison of open data policies and their implementation in two Dutch ministries, In Proceedings of the 13th Annual International Conference on Digital Government Research, pp. 84-89



## **PAPER III: Generating Value from Open Government Data**

In Proceedings of the 34th International Conference on Information Systems (ICIS),  
2013, Milan, Italy

# Generating Value from Open Government Data

Completed Research Paper

**Thorhildur Jetzek**  
Copenhagen Business School  
Copenhagen, Denmark  
tj.itm@cbs.dk

**Michel Avital**  
Copenhagen Business School  
Copenhagen, Denmark  
michel@avital.net

**Niels Bjørn-Andersen**  
Copenhagen Business School  
Copenhagen, Denmark  
nba.itm@cbs.dk

## Abstract

A driving force for change in society is the trend towards Open Government Data (OGD). While the value generated by OGD has been widely discussed by public bodies and other stakeholders, little attention has been paid to this phenomenon in the academic literature. Hence, we developed a conceptual model portraying how data as a resource can be transformed to value. We show the causal relationships between four contextual, enabling factors, four types of value generation mechanisms and value. We use empirical data from 61 countries to test these relationships, using the PLS method. The results mostly support the hypothesized relationships. Our conclusion is that if openness is complemented with resource governance, capabilities in society and technical connectivity, use of OGD will stimulate the generation of economic and social value through four different archetypical mechanisms: Efficiency, Innovation, Transparency and Participation.

**Keywords:** *Open government data, Public sector information, Value generating mechanisms, Openness, Social value, Economic value.*

### III.1 Introduction

Today, we generate and store more data than at any other time in history. Computing and networking capabilities combined with openness enhance the potential impact of the accumulated data, offering society an opportunity to drive massive social, political and economic change (Kundra, 2012). Thus far, there has been apparent tension between the social value driven networks of data, information and knowledge sharing, on the one hand, and the economic value driven products and services markets of ownership and exchange, on the other. However, the current trend towards openness and sharing has led to a cognitive reframing of the roles of, and relationships between, sectors, as well as a blurring of the boundaries between economic markets and social networks. These complex interdependencies are forcing us to re-think the concepts of economic and social value in a world moving towards digital interoperability, information sharing, co-creation and collaborative networks (Avital et al., 2007; Hess and Ostrom, 2006).

Having *Open Government Data* (OGD) implies that the public sector relinquishes its role as information gatekeeper in lieu of a new role as information provider (Davies, 2010). The number of OGD initiatives has grown from two to over three hundred in the period 2009-2013, and membership in the Open Government Partnership (OGP) has gone from eight to fifty-nine countries in just under two years. Over 280 government data catalogs have been published and over a million datasets have been released by governments around the world, spawning new businesses and social projects. OGD is commonly seen as a driver of efficiency and a vehicle for increasing transparency, citizen participation and innovation in society. Despite the potential significance of OGD, emphasized by an abundance of anecdotal evidence, we could not identify many studies on how OGD will contribute to value generations. To-date, the economic and social impact of open-data policies remains largely unclear, and there are scant empirical data available on the effects of the various policy approaches, thus leaving policy makers without the facts they need to assess and improve these policies (Huijboom and Van den Broek, 2011; Zuiderwijk et al., 2012).

Accordingly, this paper addresses the question: *How can use of OGD stimulate value generation?* Building on a synthesis of the OGD literature and established theories of value generation, we have developed a conceptual model that portrays the role of the value generation mechanisms for the generation of value from OGD. In this paper we suggest a nomological network of constructs that together can explain how OGD as a resource can be transformed to value. The model shows how four enabling factors: *Openness, Data governance, Capabilities* and *Technical connectivity* positively affect

four value generation mechanisms: *Efficiency*, *Innovation*, *Participation* and *Transparency*, all of which possess the ability to generate economic and social value, however differently. We conduct a macro level analysis on the impacts of nationwide OGD initiatives. We use empirical data from 61 countries collected from several respectable open data-sources to test these relationships, using the partial least squares (pls) method. The results mostly support our hypotheses on the relationships between the enabling factors, value generation mechanisms and resulting value; however, we did not find significant impacts from openness on efficiency and transparency mechanisms.

Our research has both theoretical and practical implications. The identified constructs and relationships in our conceptual model can be further used to theorize on the implications from the use of OGD in theoretical and practical contexts. We propose that OGD is a common resource, governed by the public sector. As such, it needs to be managed, monitored and protected as any other resource, with the aim of maximizing the potential value generation possibilities now and in the future. Our contribution to practice is the identification of the constructs and items that have the ability to stimulate value generation and the description of the causal relationships between these factors, the different value generation mechanisms and value.

### **III.2 Theoretical background and model development**

#### **Open Government Data**

Open data can be defined as data that are freely accessible online, available without technical restrictions to re-use, and provided under open access license that allows the data to be re-used without limitation, including across different ‘fields of endeavor’ (e.g., commercial and non-commercial alike) (OKF, 2012). The concept of Open Government Data (OGD) refers to government data defined as "data and information produced or commissioned by government or government controlled entities" (OKF, 2012) that are opened up for use and re-use by public and private agents alike. In the currently used terminology, OGD does not include data that are subject to valid privacy, security or privilege limitations, as governed by other statutes. Government data sets are an interesting subset of open data because such subsets have already been collected for specific use, have been paid for by taxpayers, are relevant and offer value beyond what is captured from the originally intended use. When opened up, government data become a common, shared resource (i.e., public good) that is provided by the government. Building on Nilsen (2010) and Shapiro and Varian

(1999), the typical features of OGD as a resource are as follows: 1) It is non-rivalrous; 2) It is not excludable; 3) It has high fixed costs; 4) It has (almost) zero marginal cost; 5) It offers valuable information and has high potential for re-use.

Two distinct ideologies seem to drive most of the OGD initiatives today: the ‘re-use of data’ perspective and the ‘Open Government’ perspective. We thus reviewed the respective tensions and contributions of these two unique streams. The literature on re-use of OGD is mostly focused on the economic value of government data, while the literature on Open Government is in a higher grade directed towards government policy and centered on how use of OGD can contribute to the generation of social value in collaborative settings. One of the most disruptive aspects of OGD is the transformation from a largely closed world to an open, interconnected world. In a closed world, private companies and public organizations each exist within their own administrative boundaries, and relations between them are based on traditional market behavior, rules and regulations. When governments become open, the mechanisms that affect value generation and appropriation move beyond the traditional buyer-seller relationships; thus connections between the public and the private, as well as the social and the economic dimensions begin to emerge.

Ultimately, we could identify four archetypical generative mechanisms. Two of them originate from the Open Government literature: transparency of government and citizen participation/collaboration (Bertot et al., 2010, 2012; Harrison et al., 2011; Linders and Wilson, 2012) and two from the re-use literature: efficiency and innovation (Gigler et al, 2011; Halonen, 2012; Janssen, 2011). We also identified a number of barriers to value generation and appropriation: 1) closed or inaccessible datasets, 2) lack of comprehensive data policies, 3) lack of validity, completeness and exhaustiveness of datasets, 4) insufficient metadata, as well as lack of technical and semantic interoperability, 5) lack of consistency in cross-border access regimes, 6) lack of motivation within the public sector, 7) lack of technical skills within the public sector, 8) lack of data literacy and technical ability and 9) too fragmented and disparate open data community (Bertot et al. 2010; Davies, 2010; Dawes, 2012; Halonen, 2012; Jansen, 2011; Janssen et al. 2012; Lee and Kwak, 2011; Mayer-Schönberger and Zappia, 2011; Zuiderwijk et al., 2012). In order to overcome these barriers, we propose that governments should focus on four enabling factors: Openness, resource governance, capabilities and technical connectivity.

## The multifaceted nature of value

The conceptual distinction between value generation and value appropriation has been growing in importance with the trend towards openness, technical connectivity and collaborative ventures. Value generation materializes when the utility of society's members increases after accounting for the resources used in that activity. Value appropriation materializes when an actor is able to capture a portion of the value created by an activity (Bowman and Ambrosini, 2000). The relationship between the generation and appropriation of value for multiple stakeholders in the context of alliances is increasingly being viewed to be multifaceted in nature (Sarker et al., 2012). Gil-Saura et al. (2009) suggest that value is a subjective, multidimensional construct; accordingly, it is only through a multidimensional view that we get a true picture of the value generated within alliance relationships. However, despite the recent focus on the creation of value in collaborative settings, little is known about the underlying mechanisms (Sarker et al., 2012).

Two types of value are frequently discussed: economic value, defined as the worth of a good or service as determined by the market, and social value, which is created when resources, inputs, processes or policies are combined to generate improvements in the lives of individuals or society as a whole (Emerson et al., 2001). However, the shifts toward an economy centered on information and the move to a networked Internet-based environment have caused significant attenuation of the limitations that market-based production places on the pursuit of value (Benkler, 2006). Due to these trends, the perceived divergence between the generation of social and economic value is becoming increasingly contested. In a recent attempt to amalgamate the concepts of social and economic value generation, Porter and Kramer (2011) introduced the term shared value. Shared value is based on the idea that societal needs, not only economic needs, define markets. It also recognizes that social harms frequently create internal costs for companies, such as wasted energy, health problems and/or the need for remedial training to compensate for inadequacies in education (Porter and Kramer, 2011). The generation of shared value is essentially about expanding the total pool of economic and social value.

In welfare economics, the term social welfare is used to describe a construct that reaches further than the commonly used measure of Gross Domestic Product (GDP) as an indicator of value. As GDP mainly measures market production, it ignores many of the determinants of social value such as the environment, freedom, health and education (Fleurbaey, 2009; Michaelson et al., 2009). For example, traffic jams may

increase GDP as a result of the increased use of gasoline, but obviously do not improve the quality of life. Moreover, if citizens are concerned about the quality of air, and air pollution is increasing, then statistical measures which ignore air pollution will provide an inaccurate estimate of what is happening to citizens' well-being (Stiglitz et al., 2009). Many have attempted to construct an aggregate measure of social welfare by using social indicators (Fleurbaey, 2009); however, there is still no general consensus on which indicators give the most reliable evidence.

Another value concept often seen in the e-government literature is public value. The central proposition in Moore's (1995) public value framework is that public resources should be used to increase value, not only in an economic sense but also more broadly in terms of what is valued by citizens and communities. Benington (2011) further extends Moore's (1995) definition of public value and adds the dimension of "What adds value to the public sphere" (Benington, 2011, p. 31), where the public sphere includes state, market and civil society. The dimensions of public value include: 1) economic value – adding value to the public realm through the generation of economic activity, 2) social and cultural value – adding value to the public realm by contributing to social cohesion and well-being, 3) political value – adding value to the public realm by stimulating and supporting democratic dialogue and active public participation, as well as citizen engagement and 4) ecological value – adding value to the public realm by actively promoting sustainable development and reducing pollution, waste, and global warming (Benington, 2011).

This leads us to the question of how we can specify the relationship between use of OGD as a resource and the generation of value. This is not necessarily a direct relationship. Value is produced by value generating mechanisms which reveal the instrumental pathways by which OGD is transformed to value. An identification of these mechanisms allows us to specify how and what value can be generated from OGD (Harrison et al., 2011). Moreover, as OGD is a strategic, open resource, it holds great potential for a number of stakeholders, including public sector agencies, private businesses, academia, citizens and civic organizations (Ubaldi, 2013). All of these might be able to co-create value or simultaneously appropriate the generated value. Therefore, we will view the relationship between OGD and value from a societal standpoint, where different mechanisms have the ability to generate outcomes that positively affect the latent concept of social and economic value, reflected by social welfare indicators like higher income, better health or increased wellbeing.

## Conceptual model

The conceptual model in Figure 1 illustrates the constructs and relationships that are the basis for the nomological network of OGD value generation. In the next subsections, we explain the individual constructs, and present our hypotheses regarding these relationships.

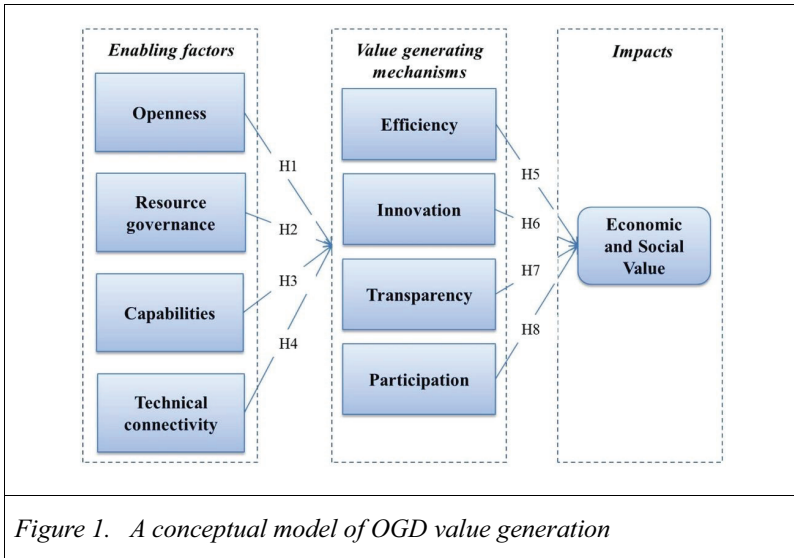


Figure 1. A conceptual model of OGD value generation

### Openness

Opening access to government data implies that public organizations participate in a process which has to consider influences, discourses and exchanges to be constructive; further, it welcomes opposing views and inputs, consequently giving up control, at least to some extent (Janssen et al., 2012). This transformation from a closed system to an open system reframes the whole context of government, having potentially far-reaching effects on both the public and the private sector. Publishing various policy documents online for increased transparency and accountability, using openness and interoperability of data to increase efficiency, enabling citizens to participate in the data generating and reviewing process and enabling access to commercially valuable information for innovation purposes are the key objectives of OGD initiatives (Davies, 2010; Karunasena et al., 2011; Vickery, 2011). Providing OGD can be seen a matter of availability, format, accessibility and license (Davies, 2010). We conceptualize



openness as a construct that has four components: 1) use of open licenses that allow for commercial re-use, 2) extent of the OGD initiative 3) availability of various types of data, and 4) usability and accessibility of available data-sources. This leads to the first hypothesis:

**H1:** *Openness positively affects the ability of society to generate value from data through the four value generation mechanisms (H1a: Efficiency, H1b: Innovation, H1c: Transparency and H1d: Participation).*

### **Resource governance**

Policies, governance mechanisms and a variety of skills and capabilities within government are needed to reap the maximum benefits from open data initiatives (Dawes, 2012). Once government has opened up, OGD become a common, shared resource, available for use within an open network of public and private stakeholders. However, this resource is still governed by the public sector as the main creator of the data. In order to ensure the equitable, efficient and sustainable use of any resource, it needs to be managed, monitored and protected (Hess and Ostrom, 2006). We conceptualize resource governance as a multi-dimensional construct that encompasses all aspects of data management, as well as the leadership and skills needed to ensure the sustainable and equitable use of the resource.

Data management must ensure that the data are of sufficient quality. Data quality can be defined as a multi-faceted term comprising utility, objectivity and integrity (Batini and Scannapieco, 2006). Unknown, inconsistent or unsatisfactory quality of OGD leads to substantial risks for validity and relevance. As errors in data become most evident when the data are used, governments could utilize the power of the network to get feedback for improvement of data, thus saving internal costs while improving the quality of the resource. It is important to give the correct context to the data, as government data are in many cases collected or created for specific purposes, and thus could be misleading if taken out of that context (Dawes, 2012). Accordingly, use could be stimulated if more information about the way open data are collected and processed were to be provided by including metadata (Zuiderwijk et al., 2012). The security of information must also be considered. A holistic approach to data management begins with an understanding of the information life cycle—the collection, updating, processing, and eventual deletion of personal information—and the adoption of a technology framework that enables governments to set controls which safeguard individuals' privacy (Lampri, 2012).

Governance must ensure the sustainability of the resource and the initiative (Hess and Ostrom, 2006). Sustainable systems are those that meet current needs of many individuals involved in producing, deciding and using a common resource without compromising the ability of future generations utilizing the resource (Hess and Ostrom, 2006). Ongoing analysis of more than 900 major change initiatives in the public sector indicates that 61 percent of those initiatives do not yield the hoped-for impact, and that a major factor in such cases is a lack of the skills, mind-sets and behaviors critical to sustaining change (McKinsey, 2012). Developing and maintaining e-Government projects is a continuous process of policy development, investment planning, innovation, learning and change management. This process must fit with, and respond to, a dynamic development strategy that supports evolving national goals and creates sustained institutional reforms. Key to further benefits, whether economic or social, is managerial leadership and political support (Bertot et al., 2010, Heeks, 2003; McKinsey, 2012). Releasing OGD requires a certain level of information and ICT capabilities by civil servants (Gigler et al., 2011). Therefore, an emphasis on public officials' data literacy and technological skills is highly relevant for efficient dissemination of OGD data.

***H2:** Resource governance increases the long term intrinsic value of OGD and therefore positively affects the ability of society to generate value from OGD through the four value generation mechanisms (H2a: Efficiency, H2b: Innovation, H2c Transparency, H2d: Participation).*

## **Capabilities**

As OGD is a common, shared resource, the generative ability of the value generation mechanisms is dependent on certain capabilities in society. A capability can be defined as a measure of the ability of an entity to achieve an objective. In our case it is the collective ability of individuals and organizations to use and re-use OGD. Following Hess and Ostrom (2006) we emphasize the need for equitable use of the data resource. Citizen's access to the Internet and their ability to utilize the provided information are considered important for ensuring equitable dissemination (Bertot et al., 2010; Gurstein, 2011; Halonen, 2012). The digital divide can be broadly defined as the gap between those who have access to technologies and those who do not; however, there are in fact multiple divides that can exist, of which access to the ICTs is but one. Those issues include technology literacy as the ability to understand and use technology and the ability of persons with disabilities to access the content through adaptive technologies (Bertot et al., 2010). Without the capabilities to access, use and make

sense of data, the generative ability of the transparency and participation mechanisms becomes limited, and innovation opportunities and efficiency gains might be forgone. We conceptualize capabilities as a multi-dimensional construct that encompasses equitable access opportunities as well as technology and data literacy.

*H3: Capabilities positively affect the ability of society to generate value from OGD through the four value generation mechanisms (H3a: Efficiency, H3b: Innovation, H3c: Transparency, H3d: Participation).*

### **Technical connectivity**

The technological backbone of any OGD initiative is an infrastructure that facilitates data exchange between government agencies and the public (Gigler et al., 2011; McKinsey, 2011). Furthermore, in order to enable effective data use, accessibility of the infrastructure is important (Gurstein, 2011). The ability to store and aggregate data and then use the derived data sets to perform analysis has increased in conjunction with recent technical developments (McKinsey, 2011). To enable value generation from OGD, public and private organizations will have to deploy technologies that can help them integrate, analyze, visualize and consume the growing torrent of available data (McKinsey, 2011). For instance, schematic heterogeneity and lack of consistency can hinder value generation by decreasing usability and complicating access as well as integration of the data. Due to the decoupling of data from its original creation context and the increasing need to simultaneously analyze structured, semi-structured and unstructured data, semantic interoperability, identity resolution and ontologies are becoming central methodologies to ensure consistency and meaningful results while allowing third parties the ability to connect different data-sources (Alani et al., 2007). While the diffusion of these technologies is in early stages, Gartner has named semantic technologies as one of the top technology trends impacting information infrastructure in 2013.

Technical connectivity is conceptualized as a construct that describes the availability of technologies that allow users to store, access, combine and analyze the data. This includes availability of the infrastructure and use of semantic technologies, as well as data management; analytics and discovery software, plus the use of multiple platforms to enable general access to content. We propose that technological connectivity positively influences efficiency through increased interoperability and innovation by introducing new ways to transform data. Furthermore we propose that technical connectivity has the ability to influence transparency by enabling accessibility and

sense-making. Finally, it facilitates e-participation by offering new channels for participation.

*H4: Technical connectivity positively affects the ability of society to generate value from OGD through the four value generation mechanisms (H4a: Efficiency, H4b: Innovation, H4c: Transparency, H4d: Participation).*

## **Efficiency**

The importance of the efficient use of public resources for economic growth and stability, as well as for general well-being has been brought to the forefront by a number of developments over the past decades (Afonso et al., 2010). As a consequence of increasing government intervention in affairs such as child care, education, and health services, public sectors have faced mounting difficulties in managing efficiently the administrative bureaucracy. Increasing cross-boundary interactions and higher levels of information exchange between citizens and government have increased the total amount of government data collected and stored. These trends call for more efficient processing of data in order to provide the expected services (Cordella, 2007). The two main goals of digitization in government are to increase agency efficiency and offer benefits to citizens (Axelsson et al., 2013). Efficiency of public sector organizations can be gained by cutting processing costs, making strategic connections between and among government agencies, and creating empowerment (European Commission, 2006). We propose that the public sector can use openness, for instance by allowing citizens to access and manage their own data, to deliver public services more efficiently while safeguarding the quality of services. As a result, resources can be moved from non-value adding tasks to value-adding tasks, positively affecting the generation of value.

The aim of efficiency is to improve resource allocation so that waste is minimized and the outcome value is maximized, given the same amount of resources. Efficiency gains can be the result of a decrease in transaction costs. Transaction costs are a part of the administrative burden of regulation for the private sector and are often consequences of the complexity and the uncertainty of the economic system. Different types of transaction costs include search and information costs, bargaining costs, monitoring costs and enforcement costs. Use of technology, standardized work procedures and free flow of information has been shown to reduce the need for normative, rule-based mechanisms of coordination (Cordella, 2007). Following this rationale, e-government is often described as the right move to implement the changes that are needed to increase the efficiency of public organizations' performance (Cordella, 2007). Openly

sharing data can reduce search costs and make bargaining and monitoring easier, as well as enable faster and easier access to information, increased automation, standardization and interoperability.

An example case where OGD is used to increase efficiency is the Danish Basic Data Program (Digitaliseringsstyrelsen, 2012). The aim is to improve the quality of all basic data registers, create a common data model for all the basic registers in Denmark and implement a common platform where both public and private users can get access to the data. As a result of these changes, coordination should get easier and the possibility for automated business processes across authorities should be increased. Furthermore, as data will be freely available online, transaction costs related to user support and billing will be diminished. The total annual savings for the public sector are projected to be around €35 million (Digitaliseringsstyrelsen, 2012). Moreover, integrated government data of better quality will also benefit private industries, such as real estate dealers, insurance companies, the financial sector and the telecom industry, which previously had to spend significant resources on creating usable information from heterogeneous data-sources. The cost-savings for the private industry are estimated to be around €65 million per annum when the program is fully implemented.

*H5: Efficiency positively affects value through decreased transaction costs, enabling better resource allocation and increased quality of public services.*

## **Innovation**

The Oslo Manual defines innovation as the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organization or external relations (OECD, 2005). Recent technological developments have provided firms with the ability to collect, manage and use different types of data in multiple ways to innovate, and subsequently create value (Koski, 2013). By 2015, innovation around data is projected to help create 4.4 million IT jobs globally (Gartner, 2011). We define data-driven innovation as an iterative process initiated by the perception of new markets and/or new service opportunities for a technology based invention, based solely or in part on OGD, thus leading to the development, production, and marketing of the process, product or service and the creation of new businesses. We propose that the recent developments in technical connectivity as well as in data analytics technologies and the increasing wealth of OGD available are the foundation for innovation from OGD. Moreover, the governance of the data and user's capabilities to access and transform the data is proposed to be influencing factors.

Following Schumpeter (1934), we assume that innovation can have economy-wide effects. Innovation is the source of value creation in Schumpeter's economic theory, bringing about novel combinations of resources, new production methods, as well as new products and services, which, in turn, lead to the transformation of markets and industries, thus increasing value. Numerous studies have confirmed the relationship between macro-level business innovation and economic value (most often conceptualized as economic growth). The social impacts of new innovations have, however, been much less discussed and analyzed, with the possible exception of Simon Kuznets (1974), who separated economic and non-economic consequences of technological innovations. Pol and Ville (2009) discuss the concept of social innovation which in their definition is generally directed at improving the quality and/or quantity of life. Social and business innovation can overlap as business innovation, while mostly dealing with profitable new ideas, can also result in social value generation. Going forward, we conceptualize data-driven innovation as business innovation capable of positive economic and social impacts.

*H6: Data-driven innovation positively affects value through generation of new knowledge, new processes, services and products, and new businesses.*

## **Transparency**

Most definitions of transparency recognize the extent to which an entity reveals relevant information about its own decision processes, procedures, functioning and performance. However, opening access to chosen public documents does not necessarily contribute to a transparent government (Gurstein, 2011; Yu and Robinson, 2012). A government can provide open data on politically neutral topics, even as it remains deeply opaque and unaccountable (Yu and Robinson, 2012). Accordingly, we cannot conclude that open access to government data is in itself a comprehensive measure of transparency. Rather, we propose that transparency is dependent on four enabling factors: openness (quantity of disclosed data, accessibility and ease of access), resource governance (through trustworthiness of data), capabilities (the general degree to which citizens are capable of accessing and interpreting the data) and technical connectivity (access to technologies used to disseminate and make sense of data).

In order to explain how the generative mechanism of transparency enables value generation, we can utilize the concept of Information Asymmetry. Information Asymmetry describes situations where one party has more or better information than another while participating in transactions, negotiations or communications. Information asymmetry can cause all sorts of sub-optimal results and behaviors, such

as Moral Hazard, where the more informed make decisions to their own benefit, with the cost falling on others. Corruption (defined in the context of government as misuse of public power for private benefits) has long been seen as a hindrance to socio-economic development. Corruption has been shown to have negative effects on GDP growth, poverty, human development and health outcomes. It seems that corruption destroys a society's social capital and has a very negative impact on people's life satisfaction (see overview in Holmberg et al., 2009 and Rothstein, 2011). This is particularly true in developing, resource rich countries, where the politically elite often has control of resources and resource rents, as well as control over patronage and the distribution of resources, a situation known as the "resource curse" (Kolstad and Wiig, 2009).

A transparency agenda for tackling illicit financial flows (the hidden company ownership that makes such flows possible, land grabs, and the secrecy by which big oil, gas and mining corporations are doing business) was presented by the British Prime Minister, David Cameron, in the G8 meeting at the World Economic Forum held in Davos in January, 2012. The claim is that citizens in developing countries are regularly robbed of the benefits of their countries' mineral wealth through poorly negotiated or corrupt backroom deals. The G8 is leading efforts to require oil, gas and mining companies to publish key financial information for each country and project they work on. The aim is to ensure that revenues from oil, gas and mining can help developing countries to forge a path to sustainable growth, instead of fuelling conflict and corruption.

*H7: Transparency positively affects value through decreased information asymmetry and less corruption.*

## **Participation**

Public participation allows members of the public to contribute ideas and expertise so that their government can make policies with the benefit of information that is widely dispersed in society. By this definition, participation provides citizens with a voice in government (Linders and Wilson, 2011). Public participation can take two discrete forms: 1) collecting opinions (citizen engagement) and 2) collecting ideas and solutions through crowdsourcing (Linders and Wilson, 2011). It has been argued that more involved democratic participation is likely to lead to superior social outcomes because of participation's role in aggregating information and preferences (see for instance Barber, 1984). The relationship between participation and OGD is essentially twofold: In order for citizens to participate, either by voicing their opinions towards

policy making or by participating directly in public projects, they must first learn about the particular issue that is being addressed, assimilating the facts and arguments. Second, they need a platform where they can make their contributions available. The evolution of citizen participation in public administration decision-making has been facing a new phase, as many government agencies have taken advantage of internet-based applications to communicate with constituents in order to provide online application services (Kim and Lee, 2012). The idea is to lower the barriers to participation for those willing-but-unable, and to make participation more attractive to those able-but-unwilling (Axelsson et al., 2010).

Participation mechanisms generate value through the synergies created from openness and sharing, allowing the public sector to draw from a larger pool of resources and to improve its ability to solve difficult social problems. The theory of collaborative advantage proposes that the real advantage from collaboration comes from something being achieved that could not have been achieved by any one of the participating stakeholders acting alone (Vangen and Huxham, 2010). A similar theoretical argument is used in the literature on Open Innovation (Chesbrough et al., 2006) where the principal idea is that an open approach to sharing knowledge across boundaries expands the firm's innovative potential, as the firm is able to tap into a much larger pool of ideas and find such ideas faster. Therefore, collecting ideas and crowdsourcing difficult tasks should lead to improved ability to solve difficult social problems such as environmental degradation or disaster management.

A good example of citizen collaboration around the generation of data is the crowdsourcing activities that have been immensely helpful in a natural disaster such as the 2010 earthquake in Haiti (Lee and Kwak, 2011). Just a few hours after the earthquake, the OpenStreetMap (OSM) Community began tracing roads from imagery that was previously available from Yahoo, and within 48 hours, high resolution imagery taken post-earthquake became available from volunteers. Over the course of the next month, over 600 people contributed information to the OSM. OSM communities have continued to work with the Government of Haiti to further develop the OSM data.

***H8: Participation positively affects value through improved problem solving resulting from participation and collaboration of citizens with government.***



### III.3 Measurements and data collection

We used several open secondary data sources for the variables in our study, described in the following subsection. All data were collected for the year 2011, except for the indicators from the United Nations E-Government Survey 2012 (Government online services index and Infrastructure index) which are from 2012. The sample collected included 61 observations from 61 countries, limited by the number of countries represented in the Open Data Index from the World Wide Web foundation (Farhan et al., 2012). All measures, sources and item wordings or descriptions of data can be provided if requested.

#### Operationalization of research variables

Providing OGD is a matter of availability, accessibility, format and license (Davies, 2010). We conceptualize openness as a formative construct that has four components: use of open licenses, extent of OGD initiative, availability of data, and accessibility of data. All measurements come from the Web Index survey (Farhan et al., 2012). The survey consists of a detailed questionnaire submitted to experts/professionals from 61 countries worldwide and assessed by national and regional peer reviewers (Annoni et al., 2012). The indicators are: *Government use of open licenses*, *Extent of OGD initiative*, *Ease of access of government data*, and *Availability of government data*, calculated from the scores of different questions determining the online availability of different types of government data. When testing the research model, we received some insignificant and even negative (Ease of access) weights for the formative indicators. We were interested in keeping all indicators, as they represent different dimensions of openness and, from a content validity perspective, it did not seem appropriate to eliminate any of them. Thus, we constructed a composite index from the average score of all indicators (see Cenfetelli and Bassellier, 2009; Petter et al., 2007).

The Data governance construct is a formative variable with three dimensions that reflect: a) data management policies that affect the quality, relevance and usefulness of presented information, b) leadership within public sector and c) relevant skill-sets within the public sector. For the first indicator, we used the level-II sub-index from the United Nations Government *Online services index*, which reflects the general level of the quality, relevance and usefulness of online information (UN, 2012). In order to measure government leadership and motivation for using OGD and technology to initiate the mechanisms discussed earlier, we used three measures. The first two, *Importance of ICT to government vision of the future* and *Government prioritization of*

*ICT* from the World Economic Forum (WEF) (see Schwab et al., 2011), are used to indicate whether the government in question has a clear e-Government strategy and is committed to keep using information and communication technologies to improve the overall competitiveness of a country. The third measure is *Participation in Open Government Partnership* (from the OGP website), a dummy variable used to indicate the government's commitment to open government. Finally, to measure the technical skills within the public sector, we used an indicator from the World Wide Foundation that indicates the extent to which Government programs specifically focus on funding ICT training for their staff.

The Capabilities formative construct is based on three dimensions: equitable access opportunities, affordability and training. In order to measure attention to equitable dissemination of the resource, we created a measure based on data from World Wide Web Foundation which we call *Web use by disabled people*. This is based on the average score from seven different questions measuring the extent of effective and useful access to the web for people with different types of disability. To measure affordability, we use the indicator *Affordability of web access* (World Wide Web Foundation). In order to capture any kind of value from data, a measure of data related skills is needed (data management, data literacy, etc.). Therefore, we take the *Extent of staff training* (World Economic Forum) in different countries into consideration, reflecting the importance of vocational and continuous on-the-job training for ensuring a constant upgrading of workers' skills.

Technical connectivity is a reflective construct that is composed of three dimensions: a) the availability of technical and telecommunications infrastructure in the country in question b) use of different platforms to disseminate and access data and c) the firm level availability of recent technologies. The indicators used are: 1) The United Nations *Telecommunications Infrastructure Index*; which is a composite weighted average index based on six basic infrastructural indicators that define a country's ICT infrastructure capacity. These are: PC's/1000 persons; Internet users/1000 persons; Telephone Lines/1000 persons; Online population; Mobile phones/1000 persons; and TV's/1000 persons. 2) *Accessibility of digital content*, measuring accessibility of digital content via multiple platforms 3) *Firm level Technology absorption* (both from World Economic Forum's Executive Opinion Survey, 2011-2012).

The adequate measurement of public sector efficiency is a difficult empirical issue, and there is scarce literature on the subject. However, some progress has been made by shifting the focus of analysis from the number of resources used by ministry to the

services delivered or outputs achieved (Afonso et al., 2010). Quality adjustments do also present a challenge; if quality of outputs is not properly taken into account when measuring efficiency, an underestimation may result. We measure the efficiency construct based on three different indicators, all reflecting some aspect of public sector efficiency and effectiveness of output. Our first indicator is *ICT use and government efficiency* from the World Economic Forum's Global Competitiveness Report (Schwab et al., 2011). This indicator shows citizen's perceptions of government efficiency as a result of digitization. The second indicator is the World Bank's governance indicator *Government Effectiveness*. This indicator aims to measure the quality of public service delivery by covering a broad range of related concepts: red tape, quality of public schools, government stability, bureaucrats' expertise, policy consistency and ability to deliver basic infrastructure (Van de Walle, 2006). The third indicator is World Bank's *Ease of doing business index* (International Bank for Reconstruction and Development). This indicator documents various efficiency and effectiveness impacts connected to the life cycle of business, such as the number of procedures to start a business, the time and cost of achieving a regulatory goal or complying with regulation and disclosure.

We model Innovation as a reflective variable with two indicators. For the first one, we used a measure from the World Wide Web Foundation that measures the direct effects of OGD on the creation of new products and services: *Creation of new applications and services based on government data*. However, anecdotal evidence shows how government data and other data are combined and analyzed, resulting in insight and knowledge that may lead to new technology-based innovations further down the value ecosystem (McKinsey, 2011). Therefore, we added a measure that reflects the development of new businesses: *Business development based on the Web* from World Wide Web Foundation.

The lack of shared meaning and understanding of the transparency concept has made it difficult to operationalize (Relly et al., 2009). A transparent government should be committed to disclosure, thus that there should be low levels of information asymmetry and corruption, and citizens should have the means to act upon corrupt behavior. For disclosure, we used *Transparency of government policymaking* from World Economic Forum (Schwab et al., 2011) combined with the existence of Freedom of Information Laws (dummy variable). To measure information asymmetry and corruption, we used *Level of undocumented extra payments or bribes*, based on the average score across five components of the following World Economic Forum's Executive Opinion Survey questions: In your country, how common is it for firms to make undocumented extra

payments or bribes connected with (a) imports and exports, (b) public utilities, (c) annual tax payments, (d) awarding of public contracts and licenses and (e) obtaining favorable judicial decisions. Finally, to measure citizen's ability to act upon corrupt behavior, we used the indicator judicial *independence* from World Economic Forum.

Participation is a reflective construct, but we used only one indicator, the United Nations *e-Participation Index*, which measures 1) the use of the Internet to facilitate provision of information by governments to citizens, 2) interaction with stakeholders and 3) engagement in decision-making processes (UN, 2012). A country's e-participation index value reflects how useful these features are and how well they have been deployed by the government, compared to all other countries. The reason is that other measures of participation are typically measures of democracy and therefore too broad to capture the type of participation that is derived from the combined effects of data, openness and use of technology.

Value (or social welfare) is conceptualized as an aggregate measure of social and economic value. The challenges of constructing a global measure of welfare by using composite indicators are a much-discussed theme (Eisler, 2007; Stiglitz et al., 2009). In particular, we need to identify the key indicators and then determine the way in which these indicators can be brought together to make a coherent system. Structural Equation Models and specifically the PLS has been recommended as a means to model the statistical relations between such indicators (Trinchera and Russolillo, 2012). We follow Stiglitz et al. (2009) who recommended the following sub-indicators to measure total welfare: i. Material living standards; ii. Health; iii. Education; iv. Personal activities including work v. Political voice and governance; vi. Social connections and relationships; vii. Environment; and viii. Insecurity of an economic, as well as a physical, nature. We follow these recommendations with one exception: As our construct is reflective and as we assume that our indicators reflect the existence of a certain level of aggregate welfare or value, we do not use political voice and governance, which indicate means rather than ends. Moreover, as Gallup's Global Wellbeing index (Gallup, 2011) measures daily experiences (well-rested, shown respect, smiling/laughter, learning/interest, enjoyment, physical pain, worry, sadness, stress, and anger). We use it to reflect three of the dimensions: personal activities, social connections and relationships, and insecurity. Other reflective indicators are: 1) for economic performance, GDP/capita from the World Bank; 2) for health: UN's Human Development Index, health sub-index; 3) for education, UN's Human Development Index, education sub-index and 4) for environment, the natural resource

management index (Center for International Earth Science Information Network, 2011).

### III.4 Analysis and findings

We used the `pls`-package for the open-source software R (Sanchez, 2013). The method chosen was partial least squares (PLS), as our research is exploratory due to the emergent state of the phenomenon and we have no prior models to confirm or test. Other reasons for using the PLS-method are the use of formative constructs, the small sample size, the complexity of the structural model and the wish to identify key drivers (Hair et al., 2011; Hair et al., 2012, Ringle et al., 2012). Since PLS is based on a series of OLS regressions, it has minimum demands regarding sample size, and generally achieves high levels of statistical power (Hair et al., 2011). A common rule of thumb regarding sample size in PLS is to use ten observations per predictor, where the sample size is the largest of two possibilities: 1) the block with the largest number of indicators or 2) the dependent variable with the largest number of independent variables impacting it (Chin, 1998; Hair et al., 2011; Marcoulides and Saunders, 2006). In our case, constructs are made from a maximum of five indicators, and value has the largest structural equation with four direct paths pointing towards it.

**Table 1: All variables**

<b>Construct</b>	<b>Type</b>	<b>Measure</b>	<b>#</b>	<b>R<sup>2</sup></b>	<b>1<sup>st</sup> eigenvalues</b>
Openness	Exogenous	Formative	4		1.00
Data governance	Exogenous	Formative	3		2.17
Capabilities	Exogenous	Formative	3		2.37
Technical connectivity	Exogenous	Reflective	3		2.72
Efficiency	Endogenous	Reflective	3	0.864	2.55
Innovation	Endogenous	Reflective	2	0.820	1.65
Transparency	Endogenous	Reflective	3	0.706	2.66
Participation	Endogenous	Reflective	1	0.551	1.00
Value	Endogenous	Reflective	5	0.746	4.14

However, as Marcoulides and Saunders (2006) point out, it is necessary to consider other characteristics of the data and model in order to ensure sufficient sample size to achieve adequate statistical power. First, we built the research model according to the current knowledge, and then collected data to test the model. Next, we performed data screening. All sources had a good reputation, and the same methodology was applied to all countries for each indicator. Missing data or departures from normality influence sample size requirements of a study and potentially deteriorate power (Marcoulides and Saunders, 2006). There were no missing data and all rows showed a reasonable degree of normality (kurtosis  $< |1.5|$ , skewness  $< |1|$  except for GDP pr. capita where it was 1.22, which we solved by converting GDP to a logarithmic scale. Based on Marcoulides and Saunders (2006), it seems that a sample size of 61 gives adequate power to draw inferences for this particular model, as both factor intercorrelations and factor loadings are high; however, we have to consider that the small sample size might affect the results. More countries will be included in the next version of the Web Index (2012), and when this version becomes available, we can retest the model with a bigger sample.

We used bootstrap validation for loadings, weights and paths with 500 bootstrap samples, for which the number of cases was 100, approximately equal to the number of observations (Hair et al., 2011). All measures were standardized before running the algorithms. One of the concerns with formatively measured constructs is multicollinearity across the indicators of each constructs. High first eigenvalues can be an indicator of multicollinearity; however, all formative variable's first eigenvalues are lower than three, as shown in table 2.

<b>Construct</b>	<b>Item</b>	<b>Loading</b>	<b>Weight</b>	<b>t-value</b>	<b>VIFs</b>
Openness	Summated scale	1.000	1.000		
Resource governance	Data governance	0.903	0.518	6.34***	1.80
	Leadership	0.847	0.369	4.87***	1.82
	Skills	0.783	0.281	2.79***	1.66
Capabilities	Equitability	0.855	0.219	3.15***	2.65
	Affordability	0.910	0.454	4.61***	2.64
	Training	0.884	0.452	6.45***	1.87

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

All Variance Inflation Factors (VIFs) were below the recommended 5.00 value (Hair et al., 2011). We checked for insignificant or negative weights (Centefelli and Bassellier, 2009; Petter et al., 2007), but all weights were significant and positive.

To evaluate the reflective measures in our model, we followed the recommendations of Hair et al. (2011). Table 3 presents the results of these quality measures.

<b>Table 3: Loadings, composite reliability, convergent validity and discriminant validity</b>						
<b>Construct</b>	<b>Item</b>	<b>Load.</b>	<b>C.Alpha</b>	<b>DG.rho</b>	<b>AVE</b>	<b>MaxCorr<sup>2</sup></b>
Technical connectivity	Infrastructure	0.945	0.948	0.967	0.906	0.85
	Diffusion	0.938				
	Accessibility	0.973				
Efficiency	ICT related efficiency gain	0.863	0.910	0.944	0.849	0.83
	Government effectiveness	0.967				
	Ease of doing business	0.931				
Innovation	Innovations	0.893	0.789	0.905	0.825	0.74
	New businesses	0.923				
Transparency	Transparency of policy	0.900	0.936	0.959	0.887	0.78
	Undocumented payments	0.960				
	Judicial independence	0.963				
Participation	e-participation	1.00	1.00	1.00	1.00	0.49
Value	Education	0.940	0.948	0.960	0.829	0.82
	Level of health	0.906				
	GDP	0.954				
	Environment	0.905				
	Wellbeing	0.840				

Hair et al.'s advice regarding internal consistency reliability is that composite reliability should be higher than 0.70 (in exploratory research, 0.60 to 0.70 is considered acceptable). For indicator reliability, they recommend that indicator loadings be higher than 0.70. For convergent validity, the rule of thumb is that the average variance extracted (AVE) should be higher than 0.50. Finally, for discriminant validity, two different test are recommended: 1) the AVE of each latent construct should be higher than the construct's highest squared correlation with any other latent construct (Fornell–Larcker criterion) and 2) an indicator's loadings should be higher than all of its cross loadings, which is valid for all items.

The results from the pls-analysis are presented in Table 4 and Figure 2.

**Table 4. Path coefficients and significance**

<b>Relationship</b>	<b>Path coeff.</b>	<b>t-statistics</b>	<b>Relationship</b>	<b>Path coeff.</b>	<b>t-statistics</b>
OP->EFF	0.09	1.53*	CAP->TR	0.37	4.63***
OP->INN	0.25	2.31**	CAP->PA	0.11	0.82
OP->TR	-0.02	-0.26	TECH->EFF	0.39	7.65***
OP->PA	0.33	2.28**	TECH->INN	0.28	3.23***
GOV->EFF	0.29	4.39***	TECH->TR	0.38	6.38***
GOV->INN	0.19	2.1**	TECH->PA	0.1	0.68
GOV->TR	0.15	1.77**	EFF->VAL	0.21	3.03***
GOV->PA	0.28	1.94**	INN->VAL	0.36	3.52***
CAP->EFF	0.22	3.15***	TR->VAL	0.17	2.11**
CAP->INN	0.26	3.61***	PA->VAL	0.24	2.37**

\*p<0.1, \*\*p<0.05, \*\*\*p<0.01

We cannot conclude that openness positively influences efficiency of government (H1a), as the path is only significant at  $p < .1$ . While we see a reason for concern regarding the existence of this relationship, more evidence is needed before we conclude that there is no relationship between openness and efficiency and effectiveness of government. There are four alternative explanations for the insignificance of this relationship: 1) *Data related issues*: The indicators we used for openness are from the first issue of the Open Data Index (World Wide Web Foundation), and data collection methods are currently being reviewed (Annoni et al.,

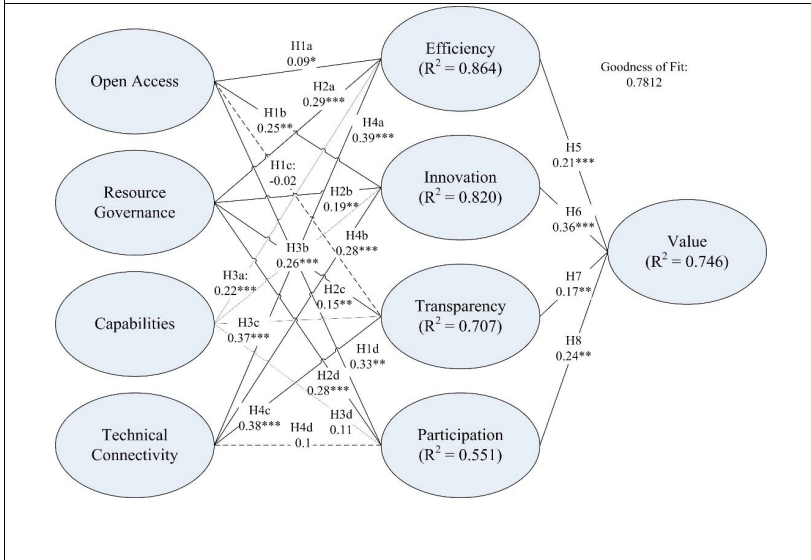


2012). The model will be re-tested when new data become available. 2) *Sample size*: With only 61 countries to test, the small sample size can lead to low accuracy of estimates and decreased statistical power. Low statistical power increases the probability of a Type II error and could lead to us failing to reject a false null hypothesis, falsely concluding that there is no relationship between openness and government efficiency. Again, the model will be retested when data for more countries become available. 3) *Misspecification*: Misspecification of the model can also lead to a Type II error. 4) *Time effect*: Due to the embryonic state of most OGD initiatives and lack of anecdotal evidence on efficiency gains from OGD, the effects from openness might not yet have materialized.

We can support the hypothesis that openness positively influences innovation mechanisms (H1b). If we look at the responses from the *Creation of new services based on government data* survey question (World Wide Web Foundation), we can see that many countries already report that there has been extensive development of new web applications and services based on government data. Surprisingly, we cannot support the hypothesis that openness positively influences transparency mechanisms (H1c). This result gives an indication that opening access to data has not (yet) helped governments to become more transparent. Rather, citizens might object to what they consider to be a cosmetic appearance to transparency. Other actions have to follow to convince citizens that transparency is really a priority of the government in question. Finally, we can support the hypothesis that openness positively influences participation mechanisms (H1d). The relationship is strong and significant, remaining robust against changes in the model during the testing phase. Thus, we can support the sentiment that citizens in countries with openness participate more, especially through government websites.

All value generation mechanisms are positively influenced by resource governance (H2a-H2d). This indicates how important it is that the public sector enjoys leadership and is highly motivated by openness. Furthermore data management policies and the necessary technical skills are important. The quality of the data is of course extremely relevant, as can be seen from the high absolute weight this indicator receives. Capabilities positively influence efficiency, innovation and transparency, but we cannot support H3d, namely, that capabilities positively affect participation. This lack of relationship is surprising, but remained robust to changes in the model in the analysis phase. The most likely explanation is that we used only one measure for participation, and UN's e-Participation index has been subject to some criticism for being too supplier oriented in the past.

Figure 2: Measurement model based on Partial Least Squares analysis



\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

All generative mechanisms except one, are positively influenced by technical connectivity (H4a, H4b and H4c). We could, however, not accept H3d. This again is surprising, given that e-Participation is defined as participation via government websites, indicating a dependency both on technical availability and capabilities. However, openness and resource governance both have a strong positive relationship with participation. Currently, only 40 per cent of 193 UN member states are leveraging social media for the benefit of e-service uptake, indicating a lack of use of pervasive technologies for participation purposes (UN, 2012). This might explain why the results indicate that the variance in public participation between countries is not explained by the general availability of technology nor the general capabilities in society to use technology. Rather, the willingness of the public sector to be open might indicate a more positive attitude towards participation and thereby encourage people to use whatever participation options there are to make their views and opinions available. As noted earlier, to increase participation, societies need to attend not only to those willing-but-unable but also to those able-but-unwilling.

The path from efficiency mechanisms to value (H5) shows the expected positive and significant coefficient. Innovation mechanisms also positively affect value (H6), where

the path has both high absolute value and is highly significant. This supports earlier findings showing the general importance of innovation for societies and more specifically highlights the importance of data-driven innovation for generating value from OGD. As expected, transparency positively influences value, supporting earlier research that shows the disruptive effects of information asymmetry and resulting corruption. Participation mechanisms also positively influence value (H8). Currently, one quarter of all UN member countries publicly commit to considering the results of e-participation in the policy-making process (UN, 2012), indicating a growing focus on the value generation possibilities of participation.

**Table 5: Effect sizes**

<b>Efficiency</b>	<b>Innovation</b>	<b>Participation</b>	<b>Transparency</b>
0.006 (weak)	0.146 (moderate)	0.111 (moderate)	0.028 (weak)

We checked the effect of each of the value generation mechanisms on Value by comparing the R<sup>2</sup> for the value construct with, and without, the variable in question, using Cohen’s f<sup>2</sup> measure (Polites and Karahanna, 2012). While participation and innovation show moderate effect sizes, efficiency and transparency both have a weak effect on Value. However, while the effect size is small, the total effect of each of the measure's efficiency and transparency is moderate, indicating that these are still important variables to consider in predicting value generation from OGD. These results indicate that efficiency and transparency in some way substitute each other as the effect size when both variables are removed is 0.15 (moderate). In a way, both efficiency and transparency have the ability to improve the public sector’s resource allocation, the first through reduced transaction costs and the second through reduced information asymmetry.

**III.5 Discussion**

Our results, both from reviewing the literature and from statistical analysis, indicate that OGD are a resource that offers the ability to increase social welfare through the generation of economic and social value. This value generation happens through a complicated network of mechanisms where the public sector, private companies, civil society and citizens all contribute to the transformation of OGD to value. While the mechanisms operate in very different ways, they are all dependent on certain factors that affect the use and usability of the data as a resource. Due to these common factors, certain synergies might arise between the mechanisms. We can support most of the

hypotheses we set forth, indicating that openness, resource governance, capabilities and technical connectivity are all important enabling factors for the system of mechanisms, while not all of them have significant influence on all the mechanisms. We can also support that all four mechanisms of efficiency, innovation, transparency and participation positively affect value.

We find that OGD has the ability to increase efficiency through decreased transaction costs. For instance, offering citizens the ability to access information via web-platforms can reduce the administrative burden of Freedom of Information (FOI) enquiries (Halonen, 2012) and openly sharing information across levels of government reduces search costs and eventually the need to re-produce data. We can statistically verify this link as the path from the efficiency construct to the value construct is significant and moderately high in absolute terms (0.21). We have also seen evidence of companies using OGD for innovative purposes and generating value, not only in monetary terms but also social value. One example is the combination of open geographic data with open data on drug prescriptions in the UK (<http://www.prescribinganalytics.com>). The visualization created from these data revealed potential savings for the National Health Service in UK of around £200 million pounds per annum, if two thirds of proprietary (expensive) statins were substituted with generic (inexpensive) versions of the same drugs. Another example is OPower, a global company, specializing in energy efficiency. They have used open data on average energy consumption patterns and big data from smart meters in homes to generate reports intended to influence consumer's energy use. Due to these reports, 15 million homes around the world have saved over 2.7 terawatt hours of energy over the last 6 years. Our statistical analysis supports the impact of data-driven innovation on value, the path coefficient was 0.36 and statistically significant.

Our results indicate that open governments value the opinions of their citizens when planning policies that influence economic growth, wellbeing, health, education and the environment and that where citizens have the opportunity to participate, the impact of those policies are improved. There is a highly significant path from the openness construct to participation (0.33) and from participation to value (0.24), supporting this link. Finally, as transparency in government is often conceptualized as open access to government data, openness of data is generally assumed to bring transparency. However, as Yu and Robinson (2012) have pointed out, governments can remain opaque even if they drastically increase technical access to data, for instance if these data-sources are not relevant for policy analysis. We conceptualized transparency as a mechanism that reduces information asymmetry and therefore adverse selection,

leading to less corruption. When conceptualized this way, we cannot confirm any link between openness and transparency. While it is too early to conclude that openness (conceptualized as increased access to government data) does not influence transparency at all, we can propose that increased openness does not automatically lead to increased transparency and there are other issues like governance, capabilities and technical connectivity that seem very relevant to the concept of transparency.

## Limitations and implications

Our study is exploratory due to the embryonic state of OGD research, and our aim was theory generation rather theory testing or confirmation. Our results have various limitations for several reasons. The sample size is small, although we have reasonable evidence to believe that the model contains enough power to draw conclusions from results. However, a bigger sample would allow us to generate more accurate results. All data on openness were taken from the Open Data Index, which was constructed for the first time in 2012 (data representing 2011). The World Wide Web Foundation that collects the data is aware of some limitations regarding methodology and is working on an improved version, which will also include more countries (Annoni et al., 2012). Comparing impacts from OGD between countries, where in many cases OGD initiatives are in their infancy, might be premature; however, we feel that our research model gives a good indication of relationships, as a basis for future research. Furthermore, most of the constructs in the study are new and need to be further validated in the future. Many of the concepts discussed are highly complex, and there still has been no consensus on how to measure many of them. Discriminant validity was marginal and some of the indicators used might be too broad to accurately measure our theoretical constructs. Future analysis with a larger sample size will enable us to conduct some more rigorous testing, for instance, multi-group analysis to search for possible unobserved heterogeneity (Sarstedt et al., 2011).

The main theoretical implications concern: a) the preliminary set of constructs we have conceptualized; b) our propositions regarding use of available, open data to measure these constructs and c) the nomological network that depicts the relationships between enabling factors, value generation mechanisms and value. There are several practical implications for public bodies planning to open their datasets for use and re-use. First, while government data as a resource offer society the ability to generate social and economic value, the value generation mechanisms are dependent on the enabling factors. 'Build it and they will come' approach to OGD is not likely to succeed, or at

least will give marginal benefits unless these factors are present. Second, different mechanisms present different routes to value generation and appropriation. It is important for OGD initiatives to be aware of what kinds of mechanisms they are hoping to encourage, and what the desired effects are. The difference between input, output and outcome needs to be clear, highlighting that mechanisms such as efficiency present a means to an end, but the end goal is likely to be the generation of value. If this relationship is well understood, it is easier to choose the right datasets, data platforms and governance procedures. Finally, we show that all four mechanisms contribute to value generation. Furthermore, the results of our analysis give an indication that there are synergies that can be exploited as the mechanisms are partly dependent on the same enabling factors. If the leaders of OGD initiatives attend to these factors, they offer both the public and the private sector the opportunity to generate value from OGD via different types of mechanisms, although full exploitation of each mechanism might require some specific considerations, which we do not elaborate further on in this paper.

### **III.6 Conclusion**

We have proposed that there are four different archetypical value generation mechanisms, each of which represents a certain type of cause-and effect relationship between OGD and value, defined as a construct that is reflected by the level of social welfare in different countries. We have also suggested that the highest level of value can emerge where possible synergies resulting from common enabling factors are exploited. We can support that all four of the identified mechanisms positively influence value, reflected in the level of education, health and wellbeing, as well as the monetary value of GDP and environmental factors. We can also support the importance of openness, as the indirect path from the construct openness to value is moderately high (0.18) and statistically significant.

We propose that openness is in itself an important enabler to the creation of value from data, as openness enables both the generation and appropriation of value, not only by the organization that produces the data but also by external stakeholders. However, while openness might be the necessary condition in this context, there are other important factors that need to be attended to if the goal is to maximize utilization of the data resource. Governments should consider the sustainability of their initiatives to minimize risk for external users, and should increase quality and usability of their data by acquiring the right skills and focus on data governance. The value generating

mechanisms also require the use of software, technical platforms and telecommunications infrastructure, as well as the general capability of individuals to access and make sense of the data. In conclusion, this study extends our understanding of the implications from opening government data to the public. It lays the foundation for further study of the interplay between enabling factors, value generation mechanisms and the resulting value. Such studies will enable governments to plan more effective strategies in order to maximize the value generation that is stimulated by use of OGD and the subsequent appropriation of value by all members of society.

### III References

- Afonso, A., Schuknecht, L. and Tanzi, V. 2010. "Public sector efficiency: evidence for new EU member states and emerging markets." *Applied Economics*, 42(17), pp. 2147-2164.
- Alani, H.; Dupplaw, D.; Sheridan, J.; O'Hara, K.; Darlington, J.; Shadbolt, N. and Tullo, C. 2007. "Unlocking the potential of public sector information with Semantic Web technology," in Proceedings of the 6th International Semantic Web Conference (ISWC), 11-15 Nov 2007, Busan, Korea.
- Annoni, P., Weziak-Bialowolska, D. and Nardo, M.2012. Annoni P., Weziak-Bialowolska D. and Nardo M. Assessment of the Web Index, Luxembourg: European Commission, Joint Research Centre.
- Avital, M., Lyytinen, K., Iacono, S., Kraemer, K.L., Sambamurthy, V., and Sawyer, S. 2007. "Data Rich and Data Poor Scholarship: Where Does IS Research Stand?" in Proceedings of the International Conference on Information Systems (ICIS), Montreal, Canada.
- Axelsson, K., Melin, U. and Lindgren, I. 2010. "Exploring the Importance of Citizen Participation and Involvement in E-government Projects - Practice, Incentives and Organization," *Transforming Government: People, Process and Policy*, (4:4), pp. 299-321.
- Axelsson, K., Melin, U. and Lindgren, I. 2013. "Public e-services for agency efficiency and citizen benefit - Findings from a stakeholder centered analysis," *Government Information Quarterly* (30), pp. 10-22.
- Batini, C. and Scannapieco, M. 2006. *Data Quality, Concepts, Methods and Techniques*, Berlin Heidelberg: Springer-Verlag.
- Benington, J. 2011. "From Private Choice to Public Value?," In In Benington, J. and Moore, M. (eds.) *Public Value: Theory and Practice*, Houndmills, Basingstoke: Palgrave MacMillan.
- Barber, B. (1984). *Strong Democracy: Participatory Politics for a New Age*. Los Angeles: University of California Press
- Benkler, Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven CT: Yale University Press.
- Bertot, J.C., Jaeger, P.T. and Grimes, J.M. 2010. "Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies," *Government Information Quarterly* (27), pp. 264-271.
- Bertot, J.C., Jaeger, P.T, and Grimes, J.M. 2012. "Promoting transparency and accountability through ICTs, social media, and collaborative e-government," *Transforming Government: People, Process and Policy*, 6:1, pp. 78-91
- Bowman, C. and Ambrosini, V. 2000. "Value Creation Versus Value Capture: Towards a Coherent Definition of Value in Strategy," *British Journal of Management*, (11:1), 1-15.
- Cenfetelli, R. T., and Bassellier, G. 2009. "Interpretation of Formative Measurement in Information Systems Research," *MIS Quarterly* (33:4), pp. 689-707.
- Center for International Earth Science Information Network (CIESIN). 2011. Natural Resource Management Index (NRM), NY: NASA Socioeconomic Data and Applications Center (SEDAC).



- Chesbrough, H., Vanhaverbeke W. and West, J. (eds.). 2006. *Open Innovation: Researching a New Paradigm*. Oxford: Oxford University Press.
- Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research*, G. A. Marcoulides (ed.), Mahwah, NJ: Lawrence Erlbaum Associates, pp. 295-336.
- Cordella, A. 2007. "E-government: towards the e-bureaucratic form?" *Journal of Information technology*, 22(2007), pp. 265-274.
- Davies, T. 2010. "Open data, democracy and public sector reform: A look at OGD use from data.gov.uk," Available at <http://tinyurl.com/7joks46>
- Dawes, S. S. 2012. "A Realistic Look at Open Data," Center for Technology in Government, University at Albany/SUNY Available at [http://www.w3.org/2012/06/pmod/pmod2012\\_submission\\_38.pdf](http://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf)
- Diamantopoulos, A., and Siguaw, J. A. 2006. "Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison of Empirical Illustration," *British Journal of Management* (17), pp. 263-282.
- Digitaliseringsstyrelsen. 2012. Basic Data. Available at <http://tinyurl.com/dxn2jhj>
- Dutta, S. and Bilbao-Osorio, B. (eds.). 2012. *The Global Information Technology Report 2012: Living in a Hyperconnected World*. Geneva: World Economic Forum and ISEAD.
- Economist Intelligence Unit. 2011. *Democracy index 2011: Democracy under stress*.
- Eisler, Riane. 2007. *Real Wealth of Nations*. San Francisco, CA: Berrett-Koehler Publishers.
- European Commission. 2006. "eGovernment Economics Project (eGEP)," Measurement Framework Final. DG Information Society and Media, European Commission.
- Farhan, H., D'Agostino, D. and Worthington, H. 2012. *Web Index 2012*, World Wide Web Foundation.
- Fleurbaey, Marc. 2009. "Beyond GDP: The Quest for a Measure of Social Welfare." *Journal of Economic Literature* 47:4, pp. 1029-75.
- Gallup. 2011. "The Behavioral Economics of GDP Growth," Washington: Gallup Global Wellbeing.
- Gartner. 2011. "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data," Press release, [www.gartner.com/it/page.jsp?id=1731916](http://www.gartner.com/it/page.jsp?id=1731916).
- Gigler, S; Custer, S. and Rahmetulla, H. 2011. "Realizing the Vision of OGD- Opportunities, Challenges and Pitfalls," World Bank Report.
- Gil-Saura, I., Frasset-Deltoro, M., and Cervera-Taulet, A. 2009. "The Value of B2B Relationships," *Industrial Management & Data Systems* (109:5), pp. 593-609.
- Gurstein, M. 2011. "Open data: Empowering the empowered or effective data use for everyone?," *First Monday* (16), Available at <http://tinyurl.com/dylszvv>
- Hair, J.F., Ringle, C.F., and Sarstedt, M. 2011. "PLS-SEM: Indeed a Silver Bullet," *Journal of Marketing Theory and Practice* (19:2), pp. 139-151.

- Hair, J.F., Ringle, C.F., and Sartzstedt, M. 2012. "An assessment of the use of partial least squares structural equation modeling in marketing research," *Journal of the Academy of Management Science* (40), pp. 414-433
- Halonon, A. 2012. "Being Open About Data," Available at <http://tinyurl.com/c8mz5vt>
- Harrison, T.M.; Guerrero, S; Burke, G.B; Cook, M; Cresswell, A; Helbig, N.; Hrdinová, J; Pardo, T. 2011. "Open Government and E-Government: Democratic Challenges from a Public Value Perspective," in Proceedings of the 12th Annual International Conference on Digital Government Research (dg.o 2011), pp. 245-253.
- Heeks, R. 2003. "Most eGovernment-for-Development Projects Fail: How Can Risks be Reduced?," Institute for Development Policy and Management, Manchester: University of Manchester.
- Hess C. and Ostrom, E. 2006. *Understanding Knowledge as a Commons – From Theory to Practice*, Cambridge, Massachusetts: MIT Press.
- Holmberg, S., Rothstein, B. and Nasiritousi, N. 2009. "Quality of Government: What You Get." *Annual Review of Political Science* (12:2009), pp. 135-161.
- Huijboom, N., and Van den Broek, T. 2011. "Open data: an international comparison of strategies," *European Journal of e Practice*.
- International Bank for Reconstruction and Development. 2012. "Ease of Doing Business 2013: Smarter Regulations for Small and Medium-Size Enterprises," Washington DC: The World bank and International Finance Corporation.
- Jansen, K. 2011. "The influence of the PSI directive on open government data: An overview of recent developments," *Government Information Quarterly* (28:2011), pp. 446-456.
- Janssen, M., Charalabidis, Y. and Zuiderwijk, A. 2012. "Benefits, Adoption Barriers and Myths of Open Data and Open Government," *Information Systems Management* (29:4), pp. 258-268.
- Karunasena, K., Deng, H. and Singh, M. (2011) "Measuring the public value of e-government: a case study from Sri Lanka", *Transforming Government: People, Process and Policy* (5:1), pp. 81- 99.
- Kaufmann, D., Kraay, A. and Mastruzzi, M. (2010). "The Worldwide Governance Indicators : A Summary of Methodology, Data and Analytical Issues," World Bank Policy Research Working Paper No. 5430
- Kolstad, I. and Wiig, A. 2009. "Is transparency the key to reducing corruption in resource-rich countries?," *World Development* (37), pp. 521-32.
- Kim, S. and Lee, J. 2012. "E-Participation, Transparency, and Trust in Local Government," *Public Administration Review* (72:6), pp. 819-828.
- Koski, H. 2013. "ICT Outsourcing, User-Driven and Open Innovation Strategies in the Generation of New Data-Based Solutions," ETLA Working Papers No. 7. Available at: <http://pub.etla.fi/ETLA-Working-Papers-7.pdf>
- Kundra, V. 2012. "Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect," *Harvard Business Review* (Jan:2012).
- Kuznets, S. 1974. *Population, Capital and Growth: Selected Essays*. London: Heinemann Educational.

- Lampri, M. 2012. "E-Government and Data Protection in the Public Sector, in Virtual Communities, Social Networks and Collaboration," in *Annals of Information Systems*, Athina A Lazakidou, ed., New York: Springer, pp. 135-147.
- Lee, G. and Kwak, J.H. 2011. "Open Government Implementation Model: A Stage Model for Achieving Increased Public Engagement," in *The Proceedings of the 12th Annual International Conference on Digital Government Research (dg.o.2011)*, pp. 254-261.
- Linders, D.; Wilson, S.C. 2011. "What is Open Government? One Year after the Directive," in *The Proceedings of the 12th Conference on Digital Government Research (dg.o.2011)*, pp. 262-271.
- Marcoulides, G. A. and Saunders, C. 2006. "PLS: A Silver Bullet?," *MIS Quarterly* (30:2), pp. iii-ix.
- Mayer-Schönberger and Zappia, (2011). "Participation and Power: Intermediaries of Open Data," Conference draft prepared for the Berlin Symposium. Available at: <http://tinyurl.com/bo3pyl8>
- McKinsey & Company. 2011. "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute.
- McKinsey & Company. 2012. "Transforming Government through Lean Management," McKinsey Center for Government.
- Michaelson, J., Abdallah, S., Steuer, N., Thompson, S. and Marks, N. 2009. *National Accounts of Well-Being: Bringing Real Wealth onto the Balance Sheet*, London: New Economics Foundation.
- Nilsen, K. 2010. "Economic theory as it applies to Public Sector Information," *Annual Review of Information Science and Technology* (44:1), pp. 419-489.
- OECD. 2005. *Oslo Manual – Guidelines for Collecting and Interpreting Innovation Data*, Paris: OECD.
- Open Knowledge Foundation (OKF). (2012). *Open Data Handbook version 1.0.0* (revised November 14, 2012). Available at <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>
- Petter, S., Straub, D., and Rai, A. 2007. "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly* (31:4), pp. 623-656.
- Pol, E and Ville, S. 2009. "Social innovation: Buzz word or enduring term?" *Journal of Socio-Economics* (38:6), pp. 878-885.
- Polites, G.L. and Karahanna, E. 2012. "Shackled to the Status Quo: The Inhibiting Effects of Incumbent System Habit, Switching IT Costs, and Inertia on New System Acceptance," *MIS Quarterly* (36:1), pp. 21-39.
- Porter, M. E. and Kramer, M.R. 2011. "Creating Shared Value," *Harvard Business Review* (89:1), pp. 62-77.
- Ringle, C.M., Sarstedt, M. and Straub, D.W. 2012. "A Critical Look at the Use of PLS-SEM in MIS Quarterly," *MIS Quarterly* (36:1), pp. iii-xiv.
- Relly, J.E. and Sabharwal, M. 2009. "Perceptions of Transparency of Government Policymaking: A cross-national study," *Government Information Quarterly* (26), pp. 148-157.
- Rothstein, B. 2011. "Anti-corruption: the indirect 'big bang' approach," *Review of International Political Economy* (18:2), pp. 228-250.
- Sanchez, G. 2013. "PLS Path Modeling with R," Available at: <http://tinyurl.com/btfcv4p>

- Sarstedt, M., Henseler, J., and Ringle, C. M. 2011. "Multi-Group Analysis in Partial Least Squares (PLS), Path Modeling: Alternative Methods and Empirical Results," *Advances in International Marketing* (22), pp. 195-218.
- Sarker, S.; Sarker, S.; Sahaym, A. and Bjørn-Andersen, N. 2012. "Exploring Value Cocreation in ERP Vendor-Partner Relationships," *MIS Quarterly* (36:1), pp. 317-338.
- Schwab, K., Sala-i-Martin, X. and Greenhill, R. (eds.). 2011. *Global Competitiveness Report 2011-2012*, Switzerland: World Economic Forum
- Schumpeter, J.A. 1934. *The Theory of Economic Development*, Cambridge: Harvard University Press.
- Sen, A. 1999. *Development as Freedom*, New York: Anchor Books.
- Shapiro, C. and Varian, H.R. 1999. *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, Boston.
- Stiglitz, J.E., Sen, A. and Fitoussi, J. P. 2009. "Report by the Commission on the Measurement of Economic Performance and Social Progress," OECD.
- Trinchera, L. and Russolillo, G. 2010. "On the use of Structural Equation Models and PLS Path modeling to build composite indicators," Working paper no. 30, Università Degli Studi Di Macerata.
- Ubaldi, B. 2013. "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives," OECD Working Papers on Public Governance, No. 22, OECD Publishing.
- UN. 2012. *E-Government Survey 2012: E-Government for the People*. New York: United Nations
- Van De Walle, S. 2006. "The state of the world's bureaucracies," *Journal of Comparative Policy Analysis: Research and Practice* (8:4), pp. 437-448.
- Vangen, S. and Huxham, C. (2010). "Introducing the theory of collaborative advantage," In *The New Public Governance? Emerging Perspectives on the Theory and Practice of Public Governance*, Osborne, S. P. (ed.), London: Routledge, pp. 163-184.
- Vickery, G. 2011. "Review of Recent Studies on PSI Re-Use and Related Market Development," Paris: European Commission
- World Bank. 2006. *E-Leadership Institutions for the Knowledge Economy*. Washington D.C.: World Bank.
- Yu, H. and Robinson, D.G. 2012. "The New Ambiguity of 'Open Government'," Available at: <http://ssrn.com/abstract=2012489>
- Zuidervijk, A, Janssen, M., Choenni, S, Meijer, R. and Alibaks, R.S. 2012. "Socio-technical Impediments of Open Data," *Electronic Journal of e-Government* (10:2), pp. 156 – 172.

## **PAPER IV: Data-driven Innovation through Open Government Data**

Journal of Theoretical and Applied Electronic Commerce Research 9(2), pp. 100-120  
(2014)

# Data-Driven Innovation through Open Government Data

Thorhildur Jetzek<sup>1</sup>, Michel Avital<sup>2</sup>, and Niels Bjorn-Andersen<sup>3</sup>

<sup>1</sup>Copenhagen Business School, Department of IT Management, Copenhagen, Denmark, tj.itm@cbs.dk

<sup>2</sup>Copenhagen Business School, Department of IT Management, Copenhagen, Denmark, michel@avital.net

<sup>3</sup>Copenhagen Business School, Department of IT Management, Copenhagen, Denmark, nba.itm@cbs.dk

Received 19 August 2013; received in revised form 5 November 2013; accepted 20 December 2013

## Abstract

The exponentially growing production of data and the social trend towards openness and sharing are powerful forces that are changing the global economy and society. Governments around the world have become active participants in this evolution, opening up their data for access and re-use by public and private agents alike. The phenomenon of Open Government Data (OGD) has spread around the world in the last four years, driven by the widely held belief that use of OGD has the ability to generate both economic and social value. However, a cursory review of the popular press, as well as an investigation of academic research and empirical data, reveals the need to further understand the relationship between OGD and value. In this paper, we focus on *how* use of OGD can bring about new innovative solutions that can generate social and economic value. We apply a critical realist approach to a case study analysis to uncover the mechanisms that can explain how data is transformed to value. We explore the case of Opower, a pioneer in using and transforming data to induce a behavioral change that has resulted in a considerable reduction in energy use over the last six years.

**Keywords:** *Open data, Open government Data, Big data, Innovation, Value, Generative mechanisms, Critical realism, Opower*

## IV.1 Introduction

Data have become part and parcel of modern times. We only have to imagine a world without Google searches, online weather forecasts or GPS technologies to realize the current impact of data on our lives. The use of technology and the subsequent generation and utilization of digital data have become ubiquitous, virtually taken for granted. The impact of these technologies is evolving continuously with the creation of new content, connectivity, analysis software and infrastructure. We have recently observed a radical trend towards networked behavior such as crowdsourcing and co-creation, driven by (among others) the emergence of the open-source software community, the general use of social networks and increased availability of Open Government Data (OGD). One of the most disruptive aspects of these changes is the transformation from a largely dichotomous world of the market and the state, to an open, interconnected world where the traditional roles of, and relationships between, sectors are changing. These complex interdependencies are forcing us to re-think how economic and social value is generated and appropriated [3], [32].

The number of OGD initiatives has grown from two to over three hundred in the period 2009-2013, and membership in the Open Government Partnership has gone from eight to fifty-nine countries in two years. Over 280 government data catalogs have been published and over a million datasets have been released by governments around the world, spawning new businesses and social projects. OGD is commonly seen as a driver of efficiency and a vehicle for increasing transparency, citizen participation and innovation in society. The hope is that OGD will eventually lead to the generation of substantial value. For instance, the European Commission launched an Open Data Strategy for Europe in December 2011, which is expected to deliver around \$53 billion boost to the EU's economy each year [20]. The strategy resulted in EU Open Data rules being formally adopted in June 2013, including directions regarding charging rules, licensing, search on data portals and interoperability. In May 2013, U.S. president Barack Obama signed an executive order, making open and machine readable the new default for government information in the USA. It is expected that open data will bring benefits to a wide range of domains, including health, energy, education, public safety, finance and global development [75].

However, despite the potential significance of OGD, many feel that it is supported mainly by anecdotal evidence and that OGD is still far from living up to its true potential. A common assumption when opening government data is that simply supplying more data freely and in more formats will lead to more use and value

creation [30]. But opening up government data is meaningful only so far as citizens and public and private organizations have not only the opportunity, but also the motivation and ability to use the data to achieve social and economic value. Lack of appropriate governance mechanisms and a lack of insight into user's perspective can explain the gap between the promises of open data and what has actually been realized [36]. Moreover, there is uncertainty as to how the value of the data will be regarded and how it should be evaluated [30]. The economic and social impacts of open-data policies remain largely unclear, and there are relatively limited empirical data available on the effects of the various policy approaches leaving policy makers without the facts they need to assess and improve these policies [34], [80].

This paper addresses the question *How can use of open government data stimulate the generation of value?* Ultimately, the study aims to explore and illustrate *how* value is generated from the use of OGD, as well as to identify the main factors that enable value generation. For that purpose, we utilize the critical realist concept of mechanism as a causal structure that contingently generates observable outcomes. The generative mechanisms portray the instrumental pathways that stem from use of OGD in the context of social, technical and organizational factors, which result in the generation of value. Furthermore, we want to explain in more depth how value generation happens in the case of a private company using OGD. For this purpose, we use a case study method, grounded in the critical realism ontology and epistemology, as suggested in [12], [76]. Based on our findings, we propose a conceptual model for data driven innovation, building on the absorptive capacity model presented in [63] and the innovation value chain model [27]. Our main contribution is the conceptualization of data driven innovation and the nomological network that furthermore shows the causal links between the external and internal enabling factors in the context of organizational innovation from data, the innovation mechanism itself and the resulting value.

## **IV.2 Critical Realism and Open Government Data**

It is a widely held belief that use of OGD can result in the generation of considerable social and economic value. *How* this can happen, however, is not well understood. In this section we explore the unique features of OGD, discuss how these unique features can underpin some specific value propositions, and identify the mechanisms that can explain the transformation from use of data to generation of value. We build on the Critical Realism philosophy and the concept of micro-macro mechanisms, proposing that events observed in the real world can be used to uncover the underlying



mechanisms that arise from an interplay of micro-level structures, decisions and actions, and in this way contribute to explaining how value can be generated.

## The Unique Features of Open Government Data

The amount of data accumulating in our increasingly digital world is breathtaking. In the past two years alone the amount of information available in the digital universe has increased to its current rate of 2.8 ZB, a number that is expected to double every year. This increase is mainly due to the continuous digitization of nearly all media, the ubiquity of Internet access and the proliferation of mobile phones, as well as data generation from surveillance cameras and smart meters. For example, around 30 billion pieces of content are shared on Facebook every month and 235 terabytes of data were collected by the US Library of Congress in 2011 [49]. More than 30 million interconnected sensors are now deployed worldwide in areas such as security, health care, transport systems or energy control systems, and their numbers are growing by around 30% a year [49]. Smart meters collect and transmit real-time data on energy [57], and smart automobiles are now able to transmit real-time data on the state of the car's components and environment [58]. Digitization affects two important features of data: 1) By making data easily accessible to more than one person at a time, it is resulting in non-rivalry and; 2) By drastically reducing marginal costs incurred by reproduction and distribution, it is making re-use economically feasible [54], [61], [70]. Furthermore, not only have we now generated all these digital data, but in many cases they are also open for use by anyone interested, allowing for even more value generation through re-use of different stakeholders.

Open data can be defined as data that are freely accessible online, available without technical restrictions to re-use, and provided under open access license that allows the data to be re-used without limitation, including across different 'fields of endeavor' (e.g., commercial and non-commercial alike) [59]. Openness changes one important feature of digital data by making them non-excludable. Accordingly, when opened up, digital data become a shared resource; a public good or what has been termed 'digital commons' [32]. The concept of OGD refers specifically to government data defined as "data and information produced or commissioned by government or government controlled entities" [59] that are opened up for use and re-use by public and private agents alike. In the currently used terminology, OGD does not include data that are subject to valid privacy, security or privilege limitations, as governed by other statutes.

Government data sets make up an interesting subset of open data. Public bodies are among the largest creators and collectors of data in many different domains [35]. These

domains range from traffic, weather, and geographical data to statistics and data on businesses and public sector budgeting [36]. The fixed costs incurred by collecting OGD can be high, making it unfeasible for private organizations to collect such data. However, in the case of government, these data have already been collected for specific use in governmental processes. Moreover, as an integral part of public operations, they have also been paid for by taxpayers. Therefore, OGD constitute a shared resource that offers value beyond what is captured from the original intended use.

## Critical Realism as a Foundation for the Study of Open Government Data

Critical realism is becoming recognized as a viable philosophical paradigm for conducting social science research. In the Information Systems discipline, we are typically confronted with a sociotechnical environment consisting of several interacting structures, each of which has the potential to impact the existing situation to generate events. This typically includes a social structure consisting of individuals, groups, and organizations, along with a set of rules and practices, technological artifacts and discursive entities such as language and culture [76]. Critical realism based research methodologies are used to identify the mechanisms that can explain such complex interactions [24]. They allow researchers to develop and support in-depth causal explanations for the outcomes of specific sociotechnical phenomena [76]. In the context of this paper, we are interested in finding the main constructs and mechanisms that are involved when actors utilize OGD to generate value. This includes structures such as legal frameworks, different data dissemination practices, use of technology artifacts, individual capabilities and various other mechanisms that can extract value from OGD.

Critical realism interrelates ontology and epistemology, that is, it conjoins assumptions about the nature of reality, and evidentiary assessment and justification of knowledge claims. On the one hand, it posits realist ontology, that is, the existence of a world independent of researchers' knowledge of it. On the other hand, critical realism embraces a fallibilist epistemology in which human beings are unable to fully understand or observe this reality, and our knowledge of it is fallible [76]. Critical realism distinguishes between the domains of the real, actual and empirical. The real domain consists of generative mechanisms, which refer to the functional mechanics of a particular phenomenon [9] - [10]. Through enabling or hindering change, such mechanisms give rise to events in an actual domain, some that are experienced, and

some are not. Finally, the empirical domain is made up of events that are directly or indirectly observed [52]. Even though critical realists argue that there is a real world out there, they accept the possibility that one observing this world does not have full access to it or have the ability to observe equally all its aspects [78]. The potentiality of the real still exists even if it remains unexercised or unobserved. Being real does not necessarily imply observable.

Critical realism has a strong focus on causality. Causality refers to "...the relationship between an action or thing (cause) and the outcome (effect) it generates" [76]. p. 789. Often, our ability to explain a given phenomenon requires the identification of the factors and relationships which cause it to occur. Critical realists emphasize that causality is contingent - in the sense that the observed outcome of a mechanism is contextual. Consistent regularities can arise under special circumstances in closed systems. However, open systems such as the social systems are far too complex. For example, causal mechanism M1 may be in operation, tending to bring about outcome O1. But outcome O1 is also likely to be influenced by causal mechanisms M2, M3...Mn, operating somewhere else in the social system [21]. These other confounding mechanisms may prevent or alter the realization of a particular causal effect [76]. Accordingly, we need to uncover the underlying mechanisms that offer the ability to generate value – and show under what circumstances value is generated, i.e., what factors enable value to materialize. This involves finding the drivers and barriers, as well as other related mechanisms, in order to explain how the value generating mechanism produced the observed impacts, and preferably why – or why not.

## Enablers of Value Generation from Data

Critical realism helps uncover the mechanisms that can explain how the use of OGD generates value, given the multitude of factors that influence this potential value generation. In this way it will help us understand how an OGD initiative could result in an impressive generation of value in one country, while a similar initiative might have shown negligible impact in another. In this regard, the observed outcome should be explained in consideration with the interaction of appropriate pre-conditions as well as the transfactual operation of the mechanisms [31]. In other words, we build on the underlying premise of critical realism, that causal effects work as intended whether their operation is observable or whether it cannot be detected by those who attempt to examine it. Moreover, in line with the principle of equifinality, in most cases there are multiple possible sets of mechanisms which may have produced the outcomes being studied in a given research program [76]. Due to the complexity of the causal

relationship between the utilization of OGD and value, we suggest that for the benefit of supporting policy generation, investigations of the causal paths should seek to analyze how different key mechanisms are enacted and which factors can lead to a successful outcome.

Several authors have pointed out various factors that hinder the generation of value from data. The most commonly identified barriers are: a) closed or inaccessible datasets, b) lack of comprehensive data policies, c) lack of validity, completeness and exhaustiveness of datasets, d) insufficient metadata, e) lack of consistency in cross-border access regimes, f) lack of motivation within public sector, g) lack of technical and semantic interoperability, h) lack of technical ability within public and private sectors, i) the digital divide and j) too fragmented and disparate open data community [7], [18] - [19], [26], [35] - [36], [47], [79]. It is also suggested that the following four multi-dimensional macro-level factors can help nations overcome the previously mentioned barriers to value generation [37]:

- 1) **Capabilities**, conceptualized as the collective ability of individuals and organizations to use and re-use OGD, as a function of equitable access opportunities and technology and data literacy.
- 2) **Openness**, conceptualized as a function of the general availability of government data, the accessibility of available data sources and the use of open licenses.
- 3) **Resource governance**, conceptualized as a function of leadership, data governance procedures and data dissemination skills within the public sector that is intended to increase the quality and sustainability of data resources.
- 4) **Technical connectivity**, conceptualized as a function of the technical infrastructure and the diffusion of technologies that allows users to store, access and analyze the data.

## Generative Mechanisms

The concept of mechanism is a key construct and a commonly used sensemaking apparatus in many disciplines of science. Mechanisms are frequently occurring and easily recognizable causal patterns. In the context of critical realism, mechanisms are used to describe and explain causal relationship through the specifying of how central events or outcomes are produced and reproduced by the structures, actions and contextual conditions in a particular setting [74]. Following [9], [31], we define

generative mechanisms as causal structures that are capable of generating observable events, Generative mechanisms can be classified on three levels: contextual or situational mechanisms (macro–micro), action-formation mechanisms (micro–micro) and transformational mechanisms (micro–macro) [29], [31]. Macro–micro mechanisms refer to how macro level constructs enable or constrain the various micro level components that subsequently may affect the generation of value. Micro–micro mechanisms relate to how specific arrangements of individual desires, beliefs, and actions may generate a specific action [29]. Micro–macro mechanisms concern how different micro level components interact in order to produce an outcome at a macro level, i.e., how the combination of different components can enable (or hinder) the generation of value from OGD. Figure 1 portrays the OGD ecosystem as recursive relationship between the contextual (macro–micro) and transformational (micro–macro) mechanisms.

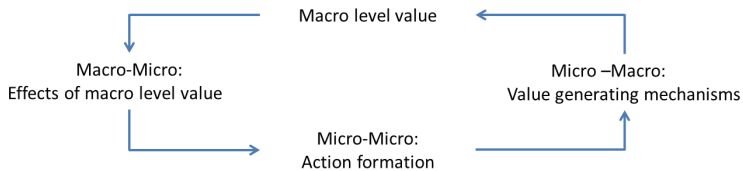


Figure 1: OGD ecosystem, extended from [31]

We model the OGD ecosystem as a recursive loop that keeps repeating indefinitely. Subsequently, in the case of OGD, the contextual or macro–micro mechanisms explain how certain outcomes that reflect the generation of macro-level value, such as the yearly Gross Domestic Product (GDP) or the general level of education within a country, enable or constrain individuals by providing motivation, as well as affecting the opportunity and ability to generate value through use of OGD.

Next, the micro-level action-formation mechanisms [29] reflect the decisions and actions which drive individuals in the process of publishing OGD and explain subsequent decisions of actors to use the data. These mechanisms make sense of individual behavior in terms of interaction between individuals and a social aggregate [29]. While the potential generation of value seems to be the most significant motivator for opening access to government data [36], there are many other micro-level factors that can enable or hinder the actual implementation of an OGD initiative [30] and the subsequent re-use of these data. Numerous barriers for the adoption of OGD policy have been identified in [36], [79], categorized into the following areas: institutional,

task complexity, use and participation, legislation, information quality and technical. For instance, one of the identified barriers stems from a risk-averse culture often found in organizations with more red tape, weak links with performance, and high involvement with elected officials [36], [79] showing the importance of strong leadership to generate . Another set of identified barriers was focused on technical issues ranging from low information quality and unavailability of a supporting infrastructure and up to the lack of standards, fragmentation and legacy systems [36].

Finally, transformational or micro–macro mechanisms explain how the different micro level components interact in order to produce an outcome at a macro level. These mechanisms explain how the actions of a number of individuals, whose interactions with one another are enabled or constrained by their social and technical environments, are transformed into some sort of a collective outcome. This type of mechanism can be used to explain how the combination of different factors enables the actions of various individuals within or outside of the public sector, allowing them to participate in the generation of value from OGD. In the following analysis, we focus on identifying different types of such *transformational mechanisms*.

### **IV.3 Value Generating Mechanisms**

In this section we create a taxonomy consisting of four different mechanisms that classifies how value can be generated from OGD. We use two dimensions to categorize these mechanisms. The first dimension reflects the extent and type of value generation from use of OGD, from low to high levels of external participation. The second dimension reflects the extent and type of OGD value appropriation options, ranging from value being appropriated by dedicated participants only to value being appropriated by society in general.

#### **Dimensions of Value**

Openness is perceived as the antidote that can counteract the tendency of technology enactment to reproduce existing rules, routines, norms and power relations, despite the new and innovative capabilities introduced by these technologies. However, this perceived premise can only be fulfilled if openness changes the nature of relationships between stakeholders and governments, and enables them to link across organizational boundaries and functions [28]. Moreover, the conceptual distinction between value generation and value appropriation has been sharpened in light of the discourse on openness, technical connectivity and collaborative ventures. Value generation

materializes when the utility of society's members increases after accounting for the resources used in that activity. Value appropriation materializes when an actor is able to capture a portion of the value created by an activity [11]. The relationship between the generation and appropriation of value of multiple stakeholders in the context of alliances is increasingly viewed as being multifaceted in nature [66]. Value can be seen as a "subjective, multidimensional construct; accordingly, it is only through a multidimensional view that we get a high-fidelity picture of the value generated within alliance relationships" [23], p. 595. However, despite the recent focus on the creation of value in collaborative settings, little is known about the underlying mechanisms [66].

Two types of value are frequently discussed in the extant literature on OGD: 1) economic value, defined as the worth of a good or service as determined by the market, most often measured relative to units of currency; and 2) social value, defined as the generated improvements in the lives of individuals or society as a whole. A recent attempt to amalgamate the concepts of economic and social value introduced the term shared value, which is meant to reconceive the intersection between society and corporate performance and involves creating economic value in a way that also creates value for society by addressing its needs and challenges [62]. Public value is another related concept in the OGD and e-government literature. The public value framework is based on the premise that public resources should be used to increase value, not only in an economic sense but also more broadly in terms of what is valued by citizens and communities [6], [53]. Based on these insights, we have developed a two-dimensional framework that is based on the *extent to which external stakeholders participate in the generation of value from OGD* as one dimension, and *the extent to which OGD initiatives are focused on generating social value* as the other dimension.

## Opening the Black Box

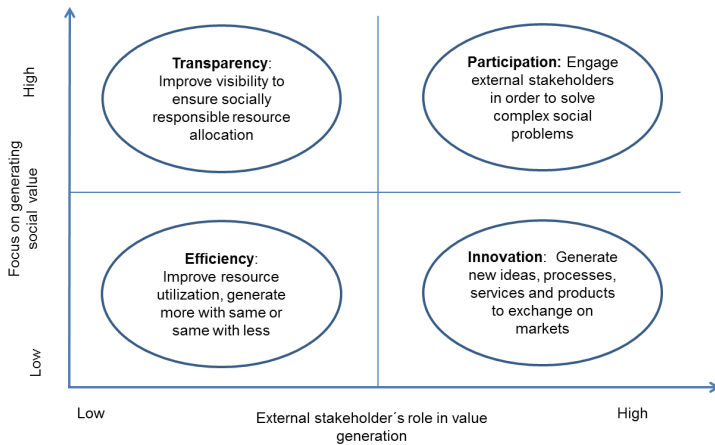
A distinction between black box explanations and mechanism-based explanations is made in [29]. The way in which two sets of events or variables are linked to one another is expressed with the mechanism,  $M: I \rightarrow M \rightarrow O$ . What characterizes black box explanations is that the link between input (I) and output (O) is assumed to be devoid of structure, or whatever structure there may be is considered to be of no interest, and thus the researcher tests only the reduced model that contains a direct link between I and O [29]. We, however, want to make a distinction between the value that is generated (outcomes observed) and the transformative mechanisms that can explain *how* this value generation materializes. We argue that identifying these mechanisms is

likely to reveal instrumental pathways by which a set of actions is related to the creation of a value [28].

Furthermore, non-commonsensual explanations require mechanisms of some generality [29]. In order to identify the main archetypes of possible transformational mechanisms that explain how OGD may be used to generate value, we conducted a wide literature search that was focused on the various operational definitions of open government data. We found that two distinct ideologies drive most open government data initiatives: the *Reuse of Data* perspective and the *Open Government* perspective. We thus reviewed the respective tensions and contributions of these two unique streams. The literature on reuse of data is mostly focused on the economic value of government data, often in connection to the European PSI-directive [35]. The literature on Open Government is mostly making reference to the Obama's 2009 Open Government Directive, and, in a more generalizable term, is directed towards examining how government policy can utilize OGD to generate additional social (or public) value in collaborative settings [8], [44].

Building on Harrison et al. [28], we ultimately adopt four general mechanisms to highlight how OGD can be used to generate value. Two mechanisms highlight the Open Government focus: *transparency of government* and *citizen participation/collaboration* [7], [28], [44] and another two highlight the reuse focus: *efficiency/effectiveness* and *innovation* [22], [26], [34]. In Figure 2, we present a framework that can be used to develop and describe different value generation strategies, each of which has the goal of stimulating one of the general mechanisms in order to generate value. The strategies are framed from the data provider perspective, which in the case of OGD is the public sector. The framework has two dimensions. One dimension, indicating the extent to which external stakeholders (i.e., actors in the private sector) generate value from the data, therefore spans the boundaries between the public and private sectors. In the right hand column of the matrix, the private sector is actively engaged in using OGD to generate firsthand value, while in the left hand column the public sector is the more active stakeholder in value generation (while the value might be appropriated by both sectors). The other dimension, indicating the extent to which OGD initiative is focused on generating social value, ranges from strategies that are focused mainly on generating economic value (with a relatively marginal social value component) to strategies that are focused mainly on generating social value (with a relatively marginal economic value component).





*Figure 2: The open government data value generation framework*

Each of the four general mechanisms has the capacity to generate a mix of social and economic value, although in different proportions. The economic value that is more apparent in the lower half of the matrix is generally measured in monetary terms. Social value generation is, on the other hand, not necessarily measurable in monetary terms; hence, it may reflect intangible value. While this framework can provide a continuous spectrum of different value generating strategies, we use it to highlight four discrete paths that can provide a foundation for four archetypical value generation strategies, thereby reducing complexity and highlighting four modalities of using OGD to generate value.

Next, we briefly review each of the general mechanisms in order to show how the use of OGD can generate value.

### **Transparency Mechanisms**

Transparency mechanisms are designed to reveal relevant information that is being generated, managed and stored by a particular entity, including information about its own decision processes, procedures and performance. However, opening access to selected public documents does not necessarily contribute to a transparent government [25], [73]. In fact, a government can pay lip service to open data by providing access to politically neutral topics, even though its operation in general remains deeply opaque and unaccountable [73]. Agents, whose access to information is increased, must also have the ability to process the information, and the opportunity and incentive to act on

that information [39]. Accordingly, we cannot conclude that open access to government data is in itself a comprehensive measure of transparency. Rather, transparency mechanisms can enable value generation if they reduce information asymmetry. Information asymmetry refers to situations in which one has more or better information than another while they are participating in transactions, negotiations or communications. Information asymmetry can cause all sorts of socially undesirable results and behaviors. One particular prevalent outcome of information asymmetry is corruption. Corruption (defined in the context of government as misuse of public power for private benefits) has long been seen as a hindrance to socio-economic development. Corruption has been shown to have negative effects on GDP growth, human development and health outcomes. It can destroy social capital and has been shown to negatively impact people's life satisfaction [33], [65]. Improved transparency can reduce information asymmetry and result in more equitable resource allocation, leading to the creation of social and economic value.

### **Participation Mechanisms**

Participation mechanisms are designed to enable and encourage public participation in government through voluntary contributions of ideas and other resources. These mechanisms provide citizens with an opportunity to influence public policy, and subsequently enhance the ties between government officials and their constituency [44]. Public participation often provides opinions through citizens' engagement as well as ideas and solutions through crowdsourcing [44]. It has been argued that more involved democratic participation is likely to lead to superior social outcomes because of participation's role in aggregating information and preferences [5]. The relationship between participation and OGD is essentially twofold: First, in order for citizens to participate, either by voicing their opinions towards policy making or by participating directly in public projects, they must have access to information about the particular issue that is being addressed. Second, their contributions lead to the generation of new data. Citizen participation in public administration decision-making has been on the rise, as many government agencies have taken advantage of Internet-based applications to communicate with constituents [38]. Participation mechanisms are designed to lower the barriers to participation for those willing-but-unable, and to make participation more attractive to those able-but-unwilling [4]. Participation mechanisms generate value through the synergies created from openness and sharing, allowing the public sector to draw from a larger pool of resources, consequently improving society's ability to solve difficult social problems.

## **Efficiency Mechanisms**

Efficiency mechanisms are designed to improve resource utilization in order to minimize waste and maximize the outcome value, using the same amount of resources. The importance of efficient use of public resources for economic growth, stability and general well-being has been brought to the forefront by a number of developments over the past decades [1]. As a consequence of increasing government intervention in affairs, such as child care, education, and health services, it is the public sectors that have faced mounting difficulties in managing efficiently the administrative bureaucracy. Moreover, increasing cross-boundary interactions and higher levels of information exchange between citizens and government have increased the total amount of government data collected and stored. These trends call for more efficient processing of data in order to provide the expected services [15]. Efficiency can be gained by cutting processing costs, making strategic connections between and among government agencies, and creating empowerment. We propose that the public sector can use open data to deliver public services more efficiently while safeguarding the quality of services, for instance, by allowing citizens to access and manage their own data or by reusing data within the public sector, thereby enabling automation of processes across governmental levels. As a result, resources can be moved from non-value adding tasks to value-adding tasks, positively affecting the generation of value.

## **Innovation Mechanisms**

Innovation mechanisms are designed to generate new or significantly improved products (goods or services), work processes, business practices, and organizational methods [56]. Recent technological developments have provided firms with the ability to collect, manage, and use different types of data in multiple ways to innovate, and subsequently create value [49]. Following [69], we assume that innovation can have economy-wide effects. Innovation brings about novel combinations of resources, new production methods, as well as new products and services, which, in turn, can lead to the transformation of markets and industries, thus increasing value [69]. Numerous studies have confirmed the relationship between macro-level business innovation and economic value. The social impacts of new innovations have, however, been much less discussed and analyzed, with the possible exception of [41], who separates economic and non-economic consequences of technological innovations. The concept of social innovation is generally directed at improving the quality and/or quantity of life [60]. Social and business innovation can, however, overlap, as business innovation - while mostly dealing with profitable new ideas - can also result in social value generation. Going forward, we view data-driven innovation as business innovation, based to a

large degree on exploiting data, and capable of generating positive economic and social impacts.

#### **IV.4 Research Design**

In general, the case study is a preferred method when (a) *how* or *why* questions are asked, (b) the investigator has little control over events, and (c) the focus is on a contemporary phenomenon within a real-life context [72]. The inquiry should rely on multiple sources of evidence where data (e.g., interview, observations, documents and archival records) should converge in a triangulating fashion. Strong triangulation of data sources is important to establish the necessary reliability and validity of a research study [72]. The case study also benefits from the prior development of theoretical propositions to guide data collection and analysis [72]. Here we apply a single-case study, spanning a period of eight years. We used a prior model of OGD value generating mechanisms to guide us in the data collection and analysis [37].

A Critical Realist-based case study is an iterative process. To abstract from specific empirical instances to mechanisms, one must distinguish general and essential conditions that underpin the phenomenon from incidental and nonessential conditions, i.e., spurious effects [12]. Accordingly, we first need to identify the structural components of the mechanism in order to understand how these components interact in order to produce the emergent outcome. Next, we need to identify and analyze the outcome tendency. And finally, we must identify the context that influences the outcome [12]. This study employed the methodological principles offered in [12], [76]. These principles include: (1) Explication of events; (2) Explication of structure and context; (3) Retroduction; (4) Empirical corroboration; (5) Validation of explanatory power. These principles do not recommend specific case study methods, but rather identify essential elements needed to derive theoretical statements of generative mechanisms [76].

The first principle suggests a “thick description of case *story* including actions and outcomes” [76]. p. 796. This is done by describing the sequence of events that links the initial conditions to the observed outcome. The second principle involves describing the structural entities, constituent parts, contextual conditions existing in the case, as well as identification of the relationships among the entities. The third principle involves using retroduction to identify and elaborate on the tendencies of contextual factors and mechanisms that may have interacted to generate the observed outcomes. Retroduction is the mode of inferential reasoning that reconstructs the conditions for

the occurrence of an empirical phenomenon and seeks to identify the generative mechanisms *that generalize beyond the immediate instance of the phenomenon* [9], [17], [67]. The fourth principle involves analytical validation of the proposed mechanism based on case data and assessment of explanatory power of each mechanism relative to alternative explanations. Finally, the fifth principle involves employing multiple approaches to support the causal analysis based on a variety of data types and sources, analytical methods, investigators and theories [76].

## Field Site

Recognizing the paucity of in-depth field studies on the *use* of OGD, our strategy was to deepen this understanding by studying one particular case in depth. We sought to identify an organization that could give us a unique and exemplary source of insights on this topic. After going through a number of cases where OGD was used to generate value, we chose the case of Opower which is a relatively mature company that has used OGD from the day it was founded. The company's innovative use of data has spurred a great deal of attention, enabling us to collect secondary data on the company and its impacts from multiple sources.. Specifically, we conducted semi-structured interviews in September 2013 with three members of Opower: a (technical) product manager, a lead analyst and an energy efficiency specialist. Interview questions were sent in beforehand and the interview was tape recorded for further analysis. Subsequent follow-up correspondence included one more employee, a senior manager, as well as in depth analysis of company documents.

As we didn't have longitudinal primary data, we used secondary data to re-create the chain of events leading to the current status, including a Harvard Business Review Teaching Case [16], an economic analysis of Opower's impact on energy efficiency [2], as well as a number of published interviews and newspaper articles. Afterwards, the key informants had an opportunity to review the event description to ensure validity [72].

Opower is an energy tech company that currently works with over 90 energy utilities servicing 22 million homes. Their main mission is to help everyone, everywhere, save energy. Founded in 2007 by two college friends, Alex Laskey and Daniel Yates, the company has now grown to over 400 employees and operates in three continents. Opower was founded on a simple premise: they want to engage the millions of people who are blissfully unaware about their energy use. To do so, they provide people with information. Not only information on their own energy consumption, but compared to other similar households, putting every customer's energy use in personal perspective.

Opower merges and analyzes utility and third party data streams to create individual customer profiles, and subsequently uses these profiles to generate personalized insights that are delivered through different channels. When provided with better information and suggestions on how to decrease energy consumption, people are empowered to take greater control of the way they use energy, regardless of age, income, education, or access to technology. In October 2013, the Opower home energy reports had helped people around the world save over three terawatt hours of energy and more than \$350 million on their energy bills. Opower uses open data to generate economic value for their investors and their utility partners, as well as social value in the form of decreased use of energy by their residential customers, leading to monetary savings for households and reduced CO<sub>2</sub> emissions.

Opower uses data from the U.S. Residential Energy Consumption Survey (RECS) to understand how households are using energy. The survey provides region-specific data on end-use energy consumption patterns, such as the type and efficiency of appliances used by the consumers, the systems and energy sources they use to heat and cool homes, among other topics. Opower combines these data with data from the U.S. Census Bureau on the mix of gas and electric heating sources in a given county, to create location-specific baselines to use when analyzing an individual's home energy consumption. On top of these data, Opower has built an analytics engine to inform and motivate customers. Their platform can store and process 15-minute interval data from smart meters as well as second-level data from millions of in-home devices at large scale and high speed. This, in turn, helps Opower deliver their Home Energy Reports with tips that are personalized for individual customers, for instance, identifying and suggesting the replacement of inefficient heating or cooling systems.

## **IV.5 Event Analysis**

In this section we develop a description of events based on multiple sources of evidence for construct validity. Following Hansen and Birkinshaw, the three main phases of the innovation process are: idea generation, idea conversion and idea diffusion. In the following analysis, after explaining the context of the case, we cluster the chain of events according to these three main phases [27].

### **Case Context**

Climate change has emerged as one of the most important economic policy issues of the early 21st century. The pollutants that contribute to global warming are commonly

known as greenhouse gas emissions. Carbon dioxide (CO<sub>2</sub>) is probably the best known greenhouse gas, representing 85% of all greenhouse gasses in the U.S. Electricity production is the largest single source of global warming pollution in the U.S., responsible for nearly 40% of greenhouse gas emissions. Traditionally, economists and policymakers have focused on relative prices as the primary force driving energy demand [2]. There are, however, three problems with price-based approaches to energy conservation. First, it has not been politically feasible to implement carbon taxes, at least in the U.S. Second, measuring the effects of an energy efficiency subsidy on energy use requires knowledge of the elasticities of demand for energy efficient goods. Third, while subsidies are in theory innocuous because they are transfers, they are in practice a large drain on increasingly-limited public funds. Spurred by these problems, interest has dramatically increased in non-price energy conservation programs that are informed by insights from behavioral science and evaluated via randomized trials [2].

A McKinsey report published in July 2009 estimated that there was a huge potential for energy-efficiency increases in the United States, and that a 23% reduction in energy usage was possible by 2020, resulting in large cost savings for the economy [48]. The report asserted that beyond the economics, efficiency represents an emissions-free energy resource. If captured at full potential, energy efficiency could abate approximately 1.1 gigatons of carbon dioxide emissions per year by 2020 and therefore serve as an important bridge to a future era of advanced low-carbon supply-side energy options. However, the study highlighted a number of barriers to the realization of significant efficiency gains, including large initial outlays of capital required to improve infrastructure, the fragmentation of efficiency opportunities, societal apathy and simple lack of awareness. While the overall potential for energy-efficiency gains was vast, it was spread out across industrial, commercial and residential buildings, making widespread cooperation difficult. Additionally, the incentive and motivation of individuals and corporations to take responsibility for improvements by themselves were seen as being low [48].

## Idea Generation

The college friends and founders of Opower, Alex Laskey and Daniel Yates, went separate ways after graduating from Harvard but became reacquainted a few years later. They soon discovered their mutual interest in preserving the environment, both of them being inclined to join the fight against climate change. In 2006 they started to discuss potential ideas. Originally, they came up with five ideas, all of which focused on reducing emissions: two non-for-profit and three for-profit. The final idea, however,

was born when Yates was examining his own energy bill and realized that he didn't understand the numbers. He could not gain understanding of whether his consumption was high or low in comparison to other households, nor what he could do about decreasing his consumption. His confusion triggered the idea of providing people with understandable and relevant information about their energy consumption, complete with tips on how to save energy. He confronted Laskey with the idea, which he immediately liked. From Laskey's previous work in political polling, he knew that just about everybody agreed when asked if saving energy was important, and yet nobody knew if they were any good at it [16].

Consequently, the founders started gathering external information on the topic of environmental sustainability. During this process they came across Robert Cialdini, Regents' Professor Emeritus of Psychology and Marketing at Arizona State University, who for over 30 years had studied how people were persuaded by social norms [16]. Cialdini and colleagues had discovered that providing high-energy consuming households with prescriptive normative information regarding the average home energy usage in their neighborhood, constructively decreased energy consumption. In contrast, for households that were initially low in their base rates of energy consumption, the same descriptive message produced a destructive boomerang effect, leading to increased levels of energy consumption. Interestingly, adding an injunctive component to the message (a smiley token) proved reconstructive by buffering this unwelcome boomerang effect, meaning that for people who were initially low in energy consumption, the same descriptive normative information combined with an injunctive message of approval led to continued consumption at the desirable low rate, rather than a significant move toward the mean. Moreover, despite concerns that normative interventions would only have an effect for a short time, the longer-term results indicated that the effects of the normative messages continued to be strong even four weeks after the initial intervention [2], [55], [68].

After reading about Cialdini and colleagues' work, the founders saw an opportunity to partner with electrical and gas utility companies to gain access to consumers' energy-usage data, and create a program to drive efficiency gains. In order to do so, they would use data on individual consumption, compare them to data on average energy consumption, and use insights from Cialdini's work on normative influences to create incentives for people to change their behavior. They approached Cialdini and offered him a role as chief scientist in their new venture to help provide the most effective normative messaging possible, which he accepted [16].



## Idea Conversion

The founders next discussed their ideas with politicians in the Texas state legislature and found that early interest in their concept was high [16]. A legislature that implemented incentives for utilities to decrease consumption was consequently passed. Afterwards, Yates and Laskey decided to approach a number of potential utilities customers to validate demand for the information-based product. In short, the founders managed to raise \$1.5 million in seed capital and then \$15 million from venture capitalists, and a startup company was born. Over the next few months, Opower worked out a four-stage framework for customer interaction; analyze, engage, measure, and sustain [2], [16]. The process began with an analysis of the customers' energy-usage habits by sifting through data from the utility, while also pulling in data from other (open) sources that provided insight into the demographics, family type and income levels of customers.

Each report contained two key features: The first was the Action Steps Module, which gave consumers information about steps they could take to conserve energy. This included recommendations to improve home energy efficiency, such as attic insulation, installation of energy-efficient lighting and replacing appliances. However, the primary suggestions centered on simple changes in usage behavior such as turning off electronics and lighting when not at home. The second feature of the home energy report was a Social Comparison Module, as shown in Figure 3. The module provides visual graphics to illustrate a customer's energy consumption relative to customers in households with roughly equivalent square footage, drawing on property data and data from the Residential Energy Consumption Survey (RECS). The Social Comparison Module, coupled with the action steps, makes it relatively easy for customers to understand what they need to do in order to reach parity with their neighbors. To reinforce *good behavior* the reports include smiley tokens alongside a good score. This idea is based on Cialdini's findings: people do not want to just save energy; they want to be acknowledged for their efforts [16].

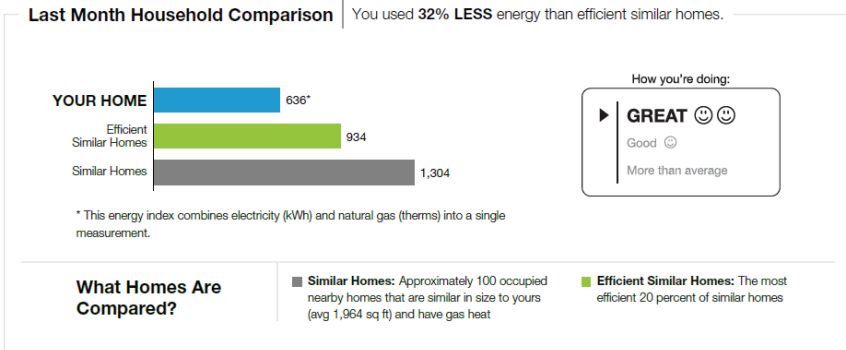


Figure 3: Social comparison module. Source: Opower

In 2007, Laskey and Yates ran a pilot study on 35,000 homes. At the start of the program, they split the households into two groups with similar demographic profiles, consumption patterns, weather, energy prices and economic conditions. The test group received the home energy reports, while the control group did not. Afterwards, they analyzed the consumers' energy-usage data and applied a rigorous methodology to measure and verify the results from the targeted reports against a control group [16]. Subsequently, Opower ran a large-scale data analysis over time with replicable results, in accordance with EPA guidelines for energy-efficiency programs. The electrical utility companies saw an almost immediate 1.5% to 3.5% savings in energy usage in those households that received the personalized report. While the numbers seemed low, a Time magazine article put the potential for these savings into perspective by translating them to the net impact of CO<sub>2</sub> emissions saved if the product would be used nationwide, stating that society would be cutting carbon and saving money at the same time. The founders were most excited about the potential scalability of the idea. The utility companies could extend the program to nearly 100% of their customer base and maintain the program's cost effectiveness. The program cost utilities just \$0.03 per kilowatt-hour of energy saved, while providing a savings of 75 gigawatt-hours per 100,000 households, far exceeding the success of other energy-efficiency programs that utility companies had implemented [16].

### Idea Diffusion and Impact

Opower's first client partner was Sacramento Municipal Utility District. In 2008, Opower (actually the company's name was Positive Energy until 2009) went to Illinois and then Minnesota, deploying their first Automatic Meter Reading (AMR) solution. In

2009, the company started servicing Colorado, Massachusetts and Virginia, implemented an online customer portal and started to use a data services layer for the first time. Also in 2009, the founders were invited to the White House to discuss with President Obama what companies in the U.S. were doing to combat global warming. Shortly thereafter, BusinessWeek named them one of 50 Tech Start-Ups to Know About and Opower was featured in a front-page story in the New York Times. In 2010, they went to California, hired their 100<sup>th</sup> employee, and deployed an advanced meter infrastructure. President Barack Obama visited the Opower headquarters in 2010, highlighting in his speech how their accomplishments were making homes more energy efficient, saving people money, generating jobs and putting America on the path to a clean energy future. At that point in time, Opower reported 100 gigawatt hours of energy saved and \$10 million saved by households. Soon after, a Time Magazine article was published, effectively disseminating the idea of how people's behavior could be modified by informing them about their choices and how this behavioral change had the ability to generate both monetary and CO<sub>2</sub> savings.

One year later Opower went overseas and set up in the UK, reporting 500 gigawatt-hours and \$50 million dollars saved! In 2011, Opower came to New York and in the same year won the two largest Smart Grid deployments in the U.S. In 2012, Opower went live with their new Energy Social App and reported one terawatt hour of energy saved. In October 2013, this number was up to three terawatt hours, which is the equivalent to around 30% of the energy produced by solar energy technologies in the U.S. every year. This translates to over 4.6 billion pounds of CO<sub>2</sub> abated and about \$350 million saved on household energy bills. An independent economic research paper from 2011 evaluated the effects of the Opower Home Energy Reports and found that the Average Treatment Effects (ATE) of Opower's programs ranged from 1.4 to 3.3% of baseline usage, with an unweighted mean ATE of 2.0% [2]. The evaluation also showed that treatment effects increased markedly as a function of pre-treatment usage, although not even the lowest consumption households increased usage in response to the treatment. The analysis concluded that this experiment showed how the simple act of informing users, a treatment that had no effect on relative prices, could persistently affect usage by as much as an 11- 20% short run price increase or a 5% long run increase would have accomplished! [2].

## **IV.6 Structural and Contextual Analysis**

In this step we need to identify the components of social and physical structure, contextual environment, along with the relationships among them [76]. The key

components are the real objects of the case, for example, persons, organizations and systems. They constitute structures, i.e., networks of objects with causal powers [12].

In the process of idea generation, the founders showed proficiency in the use of external sources [27], as they not only took the results from behavioral science to heart, but went right to the source and recruited one of the researchers that had conducted the experiments. We need to conceptualize the *openness* of the founders and their willingness to use external information in their development of an idea. A company's absorptive capacity captures the ability of a firm to recognize the value of new, external information, assimilate it and apply it to commercial ends, which is critical to its innovative capabilities [14]. Opower's primary value proposition is based on generating information and insights from various sources of data. As said earlier, their mission is to use information to empower end users to take control of their own energy use. The fact that Opower could assimilate external knowledge, as well as being able to combine externally available data with internal data for new insights, indicates the company's high level of absorptive capacity. We propose that this absorptive capacity has positively influenced data-driven innovation, which is reflected in the many innovative technical solutions Opower has built around use of these data, such as their patent pending analytics engine.

We gathered from our interviews that Opower is not dependent on the data being free of charge, at least not in the current context. However, being spared from cumbersome payment processes and complex licensing issues has significantly simplified the use of these data, and decreased transaction costs. Furthermore, while Opower has gathered an impressive amount of data on energy use from smart meters that could be aggregated to hypothesize about general use behavior, they still prefer to use the RECS survey. Partly because it comes from an independent and respectable source, and partly because it contains micro data that allow the company to manipulate very granular information and extract insights that suit their unique needs. Opower is also using other types of OGD, for instance weather data, geographic data and demographic data. The general availability of different datasets clearly benefits the company. In general, their impression is that the more availability of high quality, sufficiently granular data, the better. Regarding accessibility, Opower has experienced that links to websites containing datasets have been changed or even taken down, making it difficult to maintain a library of content across multiple years. Therefore a central repository would prove beneficial for the company.

While external data sources offer massive potential for commercial use, they are outside the control of the organization consuming them, and thus their quality may be

unknown and their supply uncertain [19], introducing a certain amount of risk for the private sector users. After analyzing which factors affected the use of OGD in the Opower case, we identified two types of risk. The first has to do with data management within the public sector, which must consider the quality of data, the management of risk, including security and privacy considerations, as well as the data stewardship processes. What Opower perceives as the biggest issue with using data supplied by the government is that the granularity is in some cases not as desired, the data are not updated frequently enough and in some cases new types of information are based on old survey data. Therefore, the quality of the data itself and the data stewardship processes are an issue for Opower. Privacy and security related considerations are still relatively unexplored as potential barriers to the generation of value from OGD, especially as the datasets released do by definition not include data that are subject to valid privacy, security or privilege limitations, as governed by other statutes. However, when combined with other types of data, such as readings from smart meters, the potential threat to privacy is certainly a risk that Opower has acknowledged. That said, worries about breach of privacy are not recognized as a barrier to value generation in the case of Opower.

The second risk involves the sustainability of the resource. While Opower is not entirely dependent on any specific governmental data-source, it would certainly be inconvenient for the company if the quality of data used would deteriorate or if data should cease to be collected due to an unsustainable business model. The sustainability of open government data as a resource is highly dependent on the attention an OGD initiative receives from politicians. The U.S. government has shown substantial interest in opening government data, highlighted by the president's executive order from May 2013, making open and machine readable the new default for government information. The executive order is to be the manifestation to the longstanding commitment to release and leverage data in support of enhanced transparency and accountability, improved government services, and a stronger economy [75].

For Opower as a data analytics company, one of the most important factors is their own technological infrastructure. Their cloud-based technical platform enables Opower to sift through vast amounts of data, amongst other readings from smart-meters that are imported every 15 minutes. They utilize a high frequency storage for storing data from 50 million homes. In addition to the storage, they also utilize an energy analytics engine that analyses 35 billion events per month and makes 1,500 calculations per second. The engine combines internal customer interaction data with data from external sources, i.e., data sourced from the utilities companies and open government

data. The data analytics engine is what gives Opower the ability to compare thousands of different homes' smart meter reads to find tiny fluctuations that indicate that certain homes are over-heating or over-cooling at certain thermostat set-points. The third major component is the interaction engine which generates 35 billion reports every month. While still highly dependent on the paper based reports, Opower is constantly evolving new channels using modern platforms, such as web-based portals and applications for smart phones and tablets.

The company's ability to analyze digital information is, of course, highly dependent on the infrastructure in the operating country, such as the availability of smart meters in homes and the general telecommunications infrastructure that allows them to access and transfer these massive amounts of data to the company's internal (cloud based) data storage. The utilities adoption of modern interaction channels is also a factor. Originally, Opower only used paper based reports, but with evolving technology has continued to develop new, more cost-effective channels. However, paper based still remains the company's main channel due to several reasons. For instance, there is a lack of correct email addresses in the utilities customer databases, and people are still being more responsive to paper based messaging than to email based. Moreover, the new technology enabled channels are dependent on the degree to which the utilities use technology to engage with customers. In a recent white paper (Is Mobile Turning into a Missed Opportunity?), Opower compares utilities' use of web-based and mobile-based technologies to that of banks, concluding that the utilities sector in the U.S. still has a long way to go to catch up.

## **IV.7 Retroduced Mechanisms**

The principle of retroduction, the core of the critical realist explanatory model, is derived from the ontological assumption of emergence and epistemological focus on explanation. The objective is to identify the most complete and logically compelling explanation of the observed events given the specific conditions of the contextual environment [76]. If there are existing mechanisms in the theoretical knowledge of a field, they are adapted to fit the specifics of the given case. However, if no existing mechanisms are adequate to explain the phenomena being studied within a specific context, a new mechanism is proposed [76]. This type of activity can be divided into two sub-steps [12]: (1) identify the contextual elements and (2) look for mechanisms. Retroduction can be seen as being some sort of a thought trial to identify and describe the elements of the causal mechanism and the contextual influences responsible for its activation [76].

## Contextual Enabling Factors

In the following sections we conceptualize the factors that we can identify as having enabled value generation from OGD in the case of Opower. Consequently, we identify which of the value generating mechanisms discussed in section 3.2 might explain how the use of OGD is being transformed to value in this case.

### **Absorptive Capacity**

The presence of valuable external sources of knowledge does not imply that the flow of new ideas and external knowledge into firms is a self-governing or easy process. External knowledge can only be assimilated and integrated with the firm's (internal) knowledge base when the firm has internal competencies that facilitate such processes. Absorptive capacity captures the ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends, thereby being critical to its innovative capabilities [14]. Modern information technologies perform an important role in the development and maintenance of a firm's absorptive capacity. Absorptive capacity has been conceptualized as a multidimensional construct that consists of three distinct yet interrelated capabilities: identification of valuable external knowledge, assimilation or transformation of valuable external knowledge, and application of assimilated external knowledge [42], [63]. While a firm's absorptive capacity is dependent upon prior related knowledge [14], the predominant theoretical view is that absorptive capacity is an organizational capability, not an asset [42], [63].

In order to conceptualize the firm's absorptive capacity, we use findings from previous research. A firm's ability to absorb valuable external knowledge depends on its level of prior related knowledge [14]. Yet, in addition to its knowledge base, a firm needs to develop structures and processes that facilitate knowledge absorption. Prior research finds that two types of organizational capabilities impact absorptive capacity: coordination capabilities and socialization capabilities. Coordination capabilities enhance knowledge exchange across intra- and inter-organizational boundaries, whereas socialization capabilities capture a firm's ability to produce a shared ideology that offers organizational members a collective identity [63]. However, empirical findings give an indication that for digitalized data-driven innovation, the firm's IT-specific absorptive capacity matters more than its general absorptive capacity [40]. IT capabilities allow firms to use technology to develop external relationships and collect knowledge from the external environment, as well as to increase knowledge application capability [63]. In the context of creating business value, we conceptualize absorptive capacity as a multi-dimensional construct, where IT capabilities and

complementary organizational capabilities (coordination and socialization) positively influence a firm's ability to identify, assimilate, transform and apply valuable external knowledge [63].

## **Openness**

The increasing importance of data, often termed big data, has been widely discussed in the media as well as in academic circles. However, one very important enabler for the generation of value from data seems to get much less attention, namely, openness. Openness has been identified as a driver of value from social technologies: the greater the degree of openness, the easier it is for an industry to capture the value from social technologies [50]. There still is, however, very little agreement about exactly what *open* means or how openness can facilitate the generation and appropriation of value from data. Big and open data share the commonality of using data in new, technology-intensive ways to gain insights [46]. It is suggested that openness is a factor that combines unrestricted availability with accessibility and technical interoperability [71]. Openness implies that organizations get the opportunity to integrate large and small volumes of data from internal and external sources to yield new insights. Accordingly, we conceptualize openness of data as a multidimensional construct, reflecting the availability of data, the technical accessibility of data as well as the legal dimension of openness (use of open licenses), [18], [22], [37].

## **Resource Governance**

When opened up, government data become a common, shared resource, available for use within an open network of public and private stakeholders. However, this resource is still governed by the public sector as the main collector of the data. Therefore, we include resource governance as a construct intended to reflect the importance of proper governance mechanisms for the ability to generate value from this resource. Based on [37], we conceptualize resource governance as a multi-dimensional construct that encompasses not only data management, but also the promotion of equitable and the sustainable provision of the data. Sustainable systems are those that meet current needs of many individuals involved in producing and using a common resource without compromising the ability of future generations to utilize the resource [32]. Ongoing analysis of more than 900 major change initiatives in the public sector indicates that 61% of these initiatives do not yield the hoped-for impact, and that a major factor in such cases is a lack of the skills, mind-sets and behaviors critical to sustaining change [51]. Resource governance must fit with, and respond to, a dynamic strategy that supports evolving national goals, and creates sustained institutional reforms. Key to



any benefits, whether economic or social, is managerial leadership and political support [7], [51].

Furthermore, data management must ensure that the data is of sufficient quality. Unknown, inconsistent or unsatisfactory quality of OGD leads to substantial risks for validity and relevance. It is also important to give the correct context to the data, as government data are in many cases collected or created for specific purposes, and thus could be misleading if taken out of that context [19]. Accordingly, use could be stimulated if more information about the way open data are collected and processed were to be provided by including metadata [80]. The security of information must also be considered. There is an ongoing debate regarding possible privacy risks in relation to the use of OGD, and even greater concerns have been voiced regarding the implementation and adoption of smart meters in homes. However, in a recent study, perceived privacy risk was not found to be a significant influence on smart meter adoption intentions [77]. This *privacy paradox* is well known in research and may result from users' perceptions regarding the sensitivity of information disclosed [77]. It is important that resource governance sufficiently addresses all concerns regarding information privacy; otherwise these concerns might adversely affect the mechanisms that can transform OGD to value.

### **Technical Connectivity**

The current trend towards a massive increase in the generation of data, as well as wider access to different kinds of data, has important implications for both public and private organizations. This trend is supported by recent advances in technology: the technical ability to manage and openly disseminate big and small datasets; the ability to analyze, mash up and make sense of different types of data; and the networking capabilities to access and link data from various sources. Research suggests that the scale and scope of changes brought on by use of data are set to expand greatly as series of technology trends accelerate and converge. To capture value from data, public and private organizations will have to deploy technologies that can help individuals and organizations to integrate, analyze, visualize and consume the growing torrent of available data [49]. Technical connectivity is conceptualized as a multi-dimensional construct that describes the availability of technologies that allows users to store, access and combine the data. The construct consists of three dimensions: 1) the infrastructure that facilitates data exchange between government agencies, private sector firms and the public 2) dissemination of software, including data organization management software, as well as analytics and discovery software and 3) access via multiple platforms, such as mobile and web-based platforms.

## Identifying Mechanisms

We can identify one macro-micro level mechanism that explains the *why* of value generation in this particular case. We can call it the motivational effects of value generation or the motivation mechanism. While the hope of generating economic value is a well-established reason for any innovation, the social value aspect plays a somewhat bigger role in this case. The Opower case illustrates how the general awareness of global warming and the prospect of generating social value by increasing energy efficiency influenced and motivated both policy makers and Opower founders. While the founders were both contemplating for-profit and non-for-profit ideas, there is one unintended benefit they received from going the for-profit way while pursuing a social-goal: The economic value generated has helped the company attract talents that might have proved difficult without the financial resources they gained. This includes sought-after technical skills that have enabled the company to evolve their use of technology for data analytics and dissemination purposes. But most employees are equally attracted to the company's mission: To help everyone, everywhere save energy. The generation of social and economic value has not only positively influenced the capabilities of the company, it has also brought awareness of the generative power of data to policy makers (referencing Obama's visit to Opower headquarters), utilities companies and the general public

Our main focus here, however, is to explain the *how* of value generation. If we start by trying to fit the case of Opower to Figure 2, we can see that the value is mostly generated by stakeholders outside of the public sector. While the public sector provides the data and is responsible for creating the incentives that make utilities willing to sacrifice a part of their income in order to achieve more energy efficiency; the main value generation happens from small changes in the energy use of 22 million households, which collectively amounts to considerable savings in energy use. This indicates that the value generating mechanisms are located on the right side of the matrix. If we look at the vertical dimension, we have to determine whether this case presents a purely social focus, a purely capital one or somewhere in between. As Opower is a privately owned company that is returning a healthy return on investment to their shareholders, they do not present a purely social, not-for-profit organization. However, Opower is a double bottom-line company, and their primary mission is to help people save energy. This mission is what drives the company and they have shown amazing results, motivating people around the world to increase energy efficiency, which have resulted in a considerable reduction in CO<sub>2</sub> emissions. The archetypical value generating mechanism that seems to be most illustrative of the

Opower case is the innovation mechanism in the upper right hand corner; however, in the case of Opower, the precise mechanism might lie closer to the middle due to the company's strong emphasis on social value generation.

In the real world, there is rarely a single mechanism that explains a certain outcome. We could say that Opower's analysis of data and clear presentation of results has increased transparency of information, which, in turn, has affected the energy use behavior of households and motivated the utilities companies to be more socially responsible in their resource allocation. Using a similar logic, use of OGD has encouraged households and utilities to be more efficient in their use of energy. However, in both cases these are second-order mechanisms that depend on Opower's innovative use of technology to transform raw data into valuable information. This indicates a potential synergy caused by interplay between different mechanisms, resulting in enhanced value generation. Finally, while Opower is certainly using government provided information, we could not identify any explicit collaboration between the public sector and the company on either the data itself or any finding based on the data. Our conclusion is that in this case the main mechanism that can explain value generation is what we call (socially motivated) *data-driven innovation*.

### **The Data-Driven Innovation Mechanism**

Firms are increasingly procuring knowledge from external sources in their innovative activities [43]. Many models have been developed to explain how firms can exploit external knowledge, ranging from simple free riding to consulting with lead users to utilization of public sources of knowledge [13]. A central part of the innovation process concerns the way firms go about organizing their search for new ideas that have commercial potential. New models of innovation have suggested that many innovative firms have changed the way they search for new ideas, adopting open search strategies that involve the use of a wide range of external actors and sources to help them achieve and sustain innovation [43]. Empirical results strongly suggest that firms with more open knowledge search strategy, having access to a larger number of information sources that can provide ideas and resources, tend to be more innovative [43], [45]. Accordingly, a lack of openness of firms to their external environment may indicate that managers over emphasize internal sources and under emphasize external sources.

The end-to-end process from idea generation to business growth can be represented as an innovation value chain, comprising three main phases: idea generation, idea conversion and idea diffusion [27]. Within these three main phases, six linking tasks

are necessary for the innovation chain [27]. The first phase represents the firms' efforts to acquire the different types of knowledge necessary for innovation, involving three linking tasks: in-house idea generation, cross-pollination and external sourcing [27]. The next phase is the process of transforming this knowledge into new products, services, business processes or behavioral innovations. This activity involves two linking tasks: screening and funding of new ideas and developing new ideas into viable products, services or businesses. This activity may again involve a combination of firms' internal and external resources. The final stage in the innovation value chain is the diffusion of firms' innovations: spreading developed ideas within and outside the company. Empirical research shows a clear causal link from knowledge sourcing through innovation to business growth and productivity [45], [64].

When data have been made available and accessible, current knowledge may be synthesized with different types of open data, and used for generating new knowledge, ideas and value propositions. Available technologies can be used to convert the data into information that is further used in product-, process- or behavioral innovations. These innovations are then diffused to society to form and establish new structures and generate different kinds of value. Finally, the new structures provide the foundation for new data and new innovation, and the loop repeats itself. Information and data are the basic currency across this whole ecosystem. The basic belief behind making the data available to external uses without restrictions to use is that the owner or custodian of data may not be best placed to understand the potential future uses of the data they hold. Waste and the destruction of value could occur if government set rules of access to information which fails to recognize the requirements of unforeseen users and uses. That is, too tight rules of engagement may unintentionally constrain the beneficial use by third parties or eventual end-users in the data-driven innovation process.

Figure 4 shows the conceptual model for the data driven innovation mechanism. The four multi-dimensional enabling factors are all capable of positively influencing the innovation mechanism, although to a different degree, depending on the contextual environment. To illustrate, while strong leadership within government might have a huge influence on the impact of an OGD initiative in the beginning stages, the influence might be less for the more mature initiatives.

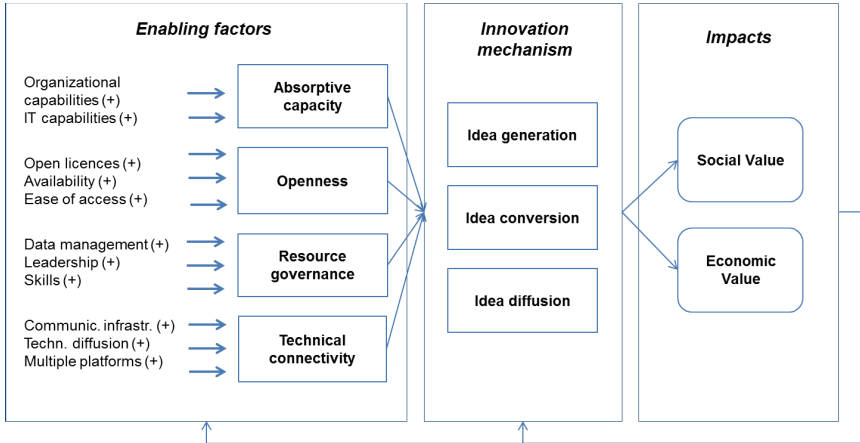


Figure 4: Conceptual model of the data driven innovation mechanism

The innovation mechanism itself is not inherently different from any generic innovation mechanism, other than in this case; the innovation is to a large degree based on the use of data. The impacts of the data-driven innovation can range from generation of internal economic value, measured as company revenue or profits, to the generation of more widely appropriated economic and social value. Although we have conceptualized innovation as business innovation, which is most often characterized by the drive towards economic value generation, the social benefits resulting from the innovation can be even more profound, as in the case of Opower. This chain of events explains how the use of a particular resource – OGD - in a specific company, and in the context of certain macro level structures, can result in the generation of macro-level value. Finally, the generation of economic and social value can further positively affect some of the micro level constructs we present in the model, resulting in a virtuous cycle of openness, innovation and value generation.

### IV.8 Empirical Corroboration

We need to ensure that the proposed mechanism of data-driven innovation offers the causal and explanatory power to explain how value is generated from the use of OGD. In an open system there are a number of mechanisms at play. After identifying a certain mechanism, we can identify others by asking how the context influences the triggering of the mechanism [12], [67]. The original driver behind the whole Opower adventure was the founder’s interest in preserving the environment. The founders

originally came up with five ideas, three non-for-profit and two for-profit. After discussing these ideas among themselves and with others, they agreed upon explaining the energy bill, an idea they termed Energy Bill 2.0. In an attempt to generalize this finding, we have suggested that a trigger for the OGD value generating mechanisms is the motivational aspect of social and economic value – rainbows and carrots. Furthermore, the value generated also supports the company’s aspirations after the fact. Opower has not encountered problems in hiring people with the right technical and data analytics skills, even though there is a reported shortage of those skills in the market. One potential reason for this is the company’s somewhat unique combination of economic and social value generation. The economic value enables them to provide good salaries and a working environment where employees get an opportunity to learn how to use the latest technologies. The social value generation attracts those that are passionate about creating an impact, changing the world for the better. We therefore suggest one contextual macro-micro mechanism: the mechanism of motivation.

However, the case highlights two other elements that have proved to be important to the process of idea generation and conversion: 1) The apparent lack of information on the energy bill and 2) Cialdini’s findings on how information about people’s neighbors had an effect on their behavior. We propose that in order to be able to utilize Cialdini’s finding and the available OGD to their benefit, the company needed to possess a good degree of absorptive capacity. Moreover, as Cialdini’s findings predict, people need information, not only about their own behavior but as compared to the behavior of others. Therefore, Opower needed access to aggregate data on consumer behavior as well as data on energy consumption in individual homes. The technical infrastructure enabled them to collect, store, merge and analyze data from different sources. These two factors together explain Opower’s ability to innovate from external datashare. Absorptive capacity can only benefit a company if it has access to valuable and re-usable external information sources. The opportunity for data-driven innovation arises from the fact that Opower had access to the data they needed from the U.S. government and that the data was of sufficient quality and sustainability (a trusted resource) for them to use.

We have suggested that the archetypical mechanism that can best explain the value generation in the case of Opower is the mechanism of innovation, where novel combinations of resources and new methods of analysis lead to the generation of a new service, which, in turn, leads to the transformation of the energy market, thus increasing value [69]. We propose that in the case of Opower, the innovation is in essence based on utilizing a specific resource, namely, data. We use the term data-

driven innovation to generalize this finding to include any innovation that is, to a large degree, based on utilizing data to generate value. The value generated for the entrepreneurs and the government has been economic in nature, i.e., the establishment of a growing company which leads to profits for the investors, new jobs for employees and higher tax payments to the government. Opower has had a 280% five-year revenue compound annual growth rate, and in October 2013 there were almost 100 available positions posted on their webpage. Furthermore, Opower has generated social value, as reduced CO2 emission contribute to environmental sustainability, which again can lead to a collective improvement in the lives of individuals, benefitting society as a whole.

## **IV.9 Discussion and Conclusion**

The aim of this research was to develop a framework for the generative mechanisms that explain *how use of OGD can stimulate the generation of value*, and use these mechanisms to explain how a private company can generate economic and social value by using OGD. We constructed a framework based on a review of the OGD and value literature, using general and established mechanisms to explain how use of OGD as a resource can result in generation of value. The framework is based on two dimensions that are intended to capture the impact of openness: First, how openness is enabling external stakeholders to contribute to the generation of value from OGD as a resource and second, how society as a whole can appropriate value in the presence of openness. Using the framework, we identified four different archetypical generative mechanisms, each of which represents a particular type of cause-and effect relationship between OGD and value. Subsequently, we used a critical realism based research approach to analyze the case of Opower. We propose that a mechanism of data-driven innovation fits the Opower case and has the ability to explain how Opower used OGD to generate value and what contextual enabling factors were involved. We suggest that our analysis has shed light on how the potential value generation motivated the idea, how the lack of information triggered the idea of innovating from data, and how the interplay of four multi-dimensional factors enabled the innovation, which ultimately led to the generation of economic and social value. Finally, we propose that the resulting value generation has, in turn, positively influenced, not only the company's own ability to generate even more value but also policymaker's interest in OGD, thus increasing our understanding of the value of data.

The contributions to knowledge are as follows: a) The two-by-two framework with the four archetypical generative mechanisms can be used to explain the complex relationship between openness, data and value, b) the illustration of the potential

virtuous cycle within the OGD ecosystem and c) the conceptual model of data-driven innovation, where we suggest a nomological network of constructs that illustrates how OGD can be used to generate value. The contribution to practice is the strategic framework which can help governments position and prioritize their strategic OGD goals and the identification of factors that can enable the generation of value from OGD through the innovation mechanism. We furthermore believe that the insights offered in this study can be applied to guide OGD initiatives that have the goal of driving innovation in the private sector. One hypothetical example of such use could be a public body planning a new OGD strategy. A discussion about the initiative's goals should include what kind of value should be generated (sought after impacts) and how. After reflecting on the different levels of interactions between stakeholders, the difference between social and economic value and value generation vs. appropriation of value, the public body in question could compare its current status (for example, strong in creating internal efficiency but weak in maintaining two-way relationships with external stakeholders) to the desired status and subsequently start a discussion on how to move from the current to the desired state.

A limitation of the study is that we use only the first four principles suggested in [76], as the fifth principle includes using multiple methods to perform validation of explanatory powers. While principle 5 is out of the scope of this particular paper, our aim is to continue to test our conceptual model using a variety of methods in future work. Furthermore, as the paper analyses only one case study, our ability to generalize our findings is somewhat limited. It is well known that OGD come in different types and structures, the use of the data varies greatly and the contextual circumstances vary between countries and even individual initiatives. More cases should be analyzed to reflect on the generalizability of this model. An interesting way to achieve this could be to follow the method used in [31] where the authors examine which *configurations* of identified enabling factors actually lead to an observed successful outcome in different cases [31]. A further limitation is that while we focus on the enabling factors and the interplay between them, as well as the mechanism of data driven innovation, we did not really analyze the possible interplay between the innovation mechanism and the other generative mechanisms, other than reflecting on how the innovation mechanisms might have triggered second-order transparency and efficiency mechanisms. We suggest that an in-depth analysis of this relationship might present an opportunity for future research.

We propose that openness is in itself an important enabler of the generation of value from data, as openness enables both the generation and appropriation of value, not only



by the organization that produces or collects the data, but also by external stakeholders. However, while openness might be seen as the necessary condition in this context, it is insufficient on its own. Governments should increase the usability and re-usability of their data by focusing on resource governance, where the aim is to ensure the quality and sustainability of the resource and to minimize risk for external users. On the demand side, companies need to possess the right capabilities as well as having access to technical platforms that enable them to link disparate data sources and transform these data to valuable information products and services. Finally, we suggest that the act of opening data and treating them as a shared resource available for use by anyone has the ability to contribute to a synergistic relationship between different mechanisms, thus contributing to the simultaneous generation of economic and social value.

### *Acknowledgments*

We would like to thank the editors of this special issue and the anonymous reviewers for their insightful comments and constructive feedback. We would also like to express our sincere appreciation to the staff at Opower for giving us valuable insight into their company. For providing general support, access to data and enlightening conversations, we would like to extend our gratitude to the Basic Data Team at the Danish Agency for Digitization, the data-veterans at KMD and our colleagues at CBS.

## **IV References**

- [1] A. Afonso, L. Schuknecht and V. Tanzi, Public sector efficiency: evidence for new EU member states and emerging markets, *Applied Economics*, vol. 42, no. 17, pp. 2147-2164, 2010.
- [2] H. Allcott, Social norms and energy conservation, *Journal of Public Economics*, vol. 95, no. 10-11, pp. 1082–1095, 2011.
- [3] M. Avital, K. Lyytinen, S. Iacono, K.L. Kraemer, V. Sambamurthy, and S. Sawyer, Data rich and data poor scholarship: Where does IS research stand? in *Proceedings of the 28th International Conference on Information Systems (ICIS)*, Montreal, 2007, pp. 1-6.
- [4] K. Axelsson, U. Melin and I. Lindgren, Exploring the importance of citizen participation and involvement in e-government projects - practice, incentives and organization, *Transforming Government: People, Process and Policy*, vol. 4, no. 4, pp. 299-321, 2010.
- [5] B. Barber, *Strong Democracy: Participatory Politics for a New Age*, Los Angeles: University of California Press, 1984.
- [6] J. Benington, From private choice to public value?, in *Public Value: Theory and Practice*, (J. Benington and M. Moore Eds.). Houndmills, Basingstoke: Palgrave MacMillan, 2011, pp. 31-49.

- [7] J.C. Bertot, P.T. Jaeger and J.M. Grimes, Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies, *Government Information Quarterly*, vol. 27, no. 3, pp. 264-271, 2010.]
- [8] J.C. Bertot, P.T. Jaeger and J.M. Grimes, Promoting transparency and accountability through ICTs, social media, and collaborative e-government, *Transforming Government: People, Process and Policy*, volume 6, no. 1, pp. 78-91, 2012.
- [9] R. Bhaskar, General introduction in *Critical Realism: Essential Readings* (M. S. Archer, R. Bhaskar, A. Collier, T. Lawson, and A. Norrie, Eds.). London: Routledge, 1998, pp. ix-xxiv.
- [10] R. Bhaskar, *A Realist Theory of Science*, London: Routledge, 2008.
- [11] C. Bowman and V. Ambrosini, Value creation versus value capture: Towards a coherent definition of value in strategy, *British Journal of Management*, vol. 11, no.1, pp. 1-15, 2000.
- [12] B. Bygstad and B.E. Munkvold, In search of mechanisms. Conducting a critical realist data analysis, in *Proceedings of Thirty Second International Conference on Information Systems (ICIS)*, Shanghai, 2011, pp. 1-15.
- [13] H. Chesbrough, Open innovation: A new paradigm for understanding industrial innovation, in *Open Innovation: Researching a New Paradigm* (H. Chesbrough, W. Vanhaverbeke and J. West, Eds.). Oxford: Oxford University Press, 2006, pp. 1-14.
- [14] V. M. Cohen and D.A. Levinthal, Absorptive-capacity: a new perspective on learning and innovation, *Administrative Science Quarterly*, vol. 35, no. 1, pp.128–152, 1990.
- [15] A. Cordella, E-government: towards the e-bureaucratic form?, *Journal of Information Technology*, vol. 22, no. 3, pp. 265-274, 2007.
- [16] A.J.C. Cuddy, K.T. Doherty and M.W. Bos, OPOWER: Increasing Energy Efficiency through Normative Influence, *Harvard Business School*, 9-911-016, 2012.
- [17] B. Danermark, M. Ekström, L. Jakobsen, and J.C. Karlsson, *Explaining Society: Critical Realism in the Social Sciences*. London: Routledge, 2002.
- [18] T. Davies. (2010, August) Open data, democracy and public sector reform: A look at OGD use from data.gov.uk. A report based on an M.S. Thesis, University of Oxford, Oxford. [Online]. Available: <http://www.opendataimpacts.net/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf>.
- [19] S.S. Dawes. (2012, June) A realistic look at open data. World Wide Web Consortium (W3C). [Online]. Available: [http://www.w3.org/2012/06/pmod/pmod2012\\_submission\\_38.pdf](http://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf)
- [20] European Commission. (2011, December) Open data: An engine for innovation, growth and transparent governance. European Commission Information Society. [Online]. Available: [http://ec.europa.eu/information\\_society/policy/psi/docs/pdfs/directive\\_proposal/2012/open\\_data.pdf](http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/open_data.pdf)
- [21] S. Fleetwood, Sketching a socio-economic model of labour markets, *Cambridge Journal of Economics*, vol. 35, no. 1, pp. 15-38, 2011.
- [22] S. Gigler, S. Custer and H. Rahmetulla. (2011, December) Realizing the vision of OGD-opportunities, challenges and pitfalls. Open Development Technology Alliance, World Bank Publications. [Online]. Available: <http://www.scribd.com/doc/75642397/Realizing-the-Vision-of-Open-Government-Data-Long-Version-Opportunities-Challenges-and-Pitfalls>

- [23] I. Gil-Saura, M. Frassetto-Deltoro and A. Cervera-Taulet, The value of B2B relationships, *Industrial Management & Data Systems*, vol. 109, no. 5, pp. 593-609, 2009.
- [24] V. Grover, K. Lyytinen, A. Srinivasan, and B. Tan, Contributing to rigorous and forward thinking explanatory theory, *Journal of the Association for Information Systems*, vol. 9, no. 2, pp. 40-47, 2008.
- [25] M. Gurstein. (2011, February) Open data: Empowering the empowered or effective data use for everyone?, *First Monday* (16). [Online]. Available: <http://tinyurl.com/dylszv>
- [26] A. Halonen. (2012) Being Open about data. Analysis of the UK open data policies and applicability of open data. The Finnish Institute in London. [Online]. Available: <http://assets.finnishinstitute.studiocoucou.com/media/W1siZiIsLjIwMTMvMDEvMzEvMTNfMDNfMTNfNjM1X2JlaW5nXzIwb3Blb18yMGFiY3V0XzIwZGF0YS5wZGYiXV0/being%2520open%2520about%2520data.pdf?sha=9b1040ca>
- [27] M. Hansen and J.M. Birkinshaw, The innovation value chain, *Harvard Business Review*, vol. 85, no. 6, pp. 2-12, 2007.
- [28] T. M. Harrison, S. Guerrero, G.B. Burke, M. Cook, A. Cresswell, N. Helbig, J. Hrdinová, and T. Pardo, Open government and E-government: Democratic challenges from a public value perspective, in *Proceedings of the 12th Annual International Conference on Digital Government Research*, New York, 2011, pp. 245-253.
- [29] P. Hedström and R. Swedberg, *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge: Cambridge University Press, pp. 1-31, 1998.
- [30] N. Helbig, A.M. Cresswell, G.B. Burke, and L. Luna-Reyes, The dynamics of opening government data: A white paper, Center for Technology in Government, The Research Foundation of State University of New York, Albany, NY, 2012.
- [31] O. Henfridsson and B. Bygstad, The generative mechanisms of digital infrastructure evolution, *MIS Quarterly*, vol. 37 no. 3, pp. 907-931, 2013.
- [32] C. Hess and E. Ostro, *Understanding Knowledge as a Commons – From Theory to Practice*. Cambridge, Massachusetts: MIT Press, 2006.
- [33] S. Holmberg, B. Rothstein and N. Nasiritousi, Quality of government: What you get, *Annual Review of Political Science*, vol. 12, no. 1, pp. 135-161, 2009.
- [34] N. Huijboom and T. Van den Broek, Open data: an international comparison of strategies, *European Journal of e Practice*, vol. 12, pp. 1-13, 2011.
- [35] K. Janssen, The influence of the PSI directive on open government data: An overview of recent developments, *Government Information Quarterly*, vol. 28, no. 4, pp. 446-456, 2011.
- [36] M. Janssen, Y. Charalabidis and A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, *Information Systems Management*, vol. 29 no. 4, pp. 258-268, 2012.
- [37] T. Jetzek, M. Avital and N. Bjørn-Andersen, Generating value from open government data, in *Proceedings of the 34th International Conference on Information Systems*, Milan, 2013, pp. 1-20.
- [38] S. Kim and J. Lee, E-Participation, transparency, and trust in local government, *Public Administration Review*, vol. 72, no. 6, pp. 819-828, 2012.

- [39] I. Kolstad and A. Wiig, Is transparency the key to reducing corruption in resource-rich countries?, *World Development*, vol. 37, no. 3, pp. 521-32, 2009.
- [40] H. Koski, ICT Outsourcing, User-Driven and Open Innovation Strategies in the Generation of New Data-Based Solutions, Helsinki, ETLA Working Papers No. 7, 2013.
- [41] S. Kuznets, *Population, Capital and Growth: Selected Essays*. London: Heinemann Educational, 1974.
- [42] P.J. Lane, B.R. Koka and S. Pathak, The reification of absorptive capacity: a critical review and rejuvenation of the construct, *Academy of Management Review*, vol. 31, no. 4, pp. 833–863, 2006.
- [43] K. Laursen and A. Salter, Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms, *Strategic Management Journal*, vol. 27, no. 2, pp. 131–150, 2006.
- [44] D. Linders and S.C. Wilson, What is open government? One year after the directive, in *The Proceedings of the 12th Conference on Digital Government Research*, (dg.o.2011), MD, USA, 2011, pp. 262-271.
- [45] J.H. Love, S. Roper and J.R. Bryson, Openness, knowledge, innovation and growth in UK business services, *Research Policy*, vol. 40, no. 10, pp. 1438-1452, 2011.
- [46] A. Marton, M. Avital and T. Blegind Jensen, Reframing open big data, in *Proceedings of 21st European Conference on Information Systems (ECIS)*, Utrecht, 2013, pp. 1-13.
- [47] V. Mayer-Schönberger and Z. Zappia (2011, October) Participation and power: Intermediaries of open data. Conference draft prepared for the Berlin Symposium. [Online]. Available: <http://tinyurl.com/bo3pyl8>
- [48] McKinsey & Company. ( 2009, July) Unlocking energy efficiency in the U.S. Economy, McKinsey & Company. [Online]. Available: [http://www.mckinsey.com/client\\_service/electric\\_power\\_and\\_natural\\_gas/latest\\_thinking/unlocking\\_energy\\_efficiency\\_in\\_the\\_us\\_economy](http://www.mckinsey.com/client_service/electric_power_and_natural_gas/latest_thinking/unlocking_energy_efficiency_in_the_us_economy).
- [49] McKinsey & Company. ( 2011, June) Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute. [Online]. Available: [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI\\_big\\_data\\_full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx)
- [50] McKinsey & Company. ( 2012, July) The social economy: Unlocking value and productivity through social technologies, McKinsey Global Institute. [Online]. Available: [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/The%20social%20economy/MGI\\_The\\_social\\_economy\\_Full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/The%20social%20economy/MGI_The_social_economy_Full_report.ashx)
- [51] McKinsey & Company. (2012, December) Transforming Government through Lean Management, McKinsey Global Institute. [Online] Available: [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/The%20social%20economy/MGI\\_The\\_social\\_economy\\_Full\\_report.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/The%20social%20economy/MGI_The_social_economy_Full_report.ashx)
- [52] K.D. Miller and E.W.K. Tsang, Testing management theories: critical realist philosophy and research methods, *Strategic Management Journal*, vol. 32, no. 2, pp. 139–158, 2011.

- [53] M. H. Moore, *Creating Public Value Strategic Management in Government*, USA, Harvard University Press, 1995.
- [54] K. Nilsen, Economic theory as it applies to public sector Information, *Annual Review of Information Science and Technology*, vol. 44, no. 1, pp. 419-489, 2010.
- [55] J.M. Nolan, P.W. Schultz, R.B. Cialdini, N.J. Goldstein and V.Griskevicius, Normative social influence is underdetected, *Personality and Social Psychology Bulletin*, vol. 34, no. 7, pp. 913–923, 2008.
- [56] OECD, *Oslo Manual – Guidelines for Collecting and Interpreting Innovation Data*, Paris: OECD, 2005.
- [57] OECD. (2012, January) *ICT Applications for the smart grid: Opportunities and policy implications*, OECD Publishing, OECD Digital Economy Papers no. 190. [Online]. Available: <http://www.oecd-ilibrary.org/content/workingpaper/5k9h2q8v9b1n-en>
- [58] OECD. (2012, January) *Machine-to-machine communications: Connecting billions of devices*, OECD Publishing, OECD Digital Economy Papers no. 192. [Online]. Available: [http://www.oecd-ilibrary.org/science-and-technology/machine-to-machine-communications\\_5k9gsh2gp043-en](http://www.oecd-ilibrary.org/science-and-technology/machine-to-machine-communications_5k9gsh2gp043-en)
- [59] Open Knowledge Foundation. (2012, November) *Open data handbook documentation 1.0.0*. Open Data. [Online]. Available: <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>
- [60] E. Pol and S. Ville, Social innovation: Buzz word or enduring term?, *Journal of Socio-Economics*, vol. 38, no. 6, 878–885, 2009
- [61] R. Pollock. (2009, May) *The economics of public sector information*, Faculty of Economics, University of Cambridge, Rufus Pollock. [Online]. Available: <http://www.rufuspollock.org/economics>
- [62] M.E. Porter and M.R. Kramer, *Creating shared value*, *Harvard Business Review*, vol. 89, no. 1, pp. 62-77, 2011.
- [63] N. Roberts, P. Galluch, M. Dinger and V. Grover, *Absorptive Capacity and Information Systems Research: Review, Synthesis, and Directions for Future Research*, *MIS Quarterly*, vol. 36, no. 2, pp. 625-648, 2012.
- [64] S. Roper, J. Du, and J.H. Love, *Modeling the innovation value chain*, *Research Policy*, vol. 37, no. 6-7, pp. 961–977, 2008.
- [65] B. Rothstein, *Anti-corruption: the indirect ‘big bang’ approach*, *Review of International Political Economy*, vol. 18 no. 2, pp. 228-250, 2011.
- [66] S. Sarker, S. Sarker, A. Sahaym and N. Bjørn-Andersen, *Exploring value cocreation in ERP vendor–partner relationships*, *MIS Quarterly*, vol. 36, no. 1, pp. 317-338, 2012.
- [67] A. Sayer, *Method in Social Science: A Realist Approach* (2nd ed.), London: Routledge, 1992.
- [68] P.W. Schultz, N.J. Nolan, R.B. Cialdini, N.J. Goldstein, and V. Griskevicius, *The constructive, destructive, and reconstructive power of social norms*, *Psychological Science*, vol. 18, no. 5, pp. 429-435, 2007.
- [69] J.A. Schumpeter, *The Theory of Economic Development*, Cambridge: Harvard University Press, 1934.

- [70] C. Shapiro and H.R. Varian, *Information Rules: A Strategic Guide to the Network Economy*, Boston: Harvard Business School Press, 1999.
- [71] Y. Tammisto and J. Lindman, Definition of open data services in software business, in *Proceedings of Software Business - Third International Conference, ICSOB, Cambridge, MA, USA, 2012*, pp. 297-303.
- [72] Yin, R.K. *Case Study Research: Design and Methods* (4th ed.). Los Angeles: Sage, 2009.
- [73] H. Yu and D.G. Robinson. (2012, February) The new ambiguity of 'open government'. *Social Science Research New York*. [Online]. Available: <http://ssrn.com/abstract=2012489>
- [74] C.K. Williams and E. Karahanna, Causal explanation in the coordinating process: A critical realist case study of federated IT governance structures, *MIS Quarterly*, vol. 37, no. 3, pp. 933-964, 2013.
- [75] White House. (2013, May) Executive order -- making open and machine readable the new default for government information. The White House. [Online]. Available: <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government>
- [76] D. Wynn Jr. and C.K Williams, Principles for Conducting critical realist case study research in information systems, *MIS Quarterly*, vol. 36, no. 3, pp. 787-810, 2012.
- [77] P. Wunderlich, D. Veit and S. Sarker, Examination of the determinants of smart meter adoption: A user perspective, in the *Proceedings of the 33rd International Conference on Information Systems, Orlando, 2012*, pp.1-19.
- [78] M. Zachariadis, S. Scott and M. Barrett, Methodological Implications of Critical Realism for Mixed-Methods Research. *MIS Quarterly*, vol. 37, no. 3, pp. 855-880.
- [79] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, and R.S. Alibaks, Socio-technical impediments of open data, *Electronic Journal of e-Government*, vol. 10, no. 2, pp. 156 – 172, 2012.
- [80] A. Zuiderwijk and M. Janssen, A comparison of open data policies and their implementation in two Dutch ministries, in *Proceedings of the 13th Annual International Conference on Digital Government Research (dg.o. 2012)*, College Park, MD, 2012, pp. 84-89.

## **PAPER V: Generating Sustainable Value from Open Data in a Sharing Society**

Presented at IFIP WG 8.6 in Aalborg, 2014

In *Creating Value for All Through IT*, edited by Birgitta Bergvall-Kåreborn and Peter Axel Nielsen, pp. 243-260. Springer Berlin Heidelberg

# Generating Sustainable Value from Open Data in a Sharing Society

Thorhildur Jetzek<sup>1</sup>, Michel Avital<sup>2</sup> and Niels Bjørn-Andersen<sup>3</sup>

1 Copenhagen Business School, tj.itm@cbs.dk

2 Copenhagen Business School, michel@avital.net

3 Copenhagen Business School, nba.itm@cbs.dk

**Abstract.** Our societies are in the midst of a paradigm shift that transforms hierarchal markets into an open and networked economy based on digital technology and information. In that context, open data is widely presumed to have a positive effect on social, environmental and economic value; however the evidence to that effect has remained scarce. Subsequently, we address the question how the use of open data can stimulate the generation of sustainable value. We argue that open data sharing and reuse can empower new ways of generating value in the sharing society. Moreover, we propose a model that describes how different mechanisms that take part within an open system generate sustainable value. These mechanisms are enabled by a number of contextual factors that provide individuals with the motivation, opportunity and ability to generate sustainable value.

**Keywords:** Sharing society, Sustainable value, Value generating mechanisms, Open data

## V.1 Introduction

The impact of the digital revolution on our societies can be compared to the ripples caused by a stone thrown in water: spreading outwards and affecting a larger and larger part of our lives with every year that passes. One of the many effects is the emergence of an already unprecedented amount of digital data that is accumulating exponentially. Moreover, a central affordance of digitization is the ability to distribute, share and collaborate, and we have thus seen an “open theme” gaining currency in recent years [18]. These trends are reflected in the explosion of Open Government Data (OGD) initiatives around the world: governments striving to open access to various data-sources and making them available for use and re-use for commercial or other purposes. However, while hundreds of national and local governments have established OGD portals and are being followed by similar initiatives by international



institutions, civil society organizations and even businesses, there is a general feeling that the open data initiatives have not yet lived up to their true potential. This feeling is not without good reason; the recent Open Data Barometer report highlights that strong evidence on the impacts of OGD is almost universally lacking [11].

This lack of evidence might, however, not be as surprising when we consider the complexity of the task at hand. How do we measure and evaluate something as complicated as the value of open data? And if we cannot show that value is generated – how can these initiatives be sustained? Our take on this dilemma is that before we even start trying to evaluate, we need to develop a deeper understanding of *to what ends* (what it is we want to accomplish with open data) and *by what means* (how this can possibly happen). And due to the embryonic nature of the open data phenomenon, research on open data impacts and affordances is still lacking (see, for instance, [59]).

Our goal with this paper is to contribute to this gap in knowledge by developing a theory on how open data can generate sustainable value. We build on the notion that the world is at an inflection point, where technological advances and boundary-crossing social challenges have come together to create a paradigm shift. This notion was very evident at the 2014 Annual Meeting of the World Economic Forum in Davos, where the shift towards the sharing economy and the current social challenges faced by our societies were repeatedly mentioned. These social challenges are numerous and urgent, and both social, environmental and economic in nature. They range from economic inequality, unemployment and poor social conditions to chronic diseases and climate change.

Given the complexity and cross-boundary nature of these challenges, a new approach is necessary. In particular, we need an approach where social and technological progress co-evolves in order to generate value [37]. We would like to contribute towards such an approach through addressing the following research question: “How can use of open data stimulate the generation of sustainable value?” The definition of *sustainable value* represents a move away from the previously dominant economic value focus and moving towards a focus on proactive, concerted efforts of businesses, government institutions and the overall community in addressing social challenges in innovative and holistic ways that generate social, environmental and economic value for all stakeholders and future generations [50].

This theory development paper is structured as follows: We first elaborate on the recent trends discussed earlier based on a review of the literature, after which we build theoretical foundations. We propose that we are experiencing a paradigm shift in how people make decisions in their quest for creating and appropriating value enabled by

the transformative power of information technology. These recent technical and social developments call for a re-interpretation of the behavioral assumptions used in some of our most prominent value theories. We proceed to discuss four value generating mechanisms that describe different paths through which the use of data can be transformed to value. Our contribution here is a framework that adds a new network-based dimension to the well-established market-based mechanisms of efficiency and innovation (exploitation and exploration). We propose that all four mechanisms interact within an ecosystem we call the *sharing society*, and that the interaction between the private and the public sectors via the different mechanisms can generate the synergies that are necessary to tackle the highly complicated social challenges we face. Finally, we visualize our theory with a nomological network that shows the relationship between the main antecedents that can enable and stimulate value generation from data, the value generating mechanisms and the resulting, sustainable value.

## **V.2 Open data empowers the sharing society**

The industrial economy was primarily based on production, where Gross Domestic Product (GDP) was the key measure of economic activity, and buying and selling goods and services on the market became the basic foundation for the value generating mechanisms. A dichotomy of the state and the market was one feature of the industrial economy [38]. The state's responsibility was to create a structure around the goods that did not fit the market mechanisms, so-called market failures. However, continuing advances in digital technologies have started to disrupt the status quo by altering the way people think, live and work, and by rearranging value pools [30]. This development has led to entirely new forms of products and services that, in many cases, are based on data collection, data re-use and data sharing. The digital revolution, including the digitization of nearly all media, the ubiquity of Internet access, the proliferation of mobile phones and the growth of the Internet of Things, has created multiple affordances which subsequently require a change in the basic assumptions that are used when discussing value maximizing behavior.

First, the digital revolution has led to an explosion in the generation and availability of data. Digitization has affected two important features of data: 1) when data become easily accessible to more than one person at a time, they acquire the feature of non-rivalry and; 2) when marginal costs incurred by re-production and distribution are drastically reduced, re-use of data becomes economically feasible [35], [44]. Second, following the advent of the "open theme", data that have traditionally been locked up

in closed repositories are now increasingly becoming open and available for use and re-use [18]. Openness has changed one important feature of digital data, namely, making them non-excludable. Accordingly, when opened up, digital data become a shared resource - a public good or what has been termed 'digital commons' [19]. Open data, particularly government data, have generated a great deal of excitement around the world for its potential to empower citizens, change how government works and improve the delivery of public services. The economic value potential from open data has also been celebrated, and recently estimated as the equivalent of 3 trillion USD per annum globally [29]. In short, computing and networking capabilities combined with openness are expected to drive massive social, political and economic change [24].

Most of the open data initiatives today are driven by governments around the world, the most important driver seemingly being their expectation that open data will stimulate the generation of considerable social, economic and environment value for their societies [22]; [58], [51]. These initiatives have either been fuelled by the political ideology of Open Government [25] or been focused on economic value generation, highlighting the potential of re-use of data for innovation and efficiency [21]. However, despite this interest, there is still not much evidence of value generation, probably as most of the open data initiatives are still in their infancy [11], [20], [59]. Thus far, most of the published material on value generation is based on predictions and hypotheses, and there is still considerable confusion regarding what needs to be done, who should do it and how (*by what means*), as well as why we are doing it and for whom (*to what ends*). For instance, the technological availability and accessibility of data need to be conceptually separated from the political openness required to drive transparency and accountability [57]. One is concerned with the usability of the resource; the other is a mechanism whereby data are being used for specific purposes.

Interestingly, in the 77 countries surveyed in the Open Data Barometer, less than one in 10 datasets (or 71 out of the 821 public datasets reviewed) were truly open, i.e., available online, in bulk, and under an explicit open license [11]. In most cases, datasets are provided in aggregate formats (often in XLS or CSV), or are not machine readable (PDF files). The Open Data Barometer report also points towards important issues of data quality and trustworthiness. Of the 113 datasets that were available in machine-readable and openly licensed form, researchers found 15 where the sustainability was questionable and 20 that were not up-to-date or published in a timely fashion [11]. Moreover, while recent developments with open data offer unprecedented access to large scale data sets on a huge variety of topics, successful use of such data requires a rather different skill set to skills encapsulated in many current views of data

literacy, currently not addressed in a school curriculum (Ridgway et al., 2013). Finally, while technology enabled services are key to harnessing the value of data, they are often limited by problems with usability, searchability, language, sufficiency of technological infrastructure and availability of computers and Internet access for many segments of the general population [4]. We conclude that governments still have some way to go: from embracing the value potential of open data to actually implementing the required value generation enabling structures.

Our aim with this paper is to contribute to an improved conceptual clarity in this discussion by presenting an open system we call the sharing society. We define the sharing society as an *open economic and social system in which information technology is leveraged to empower individuals, corporations, non-profits and governments with data that are shared, reused and transformed to sustainable value through different mechanisms.*

Following [6], we assume that the goal of the social system is to maximize utility or what we have defined as sustainable value. We use Coleman's framework as a meta-theory to explain the micro to macro level relationship between use of open data and the generation of sustainable value. Coleman's framework underscores the mediating role of individuals in linking macro-level variables such as social structure and the behavior of the social system [31]. A theory which can generate macro-level empirical generalizations as specific propositions may be thought of as "a theory of individual action, together with a theory of how these actions combine, under specific rules, to produce systemic behavior" [6], p. 20. But how is the sharing society conceptually different from the industrial economy? In the following sections, we outline the basic assumptions for our theory on how open data can generate value in the sharing society.

### **V.3 Behavioral assumptions for value generating *mechanisms***

Any theory comprises a set of assumptions from which empirical generalizations have been derived [32]. In the social sciences, a satisfactory explanation must ultimately be anchored in hypotheses or assumptions about individual behavior [12]. By making a set of explicit, behavioral assumptions, we can highlight that all macro-level, societal phenomena are inherently derived from human beliefs and actions, and that certain mechanisms mediate these actions between the initial conditions and the observed outcome. Since the core behavioral assumptions of a theory often form the foundation of its mechanistic explanations, it is crucial that these assumptions are explicitly defined and tested during the early stage of empirical research [49].

The classical economic model of strategic interaction assumes that people are (a) only concerned about bettering their own material situation, and (b) able to calculate the optimal strategy for doing so. However, new results from behavioral economics have cast doubt on these assumptions, as evidence shows that people are, in fact, concerned with unobservable payoffs such as reputation, fairness or the well-being of others [45]. Furthermore, people are also concerned with how actions taken today might affect their future well-being. Accordingly, we must consider the wider social impact as well as the sustainability of our current actions [9], [45], [47]. A number of dimensions that can make classical models of rationality somewhat more realistic, while sticking within the vein of fairly rigorous formalization are described in [46]. These include: (a) limiting the types of utility functions, (b) considering the possibility of people having a "multi-valued" utility function and (c) recognizing the costs of gathering and processing information.

The general willingness of people to generate sustainable value is not really contested, but measures have remained elusive, and hence the emphasis on material wealth in economic theory. However, while monetary based indicators offer a convenient way to measure and compare value, there is a cross-disciplinary consensus that we need a more inclusive measure of welfare or wellbeing [47]. In the field of economics, a number of initiatives propose a move beyond using Gross Domestic Product as a measure of a country's progress. Renewed attention has been given to the concept of public value in the e-government and public policy disciplines. The public value framework is based on the premise that public resources should be used to increase value, not only in an economic sense but also more broadly in terms of what is valued by citizens and communities [2], [33]. These trends have resulted in the development of new, broader measures, such as OECD's Better Life Index.

At the organizational level of analysis, the concept of shared value presents a similar ideology. The premise behind shared value is that by including the generation of social value in their strategies, private companies can improve their competitive advantage at the same time as they contribute to society [38].

While most of us would like to act for the good of society, we do not always have the means or the cognitive capability to do so. Most people have limited attention, suffer from status quo bias and choice overload, and are prone to procrastination [45]. Moreover, people typically do not apply sufficient cognitive effort to calculate an optimal strategy, resorting rather to heuristics which can be influenced by context [46]. Rationality of individuals is limited by the information they have, the cognitive limitations of their minds, and the finite amount of time they have to make a decision.

Bounded rationality is “behavior that is intendedly rational but only limitedly so” [46], p. xxiv. So construed, bounded rationality takes exception with the analytically convenient assumption of hyper-rationality but does not preclude a predominantly rational approach to the study of complex economic organization [56]. Given that people face a multi-dimensional value function with complex relationships between individual and social wellbeing and that we as individuals have limited cognitive ability to rationally choose between different options, we can conclude that collection and dissemination of different types of data and the conversion of these data to information and insight are a key resource.

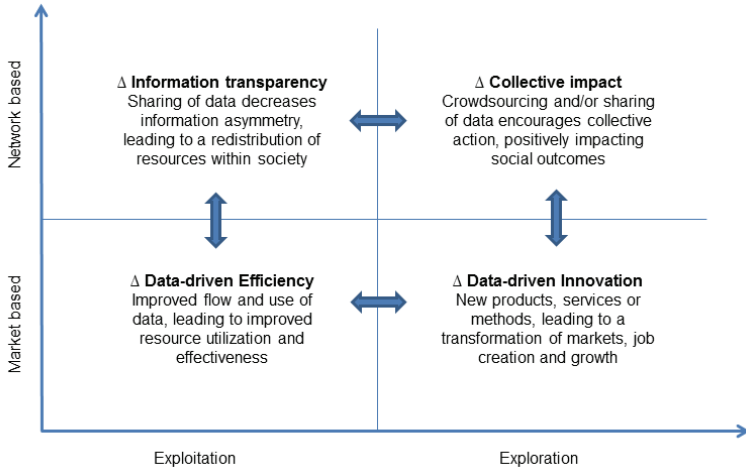
As a foundation for our theory, we make two assumptions. The first is that individuals in general want to go beyond increasing their own material wealth in their value generation efforts. We assume that individuals will strive for sustainable value, including social, environmental and economic value, for all stakeholders and future generations [50]. Sustainable value generation can be the result of collaborative efforts or it can happen through traditional market exchange, but the context in which decisions are made must provide the *motivation*, *opportunity* and *ability* for sustainable value generation to happen.

The second assumption we make is based on the notion of bounded rationality. People will be able to make decisions closer to an “optimum” if provided with the right kind of information about the situation they face. Access to the right kind of information can therefore push the boundaries of our ability to choose rationally and contribute to the generation of sustainable value. This assumption goes across all the proposed value generating mechanisms, i.e., we assume that provided with new data and information, people will choose differently, and this will lead to a sustainable value generation that will happen via different but interlinked mechanisms.

#### **V.4 A framework of four value generating mechanisms**

In order to frame the different value generation strategies, we have formulated a taxonomy/framework consisting of four archetypical value generating mechanisms. The mechanisms are organized after two dimensions, as seen in Figure 1. The x-axis dimension categorizes value mechanisms on whether data are used predominantly to do things better (exploit current resources) or to do new things (explore new opportunities). The y-axis represents our contribution to this traditional classification. The market-based mechanisms are an offspring of the traditional monetary economy, but the “new” network-based mechanisms revolve around information sharing. Market based mechanisms have traditionally been focused on the generation of economic

value, either through lower costs or increased profits and are based on classic economic theories of the likes of Williamson and Schumpeter [42], [55], [56]. However, all mechanisms in the sharing economy to some extent focus on generating sustainable value. All are important and all depend, to a certain degree, on collaboration of people, across the boundaries of organizations and sectors.



**Fig. 1.** A framework of four archetypal mechanisms that transform data to value

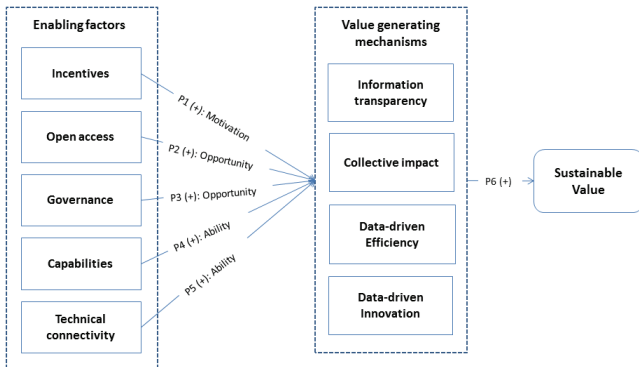
In the following list we summarize how these mechanisms work:

- Information transparency is gained if the data are available, accessible, accurate and trustworthy, and shine a light on a certain subject. The transparency mechanism generates value when individuals acquire new information that result in actions that further lead to a redistribution of resources.
- The collective impact mechanism generates value when a large group of individuals use data to collectively contribute to a common cause, positively impacting social outcomes.
- The data-driven efficiency mechanism generates value when stakeholders use data to improve productive efficiency and effectiveness, which can result in direct cost savings, saved time and effort, as well as improved quality of services.
- The data-driven innovation mechanism generates value when novel use of data leads to new innovative products, services or methods that transform markets and

industries while generating jobs, profits and multiple other affordances, thus resulting in the generation of sustainable value.

## V.5 Generating sustainable value from data in the sharing society

Figure 1 shows the four archetypical mechanisms we propose that can transform data to value. We have, however, not explained how we can *enable* this value generation to happen, depending in all circumstances on the context in which the value generation happens. The model we present in Figure 2 is based on the macro-level model in [23], but it is extended to incorporate the findings from behavioral economics that explain the role and impact of individual behavior. A class of behavioral theories, based on the Motivation-Opportunity-Ability (MOA) framework, shows how motivation, opportunity and ability to perform certain tasks impact the behavior of individuals (see, for instance, [5] and [39]). These models are commonly used in relation to social marketing and information processing [26], [40], public relations [15], knowledge sharing in workplaces [39], [43] and consumer behavior [60]. Broadly speaking, motivation can be defined as goal-directed arousal [26], [40]. In the context of this model, high motivation implies that individuals have the incentive to allocate resources to generate value from data; opportunity refers to the environmental or contextual factors that enable action; and ability represents the power or capacity to act [40].



**Fig. 2.** A model of the sharing society – transforming data to value

We propose that policy makers can influence the motivation of users by offering incentives and the opportunity for value generation not only by supporting open access to data but also by providing risk-free and sustainable, high-quality data resources for



internal and external users. Policy makers should also pay attention to the general ability of individuals to use these data by focusing on the availability of technical infrastructure as well as the general capabilities in society (although these two factors are, of course, also dependent on the provision of the market). We now briefly explain these enabling factors.

## Incentives

The organizational leadership literature has shown that employees who are aware of the positive impact their behavior has on others are more motivated to make a pro-social difference, and that inspiration, a compelling vision and intellectual stimulation, likely enhances employees' motivation to actively engage in knowledge sharing [39]. Motivation can also influence companies to change institutions in order to support social practices, as can be seen in the many nonprofit organizations recently established by for-profit companies [53].

Following [53] and [45], we maintain that people not only chase visible carrots, but they also tend to consider the bigger societal-level impact of their actions, and are willing to participate in the “larger quest for the invisible gold at the end of the rainbow.” We propose that by creating the right incentives, policy makers can motivate different stakeholders to use data to generate sustainable value. We have already seen a number of such efforts, most notably the relatively widespread Hackathons and Datapaloozas, which have been hosted by various municipal and state governments in order to encourage people from all sectors to consider the potential value of open data.

***Proposition 1:*** *Incentives positively influence the motivation of individuals to generate sustainable value from data in the sharing society, hence positively affecting each of the value generating mechanisms*

## Open access

Providing open government data can be seen as a matter of availability, format, accessibility and license [10]. Opening access to data provides everyone in society with the opportunity to use these data to generate value. Openness is a key enabler for value generation from data, as it allows various users to re-use data for different purposes and therefore unlocks the intrinsic value that data hold [29].

The basic belief behind making the data available to external uses without restrictions is that the originator, owner or custodian of information or data may not be best placed to understand the potential future uses of the data they hold. Waste and the destruction

of value could occur if government set rules of access to information which fails to recognize the requirements of unforeseen users and uses. In other words, too tight rules of engagement may unintentionally constrain the beneficial use by third parties or eventual end-users in the process of the diffusion of innovation. Moreover, openness should combine unrestricted availability of data with accessibility and technical interoperability [48]. Therefore, in addition to the general availability, an important dimension to open data is that there should be no legal or technical barriers to use and re-use. Use of open licenses and open data standards can facilitate the re-use of these data.

***Proposition 2:** Open access positively influences the opportunity for individuals to generate sustainable value from data in the sharing society, hence positively affecting each of the value generating mechanisms*

## Governance

When opened up, data become a common, shared resource, available for use within an open network of public and private stakeholders. However, this resource is still governed by the main collector or creator of the data. Therefore, we include governance as a construct that is intended to reflect the importance of data governance for the value that can be extracted from the resource. Based on [23], we conceptualize governance as a construct that describes the quality and sustainability of the data resource, where sustainability means that the common resource must meet the current needs of many individuals without compromising the ability of future generations to utilize the resource [19].

Data governance must consider data quality, data management, data policies, business process management and risk management. Data governance ensures that data can be trusted and that privacy is guarded. Making data accessible and ensuring it is fit for re-use are not insignificant challenges and such efforts raise a wide range of complex questions, including questions on how and when sharing is appropriate [18]. Without governance, the risks of using the data resource for mission critical purposes might become too high, and the data might not come at the right time or be in the right granularity to be of use for information purposes.

***Proposition 3:** Governance positively influences the opportunity of individuals to generate sustainable value from data in the sharing society, hence positively affecting each of the value generating mechanisms*

## Capabilities

As open government data are common, shared resources, the generative ability of the value generating mechanisms depends on certain capabilities in society. A capability can be defined as a measure of the ability of an entity to achieve an objective. In our case, it is the collective ability of individuals and organizations to use and re-use the data. Following [19], we emphasize the need for equitable use of the data resource. Citizen's access to the Internet and their ability to utilize the provided information are considered important for ensuring equitable dissemination [4], [16]. The digital divide can be broadly defined as the gap between those who have access to technologies and those who do not; however, there are, in fact, multiple divides that can exist, of which access to technology is but one. These issues include technology literacy as the ability to understand and use technology, as well as the ability of persons with disabilities to access the content through adaptive technologies [4]. Without the capabilities to access, use and make sense of data, the generative ability of the mechanisms becomes limited.

***Proposition 4:*** *Capabilities positively influence the ability of individuals to generate sustainable value from data, hence positively affecting each of the value generating mechanisms*

## Technical connectivity

The current trend towards a massive increase in the generation of data, as well as wider access to different kinds of data, has important implications for both public and private organizations. This trend is supported by recent advances in technology: the technical ability to manage and openly disseminate big and small datasets; the ability to analyze, mash up and make sense of different types of data; and the networking capabilities to access and link data from various sources. Research suggests that the scale and scope of changes brought on by use of data are set to expand greatly as series of technology trends accelerate and converge. To capture value from data, public and private organizations will have to deploy technologies that can help individuals and organizations to integrate, analyze, visualize and consume the growing torrent of available data [28].

Technical connectivity is conceptualized as a construct that describes the availability of technologies that allow users to store, access, combine and analyze the data. The construct consists of three dimensions: 1) the infrastructure that facilitates data exchange between government agencies, private sector firms and the public, 2) dissemination of software, including data organization management software, as well

as analytics and discovery software and 3) access via multiple platforms, such as mobile and web-based platforms.

***Proposition 5:*** *Technical connectivity positively influences the ability of individuals to generate sustainable value from data, hence positively affecting each of the value generating mechanisms*

## The generation of sustainable value

Following is a short description of the main premises of our suggested model:

- Due to the complexity and cross-boundary nature of today's social challenges, societies need to support various different value generating mechanisms.
- These mechanisms account for the way in which sustainable value can be generated from the use of data. All of the mechanisms are dependent on some sort of collaborative efforts, but these can either be network based or market based.
- All of the mechanisms are dependent on the motivation, opportunity and ability of people to generate value from data.
- While the mechanisms function independently, they can and will interact within an ecosystem of mechanisms called the sharing society, and this interaction can generate valuable synergies.
- The sharing society is defined as an open economic and social system in which information technology is leveraged to empower individuals, corporations, non-profits and government with data that are shared and reused and transformed to sustainable value through different mechanisms.

The mechanisms contribute to the generation of sustainable value by creating different types of value: increase GDP through means such as corporate profits, job creation and tax payments; contribute to society by improving the general wellbeing of individuals through means such as better health, education, social inclusion or equality; or contribute to the livability of the environment through means such as reduced emissions, less traffic congestion or access to clean water.

***Proposition 6:*** *Sustainable value is generated through a system of value generating mechanisms within the sharing society, both directly and through synergies between the mechanisms*

## V.6 Discussion: The transformative power of data

We face many social challenges in this world. While none of the challenges has simple solutions, or can be solved by individual stakeholders or groups, there are ways to overcome many of these challenges – if we utilize the power of the sharing society. The market, by itself, oftentimes lacks the incentives and appropriate models to solve many of these issues, often called market failures. The payback in most market transactions is defined by an implicit or explicit contract, and its timing occurs closely behind, or simultaneous to, the initial contract. Within areas such as public health and social issues, the monetary payback, however, is often vague, uncertain and in the distant future [40].

Market failures are thus currently considered to be either the responsibility of the state or of civil society [38]. However, in many cases the silos of government departments are poorly suited to tackling complex problems that cut across sectors and nation states. Civil society might also lack the capital, skills and resources to take promising ideas to scale [34]. We thus propose that an informal collaboration between the public and private sectors, enabled by openly sharing and re-using data and information can positively influence the generation of sustainable value.

The model of the sharing society shows the potential progress from using data under different contexts, to the mechanisms that explain how the value generation happens, to the actual impact or output – the sustainable value that is generated. We wish to suggest a potential answer to the *to what ends* question, by highlighting the social challenges we are facing today and by showing how different governments and companies have together addressed some of these challenges by using open data. We also wish to delve deeper into the question of *by what means* by pointing out that use of data can generate value through different mechanisms (although these mechanisms interact and influence each other) and by highlighting the importance of the context in which value generation happens. Thus far, there has not been much scientific evidence to support the hypothesis of a link between opening access to government data and value generation. However, the relationship between use of data, the enabling factors, the different mechanisms, and value generation and appropriation can be illustrated with anecdotal evidence. This will be provided below, where we are suggesting possible value generation potentials to four of the largest societal challenges.

## Challenge number 1: Economic inequality

One of the most acute social challenges we face today is the issue of economic inequality within or between nations. While economic inequality might primarily be seen as a violation of social justice, research shows that growing economic inequality since the mid-1970s has contributed to dysfunctional economies. It has even been linked to the recent economic crisis with devastating effects [14], [52]. Inequality is crystallized in a skewed allocation of resources, where the majority of society's resources belongs to certain groups enjoying more opportunities than do others.

To a certain degree, uneven distribution of resources is a part of the capitalistic society's market-based mechanisms. However, in some cases, resources are not allocated in a socially optimal way due to behavior such as opportunism and even corruption. Such outcomes are usually possible where there is an asymmetry of information between people, undermining accountability. Information asymmetry conceals skewed resource allocations which too often prioritize the interests of business, political and military elites over development priorities of the majority of the population.

### *Transparency and accountability to combat inequality*

Transparency refers to the absence of asymmetric information. Information transparency is a characteristic of governments, companies, organizations and individuals that are open to the clear disclosure of information, rules, plans, processes and actions.<sup>1</sup> Information transparency can contribute to improved social outcomes by generating incentives and reducing uncertainty. Increased transparency in an organization's operation increases the outside stakeholders' capacity to hold the insiders accountable and provides stakeholders with information they can use in their own decision making. The transparency mechanism is essentially network based, as it depends on multiple instances of sharing and receiving of information between stakeholders. Eventually, increased transparency should facilitate equitable and effective allocation of resources across boundaries. In the context of accountability, transparency-enhancing mechanisms involving a multitude of stakeholders throughout society can be thought of as creating millions of "auditors" [3]. This "auditing" can mobilize resources from being used where they benefit few at the cost of many, to being used in a more socially responsible manner.

One such transparency agenda for tackling poverty in the global economy was presented by the British Prime Minister, David Cameron, in the G8 meeting at the

---

<sup>1</sup> <http://www.transparency-initiative.org/about/definitions>

World Economic Forum in Davos in January, 2013. The plan was to tackle illicit financial flows, the hidden company ownership that makes such flows possible, and land grabs in developing countries. The claim is that citizens in developing countries are regularly robbed of the benefits of their countries' mineral wealth through poorly negotiated or corrupt backroom deals. Collective global action is essential to improve the transparency of land transactions, thereby attracting more responsible investment that will contribute to sustainable economic growth and reduced poverty.<sup>2</sup> In this case, open access to government data on company ownership, natural resources and taxes enables greater cross-border transparency. Interestingly, key datasets such as Land Registries and Company Registries are least likely of all the datasets reviewed in the Open Data Barometer to be available as open data, suggesting that OGD initiatives are not yet securing the release of politically important datasets that can be vital to holding governments and companies accountable [11]. However, as a part of the G8 transparency agenda, the UK has committed to establishing a publicly accessible central registry of company beneficial ownership, and is undertaking a wider review of corporate transparency.

## Challenge number 2: The Climate challenge

Climate change has emerged as one of the most important economic policy issues of the early 21st century. The pollutants that contribute to global warming are commonly known as greenhouse gas emissions. Carbon dioxide (CO<sub>2</sub>) is probably the best known greenhouse gas, representing 85% of all greenhouse gasses in the U.S. Electricity production is the largest single source of global warming pollution in the U.S., responsible for nearly 40% of greenhouse gas emissions<sup>3</sup>. A McKinsey report published in July 2009 estimated that there was a huge potential for energy-efficiency increases in the U.S., and that a 23% reduction in energy usage was possible by 2020 [27].

The study also highlighted a number of barriers to the realization of significant efficiency gains, including large initial outlays of capital required to improve infrastructure, the fragmentation of efficiency opportunities, societal apathy and simple lack of awareness. While the overall potential for energy-efficiency gains was vast, it was spread out across industrial, commercial and residential buildings, making widespread cooperation difficult. Additionally, the incentive and motivation of

---

<sup>2</sup> <https://www.gov.uk/government/publications/uk-g8-presidency-report-2013>

<sup>3</sup> <http://www.epa.gov/ghgreporting/ghgdata/reported/index.html>

individuals and corporations to take responsibility for improvements by themselves were seen as being low [27].

***Provision of information can enable collective impact***

A green growth agenda requires policy makers to examine and influence behavior in a way that collectively impacts economic, social, and environmental outcomes on multiple scales [54]. Collective impact can be defined as the commitment of a group of actors from different sectors to a common agenda for solving a complex social problem [17]. As the prevailing characterization of human decision making in policy circles has until recently been a rational economic one, a wide range of factors that affect how people make decisions has been excluded from consideration and therefore needs to be considered in predictions of human reactions to environmental conditions or proposed policy initiatives [54].

Results from behavioral science have shown that providing high-energy consuming households with prescriptive normative information regarding the average home energy usage in their neighborhood constructively decreased energy consumption. For those that were doing better than the average, adding an injunctive component to the message (a smiley token) proved reconstructive by re-affirming their “good behavior” [36], [41]. The lesson is: People do not want to just save energy. They want to get information on how they are doing, and most importantly, to be acknowledged for their efforts.

Opower is an energy tech company with a mission: to motivate everyone on earth to save energy. Opower was founded on a simple premise: to engage the millions of people who are in the dark about their energy use. To do so, they provide people with information on their own energy consumption as compared to other similar households, thereby putting every customer’s energy use in personal perspective. Opower merges and analyzes utility and open government data to create individual customer profiles, and subsequently uses these profiles to generate personalized insights that are delivered through various channels. In February 2014, the Opower home energy reports helped people around the world to save over 3.7 terawatt hours of energy and more than \$417 million on their energy bills. We propose that policy makers, civil society and private companies can use different types of collaboration platforms to influence behavior. These platforms should provide people with relevant information as well as feedback mechanisms that enable them to respond to this information.



### Challenge number 3: Efficient use of public resources

The importance of the efficient use of public resources for economic growth and stability, as well as for general well-being has been brought to the forefront by a number of developments over the past decades [1]. During the last 20 years, with the advent of the computer and the Internet as general technologies, a big portion of all processes in industrialized societies has become digitized. Research has shown that ICT has offered the capacity to reduce costs, increase the capability of machinery and increase flexibility in production planning and scheduling, thereby positively influencing productivity and efficiency.

During the initial stage of ICT adoption, many different systems were implemented, each with a specific purpose in mind, including payroll, human resources, production systems, resource planning, etc. This resulted in silos of disparate datasets in no way interconnected or integrated, causing various operational inefficiencies that included the same data being collected and hosted in different places, inability to automate processes across organizational boundaries and considerable overhead from trying to make sense of heterogeneous data sources. As an example, the costs incurred by recreating, verifying and transforming building information were estimated at \$15.8 billion in the U.S. capital facilities supply chain for 2002 alone [8].

#### *Efficient use of data sets resources free*

Increasing cross-boundary interactions and higher levels of information exchange between citizens and government due to digitization have increased the total amount of government data collected and stored. These trends call for more efficient processing of data in order to provide the expected services while still keeping costs under control [7]. Efficiency of public sector organizations can be gained by cutting processing costs, making strategic connections between and among government agencies, and creating empowerment [13]. This allows for better utilization of valuable resources, either by directing them from non-value adding to value adding tasks, or by reducing use in order to increase sustainability.

The Danish government has started an initiative where from 2012 to 2016 all basic government data will be improved in quality and context, and collection and dissemination of the data will be coordinated within the public sector. A national information infrastructure will be established for distribution of government data, with

the aim of making the administration of the basic data registers easier and more efficient. Furthermore, as data will be freely available online, costs related to user support and billing are also expected to be reduced. The total yearly savings for the public sector are projected to be around \$48 million<sup>4</sup>.

The focus of the Danish authorities is on collective savings. The business case calculated would not have been positive for individual institutions or agencies due to the large initial costs incurred by making these big changes to the data model, data quality and data distribution channels. By ensuring a positive internal business case for this initiative, the Danish authorities have increased the likelihood of the initiative being economically sustainable. Moreover, the positive external effect from this project is that integrated government data of better quality will also benefit private industries, such as real estate dealers, insurance companies, the financial sector and the telecom industry, which previously had to spend resources on creating usable information from heterogeneous data-sources. The cost-savings for the private industry are estimated to be around \$90 million per annum when the program is fully implemented.

#### Challenge number 4: Urban planning

Urbanization, the demographic transition from rural to urban, is associated with shifts from an agriculture-based economy to mass industry, technology and service. A hundred years ago, 20% of the global population lived in an urban area; by 1990, just under 40% of this population lived in cities; However, since 2010, for the first time ever, more than half of the world's population is living in urban areas. As this proportion continues to grow, by 2050, it is projected to have increased to seven out of 10 people.<sup>5</sup> With urban congestion on the rise, city planners are looking for new ways to improve transportation. According to the Texas Transportation Institute, the cost of congestion in the U.S. in 2012 was more than \$120 billion, nearly \$820 for every commuter.<sup>6</sup> Similar problems are endured by most of the world's bigger cities.

##### *Data-driven innovation to ease traffic congestion*

Following Schumpeter, we assume that innovation can have economy-wide effects. Innovation brings about novel combinations of resources, and new production methods which, in turn, can lead to the transformation of markets and industries, thus increasing value [42]. The traditional definition of innovation builds on the underlying motive to

---

<sup>4</sup> [http://www.digst.dk/ServiceMenu/English/Digitisation/~media/Files/English/Grunddata\\_UK\\_web\\_05102012\\_Publication.pdf](http://www.digst.dk/ServiceMenu/English/Digitisation/~media/Files/English/Grunddata_UK_web_05102012_Publication.pdf)

<sup>5</sup> [http://www.who.int/gho/urban\\_health/situation\\_trends/urban\\_population\\_growth\\_text/en/](http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/)

<sup>6</sup> <http://d2dt15nnpf0r.cloudfront.net/tti.tamu.edu/documents/tti-umr.pdf>

generate economic value. Numerous studies have confirmed that innovative companies generate above-average returns and that innovative nations enjoy more economic growth.

Innovation, however, can also have clear social consequences. For instance, the computer has dramatically enhanced individual productivity, learning and creativity, and the World Wide Web has enabled the connectivity that has had disruptive effects on many societies, in some cases threatening dictatorship and corruption. In order to generate sustainable value, the innovation must at the very least not have negative social or environmental consequences, and optimally lead to the simultaneous generation of social, environmental and economic value. We use the term innovation to describe the mechanism that uses market forces to allocate resources in order to create a new method, product or service that generates sustainable value.

An increasing number of governments and cities in the world have started to publish open geospatial and traffic data. Many innovative solutions are currently being developed in addition to these data. One example is INRIX, a leading provider of traffic services with the goal of solving traffic worldwide. Their traffic intelligence platform analyzes data from public and private sources, including government road sensors, official accident and incident reports and data on real-time traffic speed, crowdsourced from a large community of local drivers. The company's analysis of crowdsourced data in combination with information from traditional sources provides drivers with insights that help them choose the best way to go, minimizing the amount of time spent.

As the app used to source traffic information from individuals is available for free, INRIX's main source of income is from car-producers, GPS providers and media companies. Moreover, they have recently started to provide data and tools to public information services. In particular, the crowd-sourced data allow much faster congestion analysis than was previously possible.<sup>7</sup> This will allow public traffic engineers to measure and track congestion, thereby offering public decision makers better tools to analyze and manage transportation infrastructure for improved urban planning.

## **V.7 Conclusions**

Our aim with this paper is to generate a theory that can explain the causal connections between use of open data and the consequent generation of sustainable value. We

---

<sup>7</sup> <http://tti.tamu.edu/2010/01/11/texas-transportation-institute-teams-with-inrix/>

propose that the challenges faced by our societies urgently call for new forms of collective action between public and private stakeholders, and that we can move towards such collective action by actively sharing and re-using data across boundaries.

The first contribution of this paper is a new interpretation of the underlying behavioral assumptions. We propose that we are experiencing a paradigm shift in how people interact, enabled by the transformative power of information technology. This calls for a re-interpretation of some of the most prevalent behavioral assumptions underlying our current value theories. We propose that most people are driven not only by the wish to improve their own material situation, but also by the need for subjective well-being and the wish to be a contributing member of society. While opportunism and corruption are certainly relevant behaviors, they can be influenced by social norms. Acknowledging [46], [55], [56]; we further propose that bounded rationality is a reasonable approximation of how people behave but that the boundaries of rationality can be affected, for instance, by providing people with information.

The second contribution is a framework that adds a new dimension to the well-established market-based mechanisms of efficiency and innovation (exploitation and exploration), containing two network-based mechanisms that generate value through information transparency and collective impact. We propose that all four mechanisms interact within an ecosystem we call the sharing society, and that these interactions are capable of generating synergies in value creation. All the mechanisms are dependent on the private and public sector, together providing the motivation, opportunity and ability to generate value from data.

The third and final contribution is the nomological network, where we visualize our theory and show the main antecedents that enable data to be transformed to sustainable value via the sharing society system of mechanisms. These antecedents are supposed to reflect the context within which value generation happens, a context that can differ a great deal between countries and initiatives. We propose that the motivation, opportunity and ability of individuals to use data for value generation are influenced by: the incentives provided; the level of technical and legal openness of data; the maturity of resource (data) governance; the general data-related capabilities in society; and the technological maturity and prevalence. The motivation, opportunity and ability of individuals positively influence the different mechanisms that eventually explain how use of data is transformed to sustainable value.

This paper is limited initially moving only towards theory development; the next step is to use empirical data to test the relationships proposed here. As both open data and the sharing society are emerging phenomena, there is still not much theory to build

upon; however, we are able to borrow from established value generation theories and current research on open data and behavioral economics. Future developments and continuing research will have to testify whether or not the proposed relationships hold or not.

## V References

1. Afonso, A., Schuknecht, L. and Tanzi, V.: Public sector efficiency: evidence for new EU member states and emerging markets. *Applied Economics* 42(17), 2147-2164 (2010)
2. Benington, J.: From Private Choice to Public Value? In: Benington, J. and Moore, M. (eds.) *Public Value: Theory and Practice*, pp. 31-52. Palgrave MacMillan, Basingstoke (2011)
3. Bellver, A. and Kaufmann, D.: *Transparenting Transparency: Initial Empirics and Policy Applications*. World Bank Policy Research Working Paper (2005) <http://tinyurl.com/k77rvlj>
4. Bertot, J.C., Jaeger, P.T. and Grimes, J.M.: Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly* 27(2010), 264-271 (2010)
5. Blumberg, M., and Pringle, C. D.: The missing opportunity in organizational research: Some implications for a theory of work performance. *Academy of Management Review* 7(4), 560-569 (1982)
6. Coleman, J.C.: *Foundations of Social Theory*. Harvard University Press, Boston (1990)
7. Cordella, A.: E-government: towards the e-bureaucratic form? *Journal of Information technology* 22(2007), 265-274 (2007)
8. Curry, E., O'Donnell, J., Corry, E., Hasan, S., Keane, M. and O'Riain, S.: Linking building data in the cloud: Integrating cross-domain building data using linked data. *Advanced Engineering Informatics* 27(2013), 206-219 (2013)
9. Daly, H.E. and Cobb, J.B.: *For the Common Good: Redirecting the Economy toward Community, the Environment and a Sustainable Future*, 2nd edn. Beacon Press, Boston (1994)
10. Davies, T.: Open data, democracy and public sector reform: A look at OGD use from data.gov.uk. (2010) <http://tinyurl.com/7joks46>
11. Davies, T.: *Open Data Barometer: 2013 Global Report* (2013) <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>
12. Elster, J.: *Explaining Social Behavior*. Cambridge University Press, Cambridge (2007)
13. European Commission: *eGovernment Economics Project (eGEP), Measurement Framework Final*. DG Information Society and Media, European Commission (2006)
14. Goldberg, G.S.: Economic inequality and economic crisis: a challenge for social workers, *Social Work* 57(3), 211-224 (2012)
15. Hallahan, K.: Enhancing Motivation, Ability, and Opportunity to Process Public Relations Messages. *Public Relations Review* 26(4), 463-480 (2000).
16. Halonen, A.: *Being Open about Data* (2012) <http://tinyurl.com/c8mz5vt>

17. Hanleybrown, F., Kania, J. and Kramer, M. (2012). Channeling Change: Making Collective Impact Work, *Stanford Social Innovation Review* (2012).
18. Harrison, T.M, Pardo, T.A. and Cook, M.: Creating Open Government Ecosystems: A Research and Development Agenda. *The Future Internet* 4(4), 900-928 (2012)
19. Hess C. and Ostrom, E. *Understanding Knowledge as a Commons – From Theory to Practice*. MIT Press, Cambridge Massachusetts (2006).
20. Hujiboom, N., and Van den Broek, T.: Open data: an international comparison of strategies. *European Journal of e Practice* n° 12 (2011)
21. Jansen, K.: The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly* 28 (4), 446-456 (2011)
22. Janssen, M., Charalabidis, Y. and Zuiderwijk, A.: Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management* 29(4), 258-268 (2012)
23. Jetzek, T., Avital, M. and Bjørn-Andersen, N. *Generating Value from Open Government Data*. In: *Proceedings of the 34th International Conference on Information Systems, Milan, Italy* (2013)
24. Kundra, V.: *Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect*. Jan 2012, Joan Shorenstein Center on the Press, Politics and Public Policy, Harvard, Boston (2012)
25. Luna-Reyes, L.F., Bertot, J.C. and Mellouli, S.: Open Government, Open Data and Digital Government. *Government Information Quarterly* 31(1), 4-5 (2013)
26. MacInnis, D.J., Moorman, C. and Jaworski, B.J.: Enhancing Consumers' Motivation, Ability, and Opportunity to Process Brand Information from Ads: Conceptual Framework and Managerial Implications. *Journal of Marketing* 55(1991), 32-53, (1991)
27. McKinsey: *Unlocking energy efficiency in the U.S. Economy*. McKinsey Global Energy and Materials, July (2009)
28. McKinsey: *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, May (2011)
29. McKinsey: *Open data: Unlocking innovation and performance with liquid information*, McKinsey Global Institute, McKinsey Center for Government and McKinsey Business Technology Office (2013a)
30. McKinsey (2013b). *Disruptive technologies: Advances that will transform life, business, and the global economy*. McKinsey Global Institute (2013b)
31. Melville, N.: Information Systems Innovation for Environmental Sustainability, *MIS Quarterly* 34(1), 1-21 (2010)
32. Merton, T.: *Social Theory and Social Structure*. Free Press, New York (1949)
33. Moore, M.H. *Creating Public Value Strategic Management in Government*. Harvard University Press, Boston (1995).
34. Murray, R., Caulier-Grice, J. and Mulgan, G.: *The Open Book of Social Innovation* (2010) <http://www.nesta.org.uk/publications/open-book-social-innovation>.
35. Nilsen, K.: Economic theory as it applies to Public Sector Information. *Annual Review of Information Science and Technology* 44(1), 419-489 (2010)

36. Nolan, J.M., Schultz, P.W., Cialdini, R.B., Goldstein, N.J. and Griskevicius, V.: Normative social influence is underdetected. *Personality and Social Psychology Bulletin* 34, 913–923 (2008)
37. OECD: Fostering Innovation to Address Social Challenges. Workshop Proceedings (2011) <http://www.oecd.org/sti/inno/47861327.pdf>
38. Porter, M.E. and Kramer, M.R.: Creating Shared Value. *Harvard Business Review* 89(1), 62-77 (2011)
39. Reinholt, M. Pedersen, T. and Foss, N.J.: Why a Central Network Position isn't Enough: The Role of Motivation and Ability for Knowledge Sharing in Employee Networks. *Academy of Management Journal* 54(6), 1277-1297 (2011)
40. Rothchild, M.L.: Carrots, Sticks, and Promises: A Conceptual Framework for the Management of Public Health and Social Issue Behaviors. *Journal of Marketing* 63(4), 24-37 (1999)
41. Schultz, P.W., Nolan, N.J., Cialdini, R.B., Goldstein, N.J., and Griskevicius, V.: The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science* 18(5), 429-435 (2007)
42. Schumpeter, J.A.: *The Theory of Economic Development*. Harvard University Press, Boston (1934).
43. Siemsen, E., Roth, A.V., Balasubramanian, S.: How motivation, opportunity, and ability drive knowledge sharing: The constraining-factor model. *Journal of Operations Management* 26(2008), 426–445 (2008)
44. Shapiro, C. and Varian, H.R.: *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, Boston (1999)
45. Shogren, J.: *Behavioural Economics and Environmental Incentives*. OECD Environment Working Papers 49, OECD Publishing (2012)
46. Simon, H.: *A Behavioral Model of Rational Choice*. *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*, Wiley, New York (1957)
47. Stiglitz, J.E., Sen, A. and Fitoussi, J. P.: Report by the Commission on the Measurement of Economic Performance and Social Progress, OECD (2009)
48. Tammisto, Y. and Lindman, J.: Definition of open data services in software business. *Lecture Notes in Business Information Processing*, 114, 297-303 (2011)
49. Tsang, E. W. K.: Behavioral assumptions and theory development: The case of transaction cost economics. *Strategic Management Journal* 27(11), 999-1011 (2006)
50. van Osch, W. and Avital, M.: Generative Collectives. In *Proceedings of the International Conference on Information Systems (ICIS)*, Saint Louis, Missouri (2010)
51. van Veenstra, A.F. and van den Broek, T.A.: Opening Moves – Drivers, Enablers and Barriers of Open Data in a Semi-public Organization. In *proceedings of the 12th IFIP WG 8.5 International Conference, EGOV 2013*, Koblenz, Germany, 50-61 (2013)
52. Vestergaard J. and Wade, R.: Establishing a new Global Economic Council: governance reform at the G20, the IMF and the World Bank. *Global Policy* 3(3), 57-69 (2012)

53. von Krogh, G.; Haefliger, S.; Spaeth, S.; Wallin, M.W.: Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development. *MIS Quarterly* 36(2), 649-676 (2012)
54. Weber, E.U. and Johnson, E.J.: Psychology and Behavioral Economics Lessons for the Design of a Green Growth Strategy. White Paper for Green Growth Knowledge Platform, OECD, UNEP, World Bank (2012)
55. Williamson, O. E.: The Vertical Integration of Production: Market Failure Considerations. *The American Economic Review* 61(2), 112-123 (1971)
56. Williamson, O.E. Transaction Cost Economics: The Process of Theory Development. In: Smith, K. G., & Hitt, M. A. (eds.) *Great minds in management: The process of theory development*. Oxford University Press, Oxford, UK, pp. 485-508 (2005)
57. Yu, H. and Robinson, D.G.: The New Ambiguity of 'Open Government', 59 *UCLA Law Review* (2012)  
<http://ssrn.com/abstract=2012489>
58. Zuiderwijk, A., Janssen, M, Choenni, S, Meijer, R. and Alibaks, R.S.: Socio-technical Impediments of Open Data. *Electronic Journal of e-Government* 10(2), 156 – 172 (2012)
59. Zuiderwijk, A. and Janssen, M.: Open Data Policies, Their Implementation and Impact: A Framework for Comparison. *Government Information Quarterly* 31(1), 17–29 (2014)
60. Ölander, F. and Thøgersen, J.: Understanding of consumer behaviour as a prerequisite for environmental protection. *Journal of Consumer Policy* 18(4), 345-385 (1995)



## **PAPER VI: Managing Complexity across Multiple Dimensions of Open Government Data**

This is a single-authored manuscript in the second R&R cycle in Government Information Quarterly. I am working on improvements which are not fully reflected in this version but will hopefully be reflected in a paper published in GIQ within the year of 2015 or early year 2016

# Managing Complexity across Multiple Dimensions of Liquid Open Data:

## The Case of the Danish Basic Data Program

### **Abstract**

Current literature on open government data has uncovered a wide range of challenges related to these important initiatives. The problems encountered include lack of data quality and interoperability, problems of program governance and motivation and heterogeneous political and ideological agendas. A common open data infrastructure might resolve some of these problems; however, implementing such an infrastructure is a highly complex task. This longitudinal case study of the Danish Basic Data Program (BDP) is intended for improving our understanding of the challenges related to open data infrastructure implementation and governance. The Basic Data Program (BDP) aims to improve the quality of selected government data, make them more coherent and improve accessibility through the implementation of a common data distributing platform. This is expected to increase government efficiency and stimulate innovation. The case study describes the projects evolution and identifies the main phases of implementation, use and impacts of open data infrastructure, which are summarized in a process model. Moreover, analysis of data through three different theoretical lenses uncovered four different tensions which are identified as key challenges of an open data infrastructure implementation. These tensions are presented with four key governance strategies that were used in the BDP case.

**Keywords:** *Open data; Data infrastructure; Liquid open data; Data quality; System-of-Systems governance; Value generating mechanisms*

## VI.1 Introduction

In August 2011, under an increasing pressure for cost reduction while simultaneously having to ensure ongoing delivery of high quality public services, the Danish Agency for Digitization published their e-Government strategy for 2011-2015. The two main goals of the strategy were: a) to ensure that citizens experienced a collaborative public sector and b) to reduce the costs of service delivery (Agency for Digitization, 2011). A new but important element in the strategy was a focus on re-use of data across levels of government. While Denmark enjoys a strong tradition in maintaining legislative public registers, there was a growing need for improvement in data quality and data management in order for authorities and businesses to harvest the full potential of these data (Agency for Digitisation, 2012). Previous analysis had uncovered problems with gaps and redundancy in the data sets used across public administration, as well as a lack of clarity on who could use the data and for what purposes. Public authorities as well as private businesses had to pay for access to some of the data and in many cases suffered from cumbersome payment- and access rights procedures. These issues were causing what has been perceived as sub-optimal use of the data resource (see for instance Nielsen, 2010 and Pollock, 2008).

The value of accessible, interoperable and contextually appropriate data is becoming apparent in many areas of our societies. It is estimated that better use of data resources in the US health care could generate USD 300 billion annually and in Europe's public sector it could generate EUR 250 billion annually (McKinsey, 2011). Additionally, the growing technical and semantic ability to access and merge various disparate types of data and the use of sophisticated data analysis techniques and technologies, has the potential to vastly increase the overall value of data. The key driver behind the data value proposition, is that data need to be open and *liquid*, in the sense that they are coherent, shareable and published machine-readable formats (McKinsey, 2013). This liquidity ensures that data from different sources can be combined and data can flow readily across boundaries. If data become open and liquid, the overall economic value potential over seven key sectors could be more than USD 3 trillion annually (McKinsey, 2013). However, an increasing body of research is providing us with evidence that while the potentiality of data-driven value generation is vast, the actual execution of harnessing the value of data is still very difficult for most organizations (Conradie & Choenni, 2014; Janssen et al., 2012; Martin et al., 2014; Zuiderwijk & Janssen 2014a). A key element is missing in many current open data initiatives, namely a coherent data infrastructure.

This paper addresses the question of *How can the tensions in a multi-stakeholder open data infrastructure program be addressed through governance strategies?* I approach this question by examining the Danish Basic Data Program (BDP) that was initiated in 2012. The original aim of the program was to increase public sector efficiency; however the program soon took on a leading role in shaping the open data infrastructure in Denmark. The paper is organized as follows: First the research is motivated by discussing research findings and other evidence that show the need for better understanding of open data infrastructure implementations. Next, three different theoretical frameworks which are used to analyze the implementation of open data infrastructure in Denmark are reviewed. Thereafter the research approach is discussed, followed by a description of the case and its context. The main phases of the program are then summarized in a process model, which is the main theoretical contribution of the paper. Finally, I discuss the results from the data analysis, introducing four main governance tensions revealed by the case study data, and the four key implementation strategies that were used to resolve those tensions in the BDP case. The paper concludes by discussing the main limitations of the research approach as well as the main contributions to knowledge and practice.

## **VI.2 Motivation and Frameworks for Analysis**

The original aim of the Danish Basic Data Program (BDP) was to implement a common data layer for certain key sets of public data for improving efficiency within the public sector (see further discussion in chapter 4). The idea was that all users from public and private sectors alike would have access to the same data, effectively eliminating double entry of data as well as various shadow registers, and consequently raising the quality of the data and stimulating use. However, as the program progressed, it evolved into an *open data infrastructure* program. Infrastructure in general can be defined as basic physical and organizational structures needed for the operation of a society or enterprise, or the services and facilities necessary for an economy to function (Sullivan and Sheffrin, 2003). Open data infrastructure can be defined as *the digital infrastructure needed to enable effective data sharing and use across boundaries*. Implementing a data infrastructure is a complex initiative which amongst other requires a change in a number of socially constructed organizational and societal structures, which are often quite resilient to change. This important but non-technical side of digital infrastructure has recently been conceptualized as *soft infrastructure* (King and Uhler, 2014).

As open data policies are a recent phenomenon, applicable systematic research that identifies different policies and relates those differences to policy impacts has not emerged yet (Huijboom and van den Broek, 2011; Zuiderwijk and Janssen, 2014b). Moreover, insufficient research exists to analyze the individual steps that are required to publish open data *after* development of policy. The relationship between the context of open data initiatives (technical, social, organizational and political), various elements of open data and data platforms, and the impacts of such initiatives needs to be clarified (Zuiderwijk and Janssen, 2014b). Information science literature offers insights into specific areas such as data quality and use of metadata and computer science literature offers a sizable corpus of research on linked data principles, models and platforms. Nonetheless, these papers do not present a holistic review of the implementation and governance challenges faced by organizations which aim to publish open data. Open data infrastructures as a concept are not well represented in the literature as such, with the exception of spatial data infrastructures. While it has been proposed that open data infrastructure initiatives require a central governance organization that is willing to provide various support activities (Hofman and Rajagopal, 2014), this proposition needs more evidence for support.

This paper is intended as an input into this emerging body of research. We suggest how some of the governance tensions that arise when implementing a multi-stakeholder open data infrastructure can be alleviated via certain implementation strategies. Moreover, we propose that an important aspect of implementing an open data infrastructure, which is effectively used by multiple users for multiple purposes, is the transformation, not only from *closed to open data*, but from silo'ed to liquid data as well. A more detailed discussion of what this entails is offered in section 2.1.

## Dimensions and attributes of liquid open data

As mentioned earlier, an important aspect of implementing an open data infrastructure is the transformation from closed, silo'ed datasources to openly available data. Providing open data (supplier side view) has been proposed as being a matter of availability, accessibility, format and license (Davies, 2010). From the demand point of view, openness is proposed to combine unrestricted availability of data with accessibility and technical interoperability (Tammisto and Lindman, 2011). In practice oriented literature, the term open data is interpreted in a variety of fashions, as evidenced from the many different working definitions found online. The Open Knowledge Foundation defines open data as “data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and

sharealike” (OKF, 2015). However, this definition lacks references to the technical dimensions of open data. Alternatively, Berners Lee’s five stars of linked data specify a number of technical dimensions.<sup>8</sup> However, the five stars are not really an open data definition, but rather a maturity model that focuses on how to gradually transform data into linked data. Linked data is a method of publishing structured data so they can be interlinked and discovered through semantic queries.

An overview over multiple working definitions of open data is presented in Verhulst et al. (2014). This overview shows that the currently used definitions usually highlight 2-4 dimensions that are important for openness of data. It also shows that no two definitions are based on exactly the same dimensions. The Open Data Barometer defines “truly open” data as data that are available online, in bulk, and under an explicit open license (Davies, 2013). However, in spite of the fact that considerable efforts have been exercised to make diverse government data available to the public, less than one in ten public datasets reviewed in seventy-seven countries in 2013 could be classified as truly open according to the Open Data Barometer definition (Davies, 2013; Höchtl et al., 2014). After reviewing a number of recent publications that have analyzed which factors can influence use and value generation from data, I classified the following factors as the key barriers to use of data by multiple stakeholders within and outside of government (Conradie and Choenni, 2014; de Vries et al., 2011; Halonen, 2012; Janssen et al., 2012; Martin et al., 2014; Meyer-Schönberger and Zappia, 2011):

- 1) Availability issues: Data only available to specific groups or not available at all.
- 2) Economic issues: Too high prices limit use.
- 3) Legal issues: Not standardized licenses, unavailable licenses, laws that limit dissemination or potential use of specific sets of data.
- 4) Usability issues: Data are not published under open standards or in machine-readable formats, data quality lacking).
- 5) Discoverability issues: Links to data buried in website hierarchies, no use of metadata to facilitate search, data not linked to other data, no central repository or portal.
- 6) Accessibility issues: Lack of download possibilities, no bulk download nor web services or open API’s, not secure access, not standardized or sustainable access opportunities).

---

<sup>8</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

- 7) Interoperability issues: Disparate data originating from silo'ed databases, data not published with compatible identifiers, no data models that explain syntax and semantics available.

After reviewing various different definitions of open data and classifying common barriers, I created construct I call *Liquid open data*, consisting of seven dimensions that are important for creating an infrastructure that enables openness of data for sharing and use across boundaries. In a recent report, McKinsey (2013) interchangeably used the terms “open data” and “liquid data”, referring to data that are open, widely available, and in shareable formats. The extra word *liquid* is intended to indicate the importance of the dimensions that create the foundation for a *technical open data infrastructure*, intended to make re-use of data more efficient. Truly liquid open data are defined as *data that are available online, free-of-charge and under an open access license, published in machine-readable formats, easily discoverable, accessible and conceptually coherent*. Liquid open data can be re-used without discrimination or limitation, linked to other data and streamed across systems.

**Table 1: The Seven Dimensions of Liquid Open Data**

<b>Dimension</b>	<b>Description</b>
<i>Strategic: Availability</i>	Availability reflects the strategic importance of open data, ranging from all government data that are not subject to privacy or national security limitations being open to all by default, to government data in general not being available outside of organizational boundaries.
<i>Economic: Affordability</i>	Affordability is an economic dimension and refers to the pricing of data, ranging from data being completely free-of-charge to data being extremely expensive.
<i>Legal: Reusability</i>	Reusability is a legal dimension and depends on the type of license used for government data intended for reuse. The license can range from a type of creative commons licenses that allows anyone to use the data for whatever purposes they like, to very strict licenses that allow use for a single purpose only.
<i>Technical: Usability</i>	Usability is a technical dimension and refers to the clarity and ease with which we can interact with the data. If data are not usable, it is difficult to use them for purposes other than originally intended. Data need to be of <i>high quality</i> (fx accurate, complete and timely) and presented using standardized machine-readable data formats.

<b>Dimension</b>	<b>Description</b>
<i>Technical: Discoverability</i>	Discoverability is also a technical dimension refers to whether potential users can easily discover the data and find information about the data. Highlights use of metadata.
<i>Technical: Accessibility</i>	Accessibility refers to whether data are easily, consistently and securely accessible and downloadable or streamable. Highlights use of open standards.
<i>Technical/ conceptual: Interoperability</i>	Interoperability of data refers to data being conceptually open. Interoperable data are published in a way that enables the data to be used outside of the context within they were collected. Interoperability of data depends both on structure (syntax) and meaning (semantics) and calls for use of unique identifiers for linkability. If data are interoperable, they are also “liquid” in the sense that they can stream across systems and easily be linked to other data.

## Managing complexity – the System of Systems approach

Decision makers within government and industry are currently facing societal challenges of increasing reach and complexity, social, environmental and economic in nature. They range from economic inequality, unemployment and poor social conditions to chronic diseases and climate change. Societal problems like these do not pay attention to human-constructed boundaries and commonly transcend the boundaries between the public and the private sectors (Janssen and Estevez, 2013). Accordingly, challenges of this nature can only be effectively addressed by the coordinated actions of many different public and private agencies, operating at several different levels (Kettl et al. 2004).

However, with multiple, heterogeneous, distributed systems involved, effective analysis for decision-support quickly becomes unmanageable within the silo-context that still characterizes many organizations (DeLaurentis and Callaway, 2004). What these types of societal challenges demand is some way of systematic thinking at the basic level that creates the necessary infrastructure that allows data and information to flow across organizational and technical systems. Systems-of-Systems (SoS) can be defined as a collaborative set of systems where the components are independent dedicated systems that are separately acquired and integrated to form a single system, yet maintain a continuous operational existence independent of the collaborative



system (Rechtin and Maier, 2000). Adopting a SoS approach allows for connecting systems from multiple domains, that cross organizational boundaries, and operate at different levels (Curry, 2012). Table 2 illustrates five characteristics that can be used to identify a SoS (Boardman and Sauser, 2008):

**Table 2: The Seven Dimensions of Liquid Open Data**

<b>Characteristic</b>	<b>Description</b>
<i>Autonomy</i>	Describes the ability of individual systems to make independent choices while adhering to the common overall goal of the system
<i>Belonging</i>	Reflected in a shared mission where systems are voluntary partners in the overall system.
<i>Connectivity</i>	Describes the ability to link with the other systems and thus the need for interoperability between systems.
<i>Diversity</i>	Describes the heterogeneous roles of individual systems.
<i>Emergence</i>	Highlights the evolutionary development of a SoS.

The SoS approach strives to find a balance between allowing the necessary autonomy of specialized units so that they can fulfill their legal and strategic purposes, and integration on various levels that is necessary for the holistic approach. While considered suitable for governing complex systems, the SoS approach poses significant challenges, including the need for data and information interoperability, and collaboration and coordination between independent organizational units. In order to enable organizations and their systems to interoperate in smart government environments, we need to consider three levels of interoperability: technical, conceptual and organizational (Hjort-Madsen, 2006; Maheshwari and Janssen, 2014). Table 3 shows these three levels of interoperability.

**Table 3: Three levels of interoperability**

<b>Level</b>	<b>Three levels of interoperability (Gottschalk et al., 2009)</b>	<b>Three levels of interoperability barriers (Chen et al., 2008)</b>
<i>Technical</i>	Interoperability of systems which is concerned with the ability of two or more systems or components to exchange data and use the exchanged data	Technological barriers refer to the incompatibility of information technologies such as protocols, encoding, platforms, or infrastructures.

<b>Level</b>	<b>Three levels of interoperability (Gottschalk et al., 2009)</b>	<b>Three levels of interoperability barriers (Chen et al., 2008)</b>
<i>Conceptual</i>	Semantic interoperability which has to do with use of common standards	Conceptual barriers represent the syntactic (format) and semantic differences (interpretation of meaning) of exchanged information.
<i>Organizational</i>	Organizational interoperability is concerned with the ability of two or more units to exchange data and services to operate more effectively together.	Organizational barriers concern the organizational incompatibility of definitions of responsibility, authority, and organizational structures.

## The value generating mechanisms of open government data

Most of the open data initiatives today are driven by central and local governments around the world. One of the reasons that innovators value government data is their quality, reliability, standardization and comprehensiveness (Lakomaa and Kallberg, 2013; Jetzek et al., 2014a). In well-resourced states, few other institutions can provide such consistent data covering the whole country. This makes open government data a valuable input to economic activity (Davies, 2013). The most important driver behind open government data initiatives seems to be the expectation that open data will stimulate the generation of considerable social, economic and environment value for their societies (Janssen et al. 2012; van Veenstra & van den Broek, 2013; Zuiderwijk et al., 2012). In earlier work (Jetzek et al. 2012, 2013a; 2013b; 2014b) we have identified four distinct generative mechanisms that can explain how value is generated through use of open government data. The framework illustrated in Figure 1 classifies these mechanisms based on their ideological value generation focus as one dimension and the nature of change that happens as the other dimension.

Open data can create value through better governance through information sharing, leading to more efficient and socially responsible exploitation of current resources. Open data can also create value by enabling exploration, finding new ways of linking data across boundaries, resulting in behavioral change and new products and services. In the framework we have collapsed citizen participation and collaboration into one mechanism we call civic engagement. This mechanism describes the value that is generated when multiple stakeholders across boundaries collectively contribute to a

certain cause, either by making use of data and information or by contributing data or information. Regarding the other mechanisms we refer to Jetzek et al. (2013a, 2014b). In the context of this paper we use this framework purely to highlight the archetypical paths through which open data can contribute to value generation, and how they relate to different political agendas, driven by different types of government organizations. Moreover, I propose that each of these mechanisms requires a unique approach to the seven dimensions of liquid open data.



Figure 1: Four archetypical value generating mechanisms

### VI.3 Research Setting and Approach

This research has grown out of the author’s involvement with a Danish open data infrastructure initiative, the Danish Basic Data Program (BDP). This case was conducted in the spirit of Engaged Scholarship (van de Ven, 2007). The author was not directly involved in the program but had nevertheless access to non-classified internal documents and project events. The purpose of the research was to describe and explain and as the author was in a way “attached inside” the resulting approach can be categorized as “co-producing knowledge with collaborators.” (van de Ven, 2007, p. 27). The case study is suggested as a preferred method when (a) how or why questions are asked, (b) the investigator has little control over events, and (c) the focus is on a contemporary phenomenon within a real-life context (Yin, 2009). This approach fit

well for the BDP as the author was recording and analyzing events within an unfolding reality. Yin (2009) recommends that case study inquiry relies on multiple sources of evidence (e.g., interview, observations, documents and archival records) which should converge in a triangulating fashion. Strong triangulation of data sources is important to establish the necessary reliability and validity of a research study (Yin, 2009). The case study also benefits from the prior development of theoretical propositions to guide data collection and analysis (Yin, 2009).

The overarching research method is a longitudinal descriptive case study method, spanning a period of three and a half years. A descriptive case study is used to describe a phenomenon and the real-life context in which it occurred (Yin, 2009). I used a process view for the analysis, as the aim is to develop a process model. Accordingly, the case is structured chronologically, where events are traced over time and ordered in periods or phases (van der Blonk, 2003). The main research question is *How can the tensions in a multi-stakeholder open data infrastructure programs be addressed through governance strategies?* The next sections describe how data collection and analysis were approached.

## Data collection

The empirical data was collected between November 2011 and August 2015 through my engagement with various stakeholders directly or indirectly involved in the Danish BDP. Primary data collection was based on interviews, participation in various project-related events such as kick-off meetings for different project phases and conferences, as well as observation in a private sector company that participated in the tender for data distributing platform and was later responsible for the implementation of the technical infrastructure.

A total of 28 interviews with 22 interviewees were conducted, ranging from just under half an hour to one and a half hours. The main strategy for choosing interviewees was to identify stakeholders that were directly involved in the program and also perceived as thought leaders by other participants, as those stakeholders were in a position to influence the direction of the program. The interviews were mainly conducted in the respondents' offices and in all circumstances face-to-face. Interviews were mostly conducted in the native language, i.e. Danish, but in English for the few non-Danish speaking respondents. Certain key respondents were interviewed up to four times to document the progress of the BDP during the case study period. The key informant was the program leader which was extensively involved in dialogue with the author about the program and the main challenges and opportunities.

The interviews were all semi-structured to allow new patterns that were not identified in previous research to emerge. Key-questions were prepared beforehand; in most cases regarding the interviewee's perception of the BDP, what they saw as the main implementation barriers and how they perceived the program's primary value proposition. In some cases the interview questions focused on the "historical" aspects, to improve understanding of the social, economic and political context in which this program operated. I took extensive observational notes, both during interviews, while listening to interviews and while observing events, which were used in the data analysis. Overview over dates, number of interviewees in each interview, number of interviews with each interviewee, roles and organizations is given in Appendix A.

Additionally, I examined selected project documents. I also analyzed publications and presentations from the organizations that participated in the BDP for insights on how they perceived the program. Other material that was used to inform the analysis are policy documents, tender documents and meeting notes, besides scientific research papers on open data, interoperability and management of complex systems.

## Data analysis

All but two interviews were recorded with participants' permission. The interviews provided a deeper understanding of important events, program phases, decision points, and tensions related to the implementation process. I focused on documenting the experiences and challenges related to managing an open data infrastructure project with a special emphasis on understanding how project participants thought these data would contribute value to society and how they perceived the program itself, the main barriers and enablers and the most dominant tensions, as well as the program governance.

Time boundary validity is a form of content validity indicating the degree to which the domain is accurately reflected in the measure. If the bounded timeframe is not adequately captured, content validity is reduced because the researcher is not fully representing the domain of data in the data collection (Street and Ward, 2012). A period of three and a half years was deemed to be sufficiently long to report on a phenomenon as complex as the implementation of infrastructure, as well as sufficiently short to allow for the precision needed to explain individual events.

Figure 2 illustrates the case timeline and time boundaries.

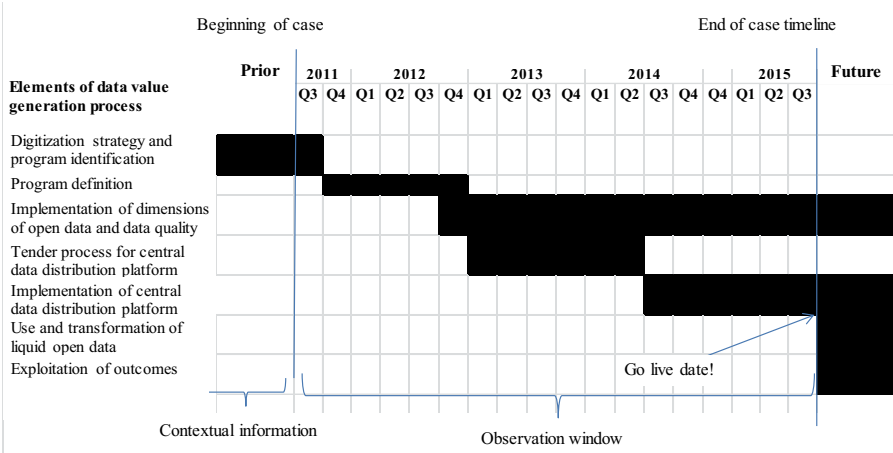


Figure 2: Case timeline

The data collection started when the Digitization strategy 2011-2015 was published in Q3 2011 and ended in Q3 2015. While the first formal interviews were conducted in August 2012, I had the opportunity to participate in program related events and receive program related documents from November 2011. I later used interviews and documents to gain insight into the historical context from which this program arises, as the context is important for explaining many of the events. The analytic approach was abductive. Accordingly, the theoretical frameworks used for analysis were developed alongside increasingly accurate documentation of the case. These frameworks or theoretical lenses (reviewed in the theory sections 2.1-2.3) were chosen in light of recurring themes in the data. After the preliminary, ongoing analysis of the data during the case timeline, I proceeded in three stages.

I identified the three biggest themes from the data. The first important theme I noticed was the multidimensionality and complexity of the open data construct itself, which led me to examining the individual dimensions of liquid open data. Without this being explicitly expressed I perceived a drive to move from organizational silos towards modular information architecture, highlighting the need for more agility and responsiveness in government processes. While not only an open data issue, one of the key ingredients in such architecture are data that are not bound to individual systems but can flow easily across systems to inform and automate as needed. Therefore I have highlighted the importance of the technical and conceptual dimensions of open data (liquidity of data).

The second theme I noticed was that this particular open data initiative did not only revolve around opening data for whatever purposes, but rather had a strong focus on two disparate mechanisms that generate value for society. One of the mechanisms is the efficiency mechanism that generates value through better use of data across levels of government. The participants that focused on efficiency as the primary goal emphasized the technical and conceptual dimensions of liquid open data. The other mechanism is the innovation mechanism that generates value through use of data as a resource for new information, products and services. The participants that had stronger focus on innovation as the primary goal made a stronger case for the strategic, economic and legal dimensions of liquid open data. The fact that different participants in the program had a different view of the importance of each of the mechanisms was the root of some tension in the program.

The third important emerging theme revolved around the governance of the program. The Danish BDP has gained an overarching commitment from seven different ministries in the central government as well as from powerful interest groups at both the regional and municipal level. The overarching goal of the program is that certain key sets of data should be used consistently across levels of government and in the private sector. Introducing interoperability between heterogeneous sets of data, owned and used by different organizations under different ministries or in different levels of government, while at the same time safeguarding the data quality and the professionalism that has identified each of the individual registers, requires a virtual line dancing between autonomy and control. Moreover, implementing an open data infrastructure requires engagement with data users, which come from both the private and public sectors. I wanted to represent the governance approach that emerged as the program leaders strived to find a suitable approach for such a complex initiative.

After identifying these three themes, I searched for frameworks or taxonomies to use for data analysis. As open data is an emerging field of research offering few conceptual frameworks suitable for data analysis, two of the frameworks chosen are based on my own work within a larger research project while the third was identified in related literature. The first framework is illustrated in figure 1. This framework served as a focusing lens in many of the interviews, and later helped reveal some important tensions in the program. The second framework is the taxonomy of seven dimensions of open data, as shown in table 1. These dimensions were applied rather later in the analytic process but are important to the case as different project groups are focusing on different dimensions of liquid open data which complicates the program governance

even further. The third framework I applied to data analysis are the characteristics of a System-of-Systems.

I analyzed the dynamics of the open data infrastructure implementation to develop the process model introduced in chapter 5. Based on all available documentary and interview material I mapped out events and key decisions. I proceeded to identify key issues or tensions and finally proceeded to construct the case narrative (Langley, 1999). The focus was on identifying the items in the previously identified taxonomies and understanding how these different themes were connected in a holistic process model.

## **VI.4 Case Context and Description**

This chapter is intended to give an indication of the economic, social and political context of the BDP and the progression of events over time. Currently, Denmark is the largest spender on social protection of the EU countries, which is almost a third (25,2%) of yearly GDP (Eurostat, 2013). Thereof care for the elderly amounts to 14,3% of GDP and the population continues to age - 23,6% of the population is was over the age of 60 in 2012 (Statistics Denmark, 2013). Administration costs in the social welfare sector in Denmark are comparatively high, and general government sector employment (excluding public corporations) as a percentage of the labor force, 26,8% in 2012, is one of the highest rates of all OECD countries (OECD, 2012). Similarly, compensation of public employees as a percentage of GDP, 19% in 2010, is the highest rate of all OECD countries (ibid). Some worry that without major changes, public finances may reach an almost unsustainable level by 2020. These numbers explain the enormous focus the Danish Ministry of Finance has on improving efficiency in the public sector. In order to protect the high level of social security, the government needed to find ways to do more with less. And one of the chosen paths towards that goal is through improved use of data.

The next few sections describe how the BDP evolved over the almost four years of observation. A process model of an open data infrastructure program is then presented in chapter 5.

### **Program identification: From strategy to action**

There had been a couple of important prior milestones on the road to more efficient reuse of basic data, before the BDP was initiated. The small Ministry of Housing, Urban and Rural Affairs had already from January 1st 2003 started to freely



disseminate data from the dwelling register, the municipal property register and the cadaster, as well as address data, demanding only that users paid a distribution fee to private sector data resellers (DECA, 2010). *“It did not take many years after starting to work with the public data in around 1993 to realize that these data could be used for many interesting things. And we felt it was a shame, you know, that the data weren’t used more”* (Interview, project manager in the Ministry of Housing, Urban and Rural Affairs, January 2014). A follow up study on the costs and benefits of freely disseminating address data was conducted in 2010. This study confirmed the success of the project, estimating the benefits over a 6 year period as EUR 57 million while the total costs over the same period were calculated as around EUR 2 million (DECA, 2010). However, the 2010 report noted that there were still problems of *“...inconsistent address information with resulting duplicate work, errors and uncertainty in other state and municipal administrative back-end systems.”* (DECA, 2010, p. 8).

Another prior milestone was the 2010 Open Data Innovation Strategy initiative (ODIS). The main goal of ODIS was to raise awareness of the potentials of public sector data, reflecting the renewed global interest in open data following Obama’s 2009 open data directive. However, while the US initiative was largely focused on transparency, participation and collaboration, these were not high on the agenda in the Denmark, where the authorities struggled with increasing demands for public services during the low-income period following the great recession. *“This transparency thing. We believed in it but it was not going to open any doors in Denmark. Because in Denmark we didn’t really have the feeling that our government wasn’t transparent enough. We never really seemed to regard that as a serious issue for Denmark...”* (Interview, project manager at Agency for Digitization, August 2012).

The Agency for Digitization had classified the information that public authorities register about citizens, companies, property, buildings, addresses etc. as *Basic Data*; core societal level reference or master data, used widely for different purposes by public and private sectors alike. Basic data were defined as including, but not being limited to: data from the person register, business register and real property register, as well as address data, place names and maps and geographic data. The common element of these basic data is that almost all other domain specific data within government have a reference to some of these data (events happen in a place, and a person or an organization is usually involved), The main idea was therefore to make sure that they were used in a more coherent manner.

While Danish society has already gained a lot of benefits from the basic data registers, the Danish authorities recognized that there were opportunities for further gains. This was reflected in the Danish Digitization Strategy for 2011-2015, which is created for the whole of public sector, including both local and central government. *“We need to be able to bring basic data into play with greater stability and capacity than we can at the moment. This means finding an efficient way to distribute data across the public and private sectors using a shared infrastructure.”* (Agency for Digitization, 2011, p. 38). The BDP was initiated to implement the data related objectives of the strategy. A total of seven ministries presented by their agencies and registers, the Danish Regions, an interest organization for the five regions in Denmark and Local Government Denmark (KL, the interest group and member authority of Danish municipalities) were all stakeholders in the BDP.

The BDP’s main goal was to create an infrastructure that would enable more efficient use of basic data across administrations and sectors (Horst et al., 2014). Three main challenges were identified: (1) while the data from the basic data registers were already used across levels of governments, they were used in different ways and in some cases the quality of the data was not sufficient for the intended use. Therefore, a number of copies and shadow registers had emerged through the years. *“The needs for property information are very different between user groups and that is reflected in how the data is presented and preserved in each register”* (Interview, product owner, KMD, August 2012). (2) As the registers had all been founded separately during different time periods to fulfill certain legal requirements for data registration, the registers were based on different (sometimes outdated) technical platforms and it was very difficult to combine data from different registers. (3) There were a number of different payment models and types of licenses that complicated the use of basic data for users in both public and private sectors.

### Program definition: Towards the goose with the golden eggs

According to the original project plan, the BDP definition phase was supposed to take place in the period October 2011 to May 2012. In January 2012 the program leaders hosted a big event, the Basic Data Camp, where in a very Danish tradition, multiple stakeholders from various public organizations came together to collectively discuss and define the program. During this meeting, the tensions between the data-driven efficiency agenda on the one hand and innovation through open data agenda on the other hand were quite apparent. The proponents of the innovation agenda wanted to set data free for all users, while the Ministry of Finance had no intention of doing so, as

this would most likely require them to replace the income lost to the registers with money from common funds. “*Within the program there was a pressure building, mainly from the Danish Geodata Agency and the Company Register, that we should also set the data free.*” (Interview, program leader Agency for Digitization, November 2013).

In February 2012, a delegation of thirty stakeholders in the BDP participated in the Danish–Dutch Key Registers Tour 2012. The objective of the visit was to discuss the main building blocks required to establish a so-called system of key registers. The Dutch strongly advised their Danish friends not to procrastinate the fundamental issues of financing and governance of the program (de Vries & Pijpker, 2013). The Danish program leaders took this challenge head on and postponed the definition phase for four months while the program group finalized their work on planning the program. To the surprise of many, the group announced in October 2012 that all data from the included key registers would be freely available to anyone to use for any purpose, excluding the data in the person register (Digitaliseringsstyrelsen, 2012).<sup>9</sup> The focus had suddenly shifted heavily towards open data, promoting the ideological agenda of innovation and growth.

The brochure Good Basic Data for Everyone – A Driver for Growth and Efficiency was published in October 2012 and marked the end of the program definition phase. The publication stated the following five principles (references to the dimensions of liquid open data are in italics in the square brackets):

- 1) In order to ensure re-use of data and to prevent double registration and shadow registers, map data, cadastral maps, central business register data, and company data will be financed by the government and released to the public and the private sectors, as is already the case with address and real property data [*Availability*]. By releasing this basic data, public authorities and private businesses alike will be able to use it freely [*Affordability*], for commercial as well as for non-commercial purposes [*Re-usability*], provided, of course, such use is lawful.
- 2) In order to enhance the quality and coherence of data [*Usability and Interoperability*] the registers of map data, real property data, address data, as well as business registers, will be expanded to include other necessary data. As a result, a number of existing registers will become redundant and therefore can be phased out.

---

<sup>9</sup> The financing of the free data is discussed in section 4.3

- 3) In order to make it possible to link data, efforts will be made to ensure that all data conforms to the same technical requirements [*Interoperability*].
- 4) In order to improve the distribution of common public-sector data, a common technical infrastructure is to be established providing for stable and efficient distribution of data; a data distributor [*Discoverability and Accessibility*].
- 5) In order to ensure efficient, effective and coordinated development and use of basic data, a cross-institutional basic-data committee is to be established [*SoS governance*].

The Danes had in a very short time frog-leaped their colleagues in the Netherlands, creating a program that would address the required elements for an open data infrastructure, simultaneously ensuring efficient internal use of basic data and opening these data up for external use.

#### Program funding: Free data does not come for free

One classic problem of making government data open is the fact that the beneficiaries of open data are seldom the same as those who render the costs of collecting the data (Fioretti, 2011; Martin et al. 2014). Many of the organizations that collected basic data in Denmark were not only accustomed to sell the data to fund their data collection efforts (or a part of them), but required by law to do so. As previously stated, the Ministry of Finance had no plans to “spend money” on subsidies to these organizations, and the financing of free basic data remained an issue in the early phases of the program definition. Moreover, the intended data quality improvements and the common infrastructure were very costly. “*Money needs to be redistributed. It is expensive, and difficult, to make these changes.*” (Interview, subprogram manager, Agency for Digitization, November, 2014).

Denmark had for many years followed the New Public Management governance policies that aim to modernize and render the public sector more efficient (Janssen & Estevez, 2013). One aspect of this type of governance is that public sector organizations should operate more like private sector organizations, which created a pressure to sell data to generate revenue if at all possible. “*The payment models were in a way contradicting the intentions of more reuse of data*” (Interview, project manager in the Ministry of Housing, Urban and Rural Affairs, January 2014). In this environment it was virtually impossible for individual organizations to open their own datasets without some kind of support from the Ministry of Finance. However, some of the BDP participants were consistently pushing the agenda of open and free public

data. *“They continued to push the agenda and then we turned it back upon them and said: “OK, tell us the costs. We don’t want to join this from an ideological perspective; we want to know the costs and you have to document the benefits”* (Interview, program leader, Agency for Digitization, November 2013).

In order to make open and free data a reality, the groups responsible for each sub-program were required to write a business case where they identified which public organizations would benefit from better data and more efficient use of data. Afterwards, the program leaders effectively transferred money from these organizations to the organizations that were responsible for collecting and delivering data. This involved a great deal of restructuring and moving responsibilities and authority between organizations across ministries and levels of government. As an indicator of the complexity of the task, the BDP group needed to change around 10 different laws in order to make these changes. What the BDP had managed to do had been tried before but without success. However, the fact that the Ministry of Finance now took the leadership of this project was one of the key success factors according to our interviewees. Moreover, the interviewees in general highlighted the important role of the program leader in making these changes possible. He had exhibited both political acumen and leadership in finding a way to make open data acceptable to the Ministry of Finance.

The business case for the whole program was calculated upfront and all monetary transfers were accepted by all main stakeholders before the project was started. The total investment was budgeted as EUR 125 million (Krieger-Røyen, 2014). Around 20% of this amount will be spent on the implementation of the data distributing platform (technical infrastructure). Around 40% is for reimbursing the participating registers for their lost income.<sup>10</sup> Around 20% of the budget is earmarked for improving data quality and around 20% is intended to cover other costs of the project. On the benefits side, the simplification and modernization of the IT architecture and the introduction of a coherent and interoperable data model is expected to reduce the need for double entry and shadow registers and reduce administration costs. These benefits are estimated to cover as around 10% of the implementation costs. Better data management processes are expected to lead to reduced costs which should cover about 10% of the costs. Finally it is estimated that efficiency gains from better and simpler work processes in the public sector will cover around 80% of the costs. After the year 2016 when the program will be fully implemented the net annual benefits to the public

---

<sup>10</sup> <http://www.slideshare.net/EUDataForum/edf2014-nicolas-lemcke-horst-ambassador-of-the-danish-basic-data-programme-agency-for-digitisation-ministry-of-finance-of-denmark-danish-basic-data>

sector are estimated as EUR 35 million while the annual benefits to the private sector are estimated as approximately EUR 65 million (Agency for Digitization, 2012).

After having crossed this important barrier the moral in the program was good and everyone was excited that open and free basic data for use across public and private sectors would finally become a reality. “*The fact that we managed to make data free really brought trust to the leadership, showing that we were able to move large sums of money around and really restructure how data collection and dissemination was financed.*” (Interview, program leader, Agency for Digitization, March 2014).

Managing complex systems – the System-of-Systems approach

The BDP implementation was officially launched in October 2012. The program consisted of seven sub-programs, and by the end of 2013, thirty-eight individual projects were being managed by the sub-programs. “*That is huge and that is vastly complicated. And of course things go wrong. It is kind of like constant firefighting.*” (Interview, program leader Agency for Digitization, November 2013). Each of the sub-programs had their own steering committee, but for daily operations and coordination, a cross-program project management organization was established, the Basic Data Secretariat. The secretariat did not have any decision power but acted only as a supporting organization. Finally, there was a board of directors with high-level officials from the biggest stakeholders (agencies, ministries and interest organization). Six of the identified subprograms had the responsibility to re-organize a special subset of the overall data structure, while the seventh subprogram was responsible for designing the architecture of a new common technical platform for data distribution and create tender documents for this platform. During the first months of the program the groups kept working more or less independently.

Moral remained good and participants still remained true to the original program vision [SoS characteristic *Belonging*] and convinced that the program had sound foundations and a potential to deliver much value to society. As time passed, however, certain governance challenges started to appear. While Denmark has a long history of collaborative government, the public sector is still fairly decentralized and a good deal of autonomy [SoS characteristic *Autonomy*] has been given to a range of public authorities (Kettl et al. 2004). This type of structure calls for a multi-level, multi-authority co-ordination and networking [SoS characteristic *Connectivity*] which demands particular skills from leaders (ibid). In the case of the BDP, individual subprograms and projects under the overall program management depended on each other; if one subprogram did not deliver on time, another subprogram could endure delays. As only a coordinating unit, the basic data secretariat lacked the authority to

really handle the issues that came up *across* subprograms. The result was that technical issues that crossed boundaries between individual subprograms had to be moved up to the board of director's level.

In early 2013, delays in the seventh subprogram that was responsible for developing the overarching architecture for the data distribution platform, write the tender documents, negotiate with vendors and oversee the implementation of the technical platform, had started to cause inconveniences in the other subprograms. This was especially true for two of the subprograms that had already started to depend on the technical infrastructure. In the summer of 2013 it became evident that the time plan for the platform needed to be delayed for a year.

“What has collided is that we have some subprograms that go across all the others but then we have the ministries [that are responsible for different subprograms] with their own finances and resources, responsible to their own laws, having ownership over their own registers. And when something goes a little bit wrong in the programs that go across [the other subprograms], you have to stop and look and say that while this part seems to have collapsed, the only thing we know is that we are held responsible for our own finances and resources. So this crossover of responsibilities is hard.” (Interview, project manager in the Ministry of Housing, Urban and Rural Affairs, January 2014)

An external review of the program was conducted to find out what changes in governance could improve coordination between individual subprograms. *“We have good coordination, we have good atmosphere. Of course we have some conflicts and tensions. There is still very strong sharing and belief in the common goal. But... I wouldn't really say people are hurting, but it has been a very tough year for some people. Because it is a huge project and many people have been working very hard for a very long time. And some things are difficult now. And I also think we have uncovered a new set of problems. Some capabilities that are lacking in our national [IT] infrastructure architecture and this lack is creating big problems for some of the subprograms.”* (Interview program leader, Agency for Digitization, November 2013)

The review report suggested that project management and coordination across subprograms needed to be strengthened, which created a need for other competencies and an increased number of personnel in the secretariat. *“It is not possible to reach our goals with consensus. We must strengthen governance, else we might end up with a program that doesn't deliver.”* (Interview, data ambassador, Agency for Digitization, September 2014). It was decided to change the program organization, as shown in figure 3, in order to ensure clear and uniform decision channels and well defined roles and responsibilities [SoS characteristic *Emergence*].

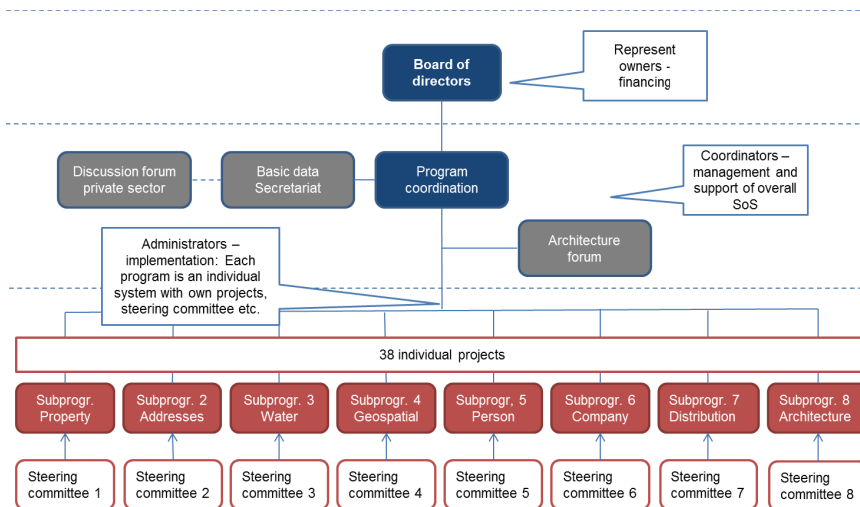


Figure 3: New organizational diagram

As each subprogram has their own budget, own steering committee and own goals [SoS Feature *Diversity*], the overall program management developed to something like a structured informal coordination, where political issues and issues that go across projects are discussed. By strengthening this coordination through different stakeholder engagement forums and through standardized procedures and deliveries, the program leaders addressed the governance related problems that had threatened to upset the strongly collaborative atmosphere of the program.

“Even though we have been doing stuff [previously] that we ought to think of as programs, we have not had the tools or the understanding of how to control a program of projects. So we are developing this understanding and these tools now and actually the BDP has been a pilot program used by our Agency to develop these tools to use across government.” (Interview, program leader Digitaliseringsstyrelsen, April 2013).

### Implementing the open data infrastructure

Prior to the BDP, users of existing basic data sources had in various contexts experienced that the data were not of sufficient quality for the purposes of the users. This perceived lack of data quality had resulted in a large number of shadow registers and double registration where individual organizations, both in the public and private sectors, enriched the data and improved their quality and coherence. A large part of the



projected efficiency gains from the BDP came from getting rid of these shadow registers, but in order to do so the quality and coherence of the data needed to be improved. Doing so represented a huge task for the BDP, including adding new data to existing registers, collapsing some registers, out-phasing others, as well as shifting responsibilities and ownership of data.

An increasing awareness of the potential of open data in society was evident over the case study period. This awareness created pressure from both public and private sectors to increase the *availability* of open data. Among other initiatives, a collective of various industry organizations created a formal document with a call for more availability of open government data.<sup>11</sup> However, due to a shortage of resources and skills, the BDP did not have the means to include more types of data in the program. Individual agencies and registers responsible for domain specific data like health data or environmental data had started to build their own data distributing platforms to cater to those pressures. There is ongoing discussion at the strategy level regarding the potential inclusion of more datasets as a part of BDP 2.0. There is also discussion about how the public sector's domain specific data can be tied into the basic data infrastructure that the BDP is implementing, so that the projected operational efficiencies will not be negated due to an emergence of multiple new, non-interoperable platforms.

In January 2013, the economic and legal dimensions of liquid open data (affordability and re-usability) were successfully implemented for geographic data from Geodata Agency and the business register from the Danish Business Authority. Making these data more *affordable* has resulted in exponential growth in data downloads. Interestingly, despite the fact that the agencies responsible for these datasets were mainly interested in the use of the data for innovation purposes, a lot of internal barriers were actually removed in the process of making data free for all, making it easier for authorities to use each other's data and thereby creating internal process efficiencies. This showcases the potential of open data to generate synergies between different value generating mechanisms (in this case innovation and efficiency). "It is crystal clear, that it is a huge feat of the Grunddata Program to get the authorities to work together. And the key to that is that there are no longer the economic barriers between us." (Interview, subprogram leader, Danish Business Authority).

The *legal dimension* was implemented for the same two sets of data, the geographic data (nautical, topographic and cadastral data) and data from the business register. This

---

<sup>11</sup>

<http://www.fsr.dk/~media/Files/Presse%20og%20nyheder/Nyheder/2015/Initiativkatalog%20til%20den%20f%C3%A6llesoffentlige%20digitaliseringsstrategi-final%20Copy.ashx>

involved changing the law to allow for uninhibited reuse of these data. A number of barriers to data reuse were overcome because of this. Previously, the municipalities could not publish their own geographic data (for instance maps) openly on their websites. In fact there was a court order prohibiting one municipality from openly disseminating their own geographic data, as by doing so the municipality was found to be in unlawful competition with commercial data providers who had purchased the geographic data and generated income by re-selling them. Due to this legal barrier, many municipalities had started to use freely available geographic data from other sources, for instance the Open Street Map (OSM). To overcome this barrier, the BDP changed the law, stating that from January 1<sup>st</sup> 2013 the municipalities could in fact freely and openly disseminate the authoritative geographic data. The three major data releases (the geographic, address and business register data) of the BDP are all using a common standardized license based on the creative Commons-licenser CC0 and CC-BY, for improved *re-usability*.

The *conceptual dimension* of liquid open data focuses on semantic and syntactic interoperability of data and is of great importance to the BDP as interoperability is a key to many of the projected efficiency and effectiveness gains. However, the need for coordination and centralized standardization efforts makes it very challenging to achieve. To address these challenges, a new subprogram was created in late 2014. Subprogram number eight is responsible for designing the overall data architecture, including common data modelling rules. To further strengthen the collaboration between the eighth subprogram and the other seven, a new architectural forum was created at the project organization and coordination level. A great deal of work has been put into creating the data modelling rules, which standardize how individual organizations model their data and thus contribute to conceptual interoperability. All the datasets included in the BDP are in the process of being modelled as a whole, complete with unique identifiers that make linking of disparate datasets possible. The data model is also made openly available which makes it easier for third party users to understand and work with the data (*usability*). Besides organizational challenges, implementation of the conceptual dimension is also demanding data-modelling skills that are incidentally in short supply in both public and private sectors.

“What we have as a goal is to create a complete data model with a complete table of contents with a list and a diagram that explains what all these data elements are, and how to they relate to each other.”...This knowledge lies on the boundary between technology and business.” (Interview, project manager, Agency for Digitization, November 2014).

The design and implementation of the technical distribution platform has perhaps been the single most challenging aspect of the BDP. The platform is a key structure in a common data infrastructure as it provides the technical foundation on which many of the technical dimensions rely. For *usability*, all data distributed via this platform will be machine readable, provided in open standard formats and of improved quality, both in terms of timeliness, completeness and accuracy. For *accessibility*, the platform will guarantee (based on the tender material) high uptime (99%) and stability of service, and open API's, web services and bulk-downloads. Finally, *discoverability* will also be improved, as the data will be available through a single point of access and making use of standardized metadata.

## User engagement, use of data and realization of benefits

The BDP offers an unprecedented opportunity for public and private organizations to use high quality, open government data across boundaries. The Danish private sector is already a big user of these authoritative data. As can be seen in a recent publication from some of the largest industry organizations in Denmark, the private sector appreciates that they have access to these data.<sup>12</sup> However, there is still work to be done: *“The private sector is already today closely linked to the public sector’s systems in many areas. Continuous access to public data is thus a critical prerequisite for a wide range of private actors. As public sector systems are developed, standards are amended and interfaces are continuously adjusted. Viewed in isolation, this occurs for good reasons, but it can be challenging [for the private sector users], as even small changes can have serious consequences, causing [private sector] systems using these data to stop.”*

A data ambassador was hired relatively early in the BDP. His responsibility was to connect with potential private sector users, keeping them informed and hearing and transmitting their point of view. The role of data ambassador seems to have been successful in creating more awareness and engagement. For instance, the open LinkedIn group he created for the program has 921 members and is relatively active. A status announcing the publication of the new data model got 54 likes, many of which came from private sector users of data which are closely following the development of the BDG. The group is used for discussing other sets of open data as well. When Copenhagen Municipality announced they were opening the city’s 3D-model data they got 312 likes, mostly from private sector users. In the governance restructuring in

<sup>12</sup>

<http://www.fsr.dk/~media/Files/Presse%20og%20nyheder/Nyheder/2015/Initiativkatalog%20til%20den%20f%C3%A6llesoffentlige%20digitaliseringsstrategi-final%20Copy.ashx>

2014, the role of data ambassador was expanded into a forum where private sector representatives from different sectors have an opportunity to meet with the program participants. The focus until now has been on large data users like the financial sector, telecommunications sector and utilities. While it is difficult to quantify the impact of this forum, this cross-sector collaboration seems to be successful and interest in and actual use of open data is increasing in Denmark.

As discussed earlier, Ministry of Finance's vision for the BDP was first and foremost to increase efficient use of a fixed set of basic (master or reference) data within the public sector, which I have referred to as the efficiency mechanism. Besides improvements in data quality, this focus has called mainly for improving interoperability of data, which was to be gained through reorganizing the data model and through improved usability and accessibility, supported by the implementation of a common data distribution platform. Perhaps as a result of this focus, the BDP's leaders managed to create a business case showing "black on white" that it is possible to realize economic gain from opening data internally by transferring money within the public sector. As discussed in section 4.3, this demanded quite a lot of reorganization, and transferring of funds across government. However, despite this success, documenting and realizing the benefits remains a challenge.

As the complexity of the chosen approach was not fully realized from the start, the program's resources came under increasing pressure as the implementation progressed. *"We are facing a shortage of key resources and competences."* (Interview, program leader, Agency for Digitization, April 2013). However, while the program leaders were aware of this shortage of skills and resources, they couldn't really acquire more resources as the realization of the benefits that were the main pillars of the business case, such as closing down shadow registers and improving process efficiency was already being threatened by delays. Their response was to keep the program lean and focused on the approved milestones; else the program would not be able to deliver the required savings. Nevertheless, as a result of the somewhat unforeseen complexity, there have still been delays and the internal savings will come later than planned. This creates a problem for the public sector beneficiaries who will have to participate in financing the program but will not be able to realize the "hard cash" benefits until later.

Alternatively to the Ministry of Finance's focus on efficiency, the Business Authority and Geodata Agency focused more on the availability, affordability and reusability dimensions, and the innovation value generating mechanism. The Geodata Agency, which is the owner of nautical, topographic and cadastral data, created a baseline measurement for use of geographic data before it was made available free-of-charge

under an open license January 1<sup>st</sup> 2013. Their ambition is to do another survey in 2016 and compare the results to the baseline from 2012. The Agency wants to measure only the effect of *free* geographic data, correcting for general market trends that have been towards increased use of geographic data for a number of years, but this is difficult to do beyond the first layer of users, as the open licenses allow all kinds of redistribution and reuse. The direct contact to the users has also decreased significantly with the advent of open data, even if the Agency requires user registration. “*Here in 2012 we had 800 users, we knew all the users.... We knew them pretty well because we were sending them a bill.*” “[*The situation before January 1<sup>st</sup> is*] *exemplified with the bonsai tree. But in 2016 we expect it to look more like a rainforest...This is going to be really difficult to measure.*” (Interview, department manager, Geodata Agency, January 2014). However, what the Geodata Agency *does* know is that the number of users of geographic data increased 10 fold in the first year after the data were set free, and the growth of users continues to increase.

Under the current regime of financial austerity in Denmark it is important to document the realized socioeconomic benefits from improving data quality and implementing liquid open data. This is however a very difficult task as there is a global lack of methodology to identify and quantify the oftentimes intangible future benefits of open data. As a part of the governance restructuring, the BDP has improved how each subprogram manages its deliveries and realized benefits. The subprograms now have a template that dictates how realized benefits, both qualitative and quantitative should be documented. This bottom-up approach makes it much easier for the program leaders to keep track of benefits that otherwise would be very hard to identify. And the standardization of how benefits are calculated implies a coordination mechanism that enables organizational interoperability and offers the potential to significantly improve how realized tangible and intangible benefits of public sector initiatives are traced across organizations.

## **VL.5 A Process Model of Open Data Supply and Value Generation**

In order to summarize the main events and constructs identified from the case, I propose the process model illustrated in figure 4 for explaining the progress from the creation of an open data strategy to implementation, use of data and evaluation of outcomes, conceptualized as four distinct phases. This process model takes departure from the supplier side of open data; however due to the unique features of the open data phenomenon which represent those of a public good, a multi-stakeholder perspective is necessary. Progressing from a data infrastructure strategy to value

generation and capture is a complex process, including many different stakeholders and recursive elements. However, for conceptual clarity I prefer to illustrate it as a linear process with a cyclical feed-back loop.

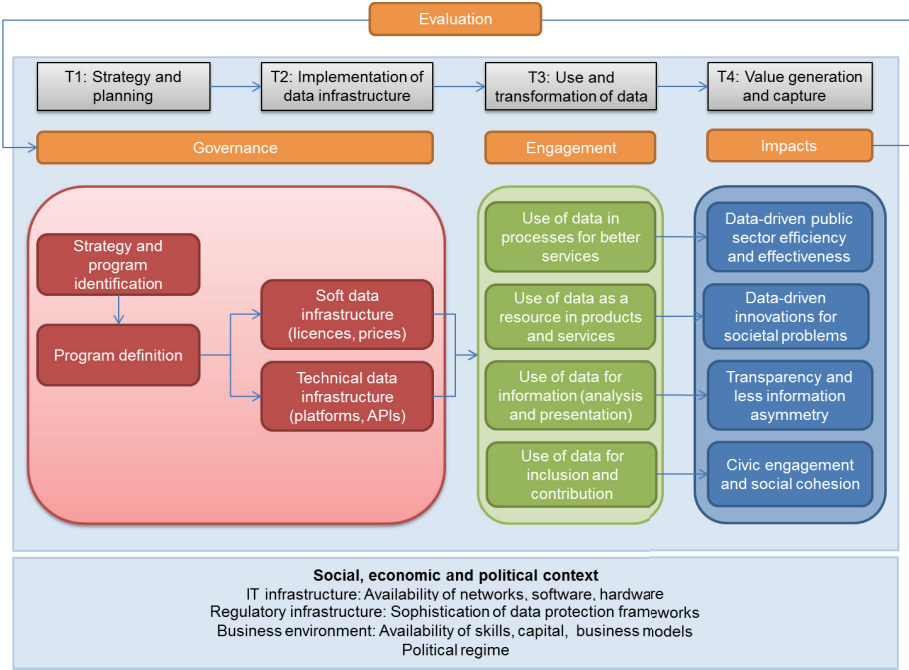


Figure 4: Process model of data-driven value generation through open data infrastructure

The first phase of the process model is focused on strategy and planning – identifying the data infrastructure program through (open) data strategy and (open) data policies, followed by a planning phase for defining the program. The second phase describes the implementation itself: Making data open in the strategic, legal and economic sense, cleaning up data and improving data quality and coherence and finally implementing a technical infrastructure for dissemination of data. This phase can be divided into soft data infrastructure implementation (changing strategy, funding, laws etc.) and technical data infrastructure implementation (improving usefulness and conceptual clarity of the data themselves and implementing platforms that support data access and discoverability of data). For successful completion of these two phases, program governance must facilitate collaboration across boundaries for multi-stakeholder

involvement. Different types of governance strategies for addressing the four most important tensions in the BDP are discussed in-depth in chapter 6.

When data become open, use of the data becomes out of control of the data custodians (Davies, 2010). As there are multiple different users and types of use, the data custodians need to engage with the data users if they want to understand how the data are used and consequently how they can improve implementation of various dimensions of liquid open data. The third phase represents the actual use of the data for various purposes and the required engagement mechanisms. Finally, the fourth phase represents the impacts of data use. The impacts could involve improved public-sector efficiency and/or effectiveness which will generate value through cost-savings and improved level of service in the public sector. Alternatively, an impact from releasing liquid open data could be data-driven innovation, some of which would hopefully address difficult societal challenges. The implementation of an open data infrastructure might result in increased transparency which in turn improves the “vision” of both the public sector itself and citizen, and reduces information asymmetry. Finally, using open data to reach and activate more citizens could result in improved levels of civic engagement and better decision making.

To support the implementation of a data infrastructure and the use of the data for value generation purposes, society needs to possess certain IT capabilities. Most countries are currently facing a shortage of skills in the area of data management, data modelling and data manipulation, analytics and visualization and Denmark is no exception, as was evident from the case data. Moreover, allowing multiple stakeholders to stream large amounts of data across boundaries calls for a high level of network connectivity. In fact, global IP traffic has increased fivefold over the past five years, and is expected to increase threefold over the next five years. Overall, IP traffic will grow at a compound annual growth rate of 23 percent from 2014 to 2019.<sup>13</sup> The business environment needs to be supportive for entrepreneurs, and new types of business models are needed to support value generation from data that are available to all and disseminated for free. Finally, there needs to be a level of political support for dissemination and use of data (for instance, countries that have restricted freedom of speech cannot hope to gain much from transparency) and the regulatory infrastructure must not only support the dissemination of open data but also offer clear rules and regulations around protection of individual privacy and use of personal data.

---

<sup>13</sup> [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI\\_Hyperconnectivity\\_WP.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html)

## VI.6 Analysis and Discussion

”We are reworking our entire data infrastructure from the bottom up. And nobody else is doing that.” (Interview program leader Digitaliseringsstyrelsen, November 2013). Because of the BDP, Denmark now has a data infrastructure model that defines how basic data are organized, how different sets of basic data are interconnected, as well as the political and legal infrastructure that guide reuse of data (Horst et al., 2014). I conclude that the BDP is in fact implementing all of the seven dimensions of liquid open data, including both the technical and the “softer” dimensions. Table 4 summarizes the results.

**Table 4: Implementation of liquid open data in the BDP case**

Dimension	Focus of the BDP	Status
<i>Strategic: Availability</i>	The BDP only focuses on the few but important datasets classified as basic data. This means that there is less focus on the Availability dimension than in many other initiatives where the focus is on opening all non-person identifiable government data.	Decided in the program definition phase.  Ongoing discussion regarding which further datasets should be classified as basic data in potential BDP 2.0.
<i>Economic: Affordability</i>	Chapter 4.3 describes how the BDP managed to make the basic data free-of-charge despite financial austerity in the public sector. Key factor is the funding model.	Decided in the program definition phase. Required changes in laws.
<i>Legal: Reusability</i>	A common license was developed based on the creative Commons-licenser CC0 and CC-BY. However, this license is currently only applied to data that are part of this particular program.	Decided in the program definition phase. Required changes in laws.



<b>Dimension</b>	<b>Focus of the BDP</b>	<b>Status</b>
<i>Technical: Usability</i>	Very strong focus on the usability dimension, especially on raising data quality. Strongest focus on data quality was for the address and property data that had already been made freely available.	Decided in the program identification phase. Raising data quality is perceived as one of the key enablers for improving efficient use of data.
<i>Technical: Discoverability</i>	The new data modeling rules are using standards which make it possible to apply linked data principles later on. The data distributing platform will provide one-stop access to the basic data and use of metadata makes web search easier.	This dimension has evolved through the implementation phase with the development of the platform architecture and modelling rules.
<i>Technical: Accessibility</i>	A strong focus on accessibility through a high-performance platform in the groups that are focused on efficient, internal use of basic data. The groups that focus more on innovation are less interested in this dimension.	The tender process for the data distributing platform influenced this dimension. The first datasets (Geodata) will be published through the platform in fall 2015.
<i>Technical/ conceptual: Interoperability</i>	According to the five stars of basic data: <sup>14</sup> Basic Data are semantically coherent and modelled accordingly to the Model Rules of Basic Data. The design of the modelling rules has mostly been the responsibility of the central program leadership, while the subprograms recognize the importance of coherent basic data.	The modelling rules are still under development and not all of the data have been modelled accordingly. The current version is 1.0.0 but version 1.1.0 is under development. <sup>15</sup>

In an infrastructure project of this range and complexity there are bound to be challenges. The interviews and observations revealed a tension between the efficiency and the innovation ideological agendas. Increased internal efficiency gains happen

<sup>14</sup> <http://www.slideshare.net/EUDataForum/edf2014-nicolas-lemcke-horst-ambassador-of-the-danish-basic-data-programme-agency-for-digitisation-ministry-of-finance-of-denmark-danish-basic-data>

<sup>15</sup> <http://www.digst.dk/Loesninger-og-infrastruktur/Grunddata/Delprogrammerne/Faelles-datamodel/Hvad-er-Grunddatadatamodellen>

mainly through better internal use of data in government information systems. This type of use demands quite an advanced high-performance, high-uptime technical solution that provides internal systems with real time data of high quality and coherence. However, for those purely focusing on stimulating innovation, it would be sufficient to provide data updated daily (at most), through APIs or web services. *“The fundamental idea behind the BDP, to create authoritative registers from which data can be used by everyone, both authorities to use in their processes but of course also the private sector, is a good one. To set data free has also been a huge success, made it possible to freely share data where there were previously thousands of barriers, both internally between public authorities and between authorities and businesses. So two very good principles there.”* *“It is also very, very important when you want to link information across [registers] to have a common data model. Great.”* *“However, there are at many unknowns and risks in this data distributing platform concept.”* (Interview, program leader, Danish Business Authority, November 2014).

One argument is that the strong focus on the technical and conceptual dimensions of liquid open data is causing delays in the program by introducing too much complexity. This in turn has created governance challenges as this level of conceptual and technical interoperability calls for a much higher level of organizational alignment than would have been needed otherwise. Some stakeholders worry that the complexity and the resulting delays have been detrimental to the use of the data for innovation purposes. Those that follow the innovation agenda would prefer that the period from planning an open data initiative to data being actually available for use was shorter. Moreover, external users have expressed their wish for increasing the number of open datasets. However, while the fact is that in 2013, use of open government data for innovation purposes was not marginal in Denmark (as well as in other leading open data countries according to the Open Data Barometer), use of open data for entrepreneurial activities has increased in Denmark. Based on quantitative data from the Open data Barometer and my own work for Open Data Barometer/Web Index, I estimated that the impact of entrepreneurial use of open data in Denmark has actually increased from 5 to 7 on an ordinal scale of 1-10 between 2013 and 2014.

It is time to revisit the research question: *How can the tensions in a multi-stakeholder open data infrastructure programs be addressed through governance strategies?* In order to answer this question, I have conceptualized the main tensions I could identify in the BDP and how they were addressed in the program. Such tensions might be inevitable in a complex infrastructure program but are in many cases not apparent at first glance - and might in fact first appear after years of observation and analysis.

Therefore, I propose that there is significant practical value for other open data infrastructure programs to be aware of these potential tensions beforehand, although they will surely depend on each program's context.

*Tension 1: Short term gains vs. Long term investment*

Although it has been implied by open data enthusiasts that governments should only worry about setting data free and let the market take care of the rest, the experiences so far have shown that the reality for government organizations is much more complicated (Zuiderwijk & Janssen, 2014b). Both public and private users want *high quality, reliable, liquid open data* that they can trust will continue to be delivered without major changes, irrespective of political landscapes and external trends. If stakeholders do not trust the quality and sustainability of data infrastructure, they will hesitate to create mission critical products, services or processes that depend on these data. This has created a chicken-and-egg type of problems for many open government data initiatives that have to balance their scarce resources in a world where demand for services is ever increasing while income generation has been stagnating or sluggish for the last 6-7 years. The imminent danger is that there will be too little infrastructure investment which will limit the growth potential of open data which will again negatively impact the willingness to invest in data infrastructure.

The BDPs strong focus on internal gains that are predicted to more than cover the required investments has to a certain degree alleviated this chicken-and-egg type of paradox. Accordingly, the first implementation strategy identified is to guarantee funding upfront with a transparent funding model that has been agreed upon by all stakeholders. Not only will this enable the program members to make the necessary long-term changes to data collection and publications, it also removes a lot of the uncertainty over the sustainability of open data initiatives that often times prohibits third party use of open data.

*Tension 2: Simplicity vs. Comprehensiveness*

Every open data initiative has to balance their ambitions for open data with the level of funding they receive. While disseminating high quality data that are liquid and open across all seven dimensions of liquid open data is a tempting idea, it is very difficult to achieve in reality. The approach chosen by the BDP was to focus on a limited number of key datasets which can be linked to most other types of data, and to develop comprehensive data modeling principles for these data. The data modeling principles are based on EU Inspire standards wherever possible and the basic units of the rules are general enough to be reused by other public sector organizations in their efforts to

publish liquid open data. Interoperability of data across different domains can thus be improved, without including too many sets of data in the first round of data modeling and data publishing.

The second implementation strategy is to start small, but create reusable data modeling rules based on open standards.

### *Tension 3: Autonomy vs. Control*

The BDP can be viewed as one indicator of a transformation of public governance in Denmark, away from New Public Management principles towards what has been termed Lean Government (Henriksen, 2006; Janssen and Estevez, 2013). Lean Government is characterized by the need to address societal problems without increasing the size of the public sector, thus requiring collaboration across boundaries (Janssen and Estevez, 2013). Interoperability is the foundation for such collaboration, including a common understanding of the basic elements of information that form the basis for seamless functioning of much of the society's most fundamental services. It has been also been proposed that information services are more likely to achieve success if they are designed toward autonomy (modular design) rather than toward top-down control because of the ability to adapt more quickly in response to dynamic changes (Montealegre et al. 2014).

Open government data do not originate from a single organization and therefore can be difficult to publish in a coherent manner. However, while each data custodian is collecting data for their own (regulatory) purposes, these data can undeniably be of much use to other organizations. The data custodians must have enough autonomy to fulfill their individual roles, while contributing towards a common goal of liquid open data, which means using a set of common standards. In order to achieve a balance between those two competing demands, the BDP chose a governance approach called System-of-Systems (SoS) governance. This style of governance seems to be well suited for a constellation of loosely coupled participants, although demanding a high level of network governance skills (Provan and Kenis, 2008). The fact that seven ministries and organizations from both regional and municipal level are collaborating this closely in the BDP has created interest in public sectors around the world.

Accordingly, the third implementation strategy concerns the adoption of a SoS approach, which has helped the BDP to steer away from the inherent silo structures that are characteristic of many governmental hierarchies and achieving coordination without losing autonomy.

#### Tension 4: Exploration vs. Exploitation

Sailing uncharted waters always brings challenges and the BDP has had its share of those. The main divergence within the program might be the view of what type of impacts are most important. For those that are mostly interested in improving data quality and efficient use of data within public sector the road to success might look different than to those that believe first and foremost that data should be set free for innovative use in the private sector. This tension has been evident in the program from the start. However, where there are tensions, synergies are also possible. The BDP case study indicates that these tensions might actually be resolved by maintaining a focus on all the seven dimensions of liquid open data, which was possible in this case due to a limited number of datasets included in the program and the funding model they applied. The infrastructural features of open data have resulted in some unexpected synergies across value generating mechanisms, reflecting the serendipitous value generation opportunities offered by open data. Infrastructural resources are considered as shared means to many ends, which satisfy the following three criteria: 1) they are non-rivalrous, 2) social demand is driven primarily by downstream productive activities, 3) the resource can be used as an input for a wide range of purposes (general purpose criteria) (Frischmann, 2012).

The fourth implementation strategy is to focus on all seven dimensions of liquid open data, as liquid open data are considered to be an infrastructural resource, capable of providing synergies between different value generating mechanisms.

### **VI.7 Contributions, Limitations and Conclusions**

This study provides three main contributions to theory: As a first contribution, I have through a synthesis between the rich data of the BDP group's experiences, the extant literature and secondary data sources, conceptualized a multi-dimensional framework for a construct called liquid open data. This framework highlights the complexity of the phenomenon of open data as an infrastructural resource. The conceptualization is also intended to contribute to conceptual clarity in an emerging field as well as to. As a second contribution, the taxonomy of four different value generating mechanisms moves the concept of value beyond the commonly utilized market- and monetary based mechanisms and thus contributes to theory on how open data as an infrastructural resource can be used to simultaneously create value through different types of mechanisms. The third and main contribution of this longitudinal case study is a process model illustrating the main phases of a government data infrastructure initiative.

As a practice oriented contribution, I propose the liquid open data construct for governments interested in understanding how different dimensions of open data influence different aspects of use. Moreover, I propose that due to their multi-stakeholder perspective, open data infrastructure implementations have the characteristics of a System-of-Systems and should be governed accordingly. Furthermore, the four identified tensions and the governance strategies employed by the BDP group might prove useful for leaders of other data infrastructure initiatives.

A limitation of the case study approach in general is that each case of open data implementation is highly dependent on a number of contextual conditions. I have endeavored to report on the context of this particular case to make it easier to compare it to other similar initiatives for improved understanding. Another limitation is that the research question asks about key strategies but these can differ a great deal between open data initiatives, depending both on context but also on the goals of the initiative in question.

However, I believe that the Danish case can nevertheless provide considerable insight into the characteristics of open data infrastructure initiatives, as well as an important glance into the possible future of open government data. The program members have, in my view, successfully managed to steer through the rough seas and might even have laid the foundation for a revolutionary change in public sector data management. Right from the start, there has been strong motivation and belief in a common cause, which has survived over the course of the program, despite multiple challenges. This belief is that when data become liquid and open, they offer the opportunity to generate a great deal of social and economic value for public and private sectors alike. I suggest that this motivational factor has played a big role in the program's success and lifted the focus towards the eventual value generation, a focus which has been suggested to be lacking in other such initiatives (Zuiderwijk and Janssen, 2014b). Finally, in a very Danish tradition, the BDP has taken the first steps toward a future where individual datasets function as Lego bricks that can be put together in plethora of ways to create much more adaptive and agile systems than was previously possible. However, time must tell if this will indeed become reality for the open government data infrastructure in Denmark.

### *Acknowledgements*

I would like to thank the reviews and senior editor for their insightful comments and constructive feedback. I express my sincerest appreciation to all the members of the Basic Data Program for providing me with such valuable insight into their important initiative. Special thanks go to program leader Jens Krieger-Røyen for his important support and for providing access to all the documents and events. I also thank all the other informants I have met with during these almost four years of research, especially my colleagues at KMD for sharing their knowledge and expertise on the subject matter. For providing financial support, I would like to thank KMD and Innovationsfonden.

## VI Appendix A: Overview over interviews

Dates	# inter- viewees	# interviews	Role	Organization	Time
August 2012	2	1	Program manager Program leader	Agency for Digitization	1:15:00
April 2013 November 2013 July 2014 January 2015	1	4	Program manager Program leader	Agency for Digitization	0:44:22 0:58:48 0:29:20 0:36:15
August 2012	1	1	Product owner	KMD	0:30:54
August 2012	1	1	Project manager	Agency for Digitization	1:11:44
August 2012 September 2014 November 2014	1	3	Product owner Project Manager	KMD Geodata Agency	1:00:03 1:16:16 0:57:35
September 2012	2	1	Product owners	KMD	
September 2012	1	1	Technical architect	Agency for Digitization	0:54:12
September 2012 January 2014	1	2	Product owner Project manager	KMD Ministry of Housing, Urban and Rural Affairs	49:01 58:56
October 2012	1	1	External		1:25:56



Dates	# inter-viewees	# interviews	Role	Organization	Time
			consultant		
November 2012	1	1	Project manager	KOMBIT	0:51:14
January 2013 January 2014	1	2	Project manager	Geodata Agency	1:25:40
February 2013	1	1	Consultant	Ministry of Housing, Urban and Rural Affairs	1:27:07
May 2013 February 2014 September 2014	1	3	Basic data ambassador	Agency for Digitization	1:06:39 0:55:32
October 2013	2	1	Project manager Programmer	Copenhagen City	0:38:26
January 2014	1	1	Project Manager	Ministry of Housing, Urban and Rural Affairs	0:53:06
February 2014	2	1	App developers	Geodata Agency	0:38:11
September 2014	1	1	Data user	KMD	0:41:01
November 2014	1	1	Project manager	Agency for digitization	0:59:46
November 2014	1	1	Project leader	Danish Business Authority	1:05:20
Total	22	28			

## VI References

- Agency for Digitization (2012). *Good basic data for everyone – a driver for growth and efficiency*. Available <http://www.digst.dk/Home/ServiceMenu/English/Digitisation/Basic%20Data>
- Agency for Digitization (2011). *eGovernment strategy 2011-2015: The digital path to future welfare*. Available <http://www.digst.dk/ServiceMenu/English/Policy-and-Strategy/eGOV-strategy>
- Boardman, J., & Sauser, B. (2008). *Systems thinking: Coping with 21st century problems*. CRC Press.
- Chen, D, Doumeingts, G. and Vernadat, F. (2008). Architectures for enterprise integration and interoperability: Past, present and future. *Computers in Industry*, 59(7), 647-659.
- Conradie, P. and Choenni, S. (2014). On the barriers for local governments releasing open data. *Government Information Quarterly*, 31(1), S10-S17.
- Curry, E. (2012). System of Systems interoperability using a linked dataspace. In the *Proceedings of the 7th International Conference on System of Systems Engineering (SoSE)*, Genoa, Italy, 101-106.
- Davies, T. (2010). Open data, democracy and public sector reform: A look at OGD use from data.gov.uk. Available at <http://tinyurl.com/7joks46> [Accessed December 2012]
- Davies, T. (2013). *Open Data Barometer: 2013 Global Report*. Available at <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>
- DECA. (2010). The value of the Danish address data: Social benefits from the 2002 agreement on procuring address data etc. free of charge. Available at [http://www.adresse-info.dk/Portals/2/Benefit/Value\\_Assessment\\_Danish\\_Address\\_Data\\_UK\\_2010-07-07b.pdf](http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf) [Accessed July 2012]
- DeLaurentis, D., & Callaway, R. K. (2004). A system-of-systems perspective for public policy decisions. *Review of Policy Research*, 21(6), 829-837.
- de Vries, M.; Kapff, L.; Achiaga, M. N.; Wauters, P.; Osimo, D.; Foley, P.; . . . & Whitehouse, D. (2011). Pricing of public sector information study (POPSIS), Brussels: European Commission, Information Society and Media Directorate General
- de Vries, M. and Pijpker, U. (2013). *The Danish Dash: A short story unravelling the Danish magic of shaping a System of Key registers in less than six months*. Available [www.pblq.nl/media/331260/danish\\_dash\\_final\\_1\\_.pdf](http://www.pblq.nl/media/331260/danish_dash_final_1_.pdf)
- Eurostat (2013). General government expenditure on social protection and health. Available [http://ec.europa.eu/eurostat/statistics-explained/index.php/General\\_government\\_expenditure\\_on\\_social\\_protection\\_and\\_health#Social\\_protection:\\_the\\_most\\_important\\_function\\_of\\_government\\_expenditure](http://ec.europa.eu/eurostat/statistics-explained/index.php/General_government_expenditure_on_social_protection_and_health#Social_protection:_the_most_important_function_of_government_expenditure)
- Fioretti, M. (2011). *Open data: Emerging trends, issues and best practices*. Available at [http://www.lem.sssp.it/WPLem/odos/odos\\_2.html](http://www.lem.sssp.it/WPLem/odos/odos_2.html) [Accessed July 2012]
- Foulonneau, M., Turki, S., Vidou, G., & Martin, S. (2014). Open data in Service design. *Electronic Journal of e-Government*, 12(2), 99-107.
- Frischmann, B. M. (2012). *Infrastructure: The Social Value of Shared Resources*. Oxford: Oxford University Press

- Gottschalk, P. (2009). Maturity levels for interoperability in digital government. *Government Information Quarterly*, 26(1), 75-81.
- Halonen, A. (2012). *Being Open About Data*. Available at <http://tinyurl.com/c8mz5vt>
- Henriksen, H. Z. (2006). Fad or investment in the future: an analysis of the demand of e-services in Danish municipalities. *Electronic Journal of e-Government*, 4(2), 19-26
- Hjørt-Madsen, K. (2006). Enterprise architecture implementation and management: A case study on interoperability. In the *Proceedings of the 39th Annual Hawaii International Conference on System Sciences HICCS*, Vol. 4, 71c-71c.
- Hofman, W., & Rajagopal, M. (2014). A technical framework for data sharing. *Journal of theoretical and applied electronic commerce research*, 9(3), 45-58
- Horst, N. L., Bjerre, S., Lind, M., & Hvingel, L. (2014). Grunddataprogrammet-en dansk infrastrukturmodel for offentlige data. *Geoforum Perspektiv*, 13(24).
- Hujiboom, N., and Van den Broek, T. 2011. Open data: an international comparison of strategies. *European Journal of e Practice*.
- Höchtel, J., Davies, T., Janssen, M. F. W. H. A., & Schieferdecker, I. (2014). Open Data: Growing Up and Getting Specific. *JeDEM: eJournal of eDemocracy and Open Government*, 6 (1), i-iii.
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258-268.
- Janssen, M. & Estevez, E. (2013) Lean government and platform-based governance - Doing more with less. *Government Information Quarterly*, 30(1), S1-S8.
- Jetzek, T., Avital, M. and Bjørn-Andersen, N. (2012). The Value of Open Government Data: A Strategic Analysis Framework. Paper presented at the 2012 *Pre-ICIS e-Government SIG Workshop*, Orlando, USA.
- Jetzek T., Avital, M. and Bjørn-Andersen, N. (2013a). The Generative Mechanisms of Open Government Data. In the *Proceedings of the 21st European Conference on Information Systems (ECIS) 2013*, Utrecht, The Netherlands.
- Jetzek, T., Avital, M. and Bjørn-Andersen, N. (2013b). Generating value from open government data. In the *Proceedings of the 34th International Conference on Information Systems (ICIS 2013)*, Milan, Italy, 1-20.
- Jetzek, T., Avital, M. and Bjørn-Andersen, N. (2014a). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2):100-120.
- Jetzek, T., Avital, M. and Bjørn-Andersen, N. (2014b). "Generating sustainable value from open data in a sharing society. " In *Creating Value for All Through IT*, edited by Birgitta Bergvall-Kåreborn and Peter Axel Nielsen, 243-260. Berlin Heidelberg :Springer
- Kettl, D. F., Pollit, Christopher, S. & James H. (2004). Towards a Danish concept of public governance: An international perspective. *Report to the Danish Forum for Top Executive Management*. Available <http://www.publicgovernance.dk/docs/0408260903.pdf>
- King, J. L., & Uhler, P. F. (2014). Soft infrastructure challenges to scientific knowledge discovery. *Communications of the ACM*, 57(9), 35-37

- Krieger-Røyen, J. (2014, june). *The Danish BDP and the governance perspective*. Available [http://inspire.ec.europa.eu/events/conferences/inspire\\_2014/pdfs/plenaries/Grunddata\\_INSPIRE\\_JRO5.pdf](http://inspire.ec.europa.eu/events/conferences/inspire_2014/pdfs/plenaries/Grunddata_INSPIRE_JRO5.pdf)
- Lakomaa, E. & Kallberg, A. (2013). Open Data as a Foundation for Innovation: The Enabling Effect of Free Public Sector Information for Entrepreneurs. *IEEE Access*, 1(2013), 558-563.
- Martin, S., Foulonneau, M., Turki, S. & Iahadjadene, M. (2014). Risk Analysis to Overcome Barriers to Open Data. *Electronic Journal of e-Government*, 11(2), 348 -359.
- Maheshwari, D., & Janssen, M., Reconceptualizing measuring, benchmarking for improving interoperability in smart ecosystems: The effect of ubiquitous data and crowdsourcing. *Government Information Quarterly*, 31(1), S84-S92.
- Mayer-Schönberger & Zappia, Z. (2011). Participation and Power: Intermediaries of Open Data. Conference draft prepared for the Berlin Symposium. Available: <http://tinyurl.com/bo3pyl8>
- McKinsey & Company. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- McKinsey & Company (2013). *Open data: Unlocking innovation and performance with liquid information*, McKinsey Global Institute, McKinsey Center for Government and McKinsey Business Technology Office.
- Montealegre, R., Hovorka, D., & Germonprez, M. (2014). A Coevolutionary View of Information Services Development: Lessons from the US National Oceanic and Atmospheric Administration. *Journal of the Association for Information Systems*, 15(9), 577-613.
- Nilsen, K. (2010), Economic theory as it applies to Public Sector Information. *Annual Review of Information Science and Technology*, 44(1), 419-489.
- OECD (2012). *Human Resources Management Country Profiles Denmark*. Available <http://www.oecd.org/gov/pem/OECD%20HRM%20Profile%20-%20Denmark.pdf>
- Open Knowledge Foundation (OKF) (2015). Available at <http://www.opendefinition.org/> and <http://opengovernmentdata.org>
- Pollock, R. (2008). *Economics of PSI*. Cambridge: University of Cambridge. Available at [http://rufuspollock.org/economics/papers/economics\\_of\\_psi.pdf](http://rufuspollock.org/economics/papers/economics_of_psi.pdf) [Accessed August 2012]
- Provan, K. G., & Kenis, P. N. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory*, 18(2), 229-252.
- Rechtin, E., & Maier, M. W. (2009). *The art of systems architecting* (3<sup>rd</sup> edition). CRC Press.
- Statistics Denmark (2013). *Denmark in figures 2013*. Available <http://www.dst.dk/pukora/epub/upload/17953/dkinfigures.pdf>
- Street, C.T. and Ward, K.W. (2012). *Improving validity and reliability in longitudinal case study timelines*. *European Journal of Information Systems*, 21(2), 160-175.
- Sullivan, A. & Sheffrin, S.M. (2003). *Economics: Principles in action*. Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Tammisto, Y. & Lindman, J. (2011). Definition of open data services in software business. Paper presented at the third International Conference on Software Business (ICSOB 2012), Cambridge, MA, USA. In Cusumano, M.A., Iyer, B., & Venkatraman, N. (eds.), *Software Business Lecture Notes in Business Information Processing*, 114(2012), 297-303.

- Van de Ven, A. H. (2007). *Engaged Scholarship - A Guide for Organizational and Social Research*. Oxford:Oxford University Press
- Van Der Blonk, H. (2003). Writing case studies in information systems research. *Journal of Information Technology*, 18(1), 45-52.
- van Veenstra, A.F. & van den Broek, T.A. (2013). Opening Moves – Drivers, Enablers and Barriers of Open Data in a Semi-public Organization. In *Proceedings of the 12th IFIP WG 8.5 International Conference, EGOV 2013*, Koblenz, Germany, 50-61
- Verhulst, S., Noveck, B. S., Caplan, R., Brown, K. & Paz, C. (2014). The Open Data Era in Health and Social Care. The Governance Lab. [online]. Available: <http://images.thegovlab.org/wordpress/wp-content/uploads/2014/06/nhs-full-report.pdf>
- Yin, R.K. (2009). *Case Study Research: Design and Methods* (4th ed.). Los Angeles: Sage.
- Zuiderwijk, A., Janssen, M., Choenni, S, Meijer, R. and Alibaks, R.S. (2012). Socio-technical Impediments of Open Data, *Electronic Journal of e-Government*, 10(2), 156 – 172.
- Zuiderwijk, A. and Janssen. M. (2014a). "Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View. " in *Open Government, Opportunities and Challenges for Public Governance*, edited by Mila Gascó-Hernández, 115-136, New York: Springer.
- Zuiderwijk, A., & Janssen, M. (2014a). The negative effects of open government data-investigating the dark side of open data. In *Proceedings of the 15th Annual International Conference on Digital Government Research*, 147-152.
- Zuiderwijk, A., & Janssen, M. (2014b). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17-29.

# **PAPER VII: Driving Sustainable Value: A Conceptual Model of Open Data as a Resource**

Unpublished manuscript

# Driving Sustainable Value

## A Conceptual Model of Open Data as a Resource

### Abstract

In spite of a recent explosion in the number of public Open Data initiatives, driven by expectations of unprecedented value generation, the mechanisms that explain how value is generated in network relationships are not yet well understood. Through analysis of secondary quantitative data for seventy-six countries in 2013, this study proposes the main constructs and relationships that can influence and explain how value generation happens through use of open government data. We propose that value generation through use of an open and free resource such as open government data is still difficult to understand for organizations and individuals as much of the resulting value is intangible. This is especially true because of the recent trend towards dissemination of valuable information through free and open platforms and services. We thus argue that we are witnessing a new class of value generating mechanisms that are not market-type mechanisms. We call them information sharing mechanisms. Moreover, we argue that open data contribute significantly to the generation of what we prefer to call sustainable value through both information sharing and market mechanisms. We suggest that having a measure such as sustainable value is imperative for societies that would like to go beyond measuring the volume of marketed economic activity and rather evaluate the degree to which society's goals are met. We thus hypothesize about the relationship between open data, two mediating variables that reflect the degree of digital content sharing on the one hand and digital commercial product and services innovation on the other, and sustainable value, as well as four important factors in our societal infrastructure that can stimulate or inhibit the value generation mechanisms. Finally, we discuss the impact of private sector accountability on the relationship between information transparency and sustainable value on the one hand and product and service innovation and sustainable value on the other.

**Keywords:** Open data, Liquid open data, Value, Mechanisms, PLS

## VII.1 Introduction

Sudden increase in produced data coupled with the trend towards openness has provided society with a valuable and unique resource, most commonly termed *open data*. Data have certainly followed mankind for a long time; the Babylonians are said to have started to register land ownership as soon as 3000 BC. But the potential to effectively access and combine vast amounts of diverse data from different sources is an emerging phenomenon. In the recent past, the few available open data sources were hard to discover and difficult to access. The efforts required to collect external data and to convert them into a well-structured form, suitable for traditional analytical techniques were simply too extensive for most users. However, due to the ever increasing digitization of human activity, the amount and diversity of available data have increased exponentially, drastically changing the landscape of the information economy. In 1986 about 1% of all stored data were digital but in 2007 this percentage had grown to 94%, and is still growing (Hilbert and Lopez, 2011). It has been estimated that 2,5 exabytes of data were generated every day in 2012 and that the total volume of data will grow around 60% per year.<sup>16</sup> Moreover, new ways of storing and analyzing data have offered the possibility to collect and connect large amounts of heterogeneous datasources and to transform them into appealing visualizations. These changes are driving a dramatic increase in the *level of information* available.

The overarching expectation is that better use through openly sharing these data will create great benefits for society, organizations, and individuals in the coming years (Beardsley et al, 2014; Kundra, 2012; European Commission, 2011b). As an example, while the genomics revolution accelerated genome mapping through gene-sequencing machines, the real revolution began when researchers started to share genomics data (King and Uhler, 2014). However, for these expectations to fully materialize, a number of factors must be in place. To exploit the opportunities offered by open data, we must move beyond technical challenges, to the challenges of "soft infrastructure": institutional factors, governance, and cultural inertia that tend to impede payoffs from the rapid evolution of techniques and methods (King and Uhler, 2014). Among the most important macro-level enabling factors for value generation through data are robust high-speed Internet networks, a workforce with relevant skills and education, and consumer trust regarding the protection of personal data and privacy (Bilbao-Osorio et al., 2014). Weak or absent data protection laws are undermining citizen confidence in open data initiatives and limited training and support for intermediaries hinders use of data for economic and social benefits (Andrade et al., 2014). The public

---

<sup>16</sup> [https://www-304.ibm.com/events/www/grp/grp004.nsf/vLookupPDFs/Tim%20Paydos%20Presentation/\\$file/Tim%20Paydos%20Presentation.pdf](https://www-304.ibm.com/events/www/grp/grp004.nsf/vLookupPDFs/Tim%20Paydos%20Presentation/$file/Tim%20Paydos%20Presentation.pdf)



sector plays an important role in facilitating value generation from data. Apart from producing, disseminating and using data themselves, the public sector must also promote and foster data-driven innovation and growth throughout (Andrade et al., 2014).

In this paper we will address the question “how can open data contribute to sustainable value generation”. For this purpose we need to construct a clear terminology, i.e. to untangle the vague concepts of open data and sustainable value. The ambiguity of the open data concept has led to a variety of interpretations, making it harder for policymakers and other users to articulate clear priorities and make cogent demands (Yu and Robinson, 2012). Thus we have conceptualized what we name *liquid open data* as a multidimensional construct consisting of 7 dimensions. Moreover, we aim to conceptualize a construct that can reflect the impact of using of liquid open data. Currently, there is not much evidence for the value generated as a result of OGD initiatives. In fact, the recent Open Data Barometer report highlights that strong evidence on the impacts of OGD is almost universally lacking (Davies, 2013). This lack of evidence is further complicated by questions of how to measure and evaluate something as complex as the value of open data.

The problem of identifying value from open or public resources is related to similar discussions in various fields of research regarding the nature of value creation and capture in the *networked economy* (see for instance Benington, 2011; Bowman and Ambrosini, 2010, Grover and Kohli, 2012; Stiglitz et al., 2009). Due to technological advances, value generating mechanisms are slowly shifting from one-to-one transactions towards many-to-many networks where the size of the networks and the interactions themselves facilitate value generation (Bowman, 2014; Viscusi et al., 2014). Hence, the predominant monetary based measurements are becoming increasingly problematic. For instance, despite the fact that creation and use of digital information is growing exponentially, the share of the information sector (software, publishing, motion picture and sound recording, broadcasting, telecom, and information and data processing services) has not grown in the national accounts since around the 1960s and is still around 4% of GDP (Brynjolfsson and Saunders, 2009). As the national accounts present the sum of value-added in market-based transactions, dissemination and use of free data and information services is not accounted for in the GDP measure. While GDP is an accurate metric of value in industrial-age industries like steel or automobiles *it will miss most of the value* in information goods (Brynjolfsson and Saunders, 2009; Brynjolfsson and Oh, 2012).

We suggest that due to the exponentially growing resource of open data, the value of data and information is becoming an increasingly important topic, highly relevant to modern societies. We propose that all societies need to be aware of the importance of mechanisms that generate value through dissemination of data and conversion of these data into information. These mechanisms are recognized by most but have not enjoyed the same interest as the market mechanisms in the existing theoretical and empirical literature. The paper is structured as follows: First we juxtapose current definitions of open data to generate a construct that can be used to measure country-level openness of government data. Next we proceed to discussing the multi-dimensional value construct, highlighting the importance of looking at a variety of indicators as much of the value generated from open data is not easily measured in currency. When open data are shared within a network of different stakeholders, value is derived from both the information that is generated and the traditional transactions between a buyer and a seller. The two following chapters describe the research model and research method and finally we proceed to discussion and conclusions.

## **VII.2 Dimensions of Open Data as a Strategic Resource**

We view data as a raw material or resource that can be used in a variety of ways. As a resource, data offer some unique attributes. First, digitization has affected two important features of data: 1) when data become easily accessible to more than one person at a time, they acquire the feature of *non-rivalry* and; 2) when marginal costs incurred by re-production and distribution are drastically reduced, re-use of data becomes economically feasible (Nilsen, 2010, Shapiro and Varian, 1999). Second, as openness implies that the data are available for use outside of the organization that produces or collects the data, openness offers the feature of *non-excludability*. Accordingly, when opened up, digital data become a shared resource - a public good or what has been termed “digital commons” (Hess and Ostrom, 2006). The economic properties of data suggest looking at data as an infrastructural resource (OECD, 2014). Infrastructural resources are considered as shared means to many ends, which satisfy the following three criteria: 1) they are non-rivalrous, 2) social demand is driven primarily by downstream productive activities, 3) the resource can be used as an input for a wide range of purposes (general purpose criteria) (Frischmann, 2012).

*Open data* can be defined as data that are freely accessible online, available without technical restrictions to re-use, and provided under open access license that allows the data to be re-used without limitation, including across different ‘fields of endeavor’ (e.g., commercial and non-commercial alike) (OKF, 2012). However, this is just one of

many definitions and interpretations of what is open data that are used in the practice oriented literature. While most of the definitions are very similar, there are important differences (Verhulst et al., 2014).

The Open Knowledge Foundation has emphasized the following principles of open data (OKF, 2012):

- Availability and Access: the data must be *available as a whole*, at no more than a *reasonable reproduction cost*, preferably by *downloading over the internet*. The data must also be available in a *convenient and modifiable form*.
- Reuse and Redistribution: the data must be provided under terms that *permit reuse and redistribution* including the intermixing with other datasets.
- Universal Participation: everyone must be able to use, reuse and redistribute - there should be *no discrimination* against fields of endeavor or against persons or groups

The Open Data Handbook highlights the importance of interoperability, which denotes the ability to interoperate different datasets<sup>17</sup>. In the scientific literature, providing open data (supplier side view) has been proposed as being a matter of availability, accessibility, format and license (Davies, 2010). From the demand point of view, openness is proposed to combine unrestricted availability of data with accessibility and technical interoperability (Tammisto and Lindman, 2012; Lindman and Tammisto, 2011). Berners-Lee set out a progressive model for open data publication, where the goal was to move towards making data machine readable in open formats, and then ultimately to complement these accessible datasets with standardized and linked datasets, supporting citizens, entrepreneurs and government itself to connect disparate data across the web (Davies, 2013).

The concept of liquid data is a recent one, referring to data that are open, widely available, and published in shareable formats (McKinsey, 2013). The concept of liquidity implies the ability of data to “stream” across systems. Building on this concept, we extend the definition of open data by emphasizing the potential to use open data across systems in a consistent manner. Table 1 gives an overview over seven different dimensions of what we define as **Liquid open data**: *data that are available online, free-of-charge and under an open access license, published in machine-readable formats, easily discoverable, accessible and conceptually coherent. Liquid open data can be re-used without discrimination or limitation, linked to other data and streamed across systems.*

---

<sup>17</sup> <http://opendatahandbook.org/>

**Table 1: The Seven Dimensions of Liquid Open Data**

<b>Dimension</b>	<b>Description</b>
<i>Availability</i>	The availability dimension reflects the strategic importance of open data, ranging from all government data that are not subject to privacy or national security limitations being open to all by default, to government data in general not being available outside of organizational boundaries.
<i>Affordability</i>	The affordability dimension refers to the pricing of data. For data to be open y should preferably be free, or at most sold at marginal costs (in which case transaction costs could easily outweigh potential income from data).
<i>Reusability</i>	The reusability dimension depends on the type of license used for government data intended for reuse. The license can range from a type of creative commons licenses that allows anyone to use the data for whatever purposes they like, to very strict licenses that allow use for a single purpose only. Reusability can also depend on the legislative framework which should not present barriers to reuse of data.
<i>Usability</i>	The usability dimension refers to the clarity and ease with which we can interact with the data out of original context. Usability requires that data are technology and context agnostic. Usable data should be provided in machine readable formats using open standards and including documentation about how data are produced and published (metadata) in order to give context.
<i>Discoverability</i>	The discoverability dimension refers to whether potential users can easily discover the data, for instance using search mechanisms.
<i>Accessibility</i>	The accessibility dimension refers to whether data are easily, consistently and securely accessible and downloadable or streamable.
<i>Interoperability</i>	The interoperability attribute reflects the idea that data should be conceptually open - meaning that data are published in a way that enables them to be used outside of the context within they were collected. Interoperability of data depends both on structure (syntax) and meaning (semantics).

Despite an increasing coherence of the concept of open data, understanding *how value is created from open data* is not at all straightforward, particularly considering the complexity of value constructs such as public value (Zuiderwijk et al., 2014a). Most of the current research has focused on how collection and dissemination of data should proceed (data policies) and the main enablers and barriers for publishing open data (Conradie and Choenni, 2014; Janssen et al., 2012; Martin et al., 2013; Zuiderwijk and Janssen 2014a; 2014b; van Veenstra and van den Broek, 2014b; Zuiderwijk et al., 2012; 2014). Moreover, a growing body of literature now studies the demand side, including how open data are used in services, use of different business models, the role of intermediaries and the formation of open data ecosystems (Foullonneau et al., 2014, Heimstädt et al., 2014, Hjalmarsson et al. (2015), Kuk and Davies 2011, Lindman et al., 2013). However, there is a gap in the literature when it comes to understanding the relationship between open data and value (Zuiderwijk et al., 2014a). In the next chapters we review the complex value generation and capture that follows the transformation of modern society from an industrialized economy to a networked economy of interrelated stakeholders that are to an increasing degree openly sharing resources and information.

### **VIII.3 Value Generation in a Networked Economy**

#### **The Multifaceted Nature of Value**

Most of the written material on open data is grounded in the assumption that making data open will result in value being generated (Ubaldi, 2013). While the specifics of how this happens are usually not clear, value generation from liquid open data is based upon being given the opportunity to reuse data beyond the organizational boundaries of the data custodian. Organizations that generate value from open data commonly use a variety of available datasources, where in many cases open government data are linked to private or proprietary data (Jetzek et al., 2014a, Jetzek et al. 2014b). There is preliminary evidence that use of open data is increasing in the private sector, and that this usage is already contributing to value generation. For instance, a recent survey of 318 C-level executives conducted by the Economist Intelligence Unit (2013) shows that 70% of the surveyed firms did collect open data, both from governments and other sources. The survey furthermore reveals that 54% of high-growth firms both collect and analyze open data, while the percentage for low-growth firms is 36%. From a macro perspective, the economic value potential from open data has been estimated as \$3 trillion USD pr. year globally over seven sectors (McKinsey, 2013).

Two types of value are frequently discussed in the open data literature: *economic value*, defined as the worth of a good or service as determined by the market; and *social value*, which is created when resources, inputs, processes or policies are combined to generate improvements in the lives of individuals or society as a whole. However, the structures of the industrial economy's markets were mainly built to facilitate the creation and capture of economic value. One feature of the industrial economy is a dichotomy of the market and the state, where the state is responsible to create a structure around so-called market failures and minimize the social and environmental harm that might result from pure market activity (Porter and Kramer, 2011). However, due to continuing advances in digital technologies and the insufficient ability of current social and economic structures to address emerging social challenges, the perceived divergence between the generation of social and economic value is becoming increasingly contested. The 2014-2015 Global Competitiveness Report highlights how policymakers, businesses, and citizens are increasingly recognizing the need for economic growth to be balanced by the creation of opportunities and benefits for all segments of the population and the longer term focus on a sustainable environment. Accordingly, the social and environmental dimensions of an economy need to be fully considered in any growth or development agenda (Schwab, 2014).

Similar ideas have infiltrated most scientific disciplines. The concept of *shared value* has been introduced to the management literature as an attempt to amalgamate the concepts of social and economic value (Porter and Kramer, 2011). Shared value is based on the idea that societal needs, not only economic needs, define markets. It also recognizes that social harms frequently create internal costs for companies, such as wasted energy, health problems and/or the need for remedial training to compensate for inadequacies in education (Porter and Kramer, 2011). The generation of shared value is essentially about expanding the total pool of economic and social value, increasing the competitiveness of businesses and society at the same time.

Another value concept increasingly used in the public administration and e-government literature is *public value*. The central proposition in Moore's (1995) public value framework is that public resources should be used to increase value, not only in an economic sense but also more broadly in terms of what is valued by citizens and communities. This definition of public value has recently been revisited to include a discussion of "What adds value to the public sphere" (Benington, 2011, p. 31), where the public sphere includes the state, the market and civil society. The dimensions of public value include: 1) economic value – adding value to the public realm through the generation of economic activity, 2) social and cultural value – adding value to the

public realm by contributing to social cohesion and well-being, 3) political value – adding value to the public realm by stimulating and supporting democratic dialogue and active public participation, as well as citizen engagement and 4) ecological value – adding value to the public realm by actively promoting sustainable development and reducing pollution, waste, and global warming (Benington, 2011).

In welfare economics, the concept *social welfare* is used to describe a construct that reaches further than the commonly used measure of Gross Domestic Product (GDP) as an indicator of value. As GDP mainly measures market production, it ignores many of the determinants of social value such as the environment, freedom, health and education (Fleurbaey, 2009; Michaelson et al., 2009). For example, traffic jams may increase GDP as a result of the increased use of gasoline, but do obviously not improve the quality of life. However, companies that develop and sell products that reduce traffic jams by providing commuters with real-time traffic information will not only contribute to GDP but also improve wellbeing through reduced stress, decreased pollution and by saving the valuable time of commuters.

Within the IS value research stream, the need to assess *intangible value* has recently been emphasized (Kohli and Grover, 2008). In today's complex organizational settings, both businesses and consumers can be seen as the arbitrators of value creation, therefore the true benefits of IT to these stakeholders may be underreported if pure financial metrics, or even ex ante market value, are overemphasized (Kohli and Grover, 2008). Accordingly, the relationship between the generation and appropriation of value for multiple stakeholders in the context of networks and alliances is increasingly being viewed to be multifaceted in nature (Sarker et al., 2012). Gil-Saura et al. (2009) suggest that value is a subjective, multidimensional construct; accordingly, it is only through a multidimensional view that we get a true picture of the value generated within network relationships.

Building upon these different but ideologically similar discussions, we introduce the concept of sustainable value. The definition of sustainable value represents a move away from the previously dominant economic value focus and moving towards a focus on proactive, concerted efforts of businesses, government institutions and the overall community in addressing social challenges in innovative and holistic ways that generate social, environmental and economic value for all stakeholders and future generations (van Osch and Avital, 2010). We view sustainable value as *a multidimensional construct, reflecting the simultaneous generation of sustainable economic, social and environmental value in society.*

## Value Generation in a Collaborative Context

As open data are a shared or common resource, they hold great potential for a number of stakeholders, including public sector agencies, private businesses, the academia, citizens and civic organizations. The relationships that describe where value is added and where value is captured, and by whom, are becoming increasingly complex in the context of shared resources that used by a network of stakeholders (Bowman and Ambrosini, 2000; Bowman, 2014). Going beyond shared resources, we see that value is increasingly being co-created by many stakeholders. Grover and Kohli (2012, p. 231) state that “co-creation represents one of the most important streams in the IT value research area that will gain greater importance as firms expand collaborative relationships with other firms. Co-creation of value happens when *different stakeholders join together to create new value that none of them could create on its own* (Grover and Kohli, 2012). However, despite the recent focus on the creation of value in collaborative settings, little is known about the underlying mechanisms (Sarker et al., 2012).

Governments seeking to find some way to provide social protection to their citizens while strengthening national economic competitiveness are becoming more interested in cross-sector co-creation of value. They are currently exploring and supporting voluntary private sector initiatives, balancing the interest of businesses of business and society without expanding the role of government (Haufler, 2001). However, if companies are to take increasing responsibility for the generation of sustainable value, they will need to perform in the interests of shareholders and wider society. Companies will be required to inform all of these stakeholders about how they are performing in the context of sustainable value generation, and shareholders and others need to be able to hold the companies to account. While competitive *markets* are very successful in proliferating products and services to satisfy changing desires for personal consumption, but are increasingly found to be less capable of catering for more complex profound and enduring social needs (Benington et al. 2011). *Networks* have been shown to be capable of supporting value co-creation due their flexibility and adaptability (Morgan et al., 2013). However, networks present major challenges in terms of governance (how to sustain clear strategic direction) and accountability (how to account to multiple stakeholders, with very different mandates (Benington et al. 2011; Morgan et al., 2013). The new patterns of networked community governance which are now emerging are lacking the support of a distinctive economic and social theory (Benington et al., 2011).



In the context of open data, current research is revealing that open data initiatives where public bodies have strong ties to the eventual data-users have been most successful (van Veenstra and van den Broek, 2014a), calling for a more structured network governance (Eckartz et al., 2014). These results are similar to those from the open source software literature (see for instance Morgan et al., 2013). Innovative activities intended to generate sustainable value should be supported by an innovation ecosystem, which the public sector should seek to establish in order to increase use of their data (Bason, 2010). Building upon these notions of collaborative creation of value as a multi-dimensional construct in the context of open data, we can define the open data ecosystem as a *distributed network built around the sharing of data across boundaries in order to create information, products and services that contribute to sustainable value generation*.

## **VII.4 The Social Mechanisms of Value Generation**

### **Data – Information – Knowledge**

Before diving deeper into the mechanisms that can explain the transformation from raw data to sustainable value, we have to clarify the important distinction between data and information. The data – information – knowledge relationship still remains at the heart of the IS discipline (Kettinger and Li, 2010). If we are to understand the relationship between raw data and value, we need to have a basic understanding of the underlying value chain - or in this context, perhaps more accurately, value network. Our understanding is that data are recorded (captured and stored) symbols and signal readings while information is a message that contains relevant meaning, implication, or input for decision and/or action (Liew, 2007). Knowledge, on the other hand, is a personal trait, which can be reflected in (1) cognition or recognition (know-what), (2) capacity to act (know-how), and (3) understanding (know-why) that resides or is contained within the mind or in the brain (Liew, 2007). In short, the ultimate purpose of knowledge is for value creation (ibid). Increasing the knowledge of members of society in general will provide society with the potential to generate sustainable value through better decision making.

Kettinger and Li (2010) conceptualize information as representing a status of conditional readiness for an action, that is generated from the interaction between the states measured in data and their relationship with future states predicted in knowledge (Kettinger and Li, 2010). „Information is the meaning produced from data based on a knowledge framework that is associated with the selection of the state of conditional

readiness for goal-directed activities.“ (Kettinger and Li, 2010, p. 415). Thus, new information forms the basis for action through the potential influence on people’s decisions, and consequently, behavior. An important role of technology is therefore to augment people’s capabilities in dealing with multiple sources of data in order to generate information as a basis for better operations or decisions (Kettinger and Li, 2010). Following this interpretation, in order to allow individuals to generate value from data, the data need to be transformed to information that can act as the catalyst for better decisions making, behavioral change and the ability to act quicker, more accurately and with optimal use of resources.

## Value Generating Mechanisms

Data only contribute to value generation when they are being used. Data buried within closed databases do not produce any value as such, and the possibilities of value generation with the level of openness and use of data across networks of stakeholders. As discussed in chapter 2.1, one of the unique features of digital data is that they are non-rivalrous. This means that unlike many of our world’s depletable resources, data can be used over and over again without negative consequences. Open data do therefore not fall victim to the „tragedy of the commons“ (Hess and Ostrom, 2006) and hence offer an unprecedented *opportunity* for value generation. Policymakers around the world have realized this and are now actively promoting the use of open data (see for instance Kundra, 2012 and EU, 2011b). However, in order to maximize social gain from data, individuals need to have, not only the opportunity, but also the *motivation* and *ability* to undertake the actions necessary to transform data to value. While open data initiatives certainly present individuals with an opportunity, without motivation or ability to use the data for value generation, this option might be of marginal value.

Why are data valuable? In his recent Ted talk, Kenneth Cukier offers a simple answer: You have more information. You can do things you couldn’t do before (Cukier, 2014). New value happens when open data are used to generate new information, support process improvements or as a resource in new products or services.

“The one mechanism that economists relate most of their analysis to – their master mechanism, so to speak – is the market.” (Hedstrom and Swedberg, 1998, p. 3). Process improvements and development of new products and services usually make use of the market mechanism, referring to the use of monetary exchange between buyers and sellers within an open and understood system of value and time trade-offs to produce the best distribution of goods and services. The market mechanism is designed to create a win-win situation between two parties, the buyer and the seller.

This mechanism is therefore essentially bilateral. The “object” that facilitates the transaction between two partners is currency, or the price of the good/service in question. We can assume that the transaction generates value for both parties, as they are free to enter the transaction and would only do so if they perceive that they will gain some kind of value from the exchange. The value generated for the buyer is to fulfill his or hers specific need while the value generated for the seller is the monetary value from the amount paid.

However, the shifts toward an economy centered on information and the move to a networked Internet-based environment have caused significant attenuation of the limitations that market-based production places on the pursuit of value (Benkler, 2006). Much of the value that is created through use of the internet is not supported by the traditional market mechanism. Another way of generating value from open data is through what we define as the *information sharing mechanism which creates a mutual gain over a network of stakeholders through the simultaneous creation, dissemination and appropriation of information*. In a way, it is the interaction itself that generates value (Bowman, 2015; Viscusi et al., 2014). The more people that participate in the creation and open sharing of information, the more value is created, as a result of reduced information asymmetries and more opportunities for knowledge generation. In short, more information results in increasing “conditional readiness for goal-directed activities” (Kettinger and Li, 2010, p. 415).

## **VII.5 Research Framework and Hypotheses Development**

Our research model is shown in Figure 1. This model is extended from prior models of Jetzek et al. (2013a; 2013b; 2014a; 2014b) and portrays a nomological network that consists of two underlying but unobservable value generation mechanisms. The first underlying mechanism is the information sharing mechanism which explains how value can be created through informing people (increasing transparency) and preparing for improved decision making and action (increasing civic engagement). While there are not financial transactions taking place, more availability of open and readily consumed information will reduce information asymmetries and might, for instance, enlighten and inspire citizens to make better decisions for instance regarding their health or the sustainability of the environment. We propose that we can find evidence for how active this mechanism is by looking for indicators that reflect the dissemination of *shared digital content*.

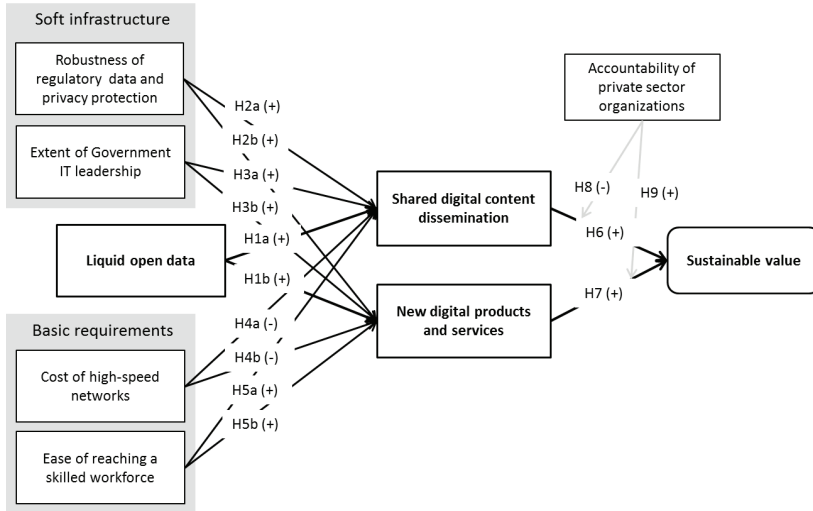


Figure 1: Research model

The second underlying but unobservable mechanism is the market mechanism that facilitates the distribution of digital products, services and processes based on open data. The market mechanism is already well known and helps new innovations infiltrate the markets and can lead to the generation of sustainable value consisting of economic, social and environmental dimensions. We propose that in the context of data-driven innovation we can find evidence for how active this mechanism is by looking for indicators that reflect the creation of new digital products and services.

We hypothesize that: (1) by opening government data, public authorities will create an opportunity for value generation in society and thus stimulate value generation through the generation of shared digital content and commercial digital products and services, facilitated by the underlying information sharing and market mechanisms. (2) The extent of new information, products and services is enabled or prohibited by the context in which value generation happens, which impacts not only the opportunity but also the motivation and ability to generate value from data. This context consists of the soft infrastructure that reflects institutional factors and governance, and the basic structures or requirements that are needed for the open data ecosystem to function properly.

## **Liquid Open Data**

We have previously discussed how publishing open data creates an opportunity for value generation through new use of data, and defined the construct liquid open data. The real estate information platform company Zillow can now be used as a real-life example to explain how this happens. Zillow is a notable example of technology startup flourishing on the availability of open data (Capgemini, 2013). Zillow's mission is to empower consumers with information and tools to make smart decisions about homes, real estate and mortgages. The company has established a successful business by creating a database built from a range of linked sources such as tax data; listing of available homes for sale or rent; mortgage information; geographic data and data on local land value and housing prices. Accordingly, open government data are one of Zillow's most valuable resources. The company offers homebuyers free information (shared content) on available properties through their website. Zillow has also built a variety of digital products and services on top of this database. One of Zillow's main offerings is their so-called "Zestimates", an advanced statistical predictive tool providing up-to-date information on home values and rental prices (Capgemini, 2013). As Zillow uses open data to create both shared content and commercial products and services, both information sharing and market mechanisms are activated in order to generate sustainable value from liquid open data.

*H1a: More openness and liquidity of data positively influences stakeholder's opportunity to share digital content*

*H1b: More openness and liquidity of data positively influences stakeholder's opportunity to develop and market new digital products and services.*

## **Robust Regulatory Data and Privacy Protection Frameworks**

Data is the new currency of the digital economy, where instead of market transactions in which partners exchange currency, people get access to free information services, often in exchange for their own data (Bean & Koeppel, 2013). We can define personal data as any data that can be attributed to an identifiable person either directly or indirectly (Beardsley et al., 2014). By definition, open government data do not include personal data. Anonymization methods are in many cases used to remove any information that could be traced back to individuals and put them at risk. Even so, many worry that open data could be traced to individuals, for instance through a combination of datasets. The protection of personal data has long been viewed as a fundamental personal right. It is important that citizens feel safe about the protection of their personal data and privacy. Governments and regulators will thus need to frame

data protection policies that safeguard the privacy of citizens (Beardsley et al., 2014). However, policymakers must also identify the appropriate balance between protecting the privacy of individuals' data and allowing for innovation in service delivery and product development (Pepper and Garrity, 2014).

The free flow of any fiat currency depends on trust. It is necessary to ensure that the same norms, principles and values that underpin physical markets are employed online, instilling confidence that publication and use of liquid open data will not result in personal data being unethically or unlawfully used. A 2011 survey by the European Union revealed that 92% of Europeans were concerned about mobile apps collecting their data without consent. 89% of those surveyed said they wanted to know when the data on their smartphone was being shared with a third party, asking for the option to give or refuse permission (European Commission, 2011a). EU citizens were also concerned about the risks linked to cyber-security. 74% of respondents agreed that the risk of becoming a victim of cybercrime had gone up (ibid). The key security issues for use of data include the reliable prevention of hacking and access by unauthorized and unwanted users to large databases and data flows. In order to ensure a healthy ecosystem where users, consumers and businesses feel safe in engaging in data-driven activities, protection against cybercrime becomes essential. The trust in legislative ability to protect people's data is likely to affect positive sentiment towards the dissemination and use of liquid open data and thus influence the motivation for engaging with open data for value generation.

We propose that a robust regulatory data protection framework will positively influence the continuing trend towards creation and dissemination of shared digital content and the creation of digital products and services, both of which rely to an increasing degree on use of personal data.

*H2a: Robust regulatory data and privacy protection frameworks increase trust, reduce perceived risks related to information sharing, and therefore positively influence stakeholder's motivation to share digital content.*

*H2b: Robust regulatory data and privacy protection frameworks increase trust, reduce perceived risks related to commercial use of data, and therefore positively influence stakeholder's motivation to develop and market new digital products and services.*

## **Digital Leadership of Government**

The relationships between where value is added and where value is captured, and by whom, are becoming increasingly complex in the context of shared resources used by a network of stakeholders (Bowman & Ambrosini, 2000; Bowman, 2014). Networks

have been shown to support value co-creation due their flexibility and adaptability (Morgan et al., 2013). However, networks present major challenges in terms of governance (Benington et al. 2011; Morgan et al., 2013). For example, it can be challenging to sustain clear strategic direction and appease multiple stakeholders, with different mandates (see discussion in Paper VI). In the context of open data, current research is revealing that ODIs, where public bodies have strong ties to the eventual data-users, have been most successful (van Veenstra and van den Broek, 2014a), calling for a more structured network governance (Eckartz et al., 2014). Moreover, innovative activities intended to generate sustainable value should be supported by an innovation ecosystem, which the public sector should establish in order to increase use of their data (Bason, 2010). In general, governments should foster an environment supportive of innovation and an ICT-friendly government policy (Bilbao-Osorio et al., 2014). Moreover, governments should have a clear vision for the internal use of data and support cross-government initiatives (Ubaldi, 2013). Strong government leadership is thus proposed as an important factor for motivating users to engage with open data.

*H3a: Digital leadership of governments positively influences stakeholder's motivation to disseminate shared digital content*

*H3b: Digital leadership of government positively influences stakeholder's motivation to develop and market new digital products and services*

### **Cost of High-Speed Networks**

The generative ability of the mechanisms that facilitate value generation through use of data depends on certain capabilities. A capability can be defined as a measure of the ability of an entity to achieve an objective. Here, these capabilities reflect the ability of individuals and organizations to create and disseminate shared digital content or develop new commercial digital products and services.

An important foundation for the ability to create and disseminate digital content, products and services is having access to high-speed network infrastructure. Improved information infrastructure has vastly increased our ability to store, retrieve, sort, filter and distribute information, thereby greatly enhancing the value of the underlying information itself (Shapiro and Varian, 1999). Due to increasing digitization and exponential growth of data, the need to provide high-speed broadband to all segments of the population has gained importance in recent years (Bilbao-Osorio et al., 2014). Unfortunately, universal connectivity and global access still remain out of reach for many citizens of the world (Avital et al., 2007). Moreover, bandwidth constraints will

increasingly become an obstacle in transmission over existing networks, due to the growing amount of supply and demand for data (Pepper and Garrity, 2014). We use a measure of the relative cost of internet bandwidth as a proxy for the general affordability and availability of network infrastructure (see table 7, appendix B for detailed measures). The higher the relative cost of using high-speed internet, the less general affordability. I propose high costs will negatively influence the ability of individuals and organizations to create and disseminate shared digital content or develop new digital products and services.

*H4a: Cost of high-speed networks negatively influences stakeholder's ability to share digital content*

*H4b: Cost of high-speed networks negatively influences stakeholder's ability to develop and market new digital products and services.*

### **Ease of Reaching a Skilled Workforce**

Another important construct that affects the ability of individuals and organizations to create and share digital content or develop new commercial digital products and services is the availability of skilled workers. The constant upgrading of workers' skills has never been more important for continuing competitiveness of our society (Bilbao-Osorio et al., 2014). If the education system does not equip students with the necessary expertise to stimulate innovation, both in and outside of organizational contexts, the capacity for innovation will suffer, both in regards to general civic engagement and product and service development. In fact, the demand for people with the deep analytical skills relevant for big data analytics might outstrip current projections of supply by 50 to 60 percent (Brown et al., 2011). We propose that societies which can easily access a workforce offering relevant technical skills, also will have more ability to generate value through creation and sharing of digital content or the development of new digital products and services.

*H5a: Ease of reaching a skilled workforce positively influences stakeholder's ability to create shared digital content*

*H5b: Ease of reaching a skilled workforce positively influences stakeholder's ability to develop and market new digital products and services.*

### **Shared Digital Content Dissemination**

Timely access and availability of previously disparate data sources can positively contribute to new knowledge, insight and awareness, ultimately creating a society equipped to make information-driven decisions. Studies show that companies who



adopt data-driven decision making are on average 5-6% more productive than their intuition-driven counterparts (Brynjolfsson et al., 2011). The energy-tech company Opower is an example of the generative power of information dissemination. Opower is a double-bottom-line company, motivated by a corporate mission to drive global energy savings. Opower merges and analyzes various types of data, including open government data, to generate reports for individual energy users outlining personalized insights on their energy use. The reports are freely available to consumers through multiple channels. The energy reports include a comparison of the individual's energy consumption with that of their "neighbors" – that is a calculated average of users with a similar profile, based on for example demographic information, house types and location based weather. Moreover, to motivate low-energy users to maintain their limited consumption, they are rewarded for their good behavior with a symbolic "smiley token" (Allcott, 2011). In May 2015, the Opower home energy reports had collectively enabled a global savings of over 7.7 terawatt hours of electricity, 12 billion pounds of CO2 emissions and a savings of approximately \$1 billion on individual energy bills. This example of people's engagement with contextualized information demonstrates how such engagement can lead to a collective generation of sustainable value.

In the not so distant past, dissemination of information happened mostly on paper and made use of the market mechanisms via channels like subscription newspapers and bookstores. At present, however, this dissemination occurs primarily through the creation of freely shared digital content and is driven by information sharing mechanisms that connect immensely diverse people on a global scale. The information sharing mechanisms are enabled by the interactive, collective intelligence of today's mobile and web-based information technologies, such as mobile apps and social media. The key components of these mechanisms are not based on financial remuneration, but on user involvement, accreditation, and tools that promote collaboration between individuals (Benkler, 2006). Value generation and appropriation through information sharing mechanisms in the context of open data emerge as a constant interaction between governments, citizens and companies, where continuous exchange of data and information is the key facilitator. For the purpose of this dissertation, it is assumed that people in general wish to make choices that benefit society, whilst safeguarding their own wellbeing. Furthermore, it is assumed that providing people with more information gives them the means to make choices that are more reflective of said wishes. Therefore, we propose that the creation and dissemination of shared digital content containing valuable information will positively influence the generation of sustainable value.

*H6: Sharing digital content positively influences the generation of sustainable value.*

### **New Digital Product and Services**

When previously unavailable data-sources are released to the public, new stakeholders are able to employ them, possibly generating innovation and an opportunity for new digital product and service creation. For example, the availability of open GPS data in the US in the 1980s stimulated the creation of many new businesses specializing in GPS and mapping services. Since then, so-called geo-services have become increasingly widespread, and are estimated to contribute annually between \$150 and \$270 billion in monetary value to global markets (Oxera, 2013). Additionally, open data from the US National Weather Service supports a private weather industry worth over \$1.5 billion per year (Kundra, 2012).

As an example of how product and service innovation through open data contributes to sustainable value, we can refer to INRIX, which is a leading provider of traffic services. INRIX's mission is to globally reduce congestion caused by traffic and in order to do so they have utilized crowdsourcing as a data collection method. Their traffic intelligence platform combines and analyzes data from public and private sources, including government road sensors, official accident and incident reports and data on real-time traffic speed, sourced from a large community of local drivers that have either downloaded an application on their smartphones, or provided information through their in-car GPS device. INRIX has developed a variety of products and services based on these data. They sell both enriched data and data-driven products and services to stakeholders such as automobile manufacturers and public traffic engineers.

While business models of many companies are transforming as they increasingly make use of information sharing mechanisms, most innovative companies still depend on the market mechanism to create revenue. These market mechanisms are driven by the forces of demand and supply, and can be used to identify business opportunities and promote new digital products, services or processes that offer the generation of sustainable value. In the specific case of data-driven innovation products that route traffic and reduce congestion, direct and indirect impacts on sustainable value can happen through: a) a reduction of unproductive time spent by commuters; b) a greener and more sustainable environment; and c) reduction in stress and pollution related health-issues. Naturally, other innovations offer different contributions to sustainable value. In general, we propose that the creation of new digital products and services will influence the generation of sustainable value in a positive way.

*H7: New digital products and services positively influence sustainable value.*

## **Accountability of Private Sector Organizations**

We are currently facing numerous and urgent economic, social, environmental societal challenges. Given the complexity and cross-boundary nature of these challenges, a new approach is necessary (Jetzek et al., 2014b; OECD, 2011). The silos of government departments are in many cases poorly equipped to tackle challenges that span sectors and nation states. Civil society often lacks the capital, skills and resources to take promising ideas to scale. In section 1.4, I discussed various concepts of value and how the dichotomy between the state and the markets was being increasingly contested (Porter and Kramer, 2011). Following Porter and Kramer (2011), I suggest that there is a role for private companies in solving these challenges and producing sustainable value.

The concepts of market failures and negative externalities have for a long while shaped company strategies that have had a tendency to exclude social and environmental considerations from their economic thinking. Thus, firms have traditionally resisted regulatory standards as being contrary to their interests (Porter and Kramer, 2011). However, continued advances in digital technologies have disrupted the status quo and altered the way many people think, live and work, rearranging value pools (McKinsey, 2013b). Governments seeking to provide social protection to their citizens, while strengthening national economic competitiveness are becoming more interested in cross-sector co-creation of value. The evolving sense of corporate social responsibility is portrayed in the number of companies that participate in collective efforts to develop international standards that go well beyond what is required by international law (Haufler, 2013). Increasingly, companies are adopting policies that address urgent social issues (ibid). However, if companies are to play a bigger role in terms of social responsibility, they must consistently perform in the interests of shareholders and wider society. To do this companies will need to make visible the precise ways in which they are contributing to sustainable value generation, allowing shareholders, government officials and the greater population to hold them accountable.

In previous sections, I have outlined a number of use cases where private companies have utilized open data to create both shared content and commercial products and services that contribute to sustainable value for society. The motives of these companies are altruistic, while simultaneously remaining economically grounded. Thus, these businesses have reconnected company success with social progress. Accordingly, I propose a moderation perspective, implying that the relationship between content dissemination and sustainable value on the one hand, and new digital products and services and sustainable value on the other hand depends on the level of

private sector accountability. However, the moderating effect of accountability of private sector organizations differs between the two proposed mechanisms.

Firstly, we propose that while dissemination of shared content is positively related to sustainable value, this effect will be more prominent in countries exhibiting less accountability of private institutions. When citizens in these countries uncover various sources of information, for instance regarding worker's rights or environmental protection, they are able to put social pressure on the corrupt or unaccountable companies. This pressure, we argue, will have an impact on the behavior of companies and positively influence the generation of sustainable value. The impact of shared content dissemination on sustainable value is, using a similar logic, suggested to be less pronounced in countries where companies already adhere to various standards and regulations, as those companies are already held accountable by shareholders and other stakeholders. To rephrase more clearly: the positive impact of increased content sharing on sustainable value is stronger in countries with less accountable private institutions.

*H8: The positive effect of shared digital content on sustainable value decreases with increased accountability of private sector institutions*

Secondly, we argue that while the generation of commercial digital products and services is positively related to sustainable value, this effect will be more prominent in countries with more accountable private sectors. When companies in general adhere to standards and regulations that hold them accountable for their actions, they are more likely to conform to social responsibility. Thus, the positive impact of new digital products and services on sustainable value is stronger in countries with more accountable private institutions.

*H9: The positive effect of commercial digital products and services on sustainable value increases with increased accountability of private sector institutions*

## **VII.6 Research Design**

The conceptual model presented in chapter 5 is based on our observation of individual open data initiatives, interviews with producers and users of open data, and a review of the popular press as well as the state-of-the-art research. However, we faced a number of barriers when confronted with the task of designing an empirical study where we would like to statistically estimate whether or not *openness* of data is a relevant enabling factor for societies aiming to stimulate value generation from data. The first barrier relates to the extreme difficulties that arise if one wants to trace the value that

governments, companies and individuals capture from using these data. Obviously we cannot know how societies would have fared without access to these data and experimentation is not an option in this context. Moreover, given the nature of open access, it is very difficult to identify all the users and usage of multiple sources of data. We can, however, compare between countries that are in different stages of opening data, therefore a correlational approach seemed sensible. The benefit of performing cross-level analysis is that it does not require us to ignore the fact that in some cases, data have been open for many years, long before the mainstream open data initiatives started. We simply look at the state of each variable in each country in the year of 2013 with the aim of uncovering whether those in the more advanced stages of open data will systematically show superior performance, conceptualized as sustainable value.

The next barrier was to figure out how to reflect the level of sustainable value that has been generated in different countries. A number of macro level analyses have attempted to estimate the overall economic value of open data (see for instance de Vries et al., 2011, Houghton, 2010, McKinsey, 2013; Vickery, 2011); however none of those analyses have attempted to capture the intangible or social dimensions of value, which are recognized to be of even more importance than the economic value (McKinsey, 2013). While most open data initiatives have highlighted the economic potential, most of them actually place more emphasis on intangible concepts like transparency and collaboration, with the underlying motive of increasing social value. Moreover, consultancies like McKinsey have highlighted the importance of various less tangible benefits from open data to consumers, like being able to make better decisions regarding where to go to school or how to get around (McKinsey, 2013). This demands that we create constructs that can reflect such unobservable phenomena as the value of good education. The ability to operationalize theoretical but unobservable phenomena has long been seen as core strength of the SEM methods. Therefore we chose to use structural equation modelling where we use different indicators to reflect the construct of sustainable value, building on earlier work like that of Stiglitz et al. (2009).

We have theorized about how open data could potentially generate value by utilizing both network and market mechanisms. We also highlight the importance of the context in which these mechanisms operate, reflecting the societal infrastructure, by including constructs such as the cost of network infrastructure, robustness of the data protection regulatory framework, availability of skills and political leadership. Because of their contextual nature, the basic structure of mechanisms is often described as a context-mechanism-outcome pattern. This calls for the use of a structural model rather than a

reduced form, and consequently, statistical methods that can estimate such models. Next, we needed to look for available data. As the context of our research is society wide use and exploitation of open data, it was a natural conclusion to go either for a multi-level model or a societal level model. As it was not feasible to collect data with surveys, given the scope and variety of open government data initiatives, we opted to collect secondary macro-level data. The first available data showing the status of open data in various countries was from 2012, but there was no available data at the organizational level for this period or later, eliminating the possibility to estimate a multi-level model.

It is not easy to measure many of the concepts we commonly discuss in relation to socio-technical systems. We opted to look for measures in available *open data* sources that could be used as proxies for the constructs in our research model. The sample chosen is a convenience sample of the 76 countries that were included in the Open data barometer 2013, published by the Open Data Institute (Davies, 2013). Table 2 provides an overview over construct definitions and origins of data, but table 8 in Appendix A shows in addition individual item measurements.

**Table 2: Construct definitions and data sources**

Construct	Definition	Source
Liquid open data	Reflects the data that are available online without technical restrictions to access,	Open Data Barometer <a href="http://www.opendataresearch.org/barometer">http://www.opendataresearch.org/barometer</a>
Robust regulatory data and privacy protection frameworks	The robustness of existing regulatory frameworks that promote privacy and data protection	Web Index <a href="http://thewebindex.org/">http://thewebindex.org/</a>
Digital leadership of government	Reflects the extent to which government is promoting ICT use	World Economic Forum’s Global Information Technology Report <a href="http://www.weforum.org/reports/global-information-technology-report-2013/">http://www.weforum.org/reports/global-information-technology-report-2013/</a>
Cost of high-speed networks	Reflects the how expensive it is to get high-speed network access for most of the	ITU - International Telecommunication Union <a href="http://www.itu.int">http://www.itu.int</a>

Construct	Definition	Source
	population	
Ease of reaching a skilled workforce	A construct measuring the availability of skilled workers in a country	WEF - World Economic Forum: Global Competitiveness Report <a href="http://www.weforum.org/reports/global-competitiveness-report-2013-2014">http://www.weforum.org/reports/global-competitiveness-report-2013-2014</a>
Shared digital content dissemination	Reflects the level of informative digital content that is freely shared online by various stakeholders	Web Index <a href="http://thewebindex.org/">http://thewebindex.org/</a>
New digital products and services	Reflects the extent to which ICT influences the creation of innovative business models, products and services and processes	WEF - World Economic Forum: Global Information Technology Report <a href="http://www.weforum.org/reports/global-information-technology-report-2013/">http://www.weforum.org/reports/global-information-technology-report-2013/</a>
Private sector accountability	Reflects the extent to which private business follow private and public rules and regulations that conduct to accountability and good practices	WEF - World Economic Forum: Global Competitiveness Report <a href="http://www.weforum.org/reports/global-competitiveness-report-2013-2014">http://www.weforum.org/reports/global-competitiveness-report-2013-2014</a>
Sustainable value	Reflects the degree of economic, social and environmental wellbeing in society.	World Bank, <a href="http://data.worldbank.org/">http://data.worldbank.org/</a> United Nations: Human Development Index <a href="http://hdr.undp.org/en/content/human-development-index-hdi">http://hdr.undp.org/en/content/human-development-index-hdi</a> Yale Center for Environmental Law & Policy (YCELP) and the Center for International Earth Science Information Network (CIESIN) at Columbia University <a href="http://epi.yale.edu/">http://epi.yale.edu/</a>

While not a random sample, our sample represents a variety of different countries in various regions of the world, as shown in table 3.

**Table 3: Regions and countries in sample**

Africa	18
Asia	13
Central America	2
Europe	20
Middle east	10
North America	2
Oceania	2
South America	9
<b>Total</b>	<b>76</b>

## VII.7 Data Analysis and Results

### Measurement Model

The method chosen was partial least squares (PLS) for various reasons. Firstly, as discussed in chapter 6.1 we wish to test a structural model, rather than the reduced version where we would directly or stepwise link use of open data and sustainable value, using some form of linear regression. While a structural model contains formulas representing the relation of every dependent variable to its independent variables, the reduced form exhibits the net or overall relation between the dependent variable and the ultimate independent variable (Tsang, 2006). The reduced model is simpler but also shallower than the structural model, as the reduced model does not include the mediating variables (ibid). Moreover, the reduced model can be derived from the structural model but not the other way round, because the task of working out the structural model from the reduced model is an inverse problem with an indefinite number of solutions. Finally, behavioral assumptions are often eliminated in the process of converting a structural model into its reduced form. Consequently, the reduced model might be less informative with respect to the related mechanism than the structural model (Tsang, 2006).

Secondly, there are various technical reasons that led us to use PLS-SEM rather than CB-SEM. PLS is the primary choice for analyzing secondary data, as it does not



emphasize model fit, as defined in the context of CB-SEM (Henseler & Sarstedt, 2013). Moreover our research is largely exploratory due to the emergent state of the open data phenomenon and the research goal is to identify the most important antecedents of the target constructs (Hair et al., 2011; Ringle et al., 2012; Sarstedt et al., 2014). Furthermore, we are using one formative variable and are estimating both mediating and moderating relationships, indicating high model complexity (Sarstedt et al., 2014). Finally, we have to work with a small set of cross-country data, including data that are not normally distributed. Since PLS is based on a series of OLS regressions, it has minimum demands regarding sample size, and generally achieves high levels of statistical power (Hair et al., 2011). A common rule of thumb regarding sample size in PLS is to use ten observations per predictor, where the sample size is the larger of two possibilities: 1) the block with the largest number of indicators or 2) the dependent variable with the largest number of independent variables impacting it (Chin, 1998; Hair et al., 2011; Marcoulides and Saunders, 2006). In our case, constructs are made from a maximum of five indicators, and the mediating variables each have five direct paths pointing towards it.

However, as Marcoulides and Saunders (2006) point out, it is necessary to consider other characteristics of the data and model in order to ensure sufficient sample size to achieve adequate statistical power. First, we built the research model according to current knowledge, and afterwards collected data to test the model. Next, we performed data screening. All sources had a good reputation, and the same methodology was applied to all countries for each indicator. Missing data or departures from normality influence sample size requirements of a study and potentially deteriorate power (Marcoulides and Saunders, 2006). There were no missing data and all rows showed a reasonable degree of normality (kurtosis  $\leq |1.15|$ , skewness  $\leq |0.70|$ ) except for the following measures: 1) Open data licenses and bulk download both exhibited a showed some of kurtosis (3,15 and 6,58 respectively) and skewness (2,07 and 2,27 respectively), indicating a long tail with the bulk of the measures clustering in the lower end of the tail. 2) Both cost measures had very high kurtosis (15,52 and 68,37 respectively) and skewness (3,63 and 8, 12 respectively) indicating the same phenomenon but more pronounced. As the PLS method does not require all indicators to be normally distributed, we kept these four measure. Table 9 in Appendix B shows averages and other descriptive statistics for all indicators. We conclude that a sample size of 76 gives adequate power to draw inferences for this particular model, as both factor inter correlations and factor loadings are high (see table 4 and table 10 in Appendix C).

We used the SmartPLS 3.2 software to estimate and evaluate the path model, using the path weighting scheme. Recommendations by Hair et al. (2011; 2012) and Sarstedt et al. (2014) provide guidance to evaluate the PLS estimates for the overall model. As shown in table 3, the  $R^2$  for all endogenous constructs are substantial or close to that (Hair et al., 2011). In addition to the evaluation of  $R^2$  values, the predictive relevance of the model is assessed through blindfolding procedures to obtain cross-validity redundancy measures for each construct (Chin, 1998). The results suggest good predictive relevance of the model, as all  $Q^2$  are well above zero (Geisser, 1975; Stone, 1974; Hair et al., 2011). Values calculated for omission distance  $d=7$  are shown in table 4.

**Table 4: All variables**

Construct	Type	Measure	#	$R^2$	$Q^2$
Open liquid data	Exogenous	Reflective	5		
Robust regulatory data and privacy protection frameworks	Exogenous	Formative	3		
Digital leadership of government	Exogenous	Reflective	3		
Cost of high-speed networks	Exogenous	Reflective	2		
Ease of reaching a skilled workforce	Exogenous	Reflective	3		
Shared digital content dissemination	Endogenous	Reflective	2	0.743	0.805
New digital products and services	Endogenous	Reflective	2	0.843	0.673
Sustainable value	Endogenous	Reflective	4	0.775	0.665

Item reliability is deemed adequate with most reflective factor loadings exhibiting high values of well above .8, and the smallest loading of .793 still well above the commonly suggested threshold value of .7 (Hair et al., 2011). Average variance extracted (AVE) of all reflective measures is clearly above the recommended level of .5 (Fornell and Larcker, 1981), confirming satisfactory convergent validity. Composite reliability of reflective measures is good with values between .896 and .972 (see Nunnally and Bernstein, 1994) and internal consistency (Cronbach’s Alpha) is in all cases above the recommended threshold of 0.7 (Hair et al., 2011). See table 5 for individual values.

**Table 5: Quality criteria of measures; outer loadings, composite reliability, convergent validity**

<b>Variable</b>	<b>Indicators</b>	<b>Factor loading</b>	<b>AVE</b>	<b>Comp. rel.</b>	<b>Chron. Alpha</b>
Liquid open data	Reusable	0.836	0.814	0.956	0.941
	Affordable	0.928			
	Discoverable	0.970			
	Usable	0.793			
	Accessible	0.968			
Digital leadership of government	Government success in ICT promotion	0.957	0.906	0.967	0.948
	Importance of ICTs to government vision of the future	0.929			
	Government procurement of advanced technology products	0.969			
Cost of high-speed networks	Cost of fixed broadband per capita income	0.848	0.812	0.896	0.782
	Cost of bandwidth pr MB	0.951			
Ease of reaching a skilled workforce	Training of workforce	0.904	0.843	0.915	0.815
	Retaining skilled workforce	0.921			
	Attracting skilled workforce	0.908			
Shared digital content dissemination	Availability of educational content on the web	0.971	0.945	0.972	0.943
	Availability of government provided content on the web	0.974			
New digital products and	ICT enabled new services and products	0.988	0.975	0.987	0.975

Variable	Indicators	Factor loading	AVE	Comp. rel.	Chron. Alpha
services	ICT enabled new organizational forms	0.987			
Sustainable value	GDP pr. capita	0.912	0.864	0.962	0.948
	Life expectancy at birth	0.941			
	Exp. years of schooling	0.938			
	Environment performance index	0.928			

Discriminant validity of all reflective measures is established as the square root of each endogenous construct's AVE being greater than the variance shared by each construct and its opposing construct (Hair et al., 2012, 2011; Henseler et al., 2009), see Table 6 and all measures highest loadings being on their own construct, see table 10 in Appendix C.

**Table 6: Quality criteria of measures; discriminant validity (Fornell-Larker criteria)**

	OP	LS	CN	AS	CO	PR	CSR	Value
<b>OP</b>	0.902							
<b>LS</b>	0.638	0.952						
<b>CN</b>	-0.302	-0.007	0.901					
<b>AS</b>	0.533	0.720	-0.211	0.918				
<b>CO</b>	0.774	0.238	-0.479	0.503	0.973			
<b>PR</b>	0.638	0.745	-0.307	0.805	0.650	0.988		
<b>CSR</b>	0.511	0.578	-0.211	0.761	0.550	0.705	1.000	
<b>Value</b>	0.729	0.294	-0.514	0.592	0.809	0.607	0.577	0.930

Quality criteria for the formative variable is shown in table 7. One of the concerns with formatively measured constructs is multicollinearity across the indicators of each constructs. High first eigenvalues can be an indicator of multicollinearity; however, all

formative variable’s first eigenvalues are lower than three. All Variance Inflation Factors (VIFs) were below the recommended 5.00 value (Hair et al., 2011). We checked for insignificant or negative weights (Centefelli and Bassellier, 2009; Petter et al., 2007). While all weights were positive and of reasonable size (larger than 0.1), the weight for extent of cybercrime legal protection framework was not significant. However we decided to leave the indicator as removing it did not change any structural relationships in the model.

**Table 7: Quality criteria for formative variables \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$**

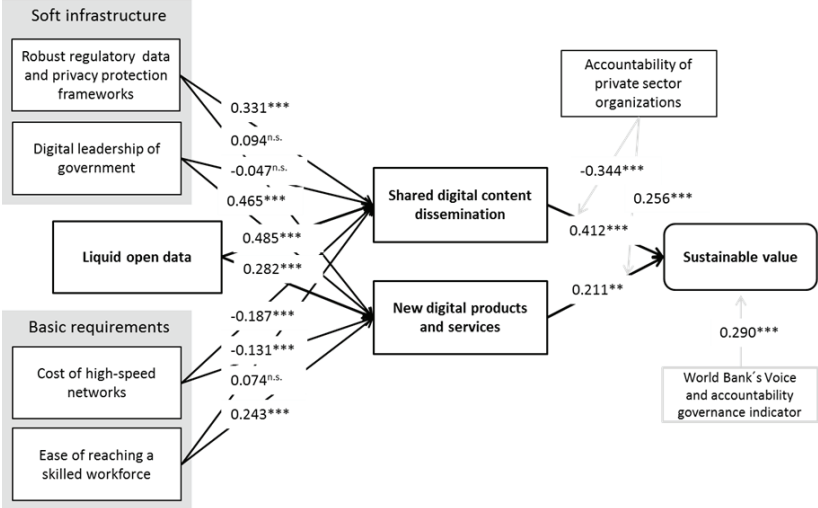
Variable	Indicators	Factor loading	Weight	t-value	VIFs
Robust data and privacy protection frameworks	Cybercrime legal protection framework	0.748	0.234	1.313 <sup>n.s</sup>	1.747
	Enforcement and protection from cybercrime	0.858	0.594	4.379* **	1.546
	Personal data protection framework	0.826	0.375	2.953* **	1.466

The SRMR Composite factor of the total model was 0.07 and the SRMR Common factor was 0.078. Ideally, for a model that fits the data, the SRMR would be “close to” 0.09 (or lower; Hu & Bentler 1999, p.27) so the model fit is good.

### Structural Model

Figure 2 provides the results of the structural model. All measures were standardized before running the algorithms. We ran the “traditional” PLS algorithm with the path weighting scheme as recommended in Smartpls 3.2. We also ran the consistent PLS algorithm, PLSc. While maintaining all strengths of PLS, the consistent version provides two key improvements. 1) Path coefficients, parameters of simultaneous equations, construct correlations, and indicator loadings are estimated consistently; 2) And the global goodness of fit of the structural model can now be assessed, which makes PLS suitable for confirmatory research (Dijkstra and Henseler, 2012). Moreover, the outcome of a Monte Carlo simulation reveals that the bias of PLSc parameter estimates is comparable to that of covariance-based structural equation modeling and shows that PLSc has advantages when using non-normally distributed

data (Dijkstra and Henseler, 2015). There were no significant differences in the parameters, although PLSc provided a slightly higher estimation for most paths. The results in figure 2 are from the PLS algorithm but table 11 in Appendix D shows differences in path estimates.



\*p<0.1, \*\*p<0.05, \*\*\*p<0.01

Figure 2: Research model results

To test for significance of relationships between latent constructs, t-values were calculated for path coefficients. We used bootstrap validation for loadings, weights and paths with 5000 bootstrap samples (Hair et al., 2011). As recommended in Smartpls 3.2, we chose the no sign changes option which is the most conservative estimation method, therefore resulting in larger standard errors and, consequently, lower t-values. There are three insignificant paths in the model, as shown in figure 4, thus we cannot support all of our previous hypotheses.

We can support hypothesis H1a as there is a significant relationship between the extent to which data are open and liquid in a country and the extent of shared digital online content dissemination. Moreover, we can also support hypothesis H6 that there is significant relationship between the availability of shared digital content and the level of sustainable value in a country. We can therefore support our hypothesis of a significant relationship between liquid open data and sustainable value, mediated through creation of digital content. This path reflects the causal relationship we have

conceptualized as the *information sharing mechanism*. Our hypothesis is that countries that seem to have a more open attitude towards sharing data and information will also be countries that are enjoying high levels of sustainable value. We cannot use the empirical data to conclude on causality as we are using cross-section data and methods, but reaffirming this as a causal chain is suggested for future research

We can also support hypothesis H1b, as we find a significant relationship between liquid open data and the creation of digital products and services. However, this relationship is weaker than the relationship between liquid open data and digital content sharing. We can also support hypothesis H7 regarding the influence of digital products and service innovation on sustainable value. It is, however, surprising that this effect is also much weaker than the similar relationship between content sharing and sustainable value. Accordingly, we can conclude that the causal relationship between liquid open data and sustainable value facilitated via the markets is weaker than the relationship facilitated via the network based information sharing. Looking for explanations, the Open Data Barometer report (Davies, 2013) has shown that although OGD policies have spread fast, the availability of truly open data remains low, with less than 7% of the datasets surveyed published both in bulk machine-readable forms, and under open licenses. There might be a “critical mass” of liquid open data needed to trigger the mechanism that transforms open data into sustainable value through product, service and process innovation. Second, in the same report, the average evidence of impact through entrepreneurial activity was scored at just 1.44 out of 10, showing that such impact is still very limited (this data was not used to test the model).

The indirect effect of liquid open data on sustainable value is 0.312 and significant, fully mediated via the shared digital content and new digital products and services. We checked for partial mediation effects by adding a direct path from liquid open data to sustainable value but the path was insignificant and didn't affect the significance of the two existing paths, indicating full mediation. The Sobel test statistic was 3.08 for the mediation effect of shared digital content (significant at  $p < 0.01$ ) and 1.79 for digital product and service innovation (significant at  $p < 0.05$ ).

We also included interaction effects as we wanted to understand the nature of relationship between the two mediating variables and sustainable value better in the context of private sector accountability, as discussed in sections 3.2 and 5.5. Both hypothesis 8 and hypotheses 9 could be confirmed. The influence of shared digital content on sustainable value is more pronounced in countries where private companies are less accountable, highlighting the need for open data and publicly available information to counter company behavior that might go against the economic, social

and environmental wellbeing of a country. On the other hand, in countries where there are already higher levels of private sector accountability, the influence of digital product and service innovation on sustainable value variable is greater. This indicates that accountable companies are more likely to produce products and services that contribute to sustainable value generation. Moderating relationships are shown in figure 3.

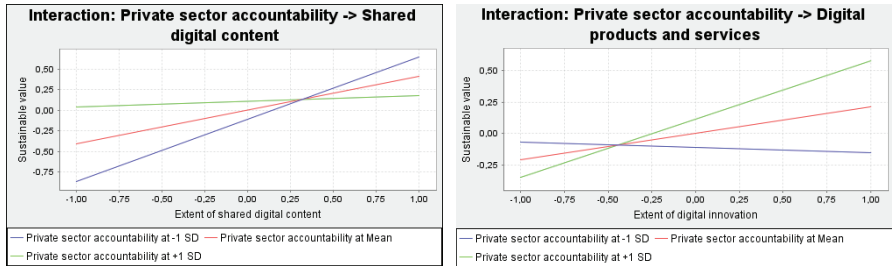


Figure 3: Moderating Relationships

We checked the effect of each of the hypothesized value generating mechanisms on sustainable value by comparing the  $R^2$  of the sustainable value construct with, and without, the variable in question, using Cohen’s  $f^2$  measure, as is shown in table 8.<sup>18</sup> We can conclude that both variables are important for explaining the variance of the sustainable value construct, but the effect of shared content is much higher than the effect of digital products and services, which is surprising. Moreover, both interaction effects are important, but the interaction between private sector accountability and shared digital content dissemination is more relevant than the interaction between private sector accountability and digital product and service innovation.

Table 8: Effect sizes -  $f^2$

Shared digital content $f^2$	Digital products and services $f^2$	Interaction: Accountability / Content $f^2$	Interaction: Accountability / Products $f^2$
0.254 (moderate/high)	0.075 (weak)	0.264 (moderate/high)	0.174 (moderate)

Looking closer at hypotheses H2a and H2b we see that there is a significant relationship between a robust regulatory privacy and data protection framework and

<sup>18</sup> The conventional value for effect size ( $f^2$ ) proposed by Cohen (1988) are 0.02 (weak), 0.15 (medium), and 0.35 (high).



the availability of shared digital content. The relationship between such a framework and digital product and service innovation was on the other hand not significant. It could be that there are two effects at play rather than one, and that those effects negate each other: The hypothesized effect is that entrepreneurs truly value the increased security and trust that a well-developed data protection framework might offer but there might be a second, reverse effect where a high level of data protection actually hinders business use of open data due to reduced possibilities. We experimented with mediators between the data protection construct and the digital innovation construct, one that measures trust and confidence of the business community (WEF Global competitiveness report, 2014) and another that measures burden of government regulation (WEF Global competitiveness report, 2014). While not included in the findings, we can confirm that there do indeed seem to be two contrasting effects, which we propose as an interesting topic for future research.

The digital leadership of government construct exhibits a highly significant and strong relationship with new digital products and services construct. However, the relationship with the shared digital content dissemination construct was not significant. While this finding also requires more in-depth analysis, out of the scope of this paper, it could be that governments that show a very prevalent ICT focus are not necessarily also those that encourage free dissemination of information (or freedom of speech for that matter). This indicates that the public sector must not only be strong in their technology orientation, encouraging the use of new technologies and showing good example with own use and interest, but also stimulate and encourage the private sector to contribute to a better society by freely generating and disseminating information on issues that are important to citizens. Drawing from our previous discussion, such information offers the potential to influence decision making and spur action towards more socially cohesive and sustainable choices and activities

Unsurprisingly, cost of high-speed networks negatively influences both shared digital content creation and digital product and service innovation. Finally, the ease of reaching a highly-skilled workforce does not seem to have a significant relationship with shared digital content creation. Offered as a possible explanation, we found that the content we included in the measure was in many cases not disseminated in a very technically advanced way. This will hopefully change with more use of data analytics and visualization techniques. The relationship between ease of reaching a skilled workforce and digital product and service innovation was moderately strong and highly significant.

## VII.8 Discussion

### Research Implications

The overarching conclusion after having estimated this model is that both the information sharing and market mechanisms contribute to the generation of sustainable value. However, to our surprise the empirically estimated path through information sharing seems to be much stronger, indicating a significant but previously unexplored relationship between open data and value. We can demonstrate how the information sharing mechanism works with the case of climate data. Government-released open climate data is currently fuelling a whole new level of discoveries within the field of environmental sustainability. Today's climate data partnerships are creating unique ventures that cross boundaries between business, government and academia. For instance, IBM offers free supercomputing hours on its World Community Grid for researchers who are studying climate change, utilizing mass quantities of open data. One of the projects making use of the grid, Harvard University's Clean Energy Project, has screened and publicly catalogued 2.3m compounds looking for new materials that could potentially double the efficiency of current carbon-based solar cells. The findings of the Harvard Clean Energy Project Database are open to the public to help further the discovery of new materials. While the value of such initiatives is not easily measured using traditional methods, it is easy to justify that these synergistic relationships and consequent generation of freely available information are contributing to sustainable value through their impact on environmental sustainability, education and health.

Our aim was to uncover the overarching mechanisms that facilitate the generation of value from data and to determine if openness of data really does stimulate generation of sustainable value across countries. Research focusing on this question is still in a nascent state (Zuiderwijk et al., 2014a). As the value generating mechanisms are generally unobservable, our model construction was aided by observation of various open data initiatives, use of prior theory and use of secondary and primary data, used to show how companies and individuals use data to generate value. We also draw on previous research on the generation of (business) value in complex multi-stakeholder relationships. Based on these theoretical perspectives we have identified new constructs that we believe add to the extant discussion, thereby extending the state-of-the art in open data research.

We have stated that we view data as a raw material or a strategic resource, to be used for the creation of information or to be used in new products, services and processes.

While market transactions still play a prominent role in enabling generation and appropriation of value, technological advances are slowly shifting from one-to-one transactions towards many-to-many network relationships with much less clear-cut relationships between value generation and value appropriation. These emerging relationships are discussed in a body of literature that focuses on *two sided markets* and the role of intermediaries for reconciling between the economic profits and the valuable network externalities that arise from information sharing. We have discussed how the intangible information sharing mechanism with its complicated many-to-many relationships between multiple and diverse stakeholders has become more relevant for societies. We have extended the concept of sustainable value which is designed to capture the impact on social welfare that is not easily measured through currency, as a great deal of the sharing of data and information over networks will never (if using current accounting methods) enter company or national accounts. We suggest that in spite of the lack of measures and the subsequent lack of awareness, the value of information is becoming increasingly important to societies.

It is always difficult to justify the choice of a dependent variable, independent of the level of analysis. Questions like do we seek money for the sake of the prestige or power it brings or for the sake of the goods and services it provides are not easily answered. Should it be the goal of government to increase total amount of gross domestic product or is the equality of the income distribution of more concern? Should companies care whether the societies they serve thrive or is that solely the responsibility of government? This paper is not written in order to supply prescriptive policy recommendations, but we do want to highlight the importance of being able to evaluate societal impact of initiatives such as OGD programs. Discussions in various fields of research regarding the nature of value creation and capture are converging to a similar ideology; that we must extend the concept of value which has previously focused on monetized aspects (income, profits, gross domestic product) to include other indicators that might lead individuals and nations to better steer their resource use towards a sustainable and humanistic future (see for instance Benington, 2011; Bowman and Ambrosini, 2010, Grover and Kohli, 2012; Stiglitz et al., 2009). We have attempted to build upon this trend in the area of information technology and data by suggesting the concept of sustainable value as a performance measure.

Our first theoretical contribution is the conceptualization of open data. We have proposed seven dimensions that capture what openness of data entails, reflecting the fact that data are not a homogenous resource, and openness itself is not a binary measure. Based on these dimensions we have used available *open data* to evaluate the

state of open data in various countries. We perceive this as a theoretical contribution, as the concept of open data has suffered from ambiguity in the past, making it harder for policy makers and researchers to analyze its implications (Yu and Robinson, 2012). We hope that our conceptualization will function as a basis for further discussion on open data and thereby move this body of research further.

Our second theoretical contribution is the proposition that openly sharing data (as a resource) and information (as a good) will significantly contribute to the generation of sustainable value, given that individuals; a) in general *want to* make decisions that lead to sustainable value and; b) that the general availability of information can create the necessary knowledge or social cohesion to enable these choices. We do not suggest that the information sharing mechanism is replacing the already well-established market mechanism, but rather that it is adding a valuable component that generates value through the interactions between multiple stakeholders. We suggest a way to compare the level of information sharing and social welfare between countries, using publicly available data that have been collected based upon state-of-the-art research that suggests how to evaluate the level of social, environmental and economic value. By using the PLS method to indicate the individual weights to each of those indicators we have a quantified estimation of the latent constructs of shared digital content and sustainable value. For policy makers, we want to highlight the importance of developing the means to, if not quantify then highlight, the importance of this “new” mechanism. This will be important for creating policies that promote the right social context for continuing generation of sustainable value.

There are several practical implications for public bodies planning to open their datasets for use and reuse. First, while offering government data as an open resource will offer society the opportunity to generate sustainable value, the value generating mechanisms are dependent on other “softer” factors that will influence the ability and motivation of individuals to use data for value generation. If the antecedents that will offer society the right combination of opportunity, ability and motivation to generate sustainable value from data are known, it will become easier to create policies that will actually stimulate those factors.

While we have simplified our previous model (Jetzek et al., 2013b) to highlight the contrast between the mechanism that operates through information sharing effects and the mechanism that uses the laws of the market, we still recognize that there are multiple different pathways that will connect the use of data to the generation of sustainable value. However, these are not easily untangled and it can be equally important to recognize and understand the overarching relationships as being

intimately familiar with the specific details. As Karl Popper said, science may be described as the art of systematic over-simplification. Most of the companies we have reviewed as successful users of data use both the network and market mechanisms, have them support each other in a variety of combinations, and are contributing in very different ways to the multifaceted concept that is sustainable value.

## Limitations and Future Research Direction

Our study is exploratory, mostly due to the embryonic state of research on the impact of open data on society. Accordingly, our aim is to generate new theory, rather than confirming or extending prior theory. This places certain limitations on our research.

First, the sample size is small as there is limited data to be found that are consistent across multiple countries and are fit for the purpose of reflecting different degrees of openness of government data. We were essentially limited by the number of observations in the Open Data Barometer. However, we have reasonable evidence to believe that the model contains enough power to draw preliminary conclusions from results. Second, the Open Data Barometer was constructed for the first time in 2013. As open data is one of the independent variables and other data were not ready for 2014 at the time of analysis, we could not introduce a time-lag into the model which would have improved our ability to draw conclusions on causality. Third, comparing impacts from open data between countries, where in many cases open data initiatives are in their infancy, might be premature. However, we feel that our research model gives a good indication of the current status, while some important questions remain for future research. Moreover, certain types of data have been open for years in a number of countries, which should already have created differences between those countries and the other countries that do not have any history of open data, if a causal chain has indeed manifested. Finally, most of the constructs in the study are new and need to be further validated in future research. Many of the concepts discussed are highly complex, and there has still not been consensus on how to measure most of them, at least at the societal level. Therefore we highlight the need for more scientific discussion on the applicability of many of those constructs.

## VII.9 Conclusion

In conclusion, this study extends our understanding of the importance of open data for societies that want to stimulate the generation of sustainable value by their citizens. It lays the foundation for further study on the interplay between different types of

mechanisms that might bring us closer to understanding how the future might look like, assuming that data, information and networks will continue to play a growing role in our societies. We have focused on a particular aspect of data, namely openness, recognizing the increasing importance of being able to conceptualize and measure the value generation that happens through free interactions between multiple loosely coupled stakeholders operating in vast networks. We hope that by showing the importance and value of these interactions we have created an increased awareness of the value of open data and the importance of sharing and co-creation among individuals, businesses and policymakers. Awareness that might lead us forward on the path towards a more sustainable, humanistic and prosperous future.

## VII References

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(10-11), 1082–1095.

Andrade, P.L., Hemery, J., Recalde, G. and Ryan, P. (2014). From Big Data to Big Social and Economic Opportunities: Which Policies Will Lead to Leveraging Data-Driven Innovation's Potential? In Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014): *The Global information technology report 2014: Rewards & Risks of Big Data*. World Economic Forum, 81-86.

Avgerou, C. (2013). Explaining Trust in IT-Mediated Elections: A Case Study of E-Voting in Brazil. *Journal of the Association for Information Systems*, 14(8), 399-419.

Avital, M., Lyytinen, K. J., King, J. L., Gordon, M. D., Granger-Happ, E., Mason, R. O., and Watson, R. T. (2007). Leveraging information technology to support agents of world benefit. *Communications of the Association for Information Systems*, 19(1), 567-588.

Bason, C. (2010). *Leading Public Sector Innovation: Co-creating for a better society*. Bristol: Policy Press.

Baumol, W. J., & Strom, R. J. (2007). Entrepreneurship & economic growth. *Strategic Entrepreneurship Journal*, 1(3-4), 233-237.

Barney, J. (1991). Firm resources & sustained competitive advantage. *Journal of management*, 17(1), 99-120.

Bean, R and Koeppe, H. (2012). Big Data Analytics: The Currency of the 21st Century Enterprise. Available at <http://www.information-management.com/news/big-data-analytics-the-currency-of-the-21st-centuryenterprise-10023139-1.html>

- Beardsley, S. Enríquez, L. Grijpink, F. Sandoval, S., Spittaels, S. & Strandell-Jansson, M. (2014) Building Trust: The Role of Regulation in Unlocking the Value of Big Data. In Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014): The Global information technology report 2014: Rewards & Risks of Big Data. World Economic Forum, 73-81.
- Benington, J. (2011). From Private Choice to Public Value? In: Benington, J., Moore, M. (eds.) *Public Value: Theory and Practice*, 31–52. Palgrave MacMillan, Basingstoke
- Benkler, Y. (2006). *The wealth of networks: How social production transforms markets & freedom*. Connecticut: Yale University Press.
- Berners-Lee, Tim (2009). The next web. Available: [http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web)
- Bertot, J.C., Jaeger, P.T. & Grimes, J.M. (2010). Using ICTs to create a culture of transparency: Egovernment & social media as openness & anti-corruption tools for societies. *Government Information Quarterly*, 27(2010), 264–271.
- Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2013). *The global information technology report 2013: Growth & Jobs in a Hyperconnected World*. World Economic Forum.
- Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014). *The global information technology report 2014: Rewards & Risks of Big Data*. World Economic Forum.
- Bhaskar, R. (1998). General introduction in *Critical Realism: Essential Readings* (M. S. Archer, R. Bhaskar, A. Collier, T. Lawson, and A. Norrie, Eds.). London: Routledge, ix-xxiv.
- Bowman, C. & Ambrosini, V. (2000). Value Creation Versus Value Capture: Towards a Coherent Definition of Value in Strategy. *British Journal of Management*, 11(1), 1-15.
- Bowman, C. & Ambrosini, V. (2010). How value is created, captured and destroyed. *European Business Review*, 22(5), 479 – 495.
- Bowman, C. (2015). The role of technology in the creation and capture of value. *Technology Analysis & Strategic Management*, 27(2), 237-248.
- Brynjolfsson, E., & Oh, J. (2012). The attention economy: measuring the value of free digital services on the Internet. In the proceedings of the 33rd International Conference on Information Systems, Orlando,
- Brynjolfsson, E., & Saunders, A. (2009). *Wired for Innovation*. MIT Press.

- Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in Numbers: How does data-driven decision-making affect firm performance? in Proceedings of the 32nd International Conference on Information Systems, Shanghai, China, paper 13.
- Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the social sciences*, 34(2), 182-210.
- Bygstad, B., & Munkvold, B. E. (2011). In search of mechanisms. Conducting a critical realist data analysis. In the Proceedings of Thirty Second International Conference on Information Systems (ICIS), Shanghai, 2011, pp. 1-15.
- Capgemini Consulting (2013). The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data
- Cas, J. (2005). Privacy in pervasive computing environments – A contradiction in terms? *Technology and Society Magazine*, IEEE, 24(1), 24–33.
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, 295(2), 295-336.
- Coleman, J.C. (1986). Social Theory, Social Research and a Theory on Action. *American Journal of Sociology*, 91(6), 1309-1335.
- Coleman, J.C. (1990). *Foundations of Social Theory*. Boston: Harvard University Press.
- Conradie, P. and Choenni, S. (2014). On the barriers for local governments releasing open data. *Government Information Quarterly*, 31(1), S10-S17.
- Curry, E., O'Donnell, J., Corry, E., Hasan, S., Keane, M. and O'Riain, S. (2013). Linking building data in the cloud: Integrating cross-domain building data using linked data. *Advanced Engineering Informatics* 27(2013), 206–219.
- Danermark, B. M. Ekström, L. Jakobsen, and J.C. Karlsson, *Explaining Society: Critical Realism in the Social Sciences*. London: Routledge, 2002.
- Davies, T. (2013). Open Data Barometer 2013 Global Report. Open Data Institute and World Wide Web Foundation. Available at <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>
- de Vries, M.; Kapff, L.; Achiaga, M. N.; Wauters, P.; Osimo, D.; Foley, P.; . . . and Whitehouse, D. (2011). Pricing of public sector information study (POPSIS), Brussels: European Commission, Information Society and Media Directorate General.
- Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, Forthcoming.



- Downward, P. & Mearman, A. (2007). Retrodution as mixed-methods triangulation in economic research: reorienting economics into social science, *Cambridge Journal of Economics*, 31(1), 77-99.
- Eckartz, S. M., Hofman, W. J. & Van Veenstra, A. F. (2014). A Decision Model for Data Sharing. In *Electronic Government* (pp. 253-264). Springer Berlin Heidelberg.
- Economic Intelligence Unit (2013). The data directive: How data is driving corporate strategy – and what still lies ahead. Available at: <http://www.wipro.com/documents/the-data-directive.pdf>
- Eisler, R. (2007). *Real Wealth of Nations*. San Francisco, CA: Berrett-Koehler Publishers
- Elster, J. (2007). *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
- Environmental Performance Index (EPI). (2014). 2014 Environmental performance index: Full report and analysis. Available at [http://epi.yale.edu/files/2014\\_epi\\_report.pdf](http://epi.yale.edu/files/2014_epi_report.pdf) [Accessed October 2 2014]
- European Commission (EU). (2011a). Special Eurobarometer 359: Attitudes on Data Protection and Electronic Identity in the European Union. Brussels: TNS Opinion & Social. Available at [http://ec.europa.eu/public\\_opinion/archives/ebs/ebs\\_359\\_en.pdf](http://ec.europa.eu/public_opinion/archives/ebs/ebs_359_en.pdf)
- European Commission (EU). (2011b). Open data: An engine for innovation, growth and transparent gov-ernance. European Commission Information Society. Available: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF>
- Fleurbaey, Marc. (2009). Beyond GDP: The Quest for a Measure of Social Welfare. *Journal of Economic Literature*, 47(4), 1029–75.
- Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39-50.
- Frischmann, B. M. (2012). *Infrastructure: The Social Value of Shared Resources*. Oxford: Oxford University Press.
- Gallaher, M.P., O'Connor, A.C., Dettbarn Jr., J.L. & Gilday, L.T. (2004). Cost Analysis of Inadequate Interoperability in the US Capital Facilities Industry. US Dept. of Commerce, Technology Administration, National Institute of Standards and Technology. Available at <http://fire.nist.gov/bfrlpubs/build04/PDF/b04022.pdf>

- Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350), 320-328.
- Gil-Saura, I., Frasset-Deltoro, M. & Cervera-Taulet, A. (2009). The Value of B2B Relationships, *Industrial Management & Data Systems*, 109(5), 593-609.
- Grover, V., & Kohli, R. (2012). Cocreating IT Value: New Capabilities and Metrics for Multifirm Environments. *MIS Quarterly*, 36(1), 225-232.
- Hilbert M. & López P. (2011).The world's technological capacity to store, communicate, and compute information. *Science*, 332 (6025), 60-65
- Hair, J. F., Ringle, C. M., and Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *The Journal of Marketing Theory and Practice*, 19(2), 139-152.
- Hair, J. F., Sarstedt, M., Ringle, C. M., and Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science*, 40(3), 414-433.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., and Sarstedt, M. (2014). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage: Thousand Oaks.
- Haufler, V. (2001). *A public role for the private sector: Industry self-regulation in a global economy*. Carnegie Endowment.
- Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge University Press.
- Heeks, R. (2003). *Most eGovernment-for-Development Projects Fail: How Can Risks be Reduced?* Institute for Development Policy and Management, Manchester: University of Manchester.
- Heimstädt, M., Saunderson, F., & Heath, T. (2014). *Conceptualizing Open Data ecosystems: A timeline analysis of Open Data development in the UK (No. 2014/12)*. Discussion Paper, School of Business & Economics: Management.
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in international marketing*, 20, 277-319.
- Henseler, J., and Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, 28(2), 565–580.
- Hess C. and Ostrom, E. (2006). *Understanding Knowledge as a Commons – From Theory to Practice*. Cambridge, Massachusetts: MIT Press

- Hitt, M. A., Ireland, R. D., Sirmon, D. G., and Trahms, C. A. (2011). Strategic entrepreneurship: Creating value for individuals, organizations, and society. *The Academy of Management Perspectives*, 25(2), 57-75.
- Hjalmarsson, A. Juell-Skielse, G., Ayele, W., Rudmark, D. and Johannesson, P. (2015). From Contest to Market Entry: A Longitudinal Survey of Innovation Barriers Constraining Open Data Services Development. In *Proceedings of the 23rd European Conference on Information Systems (ECIS) 2015, Münster, Germany*
- Houghton, J. (2011). *Costs and Benefits of Data Provision: Report to the Australian National Data Service, Centre for Strategic Economic Studies, Victoria University*
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258-268.
- Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments,” *Government Information Quarterly*, 28(4), pp. 446-456.
- Jetzek T., Avital, M. and Bjorn-Andersen, N. (2013a). The Generative Mechanisms of Open Government Data. In the proceedings of the 21st European Conference of Information Systems (ECIS), Utrecht, Holland.
- Jetzek T., Avital, M. and Bjorn-Andersen, N. (2013b). Generating Value from Open Government Data. In the proceedings of the 34th International Conference on Information Systems (ICIS), Milan, Italy.
- Jetzek, T., Avital, M. and Bjorn-Andersen, N. (2014a). Data- Driven Innovation through Open Government Data. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2), 100-120.
- Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2014b). Generating Sustainable Value from Open Data in a Sharing Society. In *Creating Value for All Through IT* (pp. 62-82). Springer Berlin Heidelberg.
- Kettinger, W. J., & Li, Y. (2010). The infological equation extended: towards conceptual clarity in the relationship between data, information and knowledge. *European Journal of Information Systems*, 19(4), 409-421.

- Kohli, R. & Grover, V. (2008). Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times. *Journal of the Association for Information Systems*, 9(1), 23-39.
- Kuk, G. and Davies, T. (2011). The Roles of Agency and Artifacts in Assembling Open Data Complementarities. In the proceedings of the 19th International Conference of Information Systems (ICIS), Shanghai, China.
- Kundra, V. (2012, Jan). Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect, *Harvard Business Review*.
- Lavie, D. (2007). Alliance portfolios and firm performance: A study of value creation and appropriation in the US software industry. *Strategic Management Journal*, 28(12), 1187-1212.
- Liew, A. (2007). Understanding data, information, knowledge and their inter-relationships. *Journal of Knowledge Management Practice*, 8(2).
- Linders, D. & Wilson, S.C. (2011). What is Open Government? One Year after the Directive, in *The Proceedings of the 12th Conference on Digital Government Research (dg.o.2011)*, 262-271.
- Lindman, J., & Tammisto, Y. (2011). Open Source & Open Data: Business Perspectives from the Frontline. In *Open Source Systems: Grounding Research* (pp. 330-333). Springer Berlin Heidelberg.
- Lindman, J., Rossi, M. & Tuunainen, V.K. (2013). Open Data Services: Research Agenda, in *Proceedings of the 47th Hawaii International Conference on System Sciences*, pp. 1239-1246.
- Lu, Y., & Ramamurthy, K. (2011). Understanding the link between information technology capability & organizational agility: An empirical examination. *MIS Quarterly*, 35(4), 931-954.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science*, 1-25.
- Marcoulides, G. A. & Saunders, C. (2006). PLS: A Silver Bullet?, *MIS Quarterly* (30:2), pp. iii-ix.
- Martin, S., Foulonneau, M., Turki, S. & Ihadjadene, M. (2014). Risk Analysis to Overcome Barriers to Open Data. *Electronic Journal of e-Government* 11(2), 348 -359.
- Mayer-Schönberger, V. & Zappia, Z. (2011, October). Participation & Power: Intermediaries of Open Data. In *1st Berlin Symposium on Internet & Society* October.

- Mayntz, R. (2004). Mechanisms in the analysis of social macro-phenomena. *Philosophy of the social sciences*, 34(2), 237-259.
- McKinsey & Company. (2012). *Transforming Government through Lean Management*, McKinsey Center for Government.
- McKinsey (2013a). *Open data: Unlocking innovation & performance with liquid information*, McKinsey Global Institute, McKinsey Center for Government & McKinsey Business Technology Office.
- McKinsey (2013b). *Disruptive technologies: Advances that will transform life, business, and the global economy*. McKinsey Global Institute.
- Melville, N. (2010). IS Innovation for Environmental Sustainability. *MIS Quarterly*, 34(1), 1-21.
- Michaelson, J., Abdallah, S., Steuer, N., Thompson, S. & Marks, N. 2009. *National Accounts of Well-Being: Bringing Real Wealth onto the Balance Sheet*. London: New Economics Foundation.
- Milberg, S. J., Smith, H. J., & Burke, S. J. 2000. Information Privacy: Corporate Management & National Regulation. *Organization Science*, 11(1), pp. 35-57.
- Moore, M.H. (1995). *Creating Public Value Strategic Management in Government*. Boston: Harvard University Press.
- Morgan, L., Feller, J., & Finnegan, P. (2013). Exploring value networks: Theorising the creation & capture of value with open source software. *European journal of information systems*, 22(5), 569-588.
- Nilsen, K.: Economic theory as it applies to Public Sector Information. (2010). *Annual Review of Information Science and Technology* 44(1), 419–489.
- Nunnally, J. C. Bernstein (1994). *Psychometric theory*.
- Open Knowledge Foundation (OKF). (2012). *Open Data Handbook version 1.0.0* (revised November 14, 2012). Available at <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>
- Oxera. (2013, January). *What is the economic impact of geo-services*. Available at [http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services\\_1.pdf](http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services_1.pdf)
- Pavlou, P. A. (2011). State of the information privacy literature: where are we now & where should we go. *MIS quarterly*, 35(4), 977-988.

- Pepper, R and Garrity, J. (2014). The Internet of Everything: How the Network Unleashes the Benefits of Big data. In Bilbao-Osorio, B., Dutta, S., & Lanvin, B. (2014): The Global information technology report 2014: Rewards & Risks of Big Data. World Economic Forum, 35-42
- Petter, S., Straub, D. & Rai, A. (2007). Specifying Formative Constructs in Information Systems Research. *MIS Quarterly*, 31(4), 623-656.
- Pol, E and Ville, S. (2009). Social innovation: Buzz word or enduring term? *Journal of Socio-Economics*, 38(6), 878–885.
- Porter, M. E. & Kramer, M.R. (2011). Creating Shared Value. *Harvard Business Review*, 89(1), 62-77.
- Rabin, M. (1998). Psychology & Economics. *Journal of Economic Literature*, 3(1), pp. 11-46.
- Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). Editor's comments: a critical look at the use of PLS-SEM in *MIS quarterly*. *MIS quarterly*, 36(1), iii-xiv.
- Sarker, S., Sarker, S., Sahaym, A., & Bjørn-Andersen, N. (2012). Exploring value cocreation in relationships between an ERP vendor and its partners: a revelatory case study. *MIS quarterly*, 36(1), 317-338.
- Sarstedt, M., Ringle, C. M., Henseler, J., & Hair, J. F. (2014). On the emancipation of PLS-SEM: A commentary on Rigdon (2012). *Long range planning*, 47(3), 154-160.
- Sayer, A. (1992). *Method in Social Science: A Realist Approach* (2nd ed.), London: Routledge.
- Schumpeter, J.A. (1934). *The Theory of Economic Development*, Cambridge. Harvard University Press
- Schwab, K. (2013). *The Global Competitiveness Report 2013–2014*. World Economic Forum.
- Schwab, K. (2013). *The Global Competitiveness Report 2014–2015*. World Economic Forum.
- Shogren, J. (2012). *Behavioural Economics & Environmental Incentives*. OECD Environment Working Papers 49. OECD Publishing.
- Simon, H. (1957). *A Behavioral Model of Rational Choice*. *Models of Man, Social & Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley.

Sine, W. D. & Lee, B. H. (2009). Tilting at windmills? The environmental movement & the emergence of the U.S. wind energy sector. *Administrative Science Quarterly*, 54(1), 123–155

Shapiro, C. & Varian, H.R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press, Boston.

Stanley D. (2003). What do we know about social cohesion?: The research perspective of the Federal Government's Social Cohesion Research Network. *Canadian Journal of Sociology*, 28(1), 5–17.

Stiglitz, J.E., Sen, A. & Fitoussi, J. P. (2009). Report by the Commission on the Measurement of Economic Performance & Social Progress, OECD.

Stone, M. (1974). Cross-validators choice & assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 111-147.

Tammisto, Y., Lindman, J. (2011). Definition of open data services in software business. In the Proceedings of Third International Conference on Software Business (ICSOB 2012), Cambridge, MA, USA, Cusumano, M.A., Iyer, B., Venkatraman, N. (eds.), vol. 114, pp. 297–303.

Thiesse, F. (2007). RFID, privacy and the perception of risk: A strategic framework. *Journal of Strategic Information Systems*, 16(2), 214–232.

Trinchera, L. & Russolillo, G. (2010). On the use of Structural Equation Models & PLS Path modeling to build composite indicators, Working paper no. 30, Università Degli Studi Di Macerata.

Tsang, E.W.K. (2006) Behavioral assumptions & theory development: The case of transaction cost economics. *Strategic Management Journal* 27(11), 999–1011.

Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. OECD Working Papers on Public Governance, No. 22, OECD Publishing.

UN Development Programme (UNDP). (2014). *Sustaining Human Progress: Reducing Vulnerability & Building Resilience*, 24 July 2014. Available at: <http://www.refworld.org/docid/514850672.html> [Accessed October 2nd 2014]

Van den Bergh, J & Antal, M. (2014). Evaluating Alternatives to GDP as Measures of Social Welfare/Progress. Working paper no 56, European Commission, European Research Area. Available at [http://www.foreurope.eu/fileadmin/documents/pdf/Workingpapers/WWWforEurope\\_WPS\\_no056\\_MS211.pdf](http://www.foreurope.eu/fileadmin/documents/pdf/Workingpapers/WWWforEurope_WPS_no056_MS211.pdf)

- van Osch, W. & Avital, M. (2010). The road to Sustainable Value: The path-dependent construction of sustainable innovation as sociomaterial practices in the car industry. In Cooperrider, D. & Avital, M. (eds.) *Positive Design & Appreciative Construction: From Sustainable Development to Sustainable Value*, 99-116
- van Veenstra, A. F., & van den Broek, T. (2014a). A Community-driven Open Data Lifecycle Model Based on Literature & Practice. In *Case Studies in e-Government 2.0* (pp. 183-198). Springer International Publishing.
- van Veenstra, A. F., & van den Broek, T. (2014b). Opening Moves – Drivers, Enablers and Barriers of Open Data. In M.A. Wimmer, M. Janssen, and H.J. Scholl (Eds.): *EGOV 2013*, 50–61, 2013.
- Venkatesh, V., Brown, S. A., and Bala, H. (2013). Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, 37(1), 21-54.
- Vickery, G. (2011). *Review of Recent Studies on PSI Re-Use and Related Market Development*, Paris: European Commission.
- Viscusi, G., Castelli, M. & Batini, C. (2014). Assessing Social Value in Open Data Initiatives: A Framework. *Future Internet*, 6(3), 498-517.
- von Krogh, G., Haefliger, S., Spaeth, S. & Wallin, M.W. (2012). Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development. *MIS Quarterly* 36(2), 649–676.
- Wade, M. & Hulland, J. (2004). Review: the resource-based view and information systems research: review, extension, and suggestions for future research. *MIS Quarterly*, 28(1), 107-142.
- Wernerfelt, B. (1984). A Resource-based View of the Firm, *Strategic Management Journal*, 5, pp. 171–180.
- Williams, C.K. & Karahanna, E. (2013). Causal explanation in the coordinating process: A critical realist case study of federated IT governance structures, *MIS Quarterly*, 37(3), 933-964.
- World Wide Web Foundation. (2013). 2013 Web Index Full Report. Available at <http://thewebindex.org/wp-content/uploads/2013/11/Web-Index-Annual-Report-2013-FINAL.pdf> [Accessed October 2nd 2014]
- York, J.G., Sarasvathy, S.D. & Wicks, A.C. (2013). An Entrepreneurial Perspective on Value Creation in Public-Private Ventures. *Academy of Management Review*, 38(2), 307-315.



- Yu, H. & Robinson, D.G. (2012). The New Ambiguity of 'Open Government'. 59 *UCLA Law Review*. Available at <http://ssrn.com/abstract=2012489>
- Zachariadis, M., Scott, S. & M. Barrett. (2013). Methodological Implications of Critical Realism for Mixed-Methods Research. *MIS Quarterly*, 37(3), 855-880.
- Zuiderwijk, A. & Janssen, M. Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View. (2014a). in *Open Government, Opportunities and Challenges for Public Governance* (M. Gascó-Hernández, Ed.). New York: Springer. 115-136
- Zuiderwijk, A. & Janssen, M. (2014b). Open data policies, their implementation and impact: A comparison framework, *Government Information Quarterly*, 31(1), 17-29.
- Zuiderwijk, A., Helbig, N. Gil-Garcia, J. R. & Janssen, M. (2014a). Special Issue on Innovation through Open Data: Guest Editors' Introduction. *Journal of theoretical and applied electronic commerce* , 9(2), 1-13.
- Zuiderwijk, A., Janssen, M., Choenni, S. and Meijer, R. (2014b). Design principles for improving the process of publishing open data. *Transforming Government: People, Process and Policy* 8(2),
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R. and Sheikh Alibaks, R. (2012). Socio-technical impediments of open data. *Electronic Journal of eGovernment*, vol. 10, no. 2, pp. 156 - 172, 2012

## VII Appendix A: Operationalization of variables

Construct	Definition	Measures	Source
Liquid open data	Liquid open data are available online without technical restrictions to access, link and stream across systems, provided for free and under an open access license	Is the data openly licensed? (aggregated from 15 data categories) [1-10]	Open Data Barometer <a href="http://www.opendataresear.ch.org/barometer">http://www.opendataresear.ch.org/barometer</a>
		Is the dataset available free of charge? (aggregated from 15 data categories) [1-10]	
		Was it easy to find information about this dataset? (aggregated from 15 data categories) [1-10]	
		Is the dataset provided in machine readable formats? (aggregated from 15 data categories) [1-10]	
		Is the dataset downloadable for bulk download (aggregated from 15 data categories) [1-10]	
Robustness of regulatory data and privacy protection	The robustness of existing regulatory frameworks that promote privacy and data protection	To what extent is there a robust legal or regulatory framework for protection of personal data in your country? [1-10]	Web Index <a href="http://thewebindex.org/">http://thewebindex.org/</a>
		To what extent does the government enforce the laws in place to protect people from crimes committed using the Internet?	
		To what extent does the law protect people from crimes committed using the Internet?	

Construct	Definition	Measures	Source
Extent of government IT leadership	Reflects the extent to which government is promoting ICT use	In your country, how successful is the government in promoting the use of information and communication technologies (ICTs)? [1 = not successful at all; 7 = extremely successful]	World Economic Forum's Global Information Technology Report <a href="http://www.weforum.org/reports/global-information-technology-report-2013/">http://www.weforum.org/reports/global-information-technology-report-2013/</a>
		To what extent does the government have a clear implementation plan for utilizing ICTs to improve your country's overall competitiveness? [1 = no plan; 7 = clear plan]	
		In your country, to what extent do government purchasing decisions foster innovation? [1 = not at all; 7 = to a great extent]	
Lack of affordable high-speed networks	Reflects the how expensive it is to get high-speed network access for most of the population	Cost of bandwidth per MB	ITU - International Telecommunication Union <a href="http://www.itu.int">http://www.itu.int</a>
		Cost of fixed broadband per capita income	
Ease of reaching a skilled workforce	A construct measuring the availability of skilled workers in a country	The general approach of companies in your country to human resources is (1 = to invest little in training and employee development, 7 = to invest heavily to attract, train, and retain employees)	WEF - World Economic Forum: Global Competitiveness Report <a href="http://www.weforum.org/reports/global-competitiveness-report-2013-2014">http://www.weforum.org/reports/global-competitiveness-report-2013-2014</a>
		Does your country retain talented people? [1 = the best and brightest leave to	

Construct	Definition	Measures	Source
		<p>pursue opportunities in other countries; 7 = the best and brightest stay and pursue opportunities in the country]</p> <p>Does your country attract talented people from abroad? [1 = not at all; 7 = attracts the best and brightest from around the world]</p>	
Shared digital content dissemination	Reflects the level of informative digital content that is freely shared online by various stakeholders	<p>Educational content creation on the web, average of the following questions:</p> <p>To what extent do CSOs use the Web to educate and inform citizens about government decision-making and public policy issues?</p> <p>To what extent do trade unions use the Web to educate and inform citizens about government decision-making and public policy issues?</p> <p>To what extent is the Web used by local organizations to disseminate environmental information and facilitate education about climate and environmental concerns?</p> <p>Governmental content creation on the web, average of the following questions:</p> <p>To what extent is</p>	Web Index <a href="http://thewebindex.org/">http://thewebindex.org/</a>

Construct	Definition	Measures	Source
		<p>information about government funded local health care services made available on the Web?</p> <p>To what extent is the Web being used by the government to inform women of their legal rights in areas such as right to equal opportunities, right to inheritance, and rights to seek legal re-dress against violence? [1-10]</p> <p>To what extent is the Web being used by the government to provide information about support that is available for victims of gender-based violence?</p> <p>To what extent does the government publish school-level information about education performance online?</p>	
Digital product and service innovation	Reflects the extent to which ICT influences the creation of innovative business models, products and services and processes	<p>To what extent are ICTs creating new business models, services and products in your country? [1 = not at all; 7 = to a significant extent]</p> <p>In your country, to what extent do ICTs enable new organizational models (e.g., virtual teams, remote working, telecommuting) within businesses? [1 = not at all; 7 = to a great extent]</p>	<p>WEF - World Economic Forum: Global Information Technology Report</p> <p><a href="http://www.weforum.org/reports/global-information-technology-report-2013/">http://www.weforum.org/reports/global-information-technology-report-2013/</a></p>

<b>Construct</b>	<b>Definition</b>	<b>Measures</b>	<b>Source</b>
Corporate Social Accountability	Reflects the extent to which private business follow private and public rules and regulations that conduct to accountability and good practices	In your country, how strong are financial auditing and reporting standards? [1 = extremely weak; 7 = extremely strong]	WEF - World Economic Forum: Global Competitiveness Report <a href="http://www.weforum.org/reports/global-competitiveness-report-2013-2014">http://www.weforum.org/reports/global-competitiveness-report-2013-2014</a>
		In your country, how would you characterize corporate governance by investors and boards of directors? [1 = management has little accountability to investors and boards; [7 = management is highly accountable to investors and boards]	
		In your country, to what extent are the interests of minority shareholders protected by the legal system? [1 = not protected at all; 7 = fully protected]	
		Strength of Investor Protection Index on a 0–10 (best) scale	
Sustainable value	Reflects the degree of economic, social and environmental wellbeing in society.	Natural logarithm of Gross domestic product pr. capita	World Bank <a href="http://data.worldbank.org/">http://data.worldbank.org/</a>
		Number of years a newborn infant could expect to live if prevailing patterns of age-specific mortality rates at the time of birth stay the same throughout the infant's life (UN)	United Nations: Human Development Index <a href="http://hdr.undp.org/en/content/human-development-index-hdi">http://hdr.undp.org/en/content/human-development-index-hdi</a>
		Number of years of schooling that a child of	United Nations: Human Development Index

Construct	Definition	Measures	Source
		school entrance age can expect (UN)	<a href="http://hdr.undp.org/en/content/human-development-index-hdi">http://hdr.undp.org/en/content/human-development-index-hdi</a>
		Measures environmental sustainability based on 67 empirical measurements	Yale Center for Environmental Law & Policy (YCELP) and the Center for International Earth Science Information Network (CIESIN) at Columbia University <a href="http://epi.yale.edu/">http://epi.yale.edu/</a>
Voice and accountability	Control variable	Voice and Accountability is an index developed by the World bank to reflects perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media	World Bank: World Governance Indicators <a href="http://info.worldbank.org/governance/wgi/">http://info.worldbank.org/governance/wgi/</a>

**Table 8: Operationalizing of variables**

## VII Appendix B: Descriptive statistics

Indicator	Mean	Standard err.	Kurtosis	Skewness	Min	Max	Count
OD1	3,69	0,301	-0,7412	0,223	0	10	76
OD2	1,17	0,275	3,155	2,065	0	10	76
OD3	3,89	0,322	-0,940	0,211	0	10	76
OD4	1,51	0,228	6,578	2,270	0	10	76
OD5	3,81	0,337	-0,756	0,378	0	10	76
DP1	6,95	0,367	-0,580	-0,633	0	10	76
DP2	5,45	0,362	-0,869	0,077	0	10	76
DP3	6,47	0,368	-0,713	-0,460	0	10	76
GL1	4,15	0,084	0,331	0,040	2,25	5,89	76
GL2	4,58	0,083	0,0396	-0,053	2,92	6,20	76
GL3	3,58	0,076	1,002	0,166	1,87	5,57	76
CN1	0,18	0,044	15,522	3,632	0,004	2,32	76
CN2	58201,4	9804,0	5,29593	2,310	228,8	391106	76
AS1	4,16	0,077	-0,785	0,131	2,94	5,64	76
AS2	3,78	0,109	-0,401	0,438	1,81	5,97	76
AS3	3,73	0,116	-0,081	0,401	1,48	6,08	76
CO1	6,63	0,296	-1,147	-0,142	1,2	10	76
CO2	5,12	0,300	-0,974	0,090	0	10	76
PR1	4,675	0,078	-0,143	-0,397	2,55	5,89	76
PR2	4,40	0,077	-0,708	-0,174	2,68	5,57	76
VAL1	9,00	0,197	-0,260	-0,639	3,73	11,52	76
VAL2	72,76	1,010	-0,764	-0,705	52,51	83,58	76
VAL3	13,85	0,339	-0,585	-0,317	7,5	19,9	76
VAL4	57,21	1,956	-1,114	-0,126	18,43	87,67	76
CSR	4,62	0,074	-0,010	0,270	3,17	6,21	76
VOICE	-0,024	0,117	-0,563	-0,491	-2,68	1,40	76

*Table 9: Descriptive statistics*



### VII Appendix C: Cross loadings

	<b>OD</b>	<b>DP</b>	<b>GL</b>	<b>CN</b>	<b>AS</b>	<b>CP</b>	<b>PR</b>	<b>Value</b>
<b>OD1</b>	<b>0.793</b>	0.407	0.183	-0.161	0.439	0.544	0.468	0.455
<b>OD2</b>	<b>0.970</b>	0.638	0.245	-0.344	0.542	0.777	0.610	0.771
<b>OD3</b>	<b>0.928</b>	0.612	0.293	-0.306	0.531	0.729	0.630	0.702
<b>OD4</b>	<b>0.968</b>	0.658	0.265	-0.348	0.532	0.795	0.635	0.784
<b>OD5</b>	<b>0.836</b>	0.427	0.203	-0.155	0.439	0.607	0.511	0.513
<b>DP1</b>	0.489	<b>0.776</b>	0.263	-0.385	0.349	0.544	0.468	0.562
<b>DP2</b>	0.542	<b>0.895</b>	0.373	-0.346	0.526	0.624	0.544	0.693
<b>DP3</b>	0.488	<b>0.765</b>	0.062	-0.247	0.200	0.601	0.382	0.543
<b>GL1</b>	0.306	0.317	<b>0.969</b>	-0.054	0.688	0.277	0.753	0.334
<b>GL2</b>	0.244	0.255	<b>0.957</b>	0.004	0.666	0.210	0.695	0.232
<b>GL3</b>	0.207	0.302	<b>0.929</b>	0.035	0.704	0.188	0.676	0.269
<b>CN1</b>	-0.166	-0.267	-0.038	<b>0.848</b>	-0.078	-0.300	-0.195	-0.266
<b>CN2</b>	-0.339	-0.406	0.012	<b>0.951</b>	-0.259	-0.514	-0.330	-0.588
<b>AS1</b>	0.600	0.520	0.591	-0.312	<b>0.921</b>	0.614	0.788	0.692
<b>AS2</b>	0.444	0.419	0.631	-0.138	<b>0.904</b>	0.394	0.686	0.459
<b>AS3</b>	0.442	0.319	0.765	-0.094	<b>0.908</b>	0.324	0.715	0.426
<b>CO1</b>	0.775	0.712	0.162	-0.486	0.480	<b>0.974</b>	0.600	0.808
<b>CO2</b>	0.730	0.695	0.306	-0.440	0.498	<b>0.971</b>	0.665	0.764
<b>PR1</b>	0.625	0.600	0.735	-0.280	0.775	0.657	<b>0.987</b>	0.657
<b>PR2</b>	0.635	0.538	0.737	-0.325	0.816	0.626	<b>0.988</b>	0.667
<b>VAL1</b>	0.646	0.660	0.220	-0.545	0.574	0.715	0.588	<b>0.912</b>
<b>VAL2</b>	0.710	0.704	0.274	-0.388	0.523	0.776	0.639	<b>0.941</b>
<b>VAL3</b>	0.689	0.739	0.311	-0.431	0.578	0.752	0.639	<b>0.938</b>
<b>VAL4</b>	0.663	0.670	0.284	-0.561	0.527	0.763	0.624	<b>0.928</b>
<b>CSR</b>	0.428	0.521	0.473	0.492	0.382	0.320	0.480	0.337

*Table 10: Cross loadings*

## VII Appendix D: Differences in parameters between PLSc and PLS

Path	PLSc coefficient	PLS coefficient	Difference
Liquid open data -> Shared content	0,525	0,485	0,04
Liquid open data -> Digital products	0,294	0,282	0,012
Regulatory protection -> Shared content	0,315	0,331	-0,016
Regulatory protection -> Digital products	0,073	0,094	-0,021
Gov't IT leadership -> Shared content	-0,049	-0,047	-0,002
Gov't IT leadership -> Digital products	0,482	0,465	0,017
Cost of network -> Shared content	-0,196	-0,187	-0,009
Cost of network -> Digital products	-0,139	-0,131	-0,008
Availability of skills -> Shared digital content	0,068	0,074	-0,006
Availability of skills -> Digital product innovation	0,241	0,243	-0,002
Shared content -> Sustainable value	0,48	0,412	0,068
Digital products-> Sustainable value	0,202	0,211	-0,009
Interaction: accountability content	-0,353	-0,344	-0,009
Interaction: accountability product	0,256	0,256	0
Voice -> Sustainable value	0,258	0,29	-0,032

*Table 11: Comparison between parameter estimation from PLSc and PLS*

## **PAPER VIII: Innovation in the Open Data Ecosystem**

### **Exploring the role of real options thinking and multi-sided platforms for sustainable value generation through open data**

In Analytics, Innovation and Excellence Driven Enterprise Sustainability, edited by Rick Edgeman, Elias G. Carayannis and Stavros Sindakis, Palgrave Studies in Democracy, Innovation and Entrepreneurship for Growth. (Forthcoming in 2015).

# Innovation in the Open Data Ecosystem

Exploring the role of real options thinking and multi-sided platforms for sustainable value generation through open data

*Thorhildur Jetzek*

*Copenhagen Business School, Department of IT Management*

*tj.itm@cbs.dk*

## **Abstract**

While open data as a phenomenon is rapidly growing up, innovation through open data is still less than expected. Research has shown that in spite of emerging new businesses models, private sector stakeholders are struggling to generate monetary income from open data. This is worrying as open data initiatives might not be sustained if there no evidence of value generation through external use of the data. We suggest that insights from two established theories, real options theory and theory of two-sided markets might help us create a more coherent picture of the complex relationships between innovation and value generation in the open data ecosystem, and even resolve what we call the open data value paradox. We propose that governments which openly publish data are providing private sector stakeholders with the equivalent of a real option. By conceptualizing the uncertain or serendipitous value of open government data as option value we might be able to stimulate activity and investment in the open data ecosystem. Moreover, we propose that by utilizing two-sided markets type of business models, private companies can use the data as a resource to provide free information and by capitalizing on the resulting positive network externalities, generate monetary income as well. Finally, we propose that governments should provide the necessary nourishment to this ecosystem in order to stimulate the generation of sustainable value.

**Keywords:** *Open data ecosystem, Open data value paradox, Sustainable value, Real options value, Two-sided markets, Multi-sided platforms*

## VIII.1 Introduction

*“The miracle is this: The more we share the more we have.”  
Leonard Nimoy 1931-2015*

Our world is at an inflection point where technological advances and boundary-crossing social challenges have come together to create a paradigm shift. Our societies are facing multiple and urgent social challenges, ranging from economic inequality, unemployment, and poor social conditions to chronic diseases and climate change. Given the complexity and cross-boundary nature of these challenges, a new approach where social and technological progress co-evolves in order to generate sustainable value is necessary (OECD 2011). While there are multiple potential paths that lead to the creation of sustainable value, we believe that *innovation through open data* is one path that is currently showing high promise for future value generation.

All scientists understand the importance of new insights for scientific advances. These insights have in the past been based on careful analysis of data collected by researchers that subsequently have disseminated their findings through the education system and scientific publications. These insights have changed our view of the world, our attitudes, beliefs, and behavior, finally impacting everyday actions to generate value for individuals and society. However, due to scientific and technological advances, we are currently on the brink of a new era where we have started to see a significant change in the pace of these processes. The digital revolution, including the digitization of nearly all media, the ubiquity of Internet access, the proliferation of mobile phones, and the growth of the Internet of Things have led to exponentially increasing amounts of data that offer a world of new information and insights. The current trend towards *open access* to these data has furthermore allowed stakeholders to cross borders and sector-based boundaries, and has completely revolutionized how public and private stakeholders are collectively addressing some of our most difficult social challenges (Bakici et al. 2013). Moreover, personalized insights and awareness can now be available to individuals in near real time, continuously impacting the way we act and interact. All of these changes require new theoretical lenses for analyzing value maximizing behavior in order for us to better understand and support the continuing evolution towards openness, sharing, and new insights that create value for all.

One of the important issues for safeguarding the ongoing supply of open data from government is the ability to make the value of open data explicit and visible to policy makers and open data users in the private sector. This is, however, not a trivial task, as open data are generally a free resource, exhibiting the features of a public good (Jetzek

et al. 2013). The openness of data guarantees that everyone can use the data for whatever purposes, and is an important feature when it comes to stimulate use that might lead to sustainable value generation (Jetzek et al. 2014b). However, the same features make it extremely difficult to trace where open data are used and for what purposes, primarily for two reasons: 1) A significant share of the value generation from data happens through information generation and network effects, so-called non-market production (Brynjolfsson and Oh 2012). 2) As open data might be used for future innovations, another important element of the value of open data is the fact that the data are available for unforeseen or “serendipitous” use in the future. While current methods of measuring market activity capture only materialized market-based transactions, these two important aspects remain unidentified.

The provision of high-quality data can require significant up-front investments (OECD 2014). In the case of governments these investment costs usually exceed the expected “private” benefits of data sharing to the organizations that collect the data. Governments must therefore look for evidence of societal level value generation from open data. Accordingly, if open data initiatives are not perceived to contribute much to value generation, we could potentially enter a downward facing spiral instead of benefitting from the synergistic relationship between information production and dissemination and entrepreneurial activity that is expected to result from innovation through open data. This could lead to a paradoxical situation that we call the *open data value paradox*. This paradox describes a situation where entrepreneurs do not use the data because the data are not usable enough, there is too much uncertainty over the sustainability of initiatives, and therefore little or no value is generated. However, the data providers are not willing to invest in the people and technology necessary to make the data more usable and sustainable unless they observe some evidence of value generation. We suggest that in order to maintain and stimulate use of open data, we need more open data platforms or data intermediaries. To justify the investment in such platforms, we propose a model that can make the sustainable value of open data more explicit to governments, businesses, and individuals, and thereby act to resolve the open data value paradox.

In this paper we propose the use of two established theories that might help make the contribution of openness of data towards the generation of sustainable value for society more explicit. The first theory we borrow from is the theory of *two-sided markets*. The business models that have been developed in these types of markets (now most commonly called Multi Sided Platforms or MSPs) are generally based on two or more sides of affiliated customers that interact via digital platforms. The intangible value

that market participants can gain from network externalities is not internalized and will drive the subsequent generation of economic value that can be used to justify the investment necessary to attract the required number of participants.

The second theory we suggest to borrow from is *real option* theory. Innovation researchers posit that an option value approach can influence the motivations of early adopters. While most company managers know they must innovate to thrive, technical innovation – for instance data analytics – is accompanied with uncertainty about the benefits of using the innovation and the irreversibility that arises from high learning and adaptation costs during deployment, as well as high switching costs after deployment (Fichman 2004). In the case of high uncertainty and irreversibility, it is fruitful to view such investments through a real options lens (ibid). Both approaches will alleviate the open data value paradox by emphasizing the factors that make value generation from open data unique, while using methods and theories that are already well established in industry and economics.

## **VIII.2 The Sustainable Value of Open Data**

Openness as an overarching philosophy or concept implies that those who embrace openness are willing to share not only raw data but also their problems, experiences, and questions. This sharing leads to the generation and dissemination of shared information, which in turn can influence how individuals view the world, make their decisions, and go about their business. Unfortunately, while the value generation that might arise from sharing data and information might seem plausible and intuitive to modern citizens, we are still struggling when we try to link these intuitive notions to the “business” world of measures, where thousands of initiatives must fight for the same funds. This struggle is relevant to governments and businesses: investment in a particular initiative must be justified as something that is inherently valuable to the owners of funds - eventually the taxpayers in the case of government. In order to sustain open data initiatives, governments (or businesses) must somehow quantify the potential social and environmental value that might very well exceed the economic potential offered by open data (McKinsey 2013).

Hundreds of national and local governments have already opened public access to various data-sources, making them available for use and re-use for commercial or other purposes. Moreover, these initiatives have been followed by international institutions, civil society organizations and even businesses, although still to a limited degree. There is certainly mounting evidence showing how innovative use of open data is contributing to sustainable value generation. However, this evidence is mostly in the

form of anecdotes and use cases and is therefore not yet cogent or rigorous enough to put a solid foundation under open government data initiatives. The recent Open Data Barometer report highlights that strong empirical evidence on the impacts of open government data initiatives is almost universally lacking (Davies 2013). This creates a risk that these initiatives will not be sustained, for example, when new governments take power or when world-wide attention to the phenomenon diminishes. Open data, as a concept within the big data technology arena, are still on their way to the top in Gartner's Hype Cycle for Big data 2014. This also indicates that within a few years, interest in the concept of open data might start declining into the trough of disillusionment (Buytendijk 2014). When that happens, the uncertainty on the extent and nature of the return on investment in open data represents a clear risk for the sustainability of these initiatives (Martin et al. 2014).

If commercial firms do not realize a return on their innovative activities, they will tend to under-invest in those activities that are either highly risky or that are easily imitated by free-riding competitors (West and Gallagher 2006). Using the same kind of logic, governments will not continue to invest in infrastructure projects such as open data platforms if they do not perceive public value from their efforts and good use being put to the infrastructure in question. In the case of open data, the main barrier to ongoing investment is the nature of value that is generated. The value of open data is to a large degree mediated through network effects without any market transactions taking place. Network effects exist when the value of a good or a service increases as more consumers use them or as more supply-side partners augment the service (Bharadwaj et al. 2013). Open data are a good example of how network effects bring superior value, as digital data are a resource, the value of which will increase the more it is used (Nilsen 2010). However, most of our predominantly used valuation methods still rely heavily on market activity and the generation of economic profits, and do not explicitly recognize the intangible value of information sharing.

For instance, the resource-based view (Barney, 1991; Wade and Hulland, 2004, Wernerfelt, 1984) predicts that firms can only sustain their competitive advantage through use of valuable resources that are neither perfectly imitable nor substitutable without great effort. Their definition clearly excludes open data as a resource upon which firms can build competitive advantage. The value chain approach (Porter, 1985) looks at how each "step" in a chain of activities within a single firm generates economic value. However, this approach misses both the value generated from network externalities as well as the intangible or social and environmental dimensions of value. Modern portfolio theory predicts that the riskier the asset, the higher return on



investment is demanded by the investor, but the return on investment is generally calculated by dividing net profit by the company's total assets, thereby excluding both the generation of intangible value and use of external resources. All of these theories will therefore drive the firm's focus towards internal assets and the generation of economic value from those assets. The resulting approach will in some cases result in products and services that contribute to solving social challenges, but unfortunately, there are also many cases where social wellbeing and economic profits will end up as opposing interests, resulting in so-called market failures.

In order to fully embrace the value of open data, we need to include value that is created through better decision making and network effects that might eventually lead to reduced CO<sub>2</sub> emissions, increased availability of clean water in areas where water is contaminated, better and more equitable use of resources or healthier citizens, to name a few. This type of value is not easily captured by looking only at current revenue streams of companies. Porter and Kramer (2011, 3) state that companies: "...continue to view value creation narrowly, optimizing short-term financial performance in a bubble while missing the most important customer needs and ignoring the broader influences that determine their longer-term success." The authors suggest a new approach, utilizing the principle of *shared value*, which prescribes that companies should continue to create economic value, but in a way that simultaneously creates value for society by addressing its needs and challenges (ibid). To extend the concept of shared value to include generation of value by not only businesses but also governments and individuals, we introduce the concept of *sustainable value*. The definition of sustainable value focuses on the proactive, concerted efforts of businesses, government institutions, and the overall community to address social challenges in innovative ways, thereby generating social, environmental, and economic value for all stakeholders and future generations (van Osch and Avital, 2010).

If private and public stakeholders wish to formally or informally collaborate on generating sustainable value, they must move beyond the market mechanism, a term that is borrowed from economics, referring to the use of monetary exchange between buyers and sellers within an open and understood system of value and time trade-offs to produce the best distribution of goods and services. While the market mechanism still functions well for distribution of goods and services, the shifts toward an economy centered on information and the move to a networked Internet-based environment have caused significant attenuation of the limitations that market-based production places on the pursuit of value (Benkler 2006). We must examine different types of mechanisms that facilitate shared or sustainable value generation, and then subsequently highlight

not only economic implications of innovation but the social and environmental implications as well.

Another mechanism has already become a foundation for generating value from open data, i.e., the network mechanism, which we define as a *mechanism that generates value from actionable insights gained through information sharing and reuse over networks*. The network mechanism refers to the actions of what we can call information creators and information consumers, but in fact it is not simple to distinguish between who creates and who consumes information. In many current business models the information consumers are also generating valuable data for platforms owners that are “crowd-sourced” to create new or improved information. However the main distinction between the market and the network mechanisms is that in the latter there is no monetary exchange and the relationships are many-to-many, instead of the traditional one-to-one relationship between buyers and sellers. We propose that intermediaries can play a valuable role in leading the market and network mechanisms together, thus creating a structure around these complex relationships that allows for synergistic value generation.

### **VIII.3 The Role of Intermediaries in Open Data Ecosystems**

Intermediaries are important in markets because of five limitations of direct transactions that can be better managed by intermediaries: Search costs, lack of privacy, incomplete information, contracting risk and pricing (Resnick, Zeckhauser, and Avery 1995; Janssen and Zuiderwijk 2014). Accordingly, intermediaries have basically four roles: a) information aggregation, 2) providing trust, 3) facilitating, and 4) matching (Bailey and Bakos 1997). The intermediary can be an agent of any kind, a government organization, an individual or a private company. Recently, the democratization of content as well as the subsequent sharing, remixing, redistribution, and re-syndication of content in newer and more useful forms has caused dramatic power shifts in the intermediary market (Bharadwaj et al. 2013). These trends have disrupted the traditional value chain of economic profits while creating new sources of value. For instance, so-called peer-to-peer internet based business models (sometimes aggregated under the heading *Sharing Economy*) have challenged various traditional intermediaries, such as taxi services, and a new type of platform intermediary has moved in to take their place (Cannon and Summers 2014). These new types of intermediaries are creating an important layer that matches demand and supply for services, utilizing economies of scale and digital technologies as well as the business models of two-sided markets.

While we suggest that data intermediaries are important for open data, we must address the question of why intermediaries are needed in this open and networked world that promises to facilitate peer-to-peer relationships. The answer lies in the still relevant transaction and search costs, as well as in the fact that datasets are getting increasingly bigger, introducing a barrier for users that cannot easily download or move these datasets. Moreover, raw data are in many cases of little or no use to end-users who do not have the capabilities or time to manipulate and process these data (Janssen and Zuiderwijk 2014). While leading countries are implementing of National Data Infrastructures, including platforms where users can directly access open data, the openness of many available government data is still surprisingly low (Davies 2013). Openness of data is not a binary construct but has many dimensions, ranging from licenses to prices to usability and technical accessibility. Making data available in current form is therefore by far not the only milestone to cross when it comes to enabling use of open data (Conradie and Choenni 2014, Janssen et al. 2012, Martin et al. 2013, Zuiderwijk and Janssen 2014a; 2014b, van Veenstra and van den Broek 2013, Zuiderwijk et al., 2012; 2014).

A recent review of the open datasets provided in Berlin, Germany showed that approximately 90% of the data provided were published in a PDF format (Martin et al. 2013). The Open Data Barometer reveals that of the 821 datasets surveyed in 2013, less than 7% of the datasets were published both in bulk machine-readable forms, and under open licenses. Only 1.2% of open data were published as linked data (Davies 2013). Moreover, there are multiple open data initiatives in most countries, where different private and civil society organizations, local governments, and state government each have their own policies and standards for open data. During these early days of open data, open data initiatives are heterogeneous in nature, licenses differ between initiatives, open data standards are still underdeveloped and underused, and there are heterogeneous formats and a lack of metadata, as well as limited network activity (Mayer-Schönberger and Zappia 2011, Martin et al. 2013). For most individuals and smaller developers, these issues come together to *create a substantial barrier to entry*, as the efforts involved in acquiring, manipulating, and analyzing these disparate data are simply too extensive, in comparison to an uncertain and potentially non-economic gain.

In most of the world, governments are already struggling with budgetary restraints and increased demand for services. Making data open is never an effortless task, and these constraints limit government's aspirations for open data, even if the potential for value generation may be clear to them. As governments may not be able to do everything on

their own, data intermediaries could play a crucial role in the open data ecosystem by facilitating data and information access for smaller organizations that may not have the capacity and capabilities to store, integrate, and analyze large and heterogeneous datasets. Intermediaries might also contribute directly to value generation by augmenting and amplifying the circulation of open data by sanitizing and curating data coming from both public and private sources. By making data easier to access, manipulate, and use, intermediaries will drive information creation and product, service, or process innovation based on these data.

Having easy, one-stop access to data services offers a value proposition for companies striving to create a competitive advantage in an increasingly data driven world (Lindman et al. 2014). However, a large share of data-driven services is provided for “free”, oftentimes in exchange for access to personal data (OECD 2014). Data intermediaries need to adapt to market conditions where users are accustomed to having free access to data, information, and information services. To enable the ongoing generation of valuable but free information, the data intermediaries must implement business models that allow them to generate economic profit by capitalizing on the positive network externalities that arise from the interactions of multiple stakeholders using the provided platforms to gain access to the services provided by these intermediaries and their affiliates.

#### **VIII.4 The Economics of Two Sided Markets**

The theory of two-sided markets – more recently referred to as multi-sided platforms - has emerged over the past decade as one of the most active areas of research in information systems, economics, and strategy. It has also drawn considerable interest from practitioners. Two-sided markets are defined as “markets in which one or several platforms enable interactions between end-users, and try to get the two (or multiple) sides “on board” by appropriately charging each side” (Rochet and Tirole 2006, 2). The importance of this approach for the analysis on how open data are used to generate sustainable value is that, unlike classical economic theory, it explicitly recognizes the value of network externalities (Katz and Shapiro 1985, 1986). The theory of two-sided markets builds on the notion that there are non-internalized externalities among end-users: “The starting point for the theory of two-sided markets by contrast is that an end-user does not internalize the welfare impact of his use of the platform on other end-users” (Rochet and Tirole 2006, 3). Two-sided markets, by playing an intermediary role, will facilitate an interaction that would not occur without them and

therefore create value for both sides through direct and indirect interactions and network effects.

A more recent definition views multi-sided platforms or MSPs as technologies, products, or services that create value primarily by enabling direct interactions between two or more customer or participant groups (Hagiu et al. 2011, Hagiu 2014). Hagiu et al (2011, 2) define MSP as “...an organization that creates value primarily by enabling *direct* interactions between two (or more) distinct *types of affiliated* customers.” It is important to note that the term “organization” is used in order to make it clear that the notion of MSP is not restricted to regular businesses, but also encompasses groups of firms, not-for-profit organizations, or even public sector entities that create a valuable interaction service (Hagiu et al. 2011). In spite of the theoretical diversity of potential MSPs, there is a certain “winner takes all” element to MSPs, resulting from the economies of scale introduced by network effects. Therefore, through fierce competition, the trend is for a relatively low number of MSPs to “own” certain domain areas. Moreover, there is an inherent chicken-and-egg problem commonly present in multi-sided marketplaces; each side depends on the other, which makes it a challenge to build up the required critical mass needed to attract the other side (Caillaud and Jullien 2003, Hagiu 2014). These elements introduce a considerable risk to any investment in a MSP.

Hagiu et al. (2011) make a clear differentiation between a MSP and a reseller: Requiring MSPs to enable direct interactions is crucial in ruling out a broad category of intermediaries that buy goods or services from suppliers and sell them to buyers. The relevant direct interactions for a given organization are only those that the MSP specifically enables, i.e., the interactions happen on or through the platform (ibid). Successful MSPs have been shown to create enormous value by reducing search costs or transaction costs (or both) for participants (Hagiu, 2014). MSPs have also been shown to provide an efficient means of information sharing. In the context of open data, the MSP is essentially a data intermediary that provides a platform where the synergy between network externalities and the profitable business opportunities can be exploited. Data MSPs can aggregate the demand of several information requestors and standardize the flow of information from large numbers of information providers from public and private sectors (Bharosa et al. 2013).

An open data MSP would enable interactions between the main groups of stakeholders, i.e., *Data producers/data owners, producers of information, products* and services (developers, scientists, analysts, or journalists) and *information consumers*. There can be different ways in which these groups can interact over a MSP, and each group can

be a mix of public and private stakeholders. Government agencies have already started to utilize the concept of multi-sided platforms in their information infrastructure ventures. Building on the notion of collaborative value generation, rather than developing an information infrastructure and demanding that businesses use it, government agencies have started to move away from the classical approach towards actively tempting businesses to partner in achieving long-term goals, thereby contributing to sustainable value generation (Bharosa et al. 2013). In the following section we present three examples of MSPs that have transformed open data to information that is openly shared, creating value for information users and simultaneously attracting paying customers. More importantly, all of these MSPs have addressed a certain societal problem, thereby benefitting the public sector as a third side of the platform.

### Example one: Opower

Climate change has emerged as one of the most important economic policy issues of the early 21st century. The pollutants that contribute to global warming are commonly known as greenhouse gas emissions. Carbon dioxide (CO<sub>2</sub>) is probably the best known greenhouse gas, representing 85% of all greenhouse gasses in the U.S. Electricity production is the largest single source of global warming pollution in the U.S., responsible for nearly 40% of greenhouse gas emissions<sup>19</sup>. After contemplating on how to address this problem, two college friends, Alex Laskey and Daniel Yates founded the company Opower (then Positive Energy) in 2007. Opower is an energy tech company with a mission to help everyone, everywhere save energy. The company has now grown to over 500 employees and operates in three continents. By the end of 2014, Opower worked with over 95 energy utilities servicing more than 50 million homes<sup>20</sup>. In February 2015, the Opower home energy reports had helped people around the world save over *six* terawatt hours of energy and more than \$700 million on their energy bills. Opower successfully went through an IPO in April 2014 and continues to contribute to decreased CO<sub>2</sub> emissions worldwide.

As utilities deploy smart grid technologies, the volume of data they produce each day increases more than 3,000-fold. Furthermore, as customers begin to interact more with their utilities online, these interactions create even more data.<sup>21</sup> Opower's MSP can store and process 15-minute interval data from smart meters from millions of in-home devices at large scale and high speed, currently spanning more than 52 million

---

<sup>19</sup> <http://www.epa.gov/ghgreporting/ghgdata/reported/index.html>

<sup>20</sup> <http://opower.com/>

<sup>21</sup> <http://opower.com/platform/data-science>

households and businesses, and growing at a rate of more than 100 billion meter reads per year. Opower’s data analytics engine sits on top of this huge repository of data. The engine runs hundreds of algorithms that process utility data, third-party data, and customer behavioral data to power millions of personalized communications with utility customers on the platform.<sup>22</sup>

Opower merges the data streams from utilities with open data from the government to create personalized energy-use profiles. In the U.S. they use data from the Residential Energy Consumption Survey (RECS) to understand how households are using energy. The survey provides region-specific data on end-use energy consumption patterns, such as the type and efficiency of appliances used by the consumers and the systems and energy sources they use to heat and cool homes, among other topics. Opower also uses data from the U.S. Census Bureau on the mix of gas and electric heating sources in a given county in order to create location-specific profiles to use when analyzing an individual’s home energy consumption.

Opower’s products are designed to enhance the utility's interactions with their customers in order to both reduce demand and improve relationships. When designing the way the utilities interact with energy users, Opower has utilized findings from behavioral science that have predicted how people react against information provided on their own use, as compared to the use of others (Cuddy et al. 2012). These results have highlighted the importance of a feedback mechanism to drive behavioral change, creating a subtle aspect of peer-pressure (Jetzek et al. 2014a). The energy reports that Opower creates for each energy user offer a component where this individual’s energy use is compared to the use of other similar households, complete with a smiley token to indicate approval of “good behavior” (Jetzek et al 2014a). When provided with better information and suggestions on how to decrease energy consumption, as well as a token of appreciation for their efforts, customers are empowered to take greater control of the way they use energy. On the other side of the platform, the utilities benefit through increased customer engagement and better targeting of specific customer segments for efficiency. Opower has also created an API to allow utility clients to run their own internal analytics programs using the data in their analytics engine. Government might be labelled as an indirect third party on the platform, as they provide open data and in return gain greater energy efficiency. Figure 1 shows the interactions, gains and costs of the three “sides” of the Opower MSP.

---

<sup>22</sup> <http://opower.com/platform/data-science>

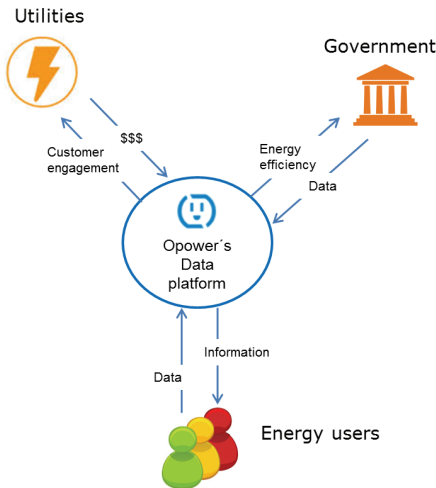


Figure 1: Opower's MSP – connecting utilities and energy users

## Example two: INRIX

According to the Texas Transportation Institute, the cost of congestion in the U.S. in 2012 was more than \$120 billion, nearly \$820 for every commuter that is said to spend over 60 hours per year on average stuck in traffic.<sup>23</sup> Similar problems are endured by most of the world's biggest cities. However, estimates suggest that since 2009, the global pool of personal geo-location data is growing yearly by 20% and by 2020 this data pool could provide USD 500 billion in value worldwide in the form of time and fuel savings, or 380 megatons (million tons) of CO<sub>2</sub> emissions saved (OECD 2014).

INRIX is a leading provider of traffic services worldwide, with the vision to solve traffic, empower drivers, inform planning, and enhance commerce<sup>24</sup>. INRIX provides historical, real-time traffic information, traffic forecasts, travel times, and travel time polygons to businesses and individuals in 40 countries (as of September 2014)<sup>25</sup>. INRIX also gathers, curates, and reports roadway incidents such as accidents, road closures, and road works. INRIX was founded by former Microsoft employees, Bryan Mistele and Craig Chapman, in July 2004. INRIX has not yet been through an IPO, but Porche recently invested \$55 million in the company, which in July 2014 employed

<sup>23</sup> <http://d2dt15nnpf0r.cloudfront.net/tti.tamu.edu/documents/tti-umr.pdf>

<sup>24</sup> <http://www.prnewswire.com/news-releases/inrix-partners-with-san-francisco-on-expanding-traffic-information-services-for-bay-area-drivers-229643681.html>

<sup>25</sup> <http://en.wikipedia.org/wiki/INRIX>



around 350 people. Figure 2 shows the interactions, gains and costs of the three “sides” of the INRIX MSP.

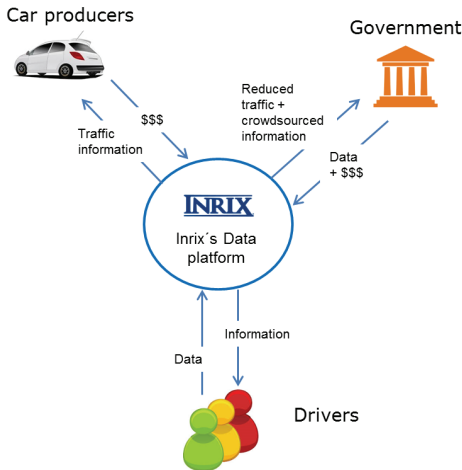


Figure 2: INRIX's MSP, connecting car producers and drivers

As of September 2014, INRIX collected data about roadway speeds from over 175 million real-time anonymous mobile phones, connected cars and other fleet vehicles equipped with GPS locator devices. They also get data from cameras and government road sensors. Moreover, INRIX keeps a database of variables that affect traffic, including open government data such as weather forecasts, special events, official accident and incident reports, school schedules, and information on road construction, which they combine with the crowd-sourced data. The data collected is processed in real-time, creating traffic speed information for major freeways, highways, and arterials across North America, as well as much of Europe, South America, and Africa. The company's analysis of data analysis provides drivers with insights that help them choose the best way to go, minimizing the amount of time spent, saving them frustration and money on gasoline. As the app used to source traffic information from individuals is available for free, INRIX's main source of income is from car-producers, GPS providers, and media companies who are interested in getting access to this type of information. Moreover, they have recently started to provide data and tools to public information services, so government also pay for access to this type of crowdsourced information. However, “in return” for providing INRIX with relevant data government gains from reduced traffic.

### Example three: Zillow

The property market is an essential sector of the economy but also one that has been a recent source of vulnerabilities and crises. While the recent recovery in global housing markets is a welcome development, societies still need to guard against another unsustainable boom, caused by overvaluation of houses followed by unsustainable levels of private debt. A detailed analysis and judgment are needed to make a call about overvaluation of real estate property.<sup>26</sup> Zillow is an online real estate database of homes across the United States. Zillow was founded in 2005 by Rich Barton and Lloyd Frink, former Microsoft executives and founders of Expedia. Zillow's mission is to empower consumers with information and tools to make smart decisions about homes, real estate, and mortgages. Currently, the company has over 30 million unique users per month scrolling through its database of more than 110 million US homes (Capgemini, 2013). Moreover, home shoppers spent more than 5 billion minutes just on Zillow apps in 2013, making 2013 a record year for mobile usage.<sup>27</sup> Zillow successfully launched its IPO in July 2011. As of March 2014, the company employed 887 people, a number that is expected to double due to a merger with competitor Trulia.

Zillow provides increased transparency in the housing market, transforming the way consumers make home-related decisions and connect with professionals.<sup>28</sup> In order to provide such transparency, Zillow has built a database from a range of linked data such as county records, tax data, listing of homes for sale or rental and mortgage information, as well as geographical data and information on local land value and house prices (Capgemini, 2013). Their website combines mapping data with information on local land value and house prices to create a service which accurately estimates the value of a house at a given address (ibid). In addition to giving value estimates of homes, Zillow offers several features, including value changes of each home in a given time frame, aerial views of homes, and prices of comparable homes in the area. Where they can access appropriate public data, they also provide basic information on a given home, such as square footage and the number of bedrooms and bathrooms. Users can even get current estimates of homes if there has been a significant change made, such as a recently remodeled kitchen. One of Zillow's main offerings is Zestimates, which uses an advanced, statistical predictive tool to provide up-to-date information on home values and rental prices.

---

<sup>26</sup> <http://www.imf.org/external/np/speeches/2014/060514.htm>

<sup>27</sup> <http://www.zillow.com/blog/zillow-mobile-2013-year-in-review-141305/>

<sup>28</sup> <http://www.zillow.com/corp/About.htm>

Zillow’s business model is straightforward: They connect people looking to buy houses with the real estate agents, mortgage lenders, and advertisers who want to reach them. In the third quarter of 2014, 74% of Zillow’s revenue came from fees that agents paid for customer leads and apartment leads, 8% came from fees that banks paid for mortgage leads, and 18% came from advertising. They have also recently started connecting renters with apartment listings, and homeowners with design ideas and contractors.<sup>29</sup> In order to make their business model work, Zillow needs to attract visitors to their site who are interested in buying houses, as well as those that offer mortgages and sales assistance. To do so, the company has built a variety of products on top of their extensive database. Figure 3 shows the interactions, gains and costs of the three “sides” of the Zillow MSP. Government provides data but gains a more efficient property market instead. Service providers pay for leads and get information about interested buyers and interested buyers and renters get access to relevant information.

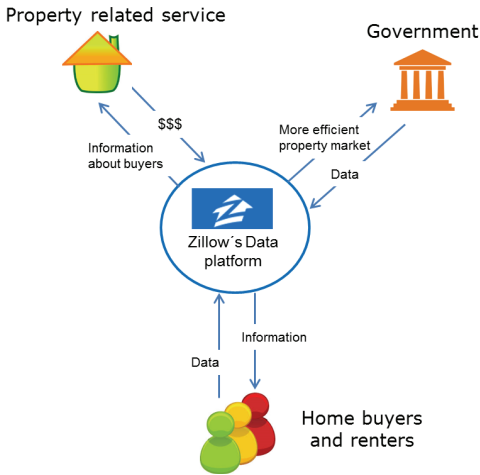


Figure 3: Zillow’s MSP, connecting home buyers/renters and service providers

<sup>29</sup> <http://priceconomics.com/the-seo-dominance-of-zillow/>

## VIII.5 Investing in Open Data MSPs: Insights from Real Options Theory

An open data ecosystem consists of many different participants: governments, academia, civil society organizations, businesses and citizens. This type of ecosystem is circular in nature, building upon a complicated network of value that is generated by different participants that are creating valuable information as well as products and services. The value appropriated is oftentimes intangible in nature and in many cases the currency is data, rather than money, as shown in the three examples above. Network externalities are created when information that is created in part based on open government data draws users to information platforms as the information is partly based on data provided by these users and becomes more valuable the more people participate. This information production and consumption activity can be utilized to create economic value which will attract even more players to the ecosystem. More use of data will also eventually benefit the data providers, even without them getting monetary reimbursements, as market participants collectively address various social challenges which can rarely be solved by governments alone. However, before this scenario can happen, the government data must be open enough and of high enough quality to be of use for entrepreneurs and the ecosystem must contribute other factors, like various important skills, technologies and low rate funds. It is important that both data providers on the supply side and prospective data users on the demand side have a relatively clear idea about potential gains from future use and what is needed in order for those benefits to be harvested.

Data providers must appropriately recognize two important elements of this ecosystem.

1) A significant share of the value generated is intangible, resulting from improved decision making and changed behavior which can impact society and the environment as a whole. 2) Those that finally appropriate the value might be far removed from those that provide the resources. Therefore, a public sector organization might invest in gathering data, ensuring the data quality and making the data open across different dimensions, but future use of these data will in a minority of cases directly benefit the organization itself. The organization depends on the yearly budget allocations from government to sustain their activities. If top-level decision makers do not sufficiently understand the complicated mechanisms that explain how much value is generated from the data, they might reduce this funding if the data are not being used, and the level of openness and quality of data might become compromised as a result. This is an *open data value paradox*, describing a situation where entrepreneurs do not use the data as the data are not usable enough and there is too much uncertainty over future

provision of open data. However, data providers are not willing to invest in the people and technology necessary to make the data more usable and sustainable unless they observe some evidence of value generation, which again depends on data being used.

In order to resolve this paradox which might lead to a downward spiraling cycle, as discussed in the introduction, we propose to use the valuation methods and ideas used in economics of real options. Uncertainty and the option holder's ability to respond to it (flexibility) are the source of value for an option. When a firm buys a traditional resource, it has already made an investment decision. Alternatively, when a firm is provided with the opportunity to use open data it has gained an *option to use* a resource, but it does not have to exercise this option right away. As contemporary firms face intense rivalry, globalization, technological change and time-to-market pressures, it is agility, defined as the ability to detect and seize market opportunities with speed and surprise, that is considered to be an imperative for business success (Sambamurthy et al. 2003). Companies know that they need to be agile and able to react quickly to external changes. Therefore, an option to use a resource, without any current commitment, must be of some worth to companies and to society as a whole. The question remains, how much is it worth?

Financial options capture a specific investment opportunity to which the holder has preferential advantage (Sandberg et al. 2014). Options theory states that a financial option provides the option holder with the right to buy (call) or sell (put) a specified quantity of an underlying financial asset at a fixed price (called a strike price or an exercise price) at or before the expiration date of the option (Black and Scholes 1973). The payoff to the option holder is dependent on the price development of the underlying financial asset which will influence whether the user will exercise the option or not. Investing in financial options can be compared to buying insurance. The maximum loss is the payment for the right to exercise the option, while the upside potential is theoretically unlimited, but depends on the price development of the underlying financial asset. Therefore, the volatility of the underlying asset price will positively influence the worth of an option, as it creates a higher probability of considerable gain.

*Real options* theory further extends financial option theory by using the same kind of logic for valuing real investment opportunities, where, unlike many other investment valuation techniques, uncertainty over future outcomes is viewed as a positive factor (Bowman and Hurry 1993). The real options valuation approach in strategic management describes how organizations position themselves to seize emergent opportunities. The theory provides insight into how tangible as well as intangible

resources can act as options that enable strategic action (Adner and Levinthal 2004). Real options are generally described as rights to future investment choices without a current obligation for full investment (Sambamurthy et al. 2003). The activation of real options is often seen as a form of incremental decision-making on investments, originating with a so-called shadow option that is present but has not been realized to become real options when exercised (Bowman and Hurry 1993). After recognizing an option as such, the holder of an option typically makes a small initial investment, holds it open until an opportunity arrives, and then exercises a choice to strike the option and capture the value inherent in that opportunity (Bowman and Hurry 1993, Gosh and Li 2013). The identification of real options is, to a significant extent, subject to contingencies such as the firm's technological capabilities, experience, and absorptive capacity, making the identification of real options virtually unique to every firm (Sarikko 2014). The value of holding an option becomes magnified especially when the options holder has preferential advantages in exploiting the opportunity provided by the option (Sambamurthy et al. 2003).

Within the Information Systems research field, real options have been used to offer a novel perspective called *digital options*. Digital options can be described as a set of IT-enabled capabilities in the form of digitized enterprise work processes and knowledge systems which create value through increased reach and richness of digitized processes and digitized knowledge (Sambamurthy et al. 2003, Overby et al. 2006). Digital options are at once a means of not only preserving the opportunity to capitalize on a new technology or practice but also of mitigating the risks induced by technological and market uncertainty (Woodard et al. 2013). While the concept of digital options has been applied in studies on ERP-systems investments, it has also received criticism for its apparent lack of detail in certain key aspects (Sarikko 2014). It has been argued that restricting digital options to process and knowledge reach and richness limits the concept's generative potential as well as its relevance to IT capabilities (Sandberg et al. 2014).

Fichman (2004) compares and contrasts IT platform valuation through the lens of Discounted Cash Flow analysis (DCF), on the one hand, and through the lens of real options valuation, on the other, showing how real options thinking will capture mechanisms that are important to the firm's competitive advantage, although the *value might be intangible and neglected through methods such as DCF* (Fichman 2004, 139). Two of the discussed real option value determinants in Fichman's model have a special relevance to open data: *susceptibility to network externalities* (the extent to which a technology increases in value to individual adopters with the size of the adoption

network) and *interpretive flexibility* (the extent to which a technology permits multiple interpretations on the part of adopters about how it should be implemented and used).

## **VIII.6 Growing sustainable value from open data**

While it is extremely difficult to predict how and for what purposes open data will be used, we propose that certain enabling factors will influence the opportunity, motivation, and ability of open data ecosystem participants to use open data for sustainable value generation (Jetzek et al. 2014b). These factors will influence how quickly data that have been opened up by the government will be put to good use, creating a bigger set of possibilities for value generation (variability) and therefore influencing the perceived option value of data. Drawing from real options theory, the option value of data is not equal to the current value of data, but rather an addition to it. The option value is based on the probability of open data being used to generate value in the future. More use creates more variability in the ways through which value will be generated and increases the network effects, thereby positively influencing the option value. As for any other risky investment, the higher the perceived future value or profitability, the more easily this opportunity will attract money and investors. To raise the option value of data, governments can thus focus on improving the opportunity, ability, and motivation of public and private stakeholders to use these data for value generation. Higher option value will motivate stakeholders to use the data, which will consequently underpin future growth of information, products and services based on these data and subsequent generation of sustainable value, thus creating a virtuous cycle of value generation and use of the data.

Drawing from the economics of two-sided markets, supported by three use cases of open data MSPs, we propose that open data intermediaries or MSPs will use data to generate information that is freely disseminated to users, and the users themselves become an additional source of data, thus creating positive network externalities that will be used to attract the “other side” to the platform. These business models will contribute to the generation of sustainable value through use of data, as the MSP business model supports the generation of intangible value that is usually not rewarded by economic profits. In the case of MSPs, this type of value generation is utilized as a tool to attract paying customers, and the two sides will continue to feed upon each other.

By extending real options thinking to the societal level, we propose that government is, by disseminating open data, essentially writing an option (or a bundle of options in the case of multiple open datasets) where more open distribution of more types of data will

increase the variability in potential outcomes and therefore raise the upside potential, while the maximum loss for the government is the investment made in making data fit for re-use (given that these data have already been collected) and the eventual loss of income from data. Of course, the decision to invest in open data in the first place also depends on the perceived option value of data. We argue that if governments recognize the option value of open data for potential users, they will be more willing to continue to provide high-quality open data to users, even if doing so does demand some further investments in people and technology.

As the value of data is dependent on network externalities, we propose that the open data real option value increases with more use of the data, but with diminishing marginal returns due to the market being at some point saturated. This effect is not directly reflected in the model below, but for those that would like to calculate the potential impact, we suggest using growth formulas, such as that of von Bertalanffy (von Bertalanffy 1938). In that case, the growth factor (K) would be dependent on the enabling factors we present below and, as suggested above, governments could influence the option value of data by focusing on these factors, as they have been found to increase use of the data (Jetzek et al. 2013). Different growth factors will lead to different outcomes, which can be assigned probabilities for a more accurate estimation of the “distribution” of possible outcomes. Of course, this valuation is based on an estimation of the “base value” (or use) of the data (as the underlying asset). The option holders (i.e., all those that can access and use the data) are influenced by these same factors, but also by their relative abilities as compared to others. Hence, the value of the option is unique to them, reflecting their own capabilities. We do not model the organizational level factors here, in order to preserve clarity of representation. We propose only that the eventual users will be influenced by the perceived value of the option they hold, and leave more detailed organizational level modelling to future research.

To identify potential determinants of sustainable value of open data, we have looked to previous research, where the most important enablers and barriers of open data have been analyzed, as well as relying on interviews and participation in an open data initiative in Denmark. The underlying assumption we make is that people are generally willing (intrinsic motivation) to use data for sustainable value generation if they are given the opportunity and they have the ability to do so. Additionally, certain structures in the economy can influence extrinsic motivation, both negatively and positively. Governments can influence extrinsic motivation of firms by reducing risk and



uncertainty, by creating an environment that encourages investors and entrepreneurs and by highlighting the opportunities present.

We have developed the following propositions around these relationships:

*P1: IT infrastructure positively influences perceived option value of open data (ability)*

Here we propose that use of data is dependent on the level of technical infrastructure in a country. Technical infrastructure facilitates data exchange between government agencies and the public and influences the ability of stakeholders to use the data. Concerning individual organizations that would like to use data, the perceived option value to them is dependent on the level of how their own IT infrastructure compares to that of other ecosystem participants.

*P2: Data-related skills positively influence perceived option value of open data (ability)*

The growth of data-related skills within a society increases the ability of stakeholders to generate value from data, and thus the probability that the data will be used for value generation. For individual organizations that would like to use data, the perceived option value to them is dependent on the skills they have as relative to those of other ecosystem participants.

*P3: Data governance positively influences perceived option value of open data (opportunity)*

Better data governance will improve data quality, so that data are accurate, complete, updated and reliable. Moreover, if governance is good, it is more likely that the open data initiative that provides the data will be sustainable. Improved data quality will, in turn, enable better and more trustworthy information to be generated from the data and therefore influencing the opportunity for value generation. The option value of data for the potential user should be higher as a result of this opportunity.

*P4: Openness of data positively influences perceived option value of open data (opportunity)*

As the openness of a set of data increases, it becomes easier to external stakeholders to access and re-use the data. Accordingly, openness creates an *opportunity* for value generation. The option value of data for the potential user should be higher as a result of this opportunity.

*P5: Uncertainty about data protection negatively influences perceived option value of open data (motivation)*

We propose that data are used more if there is *confidence and trust in the legal infrastructure* that guards individual privacy and guides those that want to use data to generate value. Less risk of data fraud will motivate data users to actively participate in responsible data use and re-use, therefore positively influencing use of data. Accordingly, uncertainty over rights and responsibilities and data ownership are likely to negatively influence the motivation to use data therefore negatively impacting the perceived option value.

*P6: Collaboration positively influences perceived option value of open data (motivation)*

We propose that data are used more if government actively engages and collaborates with external stakeholders in order to motivate private and public stakeholders to use data for various use and subsequent value generation. This collaboration can happen via public-private partnerships, hackathons or living labs or other types of formal and informal interactions between different stakeholders in the open data ecosystem.

*P7: The risk-free rate will negatively influence perceived option value of open data (motivation)*

The higher the risk-free rate, the more likely it is that money will be used for risk-less investment, rather than high risk investment. Therefore, high risk-free rate negatively influences the probability of investment and use and thereby the perceived option value of data.

*P8: Perceived option value of open data positively influences investment in MSPs.*

The higher the perceived option value of data, the more likely it is that intermediaries will invest in MSPs.

*P9: Investment in MSPs supports the generation of information, products and services based on data and therefore positively influences sustainable value*

Various stakeholders can provide information; products and services based on the data through these platforms and use the network and market mechanisms to generate valuable synergies. The more the data are used and the more synergy is created between the network mechanisms and market mechanisms that facilitate dissemination of information on the one hand, and data-driven products and services on the other, the more sustainable value will be generated and appropriated.

The model itself is presented in Figure 4.

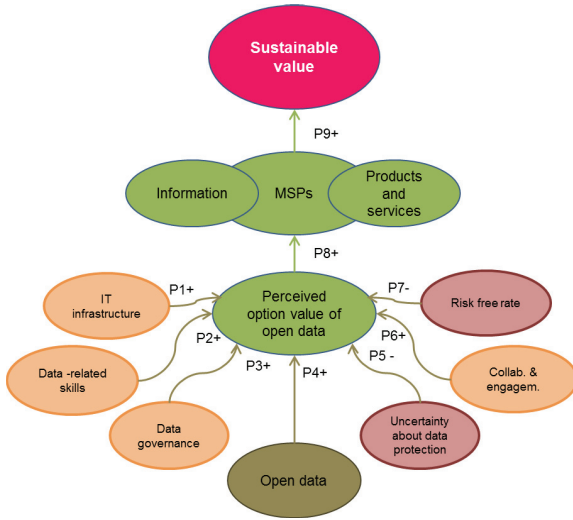


Figure 4: Model of sustainable value generation in the open data ecosystem

The various enabling factors are like the “roots” of the open data ecosystem “plant” and their main role is to provide nourishment so that the seed-like data can grow into something of value. Each of these factors will influence the opportunity, ability, or motivation of stakeholders in the ecosystem to use data for value generation, thereby contributing to a larger set of potential outcomes and increasing variability, which will positively influence the perceived option value of data. If the perceived option value is high, stakeholders will be more willing to make the investments necessary to establish MSPs as data intermediaries, despite the inherent risks. The establishment of MSPs will furthermore contribute to the generation of sustainable value, as they allow for an interaction between diverse types of affiliated stakeholders and play the network and market mechanisms against each other to create synergies that contribute to the generation of social, environmental, and economic types of value.

## VIII.7 Discussion

Our societies are changing fast, faster than many of us realize in the midst of things. The interaction between technological and social elements is a big driver in these changes, influencing not only our ability to generate value but also the way people perceive and think about value (which might at the individual level be more accurately described as *values*). We have not discussed these individual trends in depth here, but

suffice it to say that technology and network capabilities have come together to create vast amounts of data that are currently being transformed into information and used as a resource in new products and services by a multiplicity of stakeholders. This new data-driven ecosystem is highly dependent on unstructured many-to-many relationships where data and information are flowing through networks without any monetary transactions taking place, as opposed to the structured value chains of the industrial economy. Network capabilities have allowed for much more complex interactions between stakeholders, and old intermediaries have been cut out while new ones have been created. The new intermediaries are effectively playing the network mechanisms and market mechanisms against each other – using network externalities as a tool to generate the income which is necessary to sustain investments in people and technology, while simultaneously contributing to sustainable value.

In the context of this paper, we have used the term open data mostly for data generated and disseminated by *governments*, as they are currently the biggest distributors of open data in the world. We have proposed that *openness* is in itself an important enabler to the creation of sustainable value from data. Openness enables both generation and appropriation of value, not only by the organization that owns the data but also by external stakeholders. However, while openness of data might be a necessary condition for external stakeholders that want to effectively utilize the vast amounts of government data, it is insufficient on its own. Just as governments aim to provide the necessary infrastructure for efficient markets, they should be aware of the factors that are needed for a thriving data ecosystem. Such an ecosystem relies to a large degree on the generation of relevant information, which is further disseminated through network based mechanisms to generate value around society. The network mechanisms do facilitate the appropriation of value by society's stakeholders but operate under different rules than the traditional market mechanisms.

We have made a few propositions about how sustainable value can be created and how multi-sided platforms are enabling such value generation, as they are not completely bound by rent-seeking; rather, they gain from stakeholders that together are addressing complicated societal challenges, previously the responsibility of governments alone. Governments have started to realize the power of these models which thrive on *sharing* and *interactions*, and are even creating their own platforms where public sector, businesses and citizens can meet and interact to create superior sustainable value (Janssen and Estevez, 2013). However, as in any complicated market, there are various challenges present. One barrier that has been identified in prior work on MSPs is the chicken-and-egg problem, describing the need to build up a sufficient number of

participants on one side of the platform in order to attract the other side, which, in the case of open data, is usually the paying side. In the case of government provision of open data, this translates to government attracting enough users to justify the investments required for making data open. When the users come, value will be generated – but the users will not participate unless they have a current perception of the future value to be gained.

The economics of real options help us conceptualize the worth of perceived future value by building on the same ideas that underpin the financial options markets: The limited risk and the unlimited upside as well as the ability to wait and see before striking the option. Using this type of thinking might help resolve the open data value paradox. If supplying open data is conceptualized as the act of writing an option that is handed out to all market participants, we gain a tool that can help us evaluate the potential gain, viewing unpredictability and variability as a positive factor rather than as a negative one and focusing on the flexibility provided as the data are out there when the company in question needs them. Trusting that the companies will value the option they are giving, governments can focus on making data more open and create a nurturing environment for interested stakeholders, which might in turn raise the option value even further. The potential users will eventually pay back, not only by creating jobs and paying taxes but by finding innovative solutions to some of our most pressing societal problems.

### **VIII.8 Limitations, Implications and Conclusion**

The model presented in Figure 4 was created as a part of exploratory research focusing on the emerging phenomenon of open data. The goal is to uncover and visualize the complex relationship between open data, the enabling factors and barriers that can impact how much data are used and the resulting generation of sustainable value. The model has a number of limitations as such. The most obvious limitation is that a simple model will not do justice to the level of complexity in such a system. However, for conceptual clarity, we highlight only the societal level factors that are, based on our research and state of the art literature on open data, most important for our argument. Many of the constructs presented are highly dimensional, and these dimensions are also important for creating more depth in our understanding of the underlying relationships. It is important to delve deeper into the lower level mechanisms that can explain these high-level relationships, and we hope that this will be done in future research. A related limitation is that this model presents only the societal level relationships, while there are, of course, many factors at the organizational (or even

individual) level that will decide if, and what, companies decide to “strike the option” – for instance, their ability to make the necessary investments in technology and knowledge and their absorptive capacity (Jetzek et al. 2014a).

There are several implications for theory and practice. The first practical implication is for public organizations that are disseminating open data. In many cases, the necessary investments in people and technology have been based on the faith and belief that when open data as a valuable resource are made available, value generation will happen (the “*if you build it they will come*” argument). In other cases, such initiatives have been based on careful planning and business case evaluation, where the resulting (foreseen) value is transformed to a monetary equivalent and used to calculate Net Present Value (NPV). In both cases, the unforeseen or serendipitous future value generation that makes the act of open access so alluring is not explicitly evaluated, although this type of value is in many cases what drives these initiatives.

We suggest that it might prove helpful to governments to view the act of publishing open data as writing an option. Real options evaluation methods put value on flexibility and governments provide businesses with flexibility when they disseminate high quality data that the businesses can use whenever convenient. However, the company needs to wait for the right circumstances in their environment and in their organization. One easy way to visualize the potential future gains at the societal level is to create different configurations of the enabling factors that will result in low to high use of open data, resulting in a distribution of the value growth factors, from which the predicted future value of data can be calculated (a type of fuzzy set approach, see Lee and Lee 2011). This approach can make it easier to make the future value of open data more explicit, recognizing the potential without trying to foresee every possible use case that might result in tangible or intangible value generation in the future.

The second practical implication is for private firms that would like to utilize the torrents of available data to build a successful business. Not only is the business model of multi-sided platforms well suited to support free dissemination of information, it also creates an attractive marketplace for smaller companies that would like to use data to create data-driven products and services in order to make these products and services available to multiple users, without having the means to establish a platform of their own. While current MSPs range from relatively closed to relatively open, there is usually a degree of openness, as these platforms’ main function is to lead together a large number of users on both sides of the platform, benefitting both the information consumers and the information producers. The more open the platform, the more of the

value generated will benefit society - since, other things being equal, there will be more participants and more network externalities (Parker and Van Alstyne 2014).

We have contributed to knowledge about how use of open data can result in the generation of sustainable value by further extending the theory of multi-sided platforms to open data intermediaries. The use of this theory helps explain how multiple stakeholders can, through their dissemination and use of open data, create synergies that result in the simultaneous generation of social, environmental, and economic value – or what we conceptualize as sustainable value. Moreover, we suggest that by extending the simple logic that is used in economics of real options, we have theoretically supported the argument that the openness of data creates an opportunity for value generation and will accordingly positively influence perceived option value, as more variability of outcomes increases the likelihood of a favorable result. Drawing from the inherent chicken-and-egg problem that often inhibits investment in MSPs, we have looked at the open data value paradox, where the lack of use of data results in insufficient levels of openness and data governance, which then leads to less use. We propose that our model might help resolve this paradox.

In conclusion, we believe that the unique features of open data offer the potential for unprecedented generation of sustainable value. This turn of events, however, is not inevitable, and there are a number of different factors that governments and other participants in the open data ecosystem have to consider. These factors will contribute to the motivation, opportunity, and ability of individuals to use data for sustainable value generation. Our proposition is that when these factors are in place, both governments and other possible platform owners will start to perceive that the option value of data – reflecting the potential for future gain – is high enough for them to start investing in MSPs. These platforms will enable individuals to appropriate value from open data through consumption of information and information services. Smaller entrepreneurs will moreover gain the ability to create and market products and services to those individuals via these platforms. The two-sided markets business model will offer the ability to create a “Win-Win-Win” situation where governments, businesses and individuals all gain through a complex network of sharing and co-creating data, information, and information services for sustainable value. If the synergies between public/private and social/economic domains can be exploited in this way, we believe we have the potential for a quantum leap in increased productivity and social progress in the near future.

## VIII References

- Adner, Ron, and Daniel A. Levinthal. "What is not a real option: Considering boundaries for the application of real options to business strategy." *Academy of Management Review* 29(1):74-85.
- Bailey, Joseph P. and Yannis Bakos. 1997. "An exploratory study of the emerging role of electronic intermediaries." *International Journal of Electronic Commerce* (1997): 7-20.
- Bakici, Tuba, Esteve Almirall, and Jonathan Wareham. 2013. "The role of public open innovation intermediaries in local government and the public sector." *Technology Analysis & Strategic Management* 25(3): 311-327.
- Barney, Jay. 1991. "Firm resources and sustained competitive advantage." *Journal of management* 17(1): 99-120.
- Benkler, Yochai. 2006. *The wealth of networks: How social production transforms markets and freedom*. New Haven: Yale University Press, 2006.
- Bharadwaj, Anandhi, Omar A. El Sawy, Paul A. Pavlou, and N. Venkatraman. 2013. "Digital business strategy: Toward a next generation of insights." *MIS Quarterly* 37(2):471-482.
- Black, Fischer and Myron Scholes. 1973. "The pricing of options and corporate liabilities." *The Journal of Political Economy* 81(3):637-654.
- Bowman, Edward H. and Dileep Hurry. 1993. "Strategy through the option lens: An integrated view of resource investments and the incremental-choice process." *Academy of Management Review* 18(4):760-782
- Brynjolfsson, Erik, and JooHee Oh. 2012. "The attention economy: measuring the value of free digital services on the Internet," in the *Proceedings of the 33rd International Conference on Information Systems (ICIS)*, Orlando 2012.
- Buytendijk, Frank. 2014. "Hype Cycle for Big Data, 2014." <https://www.gartner.com/doc/2814517/hype-cycle-big-data->
- Caillaud Bernard and Bruno Jullien. 2003. "Chicken and Egg: Competition Among Intermediation Service Providers." *RAND Journal of Economics* 34:309-328.
- Cannon, Sarah and Lawrence H. Summers. 2014. "How Uber and the Sharing Economy Can Win Over Regulators *Harvard Business Review*." <https://hbr.org/2014/10/how-uber-and-the-sharing-economy-can-win-over-regulators/>
- Chesbrough, Henry W. 2003. "The Era of Open Innovation." *Sloan Management Review* 44(3):35-41.
- Conradie, Peter and Sunil Choenni. 2014. "On the barriers for local governments releasing open data." *Government Information Quarterly* 31(1):10-17.
- Cuddy, Amy J.C., Kyle T. Doherty, and Maarten W. Bos. 2012. "OPOWER: Increasing Energy Efficiency through Normative Influence." *Harvard Business School*, 9-911-016 [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2017957](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2017957)
- Davies, Tim. 2013. "Open Data Barometer 2013 Global Report." *Open Data Institute and World Wide Web Foundation*. <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>



- Ghosh, Suvankar, and Xiaolin Li. "A Real Options Model for Generalized Meta-Staged Projects-Valuing the Migration to SOA." *Information Systems Research* 24(4):1011-1027.
- Hagiu, Andrei and Julian Wright. 2011. "Multi-Sided Platforms." Harvard Business School Working Paper. <http://hbswk.hbs.edu/item/6681.html>
- Janssen, Marijn, Yannis Charalabidis and Anneke Zuiderwijk. 2012. "Benefits, adoption barriers and myths of open data and open government." *Information Systems Management* 29(4):258-268.
- Janssen, Marijn and Elsa Estevez. 2013. "Lean government and platform-based governance - Doing more with less." *Government Information Quarterly* 30(1):S1-S8.
- Janssen, Marijn and Anneke Zuiderwijk. 2014. "Infomediary Business Models for Connecting Open Data Providers and Users." *Social Science Computer Review* 32(5):694-711.
- Jetzek Thorhildur, Michel Avital, and Niels Bjørn-Andersen. (2013). "Generating Value from Open Government Data," in the Proceedings of the International Conference on Information Systems (ICIS), Milan 2013.
- Jetzek, Thorhildur, Michel Avital and Niels Bjørn-Andersen. 2014a. "Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research* 9(2):100-120.
- Jetzek, Thorhildur, Michel Avital and Niels Bjørn-Andersen. 2014b. "Generating sustainable value from open data in a sharing society. " In *Creating Value for All Through IT*, edited by Birgitta Bergvall-Kåreborn and Peter Axel Nielsen, 243-260. Berlin Heidelberg: Springer.
- Katz, Michael L. and Carl Shapiro. "Network externalities, competition, and compatibility." *The American economic review* (1985):424-440.
- Katz, Michael L., and Carl Shapiro. "Technology adoption in the presence of network externalities." *The Journal of Political Economy* (1986):822-841.
- Lee, Young-Chan, and Seung-Seok Lee. 2011. "The valuation of RFID investment using fuzzy real option." *Expert Systems with Applications* 38(10):12195-12201.
- Lindman, Juho, Tomi Kinnari, and Matti Rossi. 2014. "Industrial open data: Case studies of early open data entrepreneurs," in *System Sciences (HICSS), 47th Hawaii International Conference on System Science*, 739-748, IEEE.
- Michael E. Porter and Mark R. Kramer. 2011. "Creating Shared Value." *Harvard Business Review* 89(1):62-77.
- Michael E. Porter. 1985. *Competitive Advantage: Creating and Sustaining Superior Performance*. New York.: Simon and Schuster.
- Martin, Sébastien, Muriel Foulonneau, Slim Turki, and Madjid Ihadjadene. 2014. "Risk Analysis to Overcome Barriers to Open Data." *Electronic Journal of e-Government* 11(2):348 -359.
- Mayer-Schönberger, Victor and Zarino Zappia. 2011. "Participation & Power: Intermediaries of Open Data." In *1st Berlin Symposium on Internet & Society*, October.
- McKinsey. 2013. "Open data: Unlocking innovation & performance with liquid information." McKinsey Global Institute, McKinsey Center for Government & McKinsey Business Technology Office.
- Nilsen, Kirsti. 2010. "Economic theory as it applies to Public Sector Information." *Annual Review of Information Science and Technology* 44(1):419-489.

- OECD. 2011. "Fostering Innovation to Address Social Challenges." <http://www.oecd.org/sti/inno/47861327.pdf>
- OECD. 2014. "Data-driven Innovation for Growth and Well-being. Interim Synthesis Report." <http://www.oecd.org/sti/inno/data-driven-innovation-interim-synthesis.pdf>
- Overby, Eric, Anandhi Bharadwaj, and V. Sambamurthy. 2006. "Enterprise Agility and the Enabling Role of Information Technology." *European Journal of Information Systems* 15(2):120–131.
- Parker, Geoffrey and Marshall Van Alstyne. 2014. "Innovation, Openness and Platform Control." Working Paper Series, SSRN [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1079712](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1079712)
- Resnick, Paul, Richard Zeckhauser, and Chris Avery. 1995. "Roles for electronic brokers." In *Toward a Competitive Telecommunication Industry: Selected Papers from the 1994 Telecommunications Policy Research Conference*, 289-304, Mahwah, NJ: Lawrence Erlbaum Associates.
- Rochet, Jean-Charles and Jean Tirole. 2006. "Two-sided markets: A progress report." *The RAND Journal of Economics* 37(3):645-667.
- Sambamurthy, V., Anandhi Bharadwaj, and Varun Grover. 2003. "Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms." *MIS Quarterly* 27(2):237–263.
- Sandberg, Johan, Lars Mathiassen, and Nannette Napier, N. 2014. "Digital Options Theory for IT Capability Investment." *Journal of the Association for Information Systems* 15(7):422-453.
- Sarikko, Ted. 2014. "Here Today, Here Tomorrow: Considering Options Theory in Digital Platform Development." In *Creating Value for All Through IT*, edited by Birgitta Bergvall-Kåreborn and Peter Axel Nielsen, 243-260. Berlin Heidelberg :Springer.
- van Osch, Wietske and Michel Avita. 2010. "The road to Sustainable Value: The path-dependent construction of sustainable innovation as sociomaterial practices in the car industry." *Advances in Appreciative Inquiry* 3(2010): 99-116.
- van Veenstra, Anne Fleur and Tijs A. van den Broek. 2013. "Opening Moves – Drivers, Enablers and Barriers of Open Data, in the Proceedings of 12th IFIP WG 8.5 International Conference on Electronic Government (EGOV 2013), edited by Maria A. Wimmer, Marijn Janssen, and Hans Jochen Scholl, 50-61, Koblenz, Germany.
- von Bertalanffy, Ludwig. 1938. "A quantitative theory of organic growth (inquiries on growth laws. II)." *Human biology* 10(2):181-213.
- Wade, Michael, and John Hulland. 2004. "Review: the resource-based view and information systems research: review, extension, and suggestions for future research." *MIS Quarterly* 28(1):107-142.
- Wernerfelt, Birger. 1984. "A resource-based view of the firm." *Strategic Management Journal* 5(2):171–180
- West, Joel and Scott Gallagher. 2006. "Challenges of open innovation: the paradox of firm investment in open-source software." *R&D Management* 36(3):1-32.
- Woodard, C. Jason, Narayan Ramasubbu, F. Ted Tschang, and Vallabh Sambamurthy. 2013. "Design capital and design moves: the logic of digital business strategy." *MIS Quarterly* 37(2):537-564.
- Zuiderwijk, Anneke and Marijn Janssen. 2014a. "Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View." in *Open Government*,

Opportunities and Challenges for Public Governance, edited by Mila Gascó-Hernández, 115-136, New York: Springer.

Zuiderwijk, Anneke and Marijn Janssen, M. (2014b). "Open data policies, their implementation and impact: A comparison framework." *Government Information Quarterly* 31(1):17-29.

Zuiderwijk, Anneke, Marijn Janssen, Sunil Choenni, and Robert Meijer. 2014. "Design principles for improving the process of publishing open data." *Transforming Government: People, Process and Policy* 8(2):

Zuiderwijk, Anneke, Marijn Janssen, Sunil Choenni, Robert Meijer, and R. Sheikh Alibaks. (2012). "Socio-technical impediments of open data." *Electronic Journal of eGovernment* 10(2):156-172.



**TITLER I PH.D.SERIEN:****2004**

1. Martin Grieger  
*Internet-based Electronic Marketplaces and Supply Chain Management*
2. Thomas Basbøll  
*LIKENESS  
A Philosophical Investigation*
3. Morten Knudsen  
*Beslutningens vaklen  
En systemteoretisk analyse af moderniseringen af et amtskommunalt sundhedsvæsen 1980-2000*
4. Lars Bo Jeppesen  
*Organizing Consumer Innovation  
A product development strategy that is based on online communities and allows some firms to benefit from a distributed process of innovation by consumers*
5. Barbara Dragsted  
*SEGMENTATION IN TRANSLATION AND TRANSLATION MEMORY SYSTEMS  
An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process*
6. Jeanet Hardis  
*Sociale partnerskaber  
Et socialkonstruktivistisk casestudie af partnerskabsaktørers virkelighedsopfattelse mellem identitet og legitimitet*
7. Henriette Hallberg Thygesen  
*System Dynamics in Action*
8. Carsten Mejer Plath  
*Strategisk Økonomistyring*
9. Annemette Kjærgaard  
*Knowledge Management as Internal Corporate Venturing*
10. Knut Arne Hovdal  
*De professionelle i endring  
Norsk ph.d., ej til salg gennem Samfundslitteratur*
11. Søren Jeppesen  
*Environmental Practices and Greening Strategies in Small Manufacturing Enterprises in South Africa  
– A Critical Realist Approach*
12. Lars Frode Frederiksen  
*Industriel forskningsledelse  
– på sporet af mønstre og samarbejde i danske forskningsintensive virksomheder*
13. Martin Jes Iversen  
*The Governance of GN Great Nordic  
– in an age of strategic and structural transitions 1939-1988*
14. Lars Pynt Andersen  
*The Rhetorical Strategies of Danish TV Advertising  
A study of the first fifteen years with special emphasis on genre and irony*
15. Jakob Rasmussen  
*Business Perspectives on E-learning*
16. Sof Thrane  
*The Social and Economic Dynamics of Networks  
– a Weberian Analysis of Three Formalised Horizontal Networks*
17. Lene Nielsen  
*Engaging Personas and Narrative Scenarios – a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca*
18. S.J Valstad  
*Organisationsidentitet  
Norsk ph.d., ej til salg gennem Samfundslitteratur*

19. Thomas Lyse Hansen  
*Six Essays on Pricing and Weather risk in Energy Markets*
20. Sabine Madsen  
*Emerging Methods – An Interpretive Study of ISD Methods in Practice*
21. Evis Sinani  
*The Impact of Foreign Direct Investment on Efficiency, Productivity Growth and Trade: An Empirical Investigation*
22. Bent Meier Sørensen  
*Making Events Work Or, How to Multiply Your Crisis*
23. Pernille Schnoor  
*Brand Ethos  
Om troværdige brand- og virksomhedsidentiteter i et retorisk og diskursteoretisk perspektiv*
24. Sidsel Fabech  
*Von welchem Österreich ist hier die Rede?  
Diskursive forhandlinger og magtkampe mellem rivaliserende nationale identitetskonstruktioner i østrigske pressediskurser*
25. Klavs Odgaard Christensen  
*Sprogpolitik og identitetsdannelse i flersprogede forbundsstater  
Et komparativt studie af Schweiz og Canada*
26. Dana B. Minbaeva  
*Human Resource Practices and Knowledge Transfer in Multinational Corporations*
27. Holger Højlund  
*Markedets politiske fornuft  
Et studie af velfærdens organisering i perioden 1990-2003*
28. Christine Mølgaard Frandsen  
*A.s erfaring  
Om mellemværendets praktik i en transformation af mennesket og subjektiviteten*
29. Sine Nørholm Just  
*The Constitution of Meaning – A Meaningful Constitution? Legitimacy, identity, and public opinion in the debate on the future of Europe*
- 2005**
1. Claus J. Varnes  
*Managing product innovation through rules – The role of formal and structured methods in product development*
2. Helle Hedegaard Hein  
*Mellem konflikt og konsensus – Dialogudvikling på hospitalsklinikker*
3. Axel Rosenø  
*Customer Value Driven Product Innovation – A Study of Market Learning in New Product Development*
4. Søren Buhl Pedersen  
*Making space  
An outline of place branding*
5. Camilla Funck Ellehave  
*Differences that Matter  
An analysis of practices of gender and organizing in contemporary work-places*
6. Rigmor Madeleine Lond  
*Styring af kommunale forvaltninger*
7. Mette Aagaard Andreassen  
*Supply Chain versus Supply Chain Benchmarking as a Means to Managing Supply Chains*
8. Caroline Aggestam-Pontoppidan  
*From an idea to a standard  
The UN and the global governance of accountants' competence*
9. Norsk ph.d.
10. Vivienne Heng Ker-ni  
*An Experimental Field Study on the*

- Effectiveness of Grocer Media Advertising*  
*Measuring Ad Recall and Recognition, Purchase Intentions and Short-Term Sales*
11. Allan Mortensen  
*Essays on the Pricing of Corporate Bonds and Credit Derivatives*
12. Remo Stefano Chiari  
*Figure che fanno conoscere*  
*Itinerario sull'idea del valore cognitivo e espressivo della metafora e di altri tropi da Aristotele e da Vico fino al cognitivismo contemporaneo*
13. Anders McIlquham-Schmidt  
*Strategic Planning and Corporate Performance*  
*An integrative research review and a meta-analysis of the strategic planning and corporate performance literature from 1956 to 2003*
14. Jens Geersbro  
*The TDF – PMI Case*  
*Making Sense of the Dynamics of Business Relationships and Networks*
15. Mette Andersen  
*Corporate Social Responsibility in Global Supply Chains*  
*Understanding the uniqueness of firm behaviour*
16. Eva Boxenbaum  
*Institutional Genesis: Micro – Dynamic Foundations of Institutional Change*
17. Peter Lund-Thomsen  
*Capacity Development, Environmental Justice NGOs, and Governance: The Case of South Africa*
18. Signe Jarlov  
*Konstruktioner af offentlig ledelse*
19. Lars Stæhr Jensen  
*Vocabulary Knowledge and Listening Comprehension in English as a Foreign Language*
- An empirical study employing data elicited from Danish EFL learners*
20. Christian Nielsen  
*Essays on Business Reporting*  
*Production and consumption of strategic information in the market for information*
21. Marianne Thejls Fischer  
*Egos and Ethics of Management Consultants*
22. Annie Bekke Kjær  
*Performance management i Process-innovation*  
*– belyst i et social-konstruktivistisk perspektiv*
23. Suzanne Dee Pedersen  
*GENTAGELSENS METAMORFOSE*  
*Om organisering af den kreative gøren i den kunstneriske arbejdspraksis*
24. Benedikte Dorte Rosenbrink  
*Revenue Management*  
*Økonomiske, konkurrencemæssige & organisatoriske konsekvenser*
25. Thomas Riise Johansen  
*Written Accounts and Verbal Accounts*  
*The Danish Case of Accounting and Accountability to Employees*
26. Ann Fogelgren-Pedersen  
*The Mobile Internet: Pioneering Users' Adoption Decisions*
27. Birgitte Rasmussen  
*Ledelse i fællesskab – de tillidsvalgte fornyende rolle*
28. Gitte Thit Nielsen  
*Remerger*  
*– skabende ledelseskrafter i fusion og opkøb*
29. Carmine Gioia  
*A MICROECONOMETRIC ANALYSIS OF MERGERS AND ACQUISITIONS*

30. Ole Hinz  
*Den effektive forandringsleder: pilot, pædagog eller politiker?*  
*Et studie i arbejdslederens meningstilskrivninger i forbindelse med vellykket gennemførelse af ledelsesinitierede forandringsprojekter*
31. Kjell-Åge Gotvassli  
*Et praksisbasert perspektiv på dynamiske læringsnettverk i toppidretten*  
Norsk ph.d., ej til salg gennem Samfundslitteratur
32. Henriette Langstrup Nielsen  
*Linking Healthcare*  
*An inquiry into the changing performances of web-based technology for asthma monitoring*
33. Karin Tweddell Levinsen  
*Virtuel Uddannelsespraksis*  
*Master i IKT og Læring – et casestudie i hvordan proaktiv proceshåndtering kan forbedre praksis i virtuelle læringsmiljøer*
34. Anika Liversage  
*Finding a Path*  
*Labour Market Life Stories of Immigrant Professionals*
35. Kasper Elmquist Jørgensen  
*Studier i samspillet mellem stat og erhvervsliv i Danmark under 1. verdenskrig*
36. Finn Janning  
*A DIFFERENT STORY*  
*Seduction, Conquest and Discovery*
37. Patricia Ann Plackett  
*Strategic Management of the Radical Innovation Process*  
*Leveraging Social Capital for Market Uncertainty Management*
2. Niels Rom-Poulsen  
*Essays in Computational Finance*
3. Tina Brandt Husman  
*Organisational Capabilities, Competitive Advantage & Project-Based Organisations*  
*The Case of Advertising and Creative Good Production*
4. Mette Rosenkrands Johansen  
*Practice at the top*  
*– how top managers mobilise and use non-financial performance measures*
5. Eva Parum  
Corporate governance som strategisk kommunikations- og ledelsesværktøj
6. Susan Aagaard Petersen  
*Culture's Influence on Performance Management: The Case of a Danish Company in China*
7. Thomas Nicolai Pedersen  
*The Discursive Constitution of Organizational Governance – Between unity and differentiation*  
*The Case of the governance of environmental risks by World Bank environmental staff*
8. Cynthia Selin  
*Volatile Visions: Transactions in Anticipatory Knowledge*
9. Jesper Banghøj  
*Financial Accounting Information and Compensation in Danish Companies*
10. Mikkel Lucas Overby  
*Strategic Alliances in Emerging High-Tech Markets: What's the Difference and does it Matter?*
11. Tine Aage  
*External Information Acquisition of Industrial Districts and the Impact of Different Knowledge Creation Dimensions*

## 2006

1. Christian Vintergaard  
*Early Phases of Corporate Venturing*



- A case study of the Fashion and Design Branch of the Industrial District of Montebelluna, NE Italy*
12. Mikkel Flyverbom  
*Making the Global Information Society Governable  
On the Governmentality of Multi-Stakeholder Networks*
  13. Anette Grønning  
*Personen bag  
Tilstedevær i e-mail som interaktionsform mellem kunde og medarbejder i dansk forsikringskontekst*
  14. Jørn Helder  
*One Company – One Language?  
The NN-case*
  15. Lars Bjerregaard Mikkelsen  
*Differing perceptions of customer value  
Development and application of a tool for mapping perceptions of customer value at both ends of customer-supplier dyads in industrial markets*
  16. Lise Granerud  
*Exploring Learning  
Technological learning within small manufacturers in South Africa*
  17. Esben Rahbek Pedersen  
*Between Hopes and Realities:  
Reflections on the Promises and Practices of Corporate Social Responsibility (CSR)*
  18. Ramona Samson  
*The Cultural Integration Model and European Transformation.  
The Case of Romania*
- 2007**
1. Jakob Vestergaard  
*Discipline in The Global Economy  
Panopticism and the Post-Washington Consensus*
  2. Heidi Lund Hansen  
*Spaces for learning and working  
A qualitative study of change of work, management, vehicles of power and social practices in open offices*
  3. Sudhanshu Rai  
*Exploring the internal dynamics of software development teams during user analysis  
A tension enabled Institutionalization Model; "Where process becomes the objective"*
  4. Norsk ph.d.  
Ej til salg gennem Samfundslitteratur
  5. Serden Ozcan  
*EXPLORING HETEROGENEITY IN ORGANIZATIONAL ACTIONS AND OUTCOMES  
A Behavioural Perspective*
  6. Kim Sundtoft Hald  
*Inter-organizational Performance Measurement and Management in Action  
– An Ethnography on the Construction of Management, Identity and Relationships*
  7. Tobias Lindeberg  
*Evaluative Technologies  
Quality and the Multiplicity of Performance*
  8. Merete Wedell-Wedellsborg  
*Den globale soldat  
Identitetsdannelse og identitetsledelse i multinationale militære organisationer*
  9. Lars Frederiksen  
*Open Innovation Business Models  
Innovation in firm-hosted online user communities and inter-firm project ventures in the music industry  
– A collection of essays*
  10. Jonas Gabrielsen  
*Retorisk toposlære – fra statisk 'sted' til persuasiv aktivitet*

11. Christian Moldt-Jørgensen  
*Fra meningsløs til meningsfuld evaluering.*  
*Anvendelsen af studentertilfredsheds-målinger på de korte og mellemlange videregående uddannelser set fra et psykodynamisk systemperspektiv*
12. Ping Gao  
*Extending the application of actor-network theory*  
*Cases of innovation in the telecommunications industry*
13. Peter Mejlby  
*Frihed og fængsel, en del af den samme drøm?*  
*Et phronetisk baseret casestudie af frigørelsens og kontrollens sam-eksistens i værdibaseret ledelse!*
14. Kristina Birch  
*Statistical Modelling in Marketing*
15. Signe Poulsen  
*Sense and sensibility:*  
*The language of emotional appeals in insurance marketing*
16. Anders Bjerre Trolle  
*Essays on derivatives pricing and dynamic asset allocation*
17. Peter Feldhütter  
*Empirical Studies of Bond and Credit Markets*
18. Jens Henrik Eggert Christensen  
*Default and Recovery Risk Modeling and Estimation*
19. Maria Theresa Larsen  
*Academic Enterprise: A New Mission for Universities or a Contradiction in Terms?*  
*Four papers on the long-term implications of increasing industry involvement and commercialization in academia*
20. Morten Wellendorf  
*Postimplementering af teknologi i den offentlige forvaltning*  
*Analysen af en organisations kontinuerlige arbejde med informations-teknologi*
21. Ekaterina Mhaanna  
*Concept Relations for Terminological Process Analysis*
22. Stefan Ring Thorbjørnsen  
*Forsvaret i forandring*  
*Et studie i officerers kapabiliteter under påvirkning af omverdenens forandringspres mod øget styring og læring*
23. Christa Breum Amhøj  
*Det selvskabte medlemskab om managementsstaten, dens styringsteknologier og indbyggere*
24. Karoline Bromose  
*Between Technological Turbulence and Operational Stability*  
*– An empirical case study of corporate venturing in TDC*
25. Susanne Justesen  
*Navigating the Paradoxes of Diversity in Innovation Practice*  
*– A Longitudinal study of six very different innovation processes – in practice*
26. Luise Noring Henler  
*Conceptualising successful supply chain partnerships*  
*– Viewing supply chain partnerships from an organisational culture perspective*
27. Mark Mau  
*Kampen om telefonen*  
*Det danske telefonvæsen under den tyske besættelse 1940-45*
28. Jakob Halskov  
*The semiautomatic expansion of existing terminological ontologies using knowledge patterns discovered*

- on the WWW – an implementation and evaluation*
29. Gergana Koleva  
*European Policy Instruments Beyond Networks and Structure: The Innovative Medicines Initiative*
  30. Christian Geisler Asmussen  
*Global Strategy and International Diversity: A Double-Edged Sword?*
  31. Christina Holm-Petersen  
*Stolthed og fordom  
Kultur- og identitetsarbejde ved skabelsen af en ny sengeafdeling gennem fusion*
  32. Hans Peter Olsen  
*Hybrid Governance of Standardized States  
Causes and Contours of the Global Regulation of Government Auditing*
  33. Lars Bøge Sørensen  
*Risk Management in the Supply Chain*
  34. Peter Aagaard  
*Det unikkes dynamikker  
De institutionelle mulighedsbetingelser bag den individuelle udforskning i professionelt og frivilligt arbejde*
  35. Yun Mi Antorini  
*Brand Community Innovation  
An Intrinsic Case Study of the Adult Fans of LEGO Community*
  36. Joachim Lynggaard Boll  
*Labor Related Corporate Social Performance in Denmark  
Organizational and Institutional Perspectives*
- 2008**
1. Frederik Christian Vinten  
*Essays on Private Equity*
  2. Jesper Clement  
*Visual Influence of Packaging Design on In-Store Buying Decisions*
  3. Marius Brostrøm Kousgaard  
*Tid til kvalitetsmåling?  
– Studier af indrulleringsprocesser i forbindelse med introduktionen af kliniske kvalitetsdatabaser i speciallægepraksissektoren*
  4. Irene Skovgaard Smith  
*Management Consulting in Action  
Value creation and ambiguity in client-consultant relations*
  5. Anders Rom  
*Management accounting and integrated information systems  
How to exploit the potential for management accounting of information technology*
  6. Marina Candi  
*Aesthetic Design as an Element of Service Innovation in New Technology-based Firms*
  7. Morten Schnack  
*Teknologi og tværfaglighed  
– en analyse af diskussionen omkring indførelse af EPJ på en hospitalsafdeling*
  8. Helene Balslev Clausen  
*Juntos pero no revueltos – un estudio sobre emigrantes norteamericanos en un pueblo mexicano*
  9. Lise Justesen  
*Kunsten at skrive revisionsrapporter.  
En beretning om forvaltningsrevisionsens beretninger*
  10. Michael E. Hansen  
*The politics of corporate responsibility: CSR and the governance of child labor and core labor rights in the 1990s*
  11. Anne Roepstorff  
*Holdning for handling – en etnologisk undersøgelse af Virksomheders Sociale Ansvar/CSR*

12. Claus Bajlum  
*Essays on Credit Risk and Credit Derivatives*
13. Anders Bojesen  
*The Performative Power of Competence – an Inquiry into Subjectivity and Social Technologies at Work*
14. Satu Reijonen  
*Green and Fragile  
A Study on Markets and the Natural Environment*
15. Ilduara Busta  
*Corporate Governance in Banking  
A European Study*
16. Kristian Anders Hvass  
*A Boolean Analysis Predicting Industry Change: Innovation, Imitation & Business Models  
The Winning Hybrid: A case study of isomorphism in the airline industry*
17. Trine Paludan  
*De uvidende og de udviklingsparate Identitet som mulighed og restriktion blandt fabriksarbejdere på det aftayloriserede fabriksgulv*
18. Kristian Jakobsen  
*Foreign market entry in transition economies: Entry timing and mode choice*
19. Jakob Elming  
*Syntactic reordering in statistical machine translation*
20. Lars Brømsøe Termansen  
*Regional Computable General Equilibrium Models for Denmark  
Three papers laying the foundation for regional CGE models with agglomeration characteristics*
21. Mia Reinholdt  
*The Motivational Foundations of Knowledge Sharing*
22. Frederikke Krogh-Meibom  
*The Co-Evolution of Institutions and Technology  
– A Neo-Institutional Understanding of Change Processes within the Business Press – the Case Study of Financial Times*
23. Peter D. Ørberg Jensen  
*OFFSHORING OF ADVANCED AND HIGH-VALUE TECHNICAL SERVICES: ANTECEDENTS, PROCESS DYNAMICS AND FIRMLEVEL IMPACTS*
24. Pham Thi Song Hanh  
*Functional Upgrading, Relational Capability and Export Performance of Vietnamese Wood Furniture Producers*
25. Mads Vangkilde  
*Why wait?  
An Exploration of first-mover advantages among Danish e-grocers through a resource perspective*
26. Hubert Buch-Hansen  
*Rethinking the History of European Level Merger Control  
A Critical Political Economy Perspective*
- 2009**
1. Vivian Lindhardsen  
*From Independent Ratings to Communal Ratings: A Study of CWA Raters' Decision-Making Behaviours*
2. Guðrið Weihe  
*Public-Private Partnerships: Meaning and Practice*
3. Chris Nøkkentved  
*Enabling Supply Networks with Collaborative Information Infrastructures  
An Empirical Investigation of Business Model Innovation in Supplier Relationship Management*
4. Sara Louise Muhr  
*Wound, Interrupted – On the Vulnerability of Diversity Management*

5. Christine Sestoft  
*Forbrugeradfærd i et Stats- og Livsformsteoretisk perspektiv*
6. Michael Pedersen  
*Tune in, Breakdown, and Reboot: On the production of the stress-fit self-managing employee*
7. Salla Lutz  
*Position and Reposition in Networks – Exemplified by the Transformation of the Danish Pine Furniture Manufacturers*
8. Jens Forssbæck  
*Essays on market discipline in commercial and central banking*
9. Tine Murphy  
*Sense from Silence – A Basis for Organised Action*  
*How do Sensemaking Processes with Minimal Sharing Relate to the Reproduction of Organised Action?*
10. Sara Malou Strandvad  
*Inspirations for a new sociology of art: A sociomaterial study of development processes in the Danish film industry*
11. Nicolaas Mouton  
*On the evolution of social scientific metaphors: A cognitive-historical enquiry into the divergent trajectories of the idea that collective entities – states and societies, cities and corporations – are biological organisms.*
12. Lars Andreas Knutsen  
*Mobile Data Services: Shaping of user engagements*
13. Nikolaos Theodoros Korfiatis  
*Information Exchange and Behavior*  
*A Multi-method Inquiry on Online Communities*
14. Jens Albæk  
*Forestillinger om kvalitet og tværfaglighed på sygehuse*  
*– skabelse af forestillinger i læge- og plejegrupperne angående relevans af nye idéer om kvalitetsudvikling gennem tolkningsprocesser*
15. Maja Lotz  
*The Business of Co-Creation – and the Co-Creation of Business*
16. Gitte P. Jakobsen  
*Narrative Construction of Leader Identity in a Leader Development Program Context*
17. Dorte Hermansen  
*“Living the brand” som en brandorienteret dialogisk praksis: Om udvikling af medarbejdernes brandorienterede dømmekraft*
18. Aseem Kinra  
*Supply Chain (logistics) Environmental Complexity*
19. Michael Nørager  
*How to manage SMEs through the transformation from non innovative to innovative?*
20. Kristin Wallevik  
*Corporate Governance in Family Firms*  
*The Norwegian Maritime Sector*
21. Bo Hansen Hansen  
*Beyond the Process*  
*Enriching Software Process Improvement with Knowledge Management*
22. Annemette Skot-Hansen  
*Franske adjektivisk afledte adverbier, der tager præpositionssyntagmer indledt med præpositionen à som argumenter*  
*En valensgrammatisk undersøgelse*
23. Line Gry Knudsen  
*Collaborative R&D Capabilities*  
*In Search of Micro-Foundations*

24. Christian Scheuer  
*Employers meet employees  
Essays on sorting and globalization*
25. Rasmus Johnsen  
*The Great Health of Melancholy  
A Study of the Pathologies of Performativity*
26. Ha Thi Van Pham  
*Internationalization, Competitiveness  
Enhancement and Export Performance  
of Emerging Market Firms:  
Evidence from Vietnam*
27. Henriette Balieu  
*Kontrollbegrebets betydning for kausalalternationen i spansk  
En kognitiv-typologisk analyse*
- 2010**
1. Yen Tran  
*Organizing Innovation in Turbulent  
Fashion Market  
Four papers on how fashion firms create and appropriate innovation value*
2. Anders Raastrup Kristensen  
*Metaphysical Labour  
Flexibility, Performance and Commitment in Work-Life Management*
3. Margrét Sigrún Sigurdardóttir  
*Dependently independent  
Co-existence of institutional logics in the recorded music industry*
4. Ásta Dis Óladóttir  
*Internationalization from a small domestic base:  
An empirical analysis of Economics and Management*
5. Christine Secher  
*E-deltagelse i praksis – politikernes og forvaltningens medkonstruktion og konsekvenserne heraf*
6. Marianne Stang Våland  
*What we talk about when we talk about space:*
7. Rex Degnegaard  
*Strategic Change Management  
Change Management Challenges in the Danish Police Reform*
8. Ulrik Schultz Brix  
*Værdi i rekruttering – den sikre beslutning  
En pragmatisk analyse af perception og synliggørelse af værdi i rekrutterings- og udvælgelsesarbejdet*
9. Jan Ole Similå  
*Kontraktsledelse  
Relasjonen mellom virksomhetsledelse og kontraktshåndtering, belyst via fire norske virksomheter*
10. Susanne Boch Waldorff  
*Emerging Organizations: In between local translation, institutional logics and discourse*
11. Brian Kane  
*Performance Talk  
Next Generation Management of Organizational Performance*
12. Lars Ohnemus  
*Brand Thrust: Strategic Branding and Shareholder Value  
An Empirical Reconciliation of two Critical Concepts*
13. Jesper Schlamovitz  
*Håndtering af usikkerhed i film- og byggeprojekter*
14. Tommy Moesby-Jensen  
*Det faktiske livs forbindtlighed  
Føroskratisk informeret, ny-aristotelisk ἦθος-tænkning hos Martin Heidegger*
15. Christian Fich  
*Two Nations Divided by Common Values  
French National Habitus and the Rejection of American Power*

16. Peter Beyer  
*Processer, sammenhængskraft og fleksibilitet*  
*Et empirisk casestudie af omstillingsforløb i fire virksomheder*
17. Adam Buchhorn  
*Markets of Good Intentions*  
*Constructing and Organizing Biogas Markets Amid Fragility and Controversy*
18. Cecilie K. Moesby-Jensen  
*Social læring og fælles praksis*  
*Et mixed method studie, der belyser læringskonsekvenser af et lederkursus for et praksisfællesskab af offentlige mellemledere*
19. Heidi Boye  
*Fødevarer og sundhed i senmodernismen*  
*– En indsigt i hyggefænomenet og de relaterede fødevarerpraksisser*
20. Kristine Munkgård Pedersen  
*Flygtige forbindelser og midlertidige mobiliseringer*  
*Om kulturel produktion på Roskilde Festival*
21. Oliver Jacob Weber  
*Causes of Intercompany Harmony in Business Markets – An Empirical Investigation from a Dyad Perspective*
22. Susanne Ekman  
*Authority and Autonomy*  
*Paradoxes of Modern Knowledge Work*
23. Anette Frey Larsen  
*Kvalitetsledelse på danske hospitaler*  
*– Ledelsernes indflydelse på introduktion og vedligeholdelse af kvalitetsstrategier i det danske sundhedsvæsen*
24. Toyoko Sato  
*Performativity and Discourse: Japanese Advertisements on the Aesthetic Education of Desire*
25. Kenneth Brinch Jensen  
*Identifying the Last Planner System*  
*Lean management in the construction industry*
26. Javier Busquets  
*Orchestrating Network Behavior for Innovation*
27. Luke Patey  
*The Power of Resistance: India's National Oil Company and International Activism in Sudan*
28. Mette Vedel  
*Value Creation in Triadic Business Relationships. Interaction, Interconnection and Position*
29. Kristian Tørning  
*Knowledge Management Systems in Practice – A Work Place Study*
30. Qingxin Shi  
*An Empirical Study of Thinking Aloud*  
*Usability Testing from a Cultural Perspective*
31. Tanja Juul Christiansen  
*Corporate blogging: Medarbejderes kommunikative handlekraft*
32. Malgorzata Ciesielska  
*Hybrid Organisations.*  
*A study of the Open Source – business setting*
33. Jens Dick-Nielsen  
*Three Essays on Corporate Bond Market Liquidity*
34. Sabrina Speiermann  
*Modstandens Politik*  
*Kampagnestyling i Velfærdsstaten.*  
*En diskussion af trafikcampagners styringspotentiale*
35. Julie Uldam  
*Fickle Commitment. Fostering political engagement in 'the flighty world of online activism'*

36. Annegrete Juul Nielsen  
*Traveling technologies and transformations in health care*
37. Athur Mühlen-Schulte  
*Organising Development Power and Organisational Reform in the United Nations Development Programme*
38. Louise Rygaard Jonas  
*Branding på butiksgulvet Et case-studie af kultur- og identitetsarbejdet i Kvickly*
8. Ole Helby Petersen  
*Public-Private Partnerships: Policy and Regulation – With Comparative and Multi-level Case Studies from Denmark and Ireland*
9. Morten Krogh Petersen  
*'Good' Outcomes. Handling Multiplicity in Government Communication*
10. Kristian Tangsgaard Hvelplund  
*Allocation of cognitive resources in translation - an eye-tracking and key-logging study*

## 2011

1. Stefan Fraenkel  
*Key Success Factors for Sales Force Readiness during New Product Launch A Study of Product Launches in the Swedish Pharmaceutical Industry*
2. Christian Plesner Rossing  
*International Transfer Pricing in Theory and Practice*
3. Tobias Dam Hede  
*Samtalekunst og ledelsesdisciplin – en analyse af coachingsdiskursens genealogi og governmentality*
4. Kim Pettersson  
*Essays on Audit Quality, Auditor Choice, and Equity Valuation*
5. Henrik Merkelsen  
*The expert-lay controversy in risk research and management. Effects of institutional distances. Studies of risk definitions, perceptions, management and communication*
6. Simon S. Torp  
*Employee Stock Ownership: Effect on Strategic Management and Performance*
7. Mie Harder  
*Internal Antecedents of Management Innovation*
11. Moshe Yonatany  
*The Internationalization Process of Digital Service Providers*
12. Anne Vestergaard  
*Distance and Suffering Humanitarian Discourse in the age of Mediatization*
13. Thorsten Mikkelsen  
*Personlighedens indflydelse på forretningsrelationer*
14. Jane Thostrup Jagd  
*Hvorfor fortsætter fusionsbølgen udover "the tipping point"? – en empirisk analyse af information og kognitioner om fusioner*
15. Gregory Gimpel  
*Value-driven Adoption and Consumption of Technology: Understanding Technology Decision Making*
16. Thomas Stengade Sønderskov  
*Den nye mulighed Social innovation i en forretningsmæssig kontekst*
17. Jeppe Christoffersen  
*Donor supported strategic alliances in developing countries*
18. Vibeke Vad Baunsgaard  
*Dominant Ideological Modes of Rationality: Cross functional*



- integration in the process of product innovation*
19. Throstur Olaf Sigurjonsson  
*Governance Failure and Iceland's Financial Collapse*
20. Allan Sall Tang Andersen  
*Essays on the modeling of risks in interest-rate and inflation markets*
21. Heidi Tscherning  
*Mobile Devices in Social Contexts*
22. Birgitte Gorm Hansen  
*Adapting in the Knowledge Economy Lateral Strategies for Scientists and Those Who Study Them*
23. Kristina Vaarst Andersen  
*Optimal Levels of Embeddedness The Contingent Value of Networked Collaboration*
24. Justine Grøn bæk Pors  
*Noisy Management A History of Danish School Governing from 1970-2010*
25. Stefan Linder  
*Micro-foundations of Strategic Entrepreneurship Essays on Autonomous Strategic Action*
26. Xin Li  
*Toward an Integrative Framework of National Competitiveness An application to China*
27. Rune Thorbjørn Clausen  
*Værdifuld arkitektur Et eksplorativt studie af bygningers rolle i virksomheders værdiskabelse*
28. Monica Viken  
*Markedsundersøkelser som bevis i varemerke- og markedsføringsrett*
29. Christian Wymann  
*Tattooing The Economic and Artistic Constitution of a Social Phenomenon*
30. Sanne Frandsen  
*Productive Incoherence A Case Study of Branding and Identity Struggles in a Low-Prestige Organization*
31. Mads Stenbo Nielsen  
*Essays on Correlation Modelling*
32. Ivan Häuser  
*Følelse og sprog Etablering af en ekspressiv kategori, eksemplificeret på russisk*
33. Sebastian Schwenen  
*Security of Supply in Electricity Markets*
- 2012**
1. Peter Holm Andreasen  
*The Dynamics of Procurement Management - A Complexity Approach*
2. Martin Haulrich  
*Data-Driven Bitext Dependency Parsing and Alignment*
3. Line Kirkegaard  
*Konsulenten i den anden nat En undersøgelse af det intense arbejdsliv*
4. Tonny Stenheim  
*Decision usefulness of goodwill under IFRS*
5. Morten Lind Larsen  
*Produktivitet, vækst og velfærd Industrirådet og efterkrigstidens Danmark 1945 - 1958*
6. Petter Berg  
*Cartel Damages and Cost Asymmetries*
7. Lynn Kahle  
*Experiential Discourse in Marketing A methodical inquiry into practice and theory*
8. Anne Roelsgaard Obling  
*Management of Emotions in Accelerated Medical Relationships*

9. Thomas Frandsen  
*Managing Modularity of Service Processes Architecture*
10. Carina Christine Skovmøller  
*CSR som noget særligt  
Et casestudie om styring og menings-skabelse i relation til CSR ud fra en intern optik*
11. Michael Tell  
*Fradragsbeskæring af selskabers finansieringsudgifter  
En skatterelig analyse af SEL §§ 11, 11B og 11C*
12. Morten Holm  
*Customer Profitability Measurement Models  
Their Merits and Sophistication across Contexts*
13. Katja Joo Dyppe  
*Beskatning af derivater  
En analyse af dansk skatteret*
14. Esben Anton Schultz  
*Essays in Labor Economics  
Evidence from Danish Micro Data*
15. Carina Risvig Hansen  
*"Contracts not covered, or not fully covered, by the Public Sector Directive"*
16. Anja Svejgaard Pors  
*Iværksættelse af kommunikation - patientfigurer i hospitalets strategiske kommunikation*
17. Frans Bévort  
*Making sense of management with logics  
An ethnographic study of accountants who become managers*
18. René Kallestrup  
*The Dynamics of Bank and Sovereign Credit Risk*
19. Brett Crawford  
*Revisiting the Phenomenon of Interests in Organizational Institutionalism  
The Case of U.S. Chambers of Commerce*
20. Mario Daniele Amore  
*Essays on Empirical Corporate Finance*
21. Arne Stjernholm Madsen  
*The evolution of innovation strategy Studied in the context of medical device activities at the pharmaceutical company Novo Nordisk A/S in the period 1980-2008*
22. Jacob Holm Hansen  
*Is Social Integration Necessary for Corporate Branding?  
A study of corporate branding strategies at Novo Nordisk*
23. Stuart Webber  
*Corporate Profit Shifting and the Multinational Enterprise*
24. Helene Ratner  
*Promises of Reflexivity  
Managing and Researching Inclusive Schools*
25. Therese Strand  
*The Owners and the Power: Insights from Annual General Meetings*
26. Robert Gavin Strand  
*In Praise of Corporate Social Responsibility Bureaucracy*
27. Nina Sormunen  
*Auditor's going-concern reporting  
Reporting decision and content of the report*
28. John Bang Mathiasen  
*Learning within a product development working practice:  
- an understanding anchored in pragmatism*
29. Philip Holst Riis  
*Understanding Role-Oriented Enterprise Systems: From Vendors to Customers*
30. Marie Lisa Dacanay  
*Social Enterprises and the Poor  
Enhancing Social Entrepreneurship and Stakeholder Theory*

31. Fumiko Kano Glückstad  
*Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge*
32. Henrik Barslund Fosse  
*Empirical Essays in International Trade*
33. Peter Alexander Albrecht  
*Foundational hybridity and its reproduction  
Security sector reform in Sierra Leone*
34. Maja Rosenstock  
*CSR - hvor svært kan det være?  
Kulturanalytisk casestudie om udfordringer og dilemmaer med at forankre Coops CSR-strategi*
35. Jeanette Rasmussen  
*Tweens, medier og forbrug  
Et studie af 10-12 årige danske børns brug af internettet, opfattelse og forståelse af markedsføring og forbrug*
36. Ib Tunby Gulbrandsen  
*'This page is not intended for a US Audience'  
A five-act spectacle on online communication, collaboration & organization.*
37. Kasper Aalling Teilmann  
*Interactive Approaches to Rural Development*
38. Mette Mogensen  
*The Organization(s) of Well-being and Productivity  
(Re)assembling work in the Danish Post*
39. Søren Friis Møller  
*From Disinterestedness to Engagement  
Towards Relational Leadership In the Cultural Sector*
40. Nico Peter Berhausen  
*Management Control, Innovation and Strategic Objectives – Interactions and Convergence in Product Development Networks*
41. Balder Onarheim  
*Creativity under Constraints  
Creativity as Balancing 'Constrainedness'*
42. Haoyong Zhou  
*Essays on Family Firms*
43. Elisabeth Naima Mikkelsen  
*Making sense of organisational conflict  
An empirical study of enacted sense-making in everyday conflict at work*
- 2013**
1. Jacob Lyngsie  
*Entrepreneurship in an Organizational Context*
2. Signe Groth-Brodersen  
*Fra ledelse til selvet  
En socialpsykologisk analyse af forholdet imellem selvledelse, ledelse og stress i det moderne arbejdsliv*
3. Nis Høyrup Christensen  
*Shaping Markets: A Neoinstitutional Analysis of the Emerging Organizational Field of Renewable Energy in China*
4. Christian Edelvold Berg  
*As a matter of size  
THE IMPORTANCE OF CRITICAL MASS AND THE CONSEQUENCES OF SCARCITY FOR TELEVISION MARKETS*
5. Christine D. Isakson  
*Coworker Influence and Labor Mobility  
Essays on Turnover, Entrepreneurship and Location Choice in the Danish Maritime Industry*
6. Niels Joseph Jerne Lennon  
*Accounting Qualities in Practice  
Rhizomatic stories of representational faithfulness, decision making and control*
7. Shannon O'Donnell  
*Making Ensemble Possible  
How special groups organize for collaborative creativity in conditions of spatial variability and distance*

8. Robert W. D. Veitch  
*Access Decisions in a Partly-Digital World*  
*Comparing Digital Piracy and Legal Modes for Film and Music*
9. Marie Mathiesen  
*Making Strategy Work*  
*An Organizational Ethnography*
10. Arisa Shollo  
*The role of business intelligence in organizational decision-making*
11. Mia Kaspersen  
*The construction of social and environmental reporting*
12. Marcus Møller Larsen  
*The organizational design of offshoring*
13. Mette Ohm Rørdam  
*EU Law on Food Naming*  
*The prohibition against misleading names in an internal market context*
14. Hans Peter Rasmussen  
*GIV EN GED!*  
*Kan giver-idealiteter forklare støtte til velgørenhed og understøtte relationsopbygning?*
15. Ruben Schachtenhaufen  
*Fonetisk reduktion i dansk*
16. Peter Koerver Schmidt  
*Dansk CFC-beskatning*  
*I et internationalt og komparativt perspektiv*
17. Morten Froholdt  
*Strategi i den offentlige sektor*  
*En kortlægning af styringsmæssig kontekst, strategisk tilgang, samt anvendte redskaber og teknologier for udvalgte danske statslige styrelser*
18. Annette Camilla Sjørup  
*Cognitive effort in metaphor translation*  
*An eye-tracking and key-logging study*
19. Tamara Stucchi  
*The Internationalization of Emerging Market Firms: A Context-Specific Study*
20. Thomas Lopdrup-Hjorth  
*"Let's Go Outside": The Value of Co-Creation*
21. Ana Alačovska  
*Genre and Autonomy in Cultural Production*  
*The case of travel guidebook production*
22. Marius Gudmand-Høyer  
*Stemningsindssygdommens historie i det 19. århundrede*  
*Omtydningen af melankolien og manien som bipolære stemningslidelser i dansk sammenhæng under hensyn til dannelsen af det moderne følelselivs relative autonomi.*  
*En problematiserings- og erfarings-analytisk undersøgelse*
23. Lichen Alex Yu  
*Fabricating an S&OP Process*  
*Circulating References and Matters of Concern*
24. Esben Alfort  
*The Expression of a Need*  
*Understanding search*
25. Trine Pallesen  
*Assembling Markets for Wind Power*  
*An Inquiry into the Making of Market Devices*
26. Anders Koed Madsen  
*Web-Visions*  
*Repurposing digital traces to organize social attention*
27. Lærke Højgaard Christiansen  
*BREWING ORGANIZATIONAL RESPONSES TO INSTITUTIONAL LOGICS*
28. Tommy Kjær Lassen  
*EGENTLIG SELVLEDELSE*  
*En ledelsesfilosofisk afhandling om selvledelsens paradoksale dynamik og eksistentielle engagement*

29. Morten Rossing  
*Local Adaption and Meaning Creation in Performance Appraisal*
30. Søren Obed Madsen  
*Lederen som oversætter  
Et oversættelsesteoretisk perspektiv på strategisk arbejde*
31. Thomas Høgenhaven  
*Open Government Communities  
Does Design Affect Participation?*
32. Kirstine Zinck Pedersen  
*Failsafe Organizing?  
A Pragmatic Stance on Patient Safety*
33. Anne Petersen  
*Hverdagslogikker i psykiatrisk arbejde  
En institutionsetnografisk undersøgelse af hverdagen i psykiatriske organisationer*
34. Didde Maria Humle  
*Fortællinger om arbejde*
35. Mark Holst-Mikkelsen  
*Strategieksekvering i praksis  
– barrierer og muligheder!*
36. Malek Maalouf  
*Sustaining lean  
Strategies for dealing with organizational paradoxes*
37. Nicolaj Tofte Brenneche  
*Systemic Innovation In The Making  
The Social Productivity of Cartographic Crisis and Transitions in the Case of SEET*
38. Morten Gylling  
*The Structure of Discourse  
A Corpus-Based Cross-Linguistic Study*
39. Binzhang YANG  
*Urban Green Spaces for Quality Life  
- Case Study: the landscape architecture for people in Copenhagen*
40. Michael Friis Pedersen  
*Finance and Organization:  
The Implications for Whole Farm Risk Management*
41. Even Fallan  
*Issues on supply and demand for environmental accounting information*
42. Ather Nawaz  
*Website user experience  
A cross-cultural study of the relation between users' cognitive style, context of use, and information architecture of local websites*
43. Karin Beukel  
*The Determinants for Creating Valuable Inventions*
44. Arjan Markus  
*External Knowledge Sourcing and Firm Innovation  
Essays on the Micro-Foundations of Firms' Search for Innovation*
- 2014**
1. Solon Moreira  
*Four Essays on Technology Licensing and Firm Innovation*
2. Karin Strzeletz Ivertsen  
*Partnership Drift in Innovation Processes  
A study of the Think City electric car development*
3. Kathrine Hoffmann Pii  
*Responsibility Flows in Patient-centred Prevention*
4. Jane Bjørn Vedel  
*Managing Strategic Research  
An empirical analysis of science-industry collaboration in a pharmaceutical company*
5. Martin Gylling  
*Processuel strategi i organisationer  
Monografi om dobbeltheden i tænkning af strategi, dels som vidensfelt i organisationsteori, dels som kunstnerisk tilgang til at skabe i erhvervsmæssig innovation*

6. Linne Marie Lauesen  
*Corporate Social Responsibility in the Water Sector: How Material Practices and their Symbolic and Physical Meanings Form a Colonising Logic*
7. Maggie Qiuzhu Mei  
*LEARNING TO INNOVATE: The role of ambidexterity, standard, and decision process*
8. Inger Hædt-Rasmussen  
*Developing Identity for Lawyers Towards Sustainable Lawyering*
9. Sebastian Fux  
*Essays on Return Predictability and Term Structure Modelling*
10. Thorbjørn N. M. Lund-Poulsen  
*Essays on Value Based Management*
11. Oana Brindusa Albu  
*Transparency in Organizing: A Performative Approach*
12. Lena Olaison  
*Entrepreneurship at the limits*
13. Hanne Sørum  
*DRESSED FOR WEB SUCCESS? An Empirical Study of Website Quality in the Public Sector*
14. Lasse Folke Henriksen  
*Knowing networks How experts shape transnational governance*
15. Maria Halbinger  
*Entrepreneurial Individuals Empirical Investigations into Entrepreneurial Activities of Hackers and Makers*
16. Robert Spliid  
*Kapitalfondenes metoder og kompetencer*
17. Christiane Stelling  
*Public-private partnerships & the need, development and management of trusting A processual and embedded exploration*
18. Marta Gasparin  
*Management of design as a translation process*
19. Kåre Moberg  
*Assessing the Impact of Entrepreneurship Education From ABC to PhD*
20. Alexander Cole  
*Distant neighbors Collective learning beyond the cluster*
21. Martin Møller Boje Rasmussen  
*Is Competitiveness a Question of Being Alike? How the United Kingdom, Germany and Denmark Came to Compete through their Knowledge Regimes from 1993 to 2007*
22. Anders Ravn Sørensen  
*Studies in central bank legitimacy, currency and national identity Four cases from Danish monetary history*
23. Nina Bellak  
*Can Language be Managed in International Business? Insights into Language Choice from a Case Study of Danish and Austrian Multinational Corporations (MNCs)*
24. Rikke Kristine Nielsen  
*Global Mindset as Managerial Meta-competence and Organizational Capability: Boundary-crossing Leadership Cooperation in the MNC The Case of 'Group Mindset' in Solar A/S.*
25. Rasmus Koss Hartmann  
*User Innovation inside government Towards a critically performative foundation for inquiry*

26. Kristian Gylling Olesen  
*Flertydig og emergerende ledelse i folkeskolen*  
*Et aktør-netværksteoretisk ledelsesstudie af politiske evalueringsreformers betydning for ledelse i den danske folkeskole*
27. Troels Riis Larsen  
*Kampen om Danmarks omdømme 1945-2010*  
*Omdømmearbejde og omdømmepolitik*
28. Klaus Majgaard  
*Jagten på autenticitet i offentlig styring*
29. Ming Hua Li  
*Institutional Transition and Organizational Diversity: Differentiated internationalization strategies of emerging market state-owned enterprises*
30. Sofie Blinkenberg Federspiel  
*IT, organisation og digitalisering: Institutionelt arbejde i den kommunale digitaliseringsproces*
31. Elvi Weinreich  
*Hvilke offentlige ledere er der brug for når velfærdstænkningen flytter sig – er Diplomuddannelsens lederprofil svaret?*
32. Ellen Mølgaard Korsager  
*Self-conception and image of context in the growth of the firm – A Penrosian History of Fiberline Composites*
33. Else Skjold  
*The Daily Selection*
34. Marie Louise Conradsen  
*The Cancer Centre That Never Was*  
*The Organisation of Danish Cancer Research 1949-1992*
35. Virgilio Failla  
*Three Essays on the Dynamics of Entrepreneurs in the Labor Market*
36. Nicky Nedergaard  
*Brand-Based Innovation*  
*Relational Perspectives on Brand Logics and Design Innovation Strategies and Implementation*
37. Mads Gjedsted Nielsen  
*Essays in Real Estate Finance*
38. Kristin Martina Brandl  
*Process Perspectives on Service Offshoring*
39. Mia Rosa Koss Hartmann  
*In the gray zone*  
*With police in making space for creativity*
40. Karen Ingerslev  
*Healthcare Innovation under The Microscope*  
*Framing Boundaries of Wicked Problems*
41. Tim Neerup Thomsen  
*Risk Management in large Danish public capital investment programmes*
- 2015**
1. Jakob Ion Wille  
*Film som design*  
*Design af levende billeder i film og tv-serier*
2. Christiane Mossin  
*Interzones of Law and Metaphysics*  
*Hierarchies, Logics and Foundations of Social Order seen through the Prism of EU Social Rights*
3. Thomas Tøth  
*TRUSTWORTHINESS: ENABLING GLOBAL COLLABORATION*  
*An Ethnographic Study of Trust, Distance, Control, Culture and Boundary Spanning within Offshore Outsourcing of IT Services*
4. Steven Højlund  
*Evaluation Use in Evaluation Systems – The Case of the European Commission*

5. Julia Kirch Kirkegaard  
*AMBIGUOUS WINDS OF CHANGE – OR FIGHTING AGAINST WINDMILLS IN CHINESE WIND POWER*  
*A CONSTRUCTIVIST INQUIRY INTO CHINA'S PRAGMATICS OF GREEN MARKETISATION*  
*MAPPING CONTROVERSIES OVER A POTENTIAL TURN TO QUALITY IN CHINESE WIND POWER*
6. Michelle Carol Antero  
*A Multi-case Analysis of the Development of Enterprise Resource Planning Systems (ERP) Business Practices*  
  
Morten Friis-Olivarius  
*The Associative Nature of Creativity*
7. Mathew Abraham  
*New Cooperativism: A study of emerging producer organisations in India*
8. Stine Hedegaard  
*Sustainability-Focused Identity: Identity work performed to manage, negotiate and resolve barriers and tensions that arise in the process of constructing or ganizational identity in a sustainability context*
9. Cecilie Glerup  
*Organizing Science in Society – the conduct and justification of resposable research*
10. Allan Salling Pedersen  
*Implementering af ITIL® IT-governance - når best practice konflikter med kulturen*  
*Løsning af implementerings-problemer gennem anvendelse af kendte CSF i et aktionsforskningsforløb.*
11. Nihat Misir  
*A Real Options Approach to Determining Power Prices*
12. Mamdouh Medhat  
*MEASURING AND PRICING THE RISK OF CORPORATE FAILURES*
13. Rina Hansen  
*Toward a Digital Strategy for Omnichannel Retailing*
14. Eva Pallesen  
*In the rhythm of welfare creation*  
*A relational processual investigation moving beyond the conceptual horizon of welfare management*
15. Gouya Harirchi  
*In Search of Opportunities: Three Essays on Global Linkages for Innovation*
16. Lotte Holck  
*Embedded Diversity: A critical ethnographic study of the structural tensions of organizing diversity*
17. Jose Daniel Balarezo  
*Learning through Scenario Planning*
18. Louise Pram Nielsen  
*Knowledge dissemination based on terminological ontologies. Using eye tracking to further user interface design.*
19. Sofie Dam  
*PUBLIC-PRIVATE PARTNERSHIPS FOR INNOVATION AND SUSTAINABILITY TRANSFORMATION*  
*An embedded, comparative case study of municipal waste management in England and Denmark*
20. Ulrik Hartmyer Christiansen  
*Follwoing the Content of Reported Risk Across the Organization*
21. Guro Refsum Sanden  
*Language strategies in multinational corporations. A cross-sector study of financial service companies and manufacturing companies.*
22. Linn Gevoll  
*Designing performance management for operational level*  
*- A closer look on the role of design choices in framing coordination and motivation*



23. Frederik Larsen  
*Objects and Social Actions*  
*– on Second-hand Valuation Practices*
  
24. Thorhildur Hansdottir Jetzek  
*The Sustainable Value of Open*  
*Government Data*  
*Uncovering the Generative Mechanisms*  
*of Open Data through a Mixed*  
*Methods Approach*

## TITLER I ATV PH.D.-SERIEN

### 1992

1. Niels Kornum  
*Servicesamkørsel – organisation, økonomi og planlægningsmetode*

### 1995

2. Verner Worm  
*Nordiske virksomheder i Kina  
Kulturspecifikke interaktionsrelationer ved nordiske virksomhedsetableringer i Kina*

### 1999

3. Mogens Bjerre  
*Key Account Management of Complex Strategic Relationships  
An Empirical Study of the Fast Moving Consumer Goods Industry*

### 2000

4. Lotte Darsø  
*Innovation in the Making  
Interaction Research with heterogeneous Groups of Knowledge Workers creating new Knowledge and new Leads*

### 2001

5. Peter Hobolt Jensen  
*Managing Strategic Design Identities  
The case of the Lego Developer Network*

### 2002

6. Peter Lohmann  
*The Deleuzian Other of Organizational Change – Moving Perspectives of the Human*
7. Anne Marie Jess Hansen  
*To lead from a distance: The dynamic interplay between strategy and strategizing – A case study of the strategic management process*

### 2003

8. Lotte Henriksen  
*Videndeling  
– om organisatoriske og ledelsesmæssige udfordringer ved videndeling i praksis*
9. Niels Christian Nickelsen  
*Arrangements of Knowing: Coordinating Procedures Tools and Bodies in Industrial Production – a case study of the collective making of new products*

### 2005

10. Carsten Ørts Hansen  
*Konstruktion af ledelsesteknologier og effektivitet*

## TITLER I DBA PH.D.-SERIEN

### 2007

1. Peter Kastrup-Misir  
*Endeavoring to Understand Market Orientation – and the concomitant co-mutation of the researched, the researcher, the research itself and the truth*

### 2009

1. Torkild Leo Thellefsen  
*Fundamental Signs and Significance effects  
A Semeiotic outline of Fundamental Signs, Significance-effects, Knowledge Profiling and their use in Knowledge Organization and Branding*
2. Daniel Ronzani  
*When Bits Learn to Walk Don't Make Them Trip. Technological Innovation and the Role of Regulation by Law in Information Systems Research: the Case of Radio Frequency Identification (RFID)*

### 2010

1. Alexander Carnera  
*Magten over livet og livet som magt  
Studier i den biopolitiske ambivalens*