Yeh, Chih-Liang

Conference Paper

A qualitative change of personal information in the age of big data: A Study of court case in Taiwan

26th European Regional Conference of the International Telecommunications Society (ITS): "What Next for European Telecommunications?", Madrid, Spain, 24th-27th June, 2015

Provided in Cooperation with:
International Telecommunications Society (ITS)

Suggested Citation: Yeh, Chih-Liang (2015) : A qualitative change of personal information in the age of big data: A Study of court case in Taiwan, 26th European Regional Conference of the International Telecommunications Society (ITS): "What Next for European Telecommunications?", Madrid, Spain, 24th-27th June, 2015, International Telecommunications Society (ITS), Calgary

This Version is available at:
https://hdl.handle.net/10419/127199

A qualitative change of personal information in the age of big data: A study of court case in Taiwan*

Chih-Liang Yeh
Innovation Center for Big Data and Digital Convergence
Department of Information Communication
Yuan Ze University, Taiwan
chyeh@saturn.yzu.edu.tw

Abstract

Big data is boosting the economy, transforming traditional business models and creating new opportunities through the use of business intelligence, sentiment analysis, data mining and analytics. At the same time, the data overflow presents privacy concerns that could create a regulatory backlash, restraining the data economy and stifling innovation. In fact, the principles of privacy and data protection must be balanced against additional societal values, such as economic efficiency, public health, environmental protection, law enforcement, and national security. To strike a balance between the beneficial uses of data and individual privacy, governments should address an appropriate structure of privacy law, including redefinition of "personal identifiable information" (PII) based on a risk matrix taking into account the potential consequences of re-identification. The article examines current definition of personal information in various countries and discusses the effect of qualitative change influenced by the big data. The article then analyzes a court case occurred in 2014 in Taiwan to discover the extent of indirectly PII the data protection law applies, and argues that we need a new legal framework of personal information protection in the age of bid data.

Keywords: big data, personally identifiable information, data protection, privacy law, de-identification, re-identification, number portability

* This draft is only for submitting to 26th European Regional Conference of the International Telecommunications Society, San Lorenzo de El Escorial, Spain, 24th – 27th June 2015. Do not quote without author's permission.

I. Introduction: the era of big data

We live in an age of "big data." Over the past few years, the volume of data collected and stored by business and government organizations have exploded.[1] Data are generated from online transactions, emails, videos, images, logs, search queries, health records, and social networking interactions, gathered from increasingly pervasive sensors deployed in infrastructure such as communications networks, global positioning satellites, electric grids, roads and bridges, as well as in homes and mobile phones.[2] Big data boosts the economy, transforming traditional business models and creating new opportunities through the use of business intelligence, sentiment analysis and analytics. Big data also influences national policies, such as disaster alarming system, optimization of natural resources and information infrastructure.

At the same time, the "data deluge" indicates privacy concerns that could mingle with a regulatory backlash, restraining the data economy and stifling innovation.[3] To strike a balance between beneficial uses of data and the protection of individual privacy, policy-makers must address some of the most fundamental concepts of privacy law, including the definition of "personal identifiable information" (PII), the role of consent and the principle of purpose limitation and data minimization.[4]

The tasks of ensuring data security and protecting privacy become harder and harder because the information, including individuals' health records, location data, electricity use, and online activities is multiplied and shared ever more widely around the world, raising concerns about profiling, discrimination, exclusion, and loss of

[1] Kenneth Cukier, *Data, Data Everywhere*, THE ECONOMIST, Feb. 25, 2010, http://www.economist.com/node/15557443.
[2] Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NORTHWESTERN J. TECH. & INTELL. PROP. 239, 240 (2013).
[3] *See* Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63, 63 (2012).
[4] *Id.* at 64.

control. Traditionally the methods of de-identification (anonymization, encryption, key-coding) were viewed as sharp weapons allowing organizations to reap the benefits of analytics while preserving individuals' privacy. In recent years, computer scientists have shown that even anonymized data can often be re-identified and attributed to specific individuals.[5] Paul Ohm, a law scholar, observed that re-identification technologies disrupt and undermine the faith that we have placed in anonymization.[6] The de-identified data is just a temporary state rather than a stable category.

Due to the impact of big data and the limits of the personal data protection, this article attempts to review the current definition of personal information that is stipulated in the data protection laws in various countries, such as US, EU, Japan and Taiwan. In addition, the article discusses the effect of qualitative change by big data, meaning that the anonymized personal information may be re-identified or identifiable when more data are collected and are further analyzed to do so. Thus, the article discovers four factors including the technology of de-identification, data minimization, individual control (of PII), and the possibility and difficulty to recognize the combined data. The article then analyzes a court case occurred in 2014 in Taiwan to discover the extent of *indirectly* identifiable personal information that the data protection law applies. The article argues that we need a new legal framework of personal information protection in the age of bid data.

II. The current definition of personal information

A. The big data revolution and situational awareness

[5] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010).
[6] *Id*. at 1704.

Increasingly large datasets are being mined for important predictions and often surprisingly insights. We are witnessing the latest stage of the Information Revolution that has transformed our society and our lives over the past decades. But the big data phase of the revolution not only promises but also threatens a greater scale of social change at an unimaginable speed. The increasing adoption of big data is such that all kinds of human activity, ranging from dating, voting, hiring, policing and identifying terrorists, have already become heavily influenced by big data techniques.[7] Technical definitions of big data are often narrowly constrained to describe "data that exceeds the processing capacity of conventional database systems,"[8] and technologists often use the term "3V" definition of big data as "high-volume, high velocity and high variety information assets that demand cost-effective, innovation forms of information processing for enhanced insights and decision making."[9]

The Big Data Revolution is basically about awareness. The analytics of relevant big data sets give us greater awareness of everything that let us make predictions and solve problems. For example, researchers tried to analyze mobile phone traffic logs from cell tower interactions of 680,000 local commuters in Boston and traced each individual's commute, anonymously from origin to destination, and produced one of the most detailed maps of urban traffic patterns ever constructed and uncovered previously hidden patterns in urban road usage.[10] Big data presents an attractive silver bullet to defend against terrorist attacks by

[7] *See* Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 403-05 (2014).
[8] Edd Dumbil, *What Is Big Data?: An Introduction to the Big Data Landscape*, O'REILLY (Jan. 11, 2012), http://strata.oreilly.com/2012/01/what-is-big-data.html (last visited June 1, 2015).
[9] *IT Glossary: Big Data*, GARTNER, http://www.gartner.com/it-glossary/big-data (last visited June 1, 2015).
[10] *See* Pu Wang et al., *Understanding Road Usage Patterns in Urban Areas*, 2 NATURE SCI. REP. 1, 1 (2012), http://www.nature.com/srep/2012/121220/srep01001/pdf/srep01001.pdf.

expanding the "situational awareness"[11] of security services. The term "previously hidden patterns" is coined in the section 515 of the US Homeland Security Act,[12] which requires the National Operations Center to "provide *situational awareness* and a common operating picture for the entire Federal Government…and [to] ensure that critical terrorism and disaster-related information reaches government decision-makers."[13]

But how can security service identify and catch terrorists before they attack?[14] Just let government agencies have the metadata of everything in advance so they can build up a database with identifiers, such as phone numbers. Big data allows the investigators to identify the suspected terrorists if they have access to the computer's pre-attack data to find signals and inform situational awareness. The government agencies in the incident of Boston Marathon bombing have presented a good example of using various big data analytics, such as accessing Boston cell tower traffic logs, cross-checking against surveillance video and eyewitness photography, and using tools that let officials access the metadata built into every tweet sent in Boston since 2010 that contained the word "bomb," in order to identify the criminals.[15] Such a tactic would have the potential to uncover hidden patterns that could help analysts combine with other sources of intelligence to determine if an attack was about to happen. Big data

---

[11] *See, e.g.*, PAUL M. SALMON ET AT., DISTRIBUTED SITUATION AWARENESS: THEORY, MEASUREMENT AND APPLICATION TO TEAMWORK (2009).

[12] Homeland Security Act of 2002, Pub. L. No. 107-296, § 515, 116 Stat. 2135 (*amended by* Department of Homeland Security Appropriation Act, Pub. L. No. 109-295, 120 Stat. 1355, 1409 (2006)) (codified at 6 U.S.C. § 321d(b)(1)-(2)(2012)).

[13] Homeland Security Act of 2002, § 515, 6 U.S.C. § 321d(b)(1)-(2)(2012) (emphasis added). The law defines the term "situational awareness" as "information gathered from a variety of sources that, when communicated to emergency managers and decision makers, can form the basis for incident management decisionmaking." *Id*. § 321d(a).

[14] Less than 24 hours after the incident of Boston Marathon Bombing, the FBI had compiled 10 terabytes of data for finding needles in haystacks of information that might lead to the suspects. *See* Frank Konkel, *Boston Probe's Big Data Use Hints at the Future*, FCW (Apr. 26, 2013), http://fcw.com/articles/2013/04/26/big-data-boston-bomb-probe.aspx.

[15] *Id*.

analytics could also allow the identification of groups of suspected terrorists once the identity of their phone numbers became known.

However, the public is starting to ask questions about privacy as it learns about the potential privacy invasions that big data awareness allows. Yet many of the problems that concern us about big data extend beyond narrow notions of privacy. Everyone worries about the confidential information being disclosed to unknown third parties. Furthermore, we lack the transparency needed to measure the effect of big data predictions and inferences upon us because the operations of big data themselves are covered in legal and commercial secrecy. An additional concern raised by big data is that it tilts an uneven scale in favor of government/organization and against individuals—and the big benefits of big data often come at individuals' expense.[16]

B. The limits of personal information protection

The protection of personal information derives from the privacy protection. The Charter of Fundamental Rights of the European Union[17] expresses the importance of protection of privacy and personal information. For example, article 7 of the Charter provides that everyone has the right to respect for his or her private and family life, home and communications; article 8 of the Charter states that everyone has a right to the protection of personal data concerning him or her. Traditionally the concept of privacy focuses on private and secret matters and no protection is given in an open field. In 1967, the US Supreme Court

[16] The phrase "If you're not paying for it, you're not the customer; you're the product." has become a staple in online culture. *See* Johnathan Zittrain, *Meme patrol: "When Something Online is Free, You're Not the Customer, You're the Product,"* THE FUTURE OF THE INTERNET (Mar. 21, 2012), http://blogs.law.harvard.edu/futureoftheinternet/2012/03/21/meme-patrol-when-something-online-is-free-youre-not-the-customer-youre-the-product/

[17] Charter of Fundamental Rights of the European Union, *OJ* C 364/10, 18.12.2000, http://www.europarl.europa.eu/charter/pdf/text_en.pdf.

established a renowned test of "reasonable expectation of privacy," which the legal protection is provided that a certain individual presents an actual (subjective) expectation of privacy, and such an expectation of privacy is generally recognized by society as "reasonable" (objective).[18] Both two jurisdictions have recognized the protection of personal information as the fundamental right. However, the principles to protect privacy and personal information must be balanced against additional societal values, such as national security, public health, law enforcement, environmental protection, economic efficiency, and even free speech.

The societal benefits of big data must be reconciled with increased risks to individuals' privacy. For the past four decades, the tension between data innovation and information privacy has been moderated by a set of principles broadly referred to as the Fair Information Practice (FIP) or Fair Information Practice Principles (FIPPs).[19] Evolving from the framework of 1980 OECD Privacy Guidelines,[20] the US White House in February 2012 issued a version of FIPPs in the context of a report[21] prepared by the Department of Commerce that includes seven principles of individual control, transparency, respect for context, security, access and accurate, focused collection, and accountability. The big data challenges some of the fundamental principles, such as the scope of the

[18] Katz v. United States, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

[19] FIPs are a set of internationally recognized practices for addressing the privacy of information about individuals FIPs are important because they provide the underlying policy for many national laws addressing privacy and data protection matters. For more details about FIPs/FIPPs, *see* Robert Gellman, *Fair Information Practice: A Basic History*, ver. 2.13, Feb. 11, 2015, http://bobgellman.com/rg-docs/rg-FIPShistory.pdf. *But cf.* Fred H. Cate, *The Failure of Fair Information Practice Principles*, *in* CONSUMER PROTECTION IN THE AGE OF THE INFORMATION ECONOMY 341-378 (Jane K. Winn ed., 2006).

[20] *See* OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, ORG. FOR ECON. CO-OPERATION & DEV. (Sept. 23, 1980), http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html.

[21] *See* The White House, *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy* (Feb. 2012), at 4, http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

framework (addressed by the term of "personal identifiable information" (PII)), the concepts of data minimization ("focused collection"), consent ("individual control" and "respect for context"), and the right of individual access ("access and accuracy").[22]

C. The definition of personal information

The personal information around different law jurisdictions is defined in the similar ways. According to Article 2(a) of European Union's Data Protection Directive of 1995,[23] it states that "personal data" shall mean any information relating to an identified or identifiable natural person ('data subject').[24] In Section 3(1) of German Federal Data Protection Act[25] the term "personal data" means any information concerning the personal or material circumstances of an identified or identifiable individual (data subject).[26]

According to the abovementioned 1995 Directive, an identifiable person is one who can be identified, *directly or indirectly*, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.[27] In the latest draft of EU General Data Protection Regulation,[28] article 4(1) provides the definition

---

[22] Federal Trade Commission Commissioner Julie Brill said: "Big Data's impact on privacy is requiring some new and hard thinking by all of us." *See* Julie Brill, Remarks at Fordham University School of Law: Big Data, Big Issues (Mar. 2, 2012), https://www.ftc.gov/sites/default/files/documents/public_statements/big-data-big-issues/120228fordhamlawschool.pdf.

[23] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, *OJ* L 281 (Nov. 23, 1995) ("95/46/EC").

[24] Article 2(a) of 95/46/EC.

[25] Federal Data Protection Act (Bundesdatenschutzgesetz, BDSG), promulgated on 14 January 2003 (Federal Law Gazette I p. 66), as most recently amended by Article 1 of the Act of 14 August 2009 (Federal Law Gazette I p. 2814).

[26] Section 3(1) of German Federal Data Protection Act.

[27] Article 2(a) of 95/46/EC.

[28] European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)

of "data subject," which means an identified natural person or a natural person who can be identified, *directly or indirectly,* by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person.

In Taiwan's Personal Information Protection Act,[29] the term "personal information" in Article 2(1) means the name, date of birth, ID card number, passport number, characteristics, fingerprints, marital status, family, education, occupation, medical records, medical treatment, genetic information, sexual life, health examination, criminal record, contact information, financial conditions, social activities and *other information* which may be used to identify a natural person, both *directly and indirectly*. A natural person who can be identifiable *indirectly* by using other information other than the listed information in Article 2(1) of the Act means that the agencies possessing the information can not directly identify the specific person without comparing to, combining with or connecting to other information.[30]

The issue of how to determine whether a person is identifiable, one should take account of *all the means* likely reasonably to be used either by the controller or by any other person to identify the said person.[31] When the data subject is no longer identifiable, the principle of protection shall not apply to data render anonymous.[32] Referring to the recommendation No. R (90) 19 of Committee of

---

(COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)).

[29] Personal Information Protection Act, promulgated on 26.5.2010, http://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=I0050021.

[30] Article 3 of the Enforcement Rules of the Personal Information Protection Act, promulgated on 26.9.2012, http://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=I0050022.

[31] Recital (26) of 95/46/EC.

[32] *Id*.

Ministers, Council of Europe,[33] an individual shall not be regarded as "identifiable" if identification *requires an unreasonable amount of time, cost and manpower*.[34] According to Article 2(1) of Japanese Act on the Protection of Personal Information,[35] the term "personal information" shall mean information about a living individual which can identify the specific individual by name, date of birth or other description contained in such information (including such information as will *allow easy reference* to other information and will thereby enable the identification of the specific individual).

Thus, according to the above laws and regulations, personal information can be divided into three categories by different degree of identification.

Table 1: Types of personal information

Types of personal information	Description
Identified personal information	The information belongs to a specific natural person that has been identified. *E.g.*, a person discloses his health record on the web.
Directly identifiable personal information	The unclear information that is collected, processed and used for identifying the specific natural person. *E.g.*, the names or address (contact information) is collected for identifying a specific person.
Indirectly identifiable personal information	The unclear information that is collected, processed and used but not enough to identify the specific natural person. Further information is needed for identifying the specific natural person. *E.g.*, the address of an unknown person is collected and further information (such as gender or name) is needed for further identification.

[33] Council of Europe, Recommendation No. R (90)19 of the Committee of Ministers to Member States concerning the protection of personal data used for payment and other related operations, adopted 13.9.1990.
[34] *Id.* Appendix to Recommendation, 1.2.
[35] Act on the Protection of Personal Information, Act No. 57 of May 30, 2003, http://www.japaneselawtranslation.go.jp/law/detail/?id=130&vm=04&re=02.

III. A qualitative change: reviewing the technology of de-identification and the principle of data minimization and individual control

Big data changes the landscape of what we know in the world. The effect of qualitative change influenced by the big data means that the anonymized personal information may be re-identified or identifiable when more and more data are collected and are further analyzed. The privacy or data protection law applies to the *identified* or *identifiable* personal information and the identifiable information could be direct and indirect. Big data could improve the possibility of identifying the personal information, especially "indirect" identifiable information. Thus, the article discovers four factors including the technology of de-identification, data minimization, individual control (of PII), and the possibility and difficulty to recognize the combined data.

A. The Technology of de-identification

Traditionally the methods of de-identification (anonymization, encryption, key-coding) were viewed as sharp weapons allowing organizations to reap the benefits of analytics while preserving individuals' privacy. Since the anonymized data can often be re-identified and attributed to specific individuals, the de-identified data is just a temporary state rather than a stable category. It raises the concerns about the scope of information subject to privacy law. A possible solution supported by Ohm is that all data should be treated as personally identifiable and subjected to the regulatory framework.[36] However, such an expanded definition of PII would create perverse incentives for organizations to

[36] Ohm, *supra* note 5, at 1742-43.

abandon de-identification and thus increase the risks of privacy and data security.

A further risk is that, with a vastly expanded definition of PII, the privacy framework would become unworkable. Besides, anonymized information always carries some risk of re-identification and thus creates uncertainty in the framework of PII protection. We cannot know whether the information truly corresponds to a particular individual, and the dataset becomes more anonymous as larger amounts of uncertainty are introduced.

The dichotomy between "identifiable" and "non-identifiable" data based on labelling information either "personal identifiable" or not, is unhelpful and inevitably leads to an inefficient arms race between de-identifiers and re-identifiers. PII should be defined based on a risk matrix taking into account the risk, intent, and potential consequences of re-identification,[37] and the integrity, accuracy, and value of the data may be degraded or lost, together with some of potential societal benefits.[38] A law review article written by Omer Tene and Jules Polonetsky addressed a better solution to the question: to view the identifiability of data as a continuum as opposed to the current dichotomy and to adopt an approach proposed by US Federal Trade Commission in a report,[39] which overlays the statistical probability of re-identifiability with legally enforceable organizational commitments and downstream contractual obligations not to re-identify or to attempt to re-identify.[40]

We should regard de-identification as a protective measure to be taken

---

[37] *See* Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 NYU L. REV. 1814 (2011); Omer Tene, *The Complexities of Defining Personal Data: Anonymization*, 8 DATA PROT. L. & POL'Y 6 (2011).

[38] Tene & Polonetsky, *Big Data for All*, *supra* note 2, at 258.

[39] *See generally* FTC, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE, RECOMMENDATIONS FOR BUSINESS AND POLICYMAKERS (2012).

[40] According to the FTC, "as long as (1) a given data set is not reasonably identifiable, (2) the company publicly commits not to re-identify it, and (3) the company requires any downstream users of the data to keep it in de-identified form, that data will fall outside the scope of the framework." *Id.* at 22.

under the data security and accountability principles, rather than a solution to the big data puzzle. Organizations collecting and harvesting big data would be wise to de-identify data to the extent possible while not compromising their beneficial use. In the meantime, the privacy regulation will be applicable to fairly apply to de-identified data because researchers, if any appropriate incentive, have the ability to re-link almost any piece of data to an individual.[41]

B. Data Minimization

Data minimization has been a fundamental principle of privacy law.[42] It can be effectuated in a number of different ways, including by limiting collection, use, disclosure, retention, identifiability, sensitivity, and access to personal data. It can also help guard against two privacy-related risks. First, larger data stores present a more attractive target for data thieves, both outside and inside an organization – and increases the potential harm to consumers from such an event. Second, if an organization collects and retains large amounts of data, there is an increased risk that the data will be used in a way that departs from consumers' reasonable expectations.[43] Thus, organizations are required to delete data that is no longer used for the purposes for which they were collected, and to implement restrictive policies with respect to the retention of personal data in identifiable form.

In the age of big data, it is not clear that minimizing information collection is always a practice approach to privacy. The business model of big data is

---

[41] Tene & Polonetsky, *Big Data for All*, *supra* note 2, at 259.
[42] OECD Guidelines, *supra* note 20, at ¶¶ 7-8.
[43] *See* FEDERAL TRADE COMMISSION, INTERNET OF THINGS: PRIVACY & SECURITY IN A CONNECTED WORLD (FTC STAFF REPORT) iv (2015), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.

basically against the principle of data minimization. The basic concept of big data is to collect more data for longer periods of time, and to aim precisely at unanticipated secondary use of data. Organizations today collect and retain personal data through various channels including the Internet, mobile, sensors, video, e-mail, and social networking tools. Data organizations collect data directly from individuals or third parties, and they harvest private, semi-public (e.g., Facebook), or public (e.g., the electoral roll) sources. Data minimization is simply no longer the market norm.[44]

In a big data world, the principle of data minimization should be interpreted in a different way, requiring organizations to de-identify data when possible, implement reasonable security measures, and limit uses of data to those that are acceptable from not only an individual but also a societal perspective.[45]

C. Individual control/consent

Current privacy and data protection laws are premised on a biased value judgment in favor of *individual control* over *highly beneficial uses of data*.[46] In fact, a coherent regulatory framework would be based on a risk matrix, considering the value of different uses of data against the potential risks to individual autonomy and privacy.[47] Where prospective data uses clearly outweigh privacy risks, the legitimacy of processing should be assumed even if individuals decline to consent. For instance, web analytics—the measurement, collection, analysis and reporting of online data for purposes of understanding and optimizing web usage—creates remarkable value by ensuring that products

[44] Tene & Polonetsky, *Big Data for all*, *supra* note 2, at 260.
[45] *Id.*
[46] Tene & Polonetsky, *Privacy in the Age of Big Data*, *supra* note 3, at 67.
[47] *Id*.

and services can be improved to better serve consumers. Privacy risks could be minimal because the analytics, if properly implemented, deals with statistical data, typically in de-identified form. However, requiring online users to consent to the analytics would no doubt severely reduce its application and use.[48]

Currently there are too many processing activities premised on individual consent. Yet, individuals are ill-placed to make responsible decisions about their personal information because they are confronting well-documented cognitive biases and the increasing complexity of the information ecosystem.[49] Online users see the term "privacy policy" and believe that their personal information will be protected in specific ways and assume that a website that advertises a privacy policy will not share their personal information.[50] In fact, however, privacy policies often serve more as liability disclaimers for businesses than as assurances of privacy for consumers.

This article does not argue that one should not be asked to expressly consent to the use of their information or offered an option to opt out; rather, it suggests that the merits of a given data use should be discussed as a broader societal issue. When making decisions about the need for individuals' consent and how it should be obtained, policymakers should recognize that *default rules* often prevail and determine the existence of these data uses. Disputes about whether consent should be solicited or opt-out choice provided focus solely on the mechanics of expressing consent.[51] But emphasized the concepts of consent and data minimization, with little attention to the value of data use, could jeopardize

---

[48] *Id*.
[49] *Id*.
[50] Joseph Turow, Chris Jay Hoofnagle, Deirdre K. Mulligan, Nathaniel Good & Jens Grossklags, *The Federal Trade Commission and Consumer Policy in the Coming Decade*, 3(3) I/S: J. L. & POL'Y FOR INFO. SOC'Y 723, 724 (2007).
[51] *See* Omer Tene & Jules Polonetsky, *To Track or 'Do Not Track': Advancing Transparency and Individual Control in Online Behavioral Advertising*, 13 MINN. J.L. SCI. & TECH. 282 (2012).

innovation and social advancement.

The traditional view on the legitimacy of data use had always intended to take additional societal values into account beyond privacy. For instance, national security has been regarded a higher value to override privacy protection in certain cases with the satisfaction of public interest requirements. As a result, the role of consent should be separated according to normative choices made by policymakers concerning respective data uses. In some situations, consent may not be required, while in others, consent should be assumed subject to a right of refusal. In some statutory cases, consent should be required to legitimize data use.[52]

D. The possibility and difficulty to recognize the combined data

A set in mathematics is a collection of distinct objects and can be analogous to *indirect identity*. For example, the numbers 2, 4, and 6 are distinct objects when considered separately, but when they are considered collectively they form a single set.[53] We can add more numbers in the set if the conditions are met. We can also compare different sets and pull out the same factors in another set. When the amount of data is collected more and more, the intersections of the same factors become fewer and fewer. As to the personal information, the individual can be identified when the intersection becomes only one.

The previous study of EU and Japan's data protection laws has addressed proper limits on the scope of *indirectly* identifiable personal information. The article argues that certain data or datasets are limited to apply to the PIPA.

---

[52] Tene & Polonetsky, *Big Data for all*, *supra* note 2, at 262-63.

[53] The single set can be written in a mathematical way: {2, 4, 6}. *See* Set, WIKIPEDIA, https://en.wikipedia.org/wiki/Set_%28mathematics%29.

a. The data (or datasets) that are too difficult to be identified

If we use the set of "gender" to identify the specific individual, the data in the set can be the indirectly identifiable information under the PIPA. However, it can be difficult to identify the specific person because the scope of the gender set is too broad. If we collect other data combined with the gender in order to identify the specific person, it may cost too much regarding time, money and manpower. In an opposite view, the actual or possible harm caused to the individual may be trivial, and thus the court shall adopt a narrow standard to interpret the term of "indirectly identifiable." Such data or datasets excluded from the category of PII can be acceptable because it will not be amounted to detriment to moral rights of data subject.

b. The data that does not help for identification and is not reasonable expectation of privacy

In general, the more sorts of data, it seems more possibilities to identify the specific individual; however, some sorts of data may not narrow the scope for identifying the specific individual. For example, the ID card in Taiwan contains a serial number including an English letter and 9 digits. There is a coding rule on the serial number. The English letter stands for where you are born. A is represented for Taipei City and F is for New Taipei City. The first digit stands for gender: 1 is for male and 2 is for female. Therefore, the ID number contains three datasets including birth places, genders, and ID number itself.

The inferred information (birth place and gender) in the ID card does not help for identifying the specific person while the expectation of privacy

that the data subject asserted may not be reasonable (because the information of ID number legally collected contains birth place and gender). This example can also apply to the datasets of home address and zip code and the IP address and the Internet service provider.

IV. A case study in Taiwan

Since 2002 all telecommunication operators are required by the communication regulator—National Communications Commission (NCC)—to offer number portability service ("NP") for their customers and all operators jointly established the Centralized Number Portability Database (CNPDB) to manage the NP service. Each mobile user may churn to another mobile operator without changing his or her original phone number. According to the statistics by the NCC, the effective cases of number portability have reached 33.8 million since 2005 to May 2015.[54]

Table 2: The effective cases of NP (including mobile and landline in Taiwan

Year	MNP	LNP	Total
2005	93,858	94	93,952
2006	511,358	516	511,874
2007	2,080,264	1,093	2,081,357
2008	3,318,003	3,946	3,321,949
2009	3,220,594	8,109	3,228,703
2010	3,072,746	6,629	3,079,375
2011	3,068,243	5,102	3,073,345
2012	3,452,627	6,177	3,458,804
2013	3,457,314	4,756	3,462,070
2014	6,197,019	3,250	6,200,269
Jan 2015	694,180	189	694,369
Feb 2015	670,103	158	670,261

[54] NCC, the statistics of number portability of 2015, http://www.ncc.gov.tw/chinese/files/15060/%E9%9B%BB%E8%A9%B1%E8%99%9F%E7%A2%BC%E5%8F%AF%E6%94%9C%E6%9C%8D%E5%8B%99.pdf.

Mar 2015	758,599	311	758,910
Apr 2015	724,001	308	724,309
May 2015	714,708	155	714,863
Total	33,789,009	40,793	33,829,802

Source: NCC (2015).

The CNPDB is a good source for researching the churn rate and user behaviors through certain technologies of big data. In 2012, a mobile company in Taiwan launched a new app service called "M+ messenger," which allowed users to know the specific name of mobile operator affixed to the phone numbers in the contact book of mobile phone, so as to find out the intra-network friends and enjoy the cheaper rate of calls. However, the app was concerned whether the personal information was illegal collected and used, and a customer brought a lawsuit to court for damages of privacy invasion.

A. Facts and issues

A mobile operator "T" developed communication app "M+ Messenger" (hereinafter "M+") and promoted the business in the name of T's affiliate "K" (T and K is collectively the defendants). Mobile users can install such an app in their smartphones. When the app is in operation, the name of specific mobile operators of individuals is displayed on the screen of smartphone affixed to the numbers of the contact list.

Although K is not a mobile operator (K has no obligation to join the CNPDB), the identities of mobile operators as shown on the phone are queried via its parent company T. Originally the plaintiff was the user of mobile operator "C," and then he ported his phone number to mobile operator "F." On one occasion, the plaintiff discovered that his phone number and the name of his mobile operator F were

displayed on the screen of his friend's smartphone. The plaintiff believed that the function of M+ to reveal the name of his mobile operator has harmed his privacy due to illegal collection of personal information and thus violates the Personal Information Protection Act ("PIPA"). He then brought a lawsuit to Taiwan Taipei District Court for monetary damages.[55]

The issue of the case is whether the identity of specific mobile operator affixed to the phone number is the personal information under the PIPA.

B. Reasoning of the court

The court found that the identity of mobile operator of specific phone number is the personal information under the PIPA. The reasons are as follows.

(A) Phone number is the personal information under the PIPA.

The phone number of a natural person is a kind of contact information. The phone number is a combination of digits and has no distinctiveness. However, when the number is associated with (comparing to, combining with or connecting to) other personal information, such as name, personal ID number, characteristics, and/or other social activities, it can be used to *indirectly* identify the specific natural person. Therefore, the phone number is the personal information protected under the PIPA.

(B) The identity of mobile operator which the phone number is belonged to is a sort of additional information of personal phone number, and thus under the protection of PIPA.

The identity of mobile operator which the phone number is connected to

[55] Liu v. MyMusic, Taiwan Taipei District Court Bei-Hsiao-Tzu No.103-1360 Judgment for small-amount civil action.

is a sort of additional information when the number compares to, combines with or connects to other personal information. The certain identity of mobile operator can be one of the social activity data for indirectly identifying the specific individual, and thus is contact information under the protection of PIPA. Otherwise, if anyone can collect, process or use the identity of mobile operator which the number is connected to and *indirectly* identify the specific individual by means of piecing up, comparing to, combining with or connecting to data subject's information of social activities, such an individual may be in the risk of been peeped, disturbed and annoyed by marketing matters and is against the legislative purpose for the personal information protection.

(C) The collection, transmission of personal information to the third party's phone via the M+ app without consent by the data subject is illegally use of personal information.

M+ is developed by T and is promoted by its affiliate K. T is a co-founder of CNPBD and is eligible for querying the number which is belonging to specific telecommunication operator. T queried the numbers in the contact list of mobile phone holders from the CNPDB, and transmitted the names of mobile operators affixed to the number of each contact. However, the defendants could not prove that they obtained the written consent to the fact that the name of mobile operator was disclosed and transmitted to third party's phone, and thus violates the PIPA for illegal use of personal information.

The defendants claimed that the identity of mobile operator affixed to specific individual transmitted to third party is for the purpose of

distinguishing whether they are in the intra-network or inter-network, so as to save the fee of communication, but the court found that such a purpose is beyond the specific purpose when collecting the porting numbers and storing in the CNPBD. It is not a reasonable connection to such a use and is irrelevant with public interest. The illegal use of plaintiff's personal information is detrimental to his moral rights.

(D) Distinguishing the short code "57016" that is dedicated for querying intra-network or inter-network and does not connected to other personal information of users and thus does not constitute the identity.

The communication regulator—National Communications Commission required all the telecommunication operators should provide the dedicated line "57016" for customers to query the names of operators according to phone numbers. But the "57016" short code is to provide the query service when a phone number is porting to another operator's system, and does not combine with or connect to the personal information of other natural persons and has nothing to do with identity.

C. Comments

As above reasoning delivered by court, this article addresses the following comments according to the above four reasons.

(A) The information of the name of the mobile operator in this case is an *identified* personal information

In this case, the plaintiff gave the phone number to his friend, who input the number to the contact book of a smartphone. And then the M+ app queried the phone number from the CNPDB and obtained the name of the

mobile operator, which was displayed on screen of the phone. For plaintiff's friend, the identity of plaintiff is an *identified* specific individual. Thus, it is not necessary to discuss whether the specific individual can be *indirectly identifiable* by the information of identity of mobile operator. Rather, it should discuss whether the plaintiff has the reasonable expectation of privacy on the information of identify of mobile operation affixed to the phone number.

As discussed above, the phone number was assigned by the NCC to each telecommunication operators. Before the NP service implemented, customers can easily determine which the mobile operator is by the prefix 4-digit of phone number. After the NP service, the NCC required all telecommunication operators to provide the short code "57016" service for the customer to look up the relationship between the dialing number and the customer's own phone number is whether the intra-network or the inter-network. In the meantime, each mobile operator has provided the same query service on their official webpage. No matter by dialing 57016 or looking up on the webpage, the query result is either the intra-network or the inter-network, and does not report the exact name of the mobile operator that the number belongs to.

In fact, there are only five third-generation mobile operators in Taiwan. If a customer attempts to know the exact mobile operator that the specific number belongs to, he or she may only query the number at most four times to achieve the goal. Even if the plaintiff subjectively has the expectation of privacy on the name of the mobile operator that the number belongs to, under the common knowledge of general person, such information that can be easily looked up by dialing 57016 or by browsing the webpage **is objectively not reasonable** for plaintiff's expectation of privacy. Since the phone

number is identified information, it does not cause any disturbance to the autonomy of personal information if the name of the mobile operator has been identified. If the disturbance does exists, the crucial point is not the identity of mobile operator but the phone number itself. Therefore, if anyone recognizes a specific individual and legally collects the phone number of the individual, the PIPA does not extend the protection to the collection and use of the information of mobile operators that the phone numbers belong to.

(B) As to the *identifiable* specific individual, whether the identity of the mobile operator affixed to the phone number is the personal information protected under the PIPA?

If anyone does not recognize the plaintiff and collects his phone number and keys it in the contact book of a smartphone, it is an issue about the *identifiable* personal information in terms of the scenario of M+ app. The fact of this case at hand is just the situation. As discussed above, if anyone recognizes the phone number, he or she can learn the mobile operator affixed to the number by querying the CNPBD or the webpages that operators provided. Thus, there is no reasonable expectation of privacy at all. As to the identity, the court decision found that once the information of mobile operator combining with other personal information, the specific individual may be identifiable. As a matter of fact, **if anyone has firstly recognized the phone number and then collects the information of mobile operator affixed to the number, the scope of identification will not be reduced because of more personal information added.** Accordingly, the information of mobile operator affixed to the phone number has nothing to do with the identification of specific individual. In other words, the extent of

identification “phone number + mobile operator + other information” is totally the same as “phone number + other information,” and such the information of mobile operator is not the *indirectly identifiable* information that is protected under the PIPA.

(C) As to the *identifiable* specific individual, whether the PIPA is applicable to the collection of the identity of the mobile operator without the phone number?

If anyone does not collect the phone number but the information of mobile operator, the latter information does not help identify the specific individual. For example, the number of subscribers of Fareastone Telecom, the leading mobile company in Taiwan, has reached 74 million.[56] To narrow down the scope of the dataset of mobile operator whatever by gender or age, it must cost much on time, money and manpower to identify a specific individual, and thus the harms or the possibility to cause harm to the data subject is too trivial, and the PIPA is not applicable to this case.

(D) The reasoning of this case cannot justify the differences between the M+ and the short code 57016.

As discussed above, the purposes of app M+ and the short code 57016 are the same: the prerequisite is to identify the specific phone number, and then the user queries the number from the same database—CNPBD, either from the app M+ or short code 57016, and then collects and uses the information of mobile operator. Since these two situations are not different, the judgements on the two cases should not be different. However, the court

[56] Fareastone, Monthly revenue producing customers, http://www.fetnet.net/cs/Satellite/eCorporate/ecoOperationAnalysis#

differentiated these two situations: on one hand the court found that the app M+ has combined “phone number” and “the information of mobile operator,” and then the specific individual may be indirectly identifiable if other information is further collected; on the other hand, the court found that the short code 57016 has combined “phone number” and “the information of mobile operator,” but such datasets do not connect to the personal information of other natural persons and has nothing to do with identity. The reasoning by the court is contradictory.

V. Conclusion

Big data has changed the landscape that we know what the world is, especially the fundamental principles of privacy laws or data protection laws in many legal jurisdictions, such as data minimization and individual control/consent. Although it is important to protect the personal information from illegal invasion, the beneficial use of data is important as well. Many organizations use methods of de-identification to distance data from real identities in order to protect the privacy while the technologies of re-identification are developing and thus create privacy risks and uncertainty in the framework of personal data protection.

The personal identifiable information should be defined based on a risk matrix taking into account the risk, intent, and potential consequences of re-identification, as opposed to a dichotomy between “identifiable” and “non-identifiable” data. The dichotomy between “identifiable” and “non-identifiable” data based on labelling information either personal identifiable or not is not helpful. We need a new legal framework of personal information protection in the age of bid data.

Reference

Cukier, K. (2010). Data, Data Everywhere. *The Economist*, Feb. 25, 2010, http://www.economist.com/node/15557443.

Tene, O. & Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property*, ***11***, 239-273.

Tene, O. & Polonetsky, J. (2012). Privacy in the Age of Big Data: A Time for Big Decisions. *Stanford Law Review Online*, ***64***, 63-69.

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review, **57***, 1701-1777.

Richards, N. & King, J. (2014). Big Data Ethics. *Wake Forest Law Review, **49***, 393-432.

Dumbil, E. (2012). What Is Big Data?: An Introduction to the Big Data Landscape. *O'Reilly*, Jan. 11, 2012. Retrieved from http://strata.oreilly.com/2012/01/what-is-big-data.html.

Wang, P et al. (2012). Understanding Road Usage Patterns in Urban Areas. *Nature Scientific Reports*, ***2***, 1-6. Retrieved from http://www.nature.com/srep/2012/121220/srep01001/pdf/srep01001.pdf.

Salmon, P. et at. (2009). *Distributed Situation Awareness: Theory, Measurement and Application to Teamwork*. UK: Ashgate.

Konkel, F. (2013). Boston Probe's Big Data Use Hints at the Future. *FCW*, Apr. 26, 2013. Retrieved from http://fcw.com/articles/2013/04/26/big-data-boston-bomb-probe.aspx.

Zittrain, J. (2012). Meme patrol: "When Something Online is Free, You're Not the Customer, You're the Product. *The Future of the Internet*, Mar. 21, 2012. Retrieved from http://blogs.law.harvard.edu/futureoftheinternet/2012/03/21/meme-patrol-when-something-online-is-free-youre-not-the-customer-youre-the-product/

Gellman, R. (2015). *Fair Information Practice: A Basic History*, ver. 2.13. Retrieve from http://bobgellman.com/rg-docs/rg-FIPShistory.pdf.

Cate, F. (2006). The Failure of Fair Information Practice Principles. in Winn, J. (ed) *Consumer Protection in the Age of the Information Economy*, 341-378.

White House, Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy. Feb. 2012. Retrieved from http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

Brill, J. (2012). Remarks at Fordham University School of Law: Big Data, Big Issues. Mar. 2, 2012. Retrieved from https://www.ftc.gov/sites/default/files/documents/public_statements/big-data-big-issues/120228fordhamlawschool.pdf.

Schwartz, P. & Solove, D. (2011). The PII Problem: Privacy and a New Concept of Personally Identifiable Information. *New York University Law Review*, ***86***, 1814-1894.

Tene, O. (2011). The Complexities of Defining Personal Data: Anonymization. *Data Protection Law and Policy*, ***8***, 6-7 (2011).

Federal Trade Commission (2012). *Protecting Consumer Privacy in an Era of Rapid Change, Recommendations for Business and Policymakers*.

Federal Trade Commission (2015). *Internet of Things: Privacy & Security in a Connected World.* FTC Staff Report.

Turow, J. et al. (2007). The Federal Trade Commission and Consumer Policy in the Coming Decade. *Journal of Policy for Information Society*, ***3***(3), 723-749.

Tene, O. & Polonetsky, J. (2012). To Track or 'Do Not Track': Advancing Transparency and Individual Control in Online Behavioral Advertising. *Minnesota Journal of Law Science & Technology*, *13*(1), 281-357.