

Dugmore, Keith

Working Paper

Information collected by Commercial Companies: What might be of value to Official Statistics? The case of the UK Office for National Statistics

RatSWD Working Paper, No. 151

Provided in Cooperation with:

German Data Forum (RatSWD)

Suggested Citation: Dugmore, Keith (2010) : Information collected by Commercial Companies: What might be of value to Official Statistics? The case of the UK Office for National Statistics, RatSWD Working Paper, No. 151, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin

This Version is available at:

<https://hdl.handle.net/10419/43614>

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.



German Data Forum
(RatSWD)

www.germandataforum.de

RatSWD

Working Paper Series

Working Paper

No. 151

Information collected by Commercial Companies:
What might be of value to Official Statistics?
The case of the UK Office for National Statistics

Keith Dugmore

July 2010

Working Paper Series of the German Data Forum (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/2008 Heike Solga; since 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Information collected by Commercial Companies:
What might be of value to Official Statistics?
The case of the UK Office for National Statistics

Keith Dugmore*

Demographic Decisions Ltd.(dugmore[at]demographic.co.uk)

* This report was commissioned by the UK's Office for National Statistics, and was completed in November 2009. The author would like to express his thanks to ONS for agreeing that this version can be made available on the RatSWD website.

Contents	Page
Executive summary	3
1. Introduction	5
2. ONS's priorities for seeking new information	7
3. The UK population as customers	13
4. Interviewing commercial companies – method	14
5. The information collected by commercial companies – summaries	15
6. ONS's interests, and potential sources of useful information	27
7. Conclusions and recommendations – and a postscript	32
Annex A. ONS's Specification of requirement	34
Annex B. Interview Topic Guides	35
Annex C. Acknowledgements – interviewees, and additional contacts	41

Executive summary

This report investigates whether information collected by large commercial companies about their customers might have value to the UK's Office for National Statistics. ONS's focus is on information which may improve:

- Population Estimates: providing evidence of current numbers and types of migrants, both inward and outward, and movement between areas.
- 2011 Census: resource targeting for fieldwork coverage, chasing non-responses, and quality assuring the results.

Beyond 2011, as more use is made of administrative files, commercial records may also help the development of an integrated population system.

Having made some ad hoc enquiries in recent years, ONS decided to commission a more formal initial investigation of very large consumer-facing companies across major sectors – retail, financial services, leisure, energy, water and telecoms. It is gratifying to report that they have been prepared to meet to discuss the data that they collect. Each sector has different characteristics in the ranges of products offered, sales channels used, and the detail collected about customers and their addresses. However, as one would expect, all collect comprehensive information about purchases and transactions (and often debt and fraud).

Such company customer databases usually have the disadvantages that they are not representative samples or subsets of the national population as a whole – there may be biases by region and status, and the impact of marketing campaigns, for example; also the updating of records can be patchy. However, on the plus side are the large stocks of customers (often > 10 million), the large flows of new customers, and the timeliness of the data. If viewed as potential sources of insight – rather than precise statistics – about recent trends, these may offer real value. The local incidence of debt and fraud may also be a good social indicator. Again, the energy and water industries put great store in maintaining accurate and timely address files, and these too may be of value.

When the topic was raised of making data available to ONS, it was striking that the possibility of payment was not mentioned. Rather, the emphasis was on seeking better population statistics, and, more generally, good corporate relations with government. However, issues of data security were recognised as very important, and any decision to release information would need to be handled at senior levels.

The improvement of population estimates is already a very high priority for ONS, and Census planning is now also in top gear. If commercial sources are to be investigated, this

needs to be done in the lead up to the Census in March 2011, and, if findings are positive, commercial data files also captured at the time of the Census for comparative purposes.

The report makes 6 recommendations.

- 1) **Recent migrant flows, and population stock.** Explore further with a major retailer, such as Tesco, the scope for compiling regular small area statistics.
- 2) **New, empty, and multi-occupied addresses.** Check the content and availability of the electricity industry's national meter address file, with the help of E.ON; also investigate the comprehensive address files compiled by selected regional water companies, starting with Thames Water as the area of greatest interest to ONS.
- 3) **“Difficult” addresses, subject to bad debt or fraud.** Explore financial services' pooled national files on debt and fraud, seeking the advice of Nationwide.
- 4) **Assess the potential value of customer data for additional population topics.** ONS's target topics are: internet take-up; rented addresses; income; and mobile phone usage (especially to track international movements).
- 5) **Assess the potential value of customer data for other topics.** Companies have very rich data on spending on particular products and services, and this could be a valuable supplement to the Expenditure and Food Survey.
- 6) **Consider encouraging more sharing and pooling of customer records.** This is already well developed in some sectors, especially Financial Services, and ONS should consider encouraging its extension (perhaps in alliance with other government departments) to new topics of interest, to produce more statistics “for the public good”.

1. Introduction

The Office for National Statistics is interested in information that is collected by commercial companies as possible sources of statistics, or at least valuable intelligence. Having made some ad hoc enquiries in recent years, it has now decided to make a more formal investigation of very large consumer-facing companies across major industries such as retail, financial services and utilities.

The specification for this project (see Annex A) sets four main objectives:

- Produce a statement of ONS's priorities for creating improved migration and population statistics, both in terms of timeliness and quality
- Provide an overview of relevant information that is collected by commercial companies in several different sectors
- Report observations on ONS's requirements by large commercial companies that are potential suppliers
- Present findings in a report to include: what information exists that is of interest to ONS; which targets are worth pursuing; also identify areas which have no realistic prospect of success; and recommend how companies should be pursued.

The Local Government Association has also supported this initiative. They recommended that the scope should be wide enough to appraise sources that might never be formal national or official statistics in a traditional sense but which are sufficiently robust and consistent to yield useful intelligence.

This report takes us on a journey:

- It initially reviews the ONS's priorities for new information, and focuses on population estimates, and also plans for the 2011 Census (Section 2). It is helpful to anticipate ONS's timetable between now and the publication of Census results, and to consider the potential value of both individual records and aggregate area statistics. Specific topics of interest are identified.
- Section 3 then sets the wider scene, by considering the United Kingdom population as customers who spend money with commercial companies, buying a wide range of goods and services. The UK (rather than just England and Wales) has been chosen because most commercial companies operate across the whole UK; information on spending is readily available from the Expenditure and Food Survey; and also, the report's findings might also be useful to ONS's counterparts at GRO Scotland, and NISRA.

- Turning to the potential supply of information, the method of selecting companies for interviews, and the topic guides used for interviews (see Annex B), are outlined in Section 4. It should be noted that this research relates to commercial companies who collect information from their millions of customers, not to commercial Value Added Resellers (such as Experian), whose business is to collate and resell information collected by other organisations. Information offered by such resellers may be of separate interest to the ONS.
- The information collected by commercial companies is summarised in Section 5. It is very apparent that the amount of information gathered by companies about their customers varies considerably by industry, product, and sales channel, and it is useful to set this in the context of our own personal experiences as customers of large companies.
- Section 6 gets to the heart of the matter, reviewing ONS's needs for information, and relating these to information that is collected, and which might be pursued further as sources of supply. Some of these remain at the level of the particular company, but there are also instances where the major companies in a particular industry (such as financial services and energy) pool their records to produce summaries of the entire market / nation.
- The report then summarises the main conclusions, and makes six recommendations.

The author would like to express his thanks to all the people (acknowledged in Annex C) who have made time to provide information and to offer their views.

2. ONS's priorities for seeking new information

2.1 The Project Specification

The specification requires a statement of ONS's priorities for creating improved migration and population statistics. "These will include but are not limited to the following: new topics; improvements to current subject detail; geographical detail; accuracy of data produced; and currency (i.e. timeliness)."

Subsequent discussion has confirmed and illustrated the importance of these aspects, although in many cases the interest is in information which will ultimately help ONS to produce and publish better population statistics, rather than directly providing statistics themselves. For example, information about which areas have addresses with high levels of multi-occupancy could help to target Census resource and therefore improve coverage, which would help produce more accurate statistical estimates.

2.2 The priority for this exercise: "Population statistics and Census 2011"

The document "Plans for the ONS statistical work programme 2007-8"¹ provides a good public statement of ONS's strategic priorities, identifying areas of expansion and contraction.

Amongst those being expanded is "Census 2011 and population statistics", which it has been agreed should form the primary focus for this project. To quote the work programme, "The census forms the crucial benchmark for all other social statistics, including population estimates and survey results."

Progress is currently being made in the Migration Statistics Improvement Programme that will feed into improved population statistics². This is "focussed on making improvements to the methods and data sources used to estimate the population at national and local levels during the inter-censal period. These improvements are needed to minimise the risk of divergence between the rolled forward mid-year population estimates and the 2011 Census-based population estimates, and to better understand the differences that remain."

The key strands of the research are:

"Migration and Population Statistics meeting user needs:

- at the right time
- covering the relevant populations
- measuring change accurately (national and local)
- detecting turning points

¹ <http://www.statistics.gov.uk/about/ons/downloads/ONSWP0708Commonbody.pdf>

² <http://www.statistics.gov.uk/about/data/methodology/specific/population/future/imps/updates/default.asp>

and trusted as authoritative:

- based on range of developed best up to date sources
- enhanced transparent sustainable statistical methods
- with quality measures (and low gap with 2011 Census)

by highly engaged users”

Similarly, planning for the next Census, which is to be held on 27 March 2011, is well under way. The recently published White Paper³ identifies five strategic aims, which include:

- to give the highest priority to getting the national and local population counts right
- to maximise overall response and minimise differences in response rates in specific areas and among particular population sub-groups

More detailed information is given in the ONS Centre for Demography Work Programme for 2008/9⁴

Improvements in the accuracy of both mid-year estimates and also the Census results will also provide a foundation for better measurement of population in the decade 2011-2021.

2.3 Comparing population estimates and Census results for 2011 – & working backwards

The outputs from these two major projects will be subject to public scrutiny and comparison when the results of the 2011 Census are published, and ONS staff are conscious that they should both benefit from any better information that might produce both more accurate population estimates now, and the best Census counts in 2011. Looking three years ahead, and working backwards, the following can be anticipated.

³ <http://www.official-documents.gov.uk/document/cm75/7513/7513.asp>

⁴ <http://www.statistics.gov.uk/about/data/downloads/NSCDWorkProgramme2008-09Final.pdf>

Table 2.3. Population estimates and 2011 Census results – the timeline

Date & Event	Activities
<p>Late 2012: publication of the first Census results</p>	<p>Census results (estimates) for March 2011.</p> <p>Users will be interested in national and local comparisons with other sources, for example:</p> <ul style="list-style-type: none"> • ONS's Mid Year and Short Term Migrant Estimates for mid 2011. • Administrative file statistics captured for ONS by government organisations relating to the end of March 2011. <p>Differences will need to be understood and explained. They may also lead to the question: "Did you consider ALL the relevant data sources that could have helped you to maximise the quality of your updated estimates, and also the coverage of the Census?"</p>
<p>Early 2012: first assessment of the Census database, Quality Assurance, and imputation of missing records and variables</p>	<p>ONS creates an initial Census database (before any imputation of missing households, persons, and variables).</p> <p>Comparison with administrative sources – at aggregate (National / Local Authority / Output Area) and potentially at individual record levels.</p> <p>Impute missing households, persons, and variables using Census Coverage Survey and Census information.</p> <p>A contingency of imputation also informed by admin sources Note: there will be no adjustment of Census results to MYEs and STMEs, to avoid circularity.</p>
<p>Between now and 2011: assessment of novel data sources for:</p> <p>a) Tracking current population change</p> <p>b) Census coverage – planning, and chasing up responses</p>	<p>Customer data collected by commercial companies offers only partial coverage of populations (unless it is pooled by all companies in a particular market), but might be of value for:</p> <p>a) Population estimates (which are created from aggregate area statistics): the identification of any very recent apparent changes in national and local customers / populations, the recent movements of Polish migrants being one obvious example.</p> <p>b) Census planning (which needs information about addresses, as well as local areas, to target efforts to maximise response, and to identify what has changed since 2001): information about individual addresses, households and persons, and local area hotspots with particular social characteristics.</p>

In summary, information collected by commercial companies may possibly have value for:

1. Population Estimates: providing evidence of current numbers and types of migrants, both inward and outward, and movement between areas.
2. 2011 Census: resource targeting for fieldwork coverage, chasing non-responses, and quality assuring the results.
3. Beyond 2011: as more use is made of administrative records, working towards an integrated population system.

2.4 What types of additional information may be of value?

Commercial companies collect information at the individual level. This may be about individual households and persons / customers, and their characteristics and behaviour; and also customers' addresses and their characteristics (which is relevant for planning Census coverage). Many companies have millions of customers, and continuously amend their databases, adding new customers, sometimes flagging lapsed ones, and also recording existing customers' changes of address.

ONS might potentially be interested in such individual records, and also in aggregated area statistics of addresses, households or people – at postcode, Output Area, or higher levels, up to national distributions. Such aggregate statistics may have the advantages of helping detect patterns, and they may also raise fewer concerns (real or imagined) about data protection and confidentiality amongst the supplying organisations.

Unless the company concerned is a monopoly (such as a local water company), or there is pooling of records within the industry as a whole, such information will be partial, and have biases such as shifts in consumer tastes, and its own relative success in its market, and therefore not able to provide consistent area statistics. Rather than being used to estimate population or migration, either alone or in combination with other sources, it is most likely to be useful as 'indicator' type information:

- Spotting new trends by quarter or month – for example, changes in the sales of certain foods, or the arrival of new customers with particular overseas surnames.
- Triangulation – for example, confirming a Local Authority's assertion that many people have recently arrived from Romania.
- Relationships and proxies – for example, confirmation that postcodes which ONS has already identified as potentially difficult to enumerate may also have many buy-to-let properties.

2.5 ONS's priorities: specific topics of interest

Guided by ONS's Work Programme, two discussion sessions have been held with ONS staff ranging across the Centre for Demography, Census, and Methodology Divisions (Annex C). The two main priorities are:

- Population Estimates: tracking current flows of migrants, both nationally inward and outward, and movement between areas

- 2011 Census: resource targeting for fieldwork coverage, chasing non-responses, and quality assuring the results

Turning to the specific topics of interest, ONS has already carried out substantial research into classifying types of people and types of area which are hard to count⁵. Taking this and the discussions together, interest has been expressed in the following topics (many of which are co-related):

Table 2.5 Topics of interest for both Population Estimates and Census purposes

Topic
<p>People / households (individuals and / or aggregated area statistics):</p> <ul style="list-style-type: none"> • Flows of recent in-migrants and out-migrants, + current and previous addresses. • Stock of population. <p><i>for both Population Estimates and Census purposes, with such classifications as:</i></p> <ul style="list-style-type: none"> ➤ Students. ➤ Young males (& ideally the whole population by age & sex). ➤ People using more than one residence (population definition). ➤ Non-white ethnic groups. ➤ First languages other than English. ➤ Unemployed. ➤ Single person households. ➤ Households with >6 people (the Census form has 6). ➤ Very elderly – help with the Census form. ➤ Disability – blind & partially sighted.
<p>Addresses: (individual, and / or aggregated area statistics): <i>for both Population Estimates & Census purposes:</i></p> <ul style="list-style-type: none"> • New (newly built, and also conversions, e.g. from old office or warehouse). • Empty. • Second residences (occupied for part of the week or year). • Multi-occupied: sub-divided into flats or bed-sits. • Rented. • Difficult to contact: gates and answer phones. • Communal establishments (public institutions such as hospitals, student halls, and prisons, but also some private, such as hotels, and care homes).
<p>Census fieldwork:</p> <ul style="list-style-type: none"> • Recruitment difficulties in local areas. • Unsafe areas. • Time best for making contact, e.g. after 9 PM?. • Internet take-up in local areas – % of customers who order on the web. Is the Internet better for contacting young males? • Telephone numbers for addresses – potentially expensive to obtain.
<p>Census processing:</p> <ul style="list-style-type: none"> • Data from commercial companies on Income – model onto the Census database?

5 2011 Census: Producing the Numbers. Meeting held at the RSS. 1 June 2009

2.6 ONS's statistical work programme: additional possibilities

ONS's Work Programme also identifies other priority topics. These include Measuring Societal Well-Being, which might potentially benefit from commercial data sources, for example:

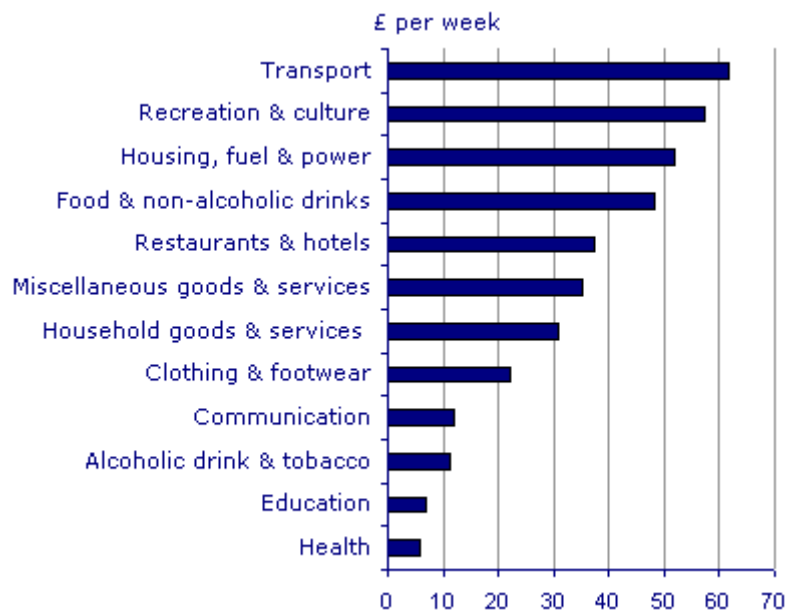
- Measurement of fuel poverty
- Profiles of customers – who buys what, e.g. “Bad” (junk food, cigarettes, alcohol), “Good” (fresh fruit & vegetables, fish, etc.). There might be scope for analysing shopping baskets – which products are bought at the same time? Also changes over time – reduced organics as the recession bites? This could add value to the Expenditure and Food Survey, especially now that it is coded with ONS's OAC geodemographic classification.

3. The UK population as customers

3.1 Where does the spending go?

Before going in pursuit of commercial companies and seeking to find out what information they gather about their customers, it is useful to consider the context of how households spend their incomes. Family Spending 2008⁶, using the 2007 Expenditure and Food Survey, shows spending across the broad categories:

Figure 3.1. Average weekly household expenditure (£459 in 2007)



These categories are broken down into much more detailed categories in Table A1 of Family Spending, and this provides a useful checklist when considering which goods and services are purchased from commercial companies, and also prompts thoughts of which suppliers are household names.

3.2 How is the money spent?

The ways in which customers spend their money varies considerably according to the goods and services involved. Spending may be regular (e.g. groceries, mortgage payments) or erratic (e.g. hotels). Some purchases remain anonymous (e.g. ad hoc snacks), whilst others can be related to an individual who has filled in a simple loyalty card or detailed financial services application form. This is explored in detail in the next Section.

6 http://www.statistics.gov.uk/downloads/theme_social/family_spending_2007/familyspending2008_web.pdf

4. Interviewing commercial companies – method

4.1 The objective

The aim is to provide an overview of relevant information that is collected by commercial companies across major industries – retail, financial services, leisure, energy, water and telecoms. Such companies each have millions of customers, and record details of these and their transactions to varying degrees.

4.2 The approach used

A target list of industries was agreed with ONS, with the aim of getting two or more of the major companies from each. This would help in drawing generalisations, and also in providing information that would be useful to ONS without revealing sensitive facts about individual companies which are not already in the public domain.

4.3 Selecting the companies

Those companies which belong to the Demographics User Group⁷ provided a good starting point, being household names, and whose staff wish to help improve the ONS's population statistics. Interviews were sought with several of these, and with some additional companies, in order to represent each of the main sectors.

4.4 The interviews

Using a qualitative approach, each interviewee was emailed a topic guide specific to their industry (see Annex B), which was used during face-to-face interviews. The questioning varied between sectors, but can be summarised under the following broad headings:

- The company's products and services; the sales channels used; whether customer details are recorded, and a single view of all each customer's activity is available; numbers active; coverage across the UK?
- Person / Household information collected; validation / accuracy?
- Address information held; addressing complications?
- Customer's purchases – transaction details?
- Broader issues: Any other useful information? Own customer analyses / segmentations? Any sharing / pooling of records with other organisations? What might ONS offer in return for help?

⁷ <http://www.demographicsusergroup.co.uk/>

5. The information collected by commercial companies – summaries

5.1 The UK context

When considering the information that is collected by commercial companies about their customers it is useful to keep in mind some headline statistics about the UK as a whole: in 2007 the total number of residents was c. 61 million (of whom 50 million were adults), and these comprised c. 25 million households (Social Trends 2009)⁸.

Therefore, even a very large company such as Tesco, which holds information about 15 million active customers, is still only covering a minority of the population, and provides a partial view of the nation as a whole. This may nonetheless have value. Also, in some industries the major companies pool information to provide a picture of the entire market (and each company's share of it), and this can be very useful.

5.2 Creating the summary profiles for major sectors

The following summaries are based on not only the interviews, but also information that is already in the public domain. In particular, application forms for loyalty cards, current accounts, and mobile phone contracts, for example, make clear the detail of the information that companies collect. Again, as customers, we receive details of our own transactions, such as credit card statements. At the aggregate level, companies also often give information on their websites and in annual reports about their numbers of customers, locations of branches, etc.

The summaries aim to provide a good understanding of the information that is collected and held by companies in different sectors, without revealing any information that is not already public, or which the interviewees wished to remain confidential.

The right-hand column of each summary provides commentary for ONS, including how the information held by companies relates to the topics of interest identified as priorities (see Table 2.5).

8 http://www.statistics.gov.uk/downloads/theme_social/Social_Trends39/Social_Trends_39.pdf

5.3 Retailers – summary profile

Sources for this profile include information from interviews with four of the largest retailers in Britain.

Industry summary	Comments for ONS
<p>The companies' products and services:</p> <ul style="list-style-type: none"> • Sales channels and products? Customers visit shops in huge numbers to buy an enormous range of products, but also order online (growing), and sometimes by catalogue, by telephone, & by mail order. • Customer details? The extent to which these are captured varies considerably. If the retailer has a loyalty card which is easily filled in and then gives significant benefits to the customer, a retailer can build a database of >10 million active records. If, however, the retailer offers a credit card, the applicant provides more information, but the numbers taking it up may be smaller, e.g. < 5m. Online customers also obviously have to provide their name and address too, for home delivery. • Single view of all each customer's activity? This is relatively simple (c.f. Financial Services) – knowledge of named customers is usually based on one main card used in the shops, plus some online orders. • Numbers of customers on database? Retailers with loyalty cards may have 10-15 million active customers. There may also be many dormant or defunct. • Coverage across the UK? The big retailers usually have UK coverage with few local gaps, although there are some regional biases. 	<p>Note:</p> <ul style="list-style-type: none"> • The contrast between Loyalty Cards (simple application, used by large numbers of customers) and Company Credit Cards (more detail required on application, and smaller numbers of customers). • Customers who simply pay with a bank credit card may be less easy to track, and those using cash remain anonymous. • Note that major retailers may appeal to different markets, e.g. affluent, older, or female.
<p>Person / Household information collected:</p> <ul style="list-style-type: none"> • Application forms? These are in the public domain, e.g. Tesco Clubcard, Boots Vantage Card, Nectar, and M&S Credit Card. The first three loyalty cards typically capture first name, surname, address and postcode, sex, date of birth, telephone number, email address, and sometimes some very basic information about the household at the time of application. Credit card applications are more detailed, including employment and income. • Updating? Not much after the initial application, except for some changes of address. • Validation / accuracy? Information given for loyalty cards is usually accepted at face value. Information for credit cards is subjected to credit checking. 	<p>People / households? (ONS interests)</p> <p>Loyalty card applications typically collect items that may directly (or indirectly, through name analysis) help to identify:</p> <ul style="list-style-type: none"> • New customers (inc. those with foreign names). • Moves: previous & new addresses. • Young males / females. • Very elderly. • Overseas names. • Single person households. • Households with >6 people. <p>Disadvantages:</p> <ul style="list-style-type: none"> • Not truly representative samples / subsets (biases by region, status, etc., and also changes in markets and marketing). • Not always validated. • Limited updating. <p>Advantages:</p> <ul style="list-style-type: none"> • Flows: timeliness, esp. large volumes

Industry summary	Comments for ONS
	<p>of recent applicants (hundreds of thousands p.a.), which could give pointers to recent in-migrants with foreign names.</p> <ul style="list-style-type: none"> • The stock: huge volumes / coverage.
<p>Address information held:</p> <ul style="list-style-type: none"> • Postcodes & PAF? The great majority of addresses are postcoded and usually checked against PAF, to ensure that Royal Mail can deliver statements and offers. • Single view of the customer's address? Sometimes separate loyalty card & online databases. 	<p>Addresses? (ONS interests)</p> <ul style="list-style-type: none"> • Multi-occupied? Some potential pointers (but not as good as utility sources). • New? Yes, if heavy recent clustering of new customers.
<p>Customers' purchases & transactions:</p> <ul style="list-style-type: none"> • Customers' purchases are fully recorded – product detail, price, place, time, payment method, etc. These amount to huge volumes of data every day. 	<p>Enormous volumes of data on purchases Up to the minute.</p> <p>Not immediately relevant to population estimation or Census planning, but a potential resource for analysing ONS's priority of Measuring Societal Well-Being".</p>
<p>Broader issues:</p> <ul style="list-style-type: none"> • Any sharing / pooling of records with other organisations? For some loyalty cards aggregate statistics are compiled and sold back to the suppliers of Fast Moving Consumer Goods (FMCG). • What might ONS offer in return for help? Improved Census coverage. More frequent Censuses. More accurate statistics, e.g. household composition. Make all ONS datasets free to use – including the EFS. <p>More widely, big companies have contact with various parts of central government: "We want to be seen to be on the side of the angels".</p> <p>Practicalities:</p> <p>"Anonymised data would be OK – but huge volumes; perhaps get government analyst to work within our office, rather than transfer files".</p> <p>"Also, if special analyses are carried out, we would like to see them in advance of publication".</p>	<p>Sharing / Pooling? Yes, some precedents already, but not as many as in Financial Services.</p> <p>What might ONS offer? A mixed wishlist! ONS might be able to make some tactical offers of help, but the strategic issue of corporate / government relations may prove to be more significant.</p> <p>Practicalities?</p> <p>Note the two interesting observations.</p> <p>It would be necessary to move quickly if information is to be sought for Census purposes, but some companies may be able to respond swiftly.</p>

5.4 Financial Services – summary profile

Sources for this profile include information from interviews with three of the largest providers of financial services in the country.

Industry summary	Comments for ONS
<p>The companies' products and services:</p> <ul style="list-style-type: none"> • Products? A wide range, often including current account (with debit card), savings, mortgages, insurance, credit card, loans, investments. • Sales channels? Branch, ATM, post, phone, internet, Independent Financial Advisers. • Single view of all each customer's activity? Best envisaged as people filling in application forms to obtain various financial products. These records are converted into a customer file (Customer 1 has Products A, B & C). • Numbers of customers on database? The big companies typically have >10m each. These vary in activity, e.g. active current accounts c.f. dormant ISAs. • Coverage across the UK? Financial services companies usually have UK coverage of branches & ATMs with few local gaps, although there are some regional biases. 	<p>Note the diversity of financial products, with the varying detail needed on their application forms, and also their subsequent frequency of transaction.</p>
<p>Person / Household information collected:</p> <ul style="list-style-type: none"> • Application forms? Some application forms need lots of detail, others little. Most application forms capture first name, surname, address and postcode, sex, date of birth, telephone number, and email address. Some also gather personal income, employment, & Census-type information. • Updating is patchy. • Validation / accuracy? Yes: internal & external validation of name & address for identity for fraud & money laundering. Also use of credit checking agencies. 	<p>People / households? (ONS interests)</p> <p>Financial Services databases may help to identify:</p> <ul style="list-style-type: none"> • New customers (inc. those with foreign names). • Moves: previous & new address. • Young males / females. • Students. • Very elderly. • Overseas origin (using names). <p>Disadvantages:</p> <ul style="list-style-type: none"> • Not truly representative samples / subsets (biases by region, status, etc.). • Limited updating. <p>Advantages:</p> <ul style="list-style-type: none"> • Validation of data at time of application. • Flows: timeliness, esp. large volumes of recent applicants (hundreds of thousands p.a.), which could give pointers to recent in-migrants with foreign names. • The stock: huge volumes / coverage.
<p>Address information held:</p> <ul style="list-style-type: none"> • Postcodes & PAF? The great majority of addresses are postcoded: the aim is to have a reliable postal address. • Single view of the customer's address? This is the aim – but created by matching records, so not perfect. 	<p>Addresses? (ONS interests)</p> <p>The following may be of interest:</p> <ul style="list-style-type: none"> • New properties. If heavy recent clustering of new customers (esp. mortgages). • Buy-to-let properties (where tenants

Industry summary	Comments for ONS
<ul style="list-style-type: none"> Additional addresses? Yes, e.g. correspondence address, buy to let property, security (property) address, etc. Overseas addresses? If short term, unaware. 	<p>may be difficult to capture in the Census).</p>
<p>Customers' transactions:</p> <ul style="list-style-type: none"> Customers' transactions are fully recorded – product, amount, channel, place, date, time, etc. Some products (credit cards, current account) give lots of transaction data in great detail. Others, e.g. mortgages & savings, give less. 	<p>Enormous volumes of transactions in great detail, some of which are frequent & recent.</p> <p>Not immediately relevant to population estimation or Census planning, but a potential resource for analysing some of ONS's other priorities (see "Broader Issues").</p>
<p>Broader issues:</p> <ul style="list-style-type: none"> Any other useful information? New housing, from mortgage applications (but should be on PAF / LA lists). <p>Fraud: Audit Commission's National Fraud Initiative, and also DWP's benefit fraud system.</p> <ul style="list-style-type: none"> Own customer analyses / segmentations? Lots of segmentation of customer behaviour, fraud detection, etc. Can deduce huge amount from transaction information, building a picture of a person's lifestyle. Any sharing / pooling of records with other organisations? There are several initiatives: <p>Pooled databases – mortgages, savings, etc (CACI, CML, & FSA).</p> <p>Pooling for Fraud prevention and Credit checking * Fraud – CIFAS. * Credit checking / Debt databases run as closed user groups.</p> <p>What might ONS offer in return for help? Best possible Census (which gives full coverage of the whole market, + breadth of questions). Better Census. Better information on Income. Easier access to other government files, e.g. bulk downloads of IDBR (ONS), Land Registry property values, Environment Agency flood risk. Crime stats. Better information for modelling / predicting fraud.</p> <p>Wariness about releasing any information (press stories, etc.), but not impossible – importance of corporate responsibility; would be decided at a very high level.</p>	<ul style="list-style-type: none"> New addresses – mortgages as a possible further source / check. Analysis of Fraud – local hotspots (type of area & people). Possible indicator of likely non-compliance with the Census. Analysis of Debt – local hotspots (type of area & people). Ditto. <p>Note that:</p> <ol style="list-style-type: none"> 1) Companies already regularly pool / share individual records for their mutual benefit – so there are precedents. 2) These are often closed user groups. 3) Of the records already pooled, the topics of potential interest include: <ul style="list-style-type: none"> • Fraud – local clusters. • Bad debt – local clusters. • Mortgages for new properties. <p>What might ONS offer? Another mixed wish list, but with the emphasis on better data and access to it (both ONS & other parts of government), plus further mention of corporate / government relations.</p>

5.5 Leisure – summary profile

This profile is based on an interview with Whitbread, a major restaurant & hotel operator (and which does not mind being identified).

Industry summary	Comments for ONS
<p>The company's products and services:</p> <p>Whitbread's Hotel & Restaurant Division operates across the UK:</p> <ul style="list-style-type: none"> • Restaurants: Beefeater (130); Brewers Fayre (130); Table Table (100); Taybarns (7). Customers pay by cash or plastic. A minority show their Nectar loyalty card, and are hence identifiable. (More information about the Nectar loyalty card is included in the Retail summary). • Costa Coffee (975). The great majority of transactions are by cash rather than card (so the person is not recorded). Doesn't take Nectar. • Premier Inn Hotels (577): Visitors register, so data could be good, but bookings can be by companies / individuals / 3rd parties – so, not a simple source of home addresses. Doesn't take Nectar. 	<p>The interview with Whitbread confirms that the Leisure industry, whilst having millions of customers, collects much less customer information than some other sectors such as financial services, and is therefore not a likely source for ONS.</p> <p>One possible exception is ONS's interest in communal establishments, and in particular, hotels, but lists of these are probably better obtained from trade directories, rather than each operator.</p> <p>The summary highlights some aspects which will be familiar to ourselves as customers.</p>

5.6 Electricity & Gas – summary profile

This profile is based on an interview with one large company, plus information which is in the public domain.

Industry summary	Comments for ONS
<p>The companies' products and services:</p> <ul style="list-style-type: none"> • Products? Only 80% of properties in the UK have gas; some properties within gas supply areas are also omitted – so it's best to use electricity information. • Payment channels? A wide range. • Single view of each customer's activity? Some, but not all companies. • Numbers of customers on database? Large companies have millions. • Coverage across the UK? Very patchy (following the privatisation of the old Regional Boards). 	<p>Note:</p> <p>All homes have electricity, but some don't have gas.</p> <p>Company's customers are often concentrated in particular regions.</p>
<p>Person / Household information collected:</p> <ul style="list-style-type: none"> • Application forms? Very simple – just name, with little or no demographic information. Obligated to accept. • Updating? Only as customers appear / disappear. • Validation / accuracy? Some use of credit checking – poor risk people are put on prepayment meters. 	<p>People / households? (ONS interests)</p> <p>Little information is collected.</p>
<p>Address information held:</p> <ul style="list-style-type: none"> • Postcodes & PAF? Postcoding – high 90's; not always corrected to PAF. • Extra address detail? Yes, sometimes by customer, & also meter locations. • Single view of the customer's address? Usually. • Additional addresses? Yes, e.g. correspondence address. • Contacting customers? Info on whether gained access in last 2 years – obliged to read the meter every 14 months. Modern meters easy to read; but c. 70% indoors. Usually read once or twice a year. Only read for current supplier – no past history (But TRANSCO make aggregate estimates). 	<p>Addresses? (ONS interests)</p> <ul style="list-style-type: none"> • Multi-occupied: sub-divided into flats or bed-sits – each with a meter (NB). • New – new connection of supply (NB). • Empty /void (NB).
<p>Customers' transactions (billing and payments):</p> <ul style="list-style-type: none"> • Customers' bills. Genuine seasonal variation, plus erratic estimation. • Fraud (by disguising identity, or bypassing the meter) – a lot of revenue is lost! • Bad Debt – also significant. 	<ul style="list-style-type: none"> • Fraud – particular properties or postcodes. • Debt / Non-payment as an indicator of "difficult" people and properties.
<p>Broader issues:</p> <ul style="list-style-type: none"> • Any other useful information? Meter readers have a crib sheet (location, etc) but patchy. Builder gets the company to install a meter, but is not billed until the customer moves in. Lists of new properties can be purchased from the big builders. • Own customer analyses / segmentations? Consumption varies by neighbourhood type – use of geodemographics. 	<p>Other useful information</p> <ul style="list-style-type: none"> • New addresses – which may be useful, especially if ONS could estimate household size from other sources.

Industry summary	Comments for ONS
<ul style="list-style-type: none"> • Any sharing / pooling of records with other organisations? Electricity: the industry has a list of addresses with meters (& addresses with multiple meters), “Gemserv”. • The Department of Energy and Climate Change publishes statistics on Domestic Energy Consumption (see NeSS website). These are based on DECC’s agreement with all electricity (& gas) suppliers to provide data for each individual meter throughout Great Britain. Metadata is available on the NeSS website. • What might ONS offer in return for help? <ul style="list-style-type: none"> ➤ Better Census. ➤ Interested in DWP list of people on benefits (data sharing). ➤ Interested in planning permissions. ➤ If Gemserv is useful, might just be a simple file dump – they wouldn’t know what to charge! ➤ Sharing addresses – OK. ➤ Sharing personal data – need DPA advice. ➤ Some companies are interested in working with government. ➤ Scope for regulator to add data provision to Licensing Conditions. <p>Postscript</p> <p>Turning to the distribution (c.f. retailing) of electricity, the National Grid has Distribution Network Operator companies, which are regional monopolies (NB). These connect customers to the network, so may be a source of addresses which offers complete coverage for a given area. It has not been established how this relates to Gemserv (see above).</p>	<p>Sharing / Pooling</p> <ul style="list-style-type: none"> • National (NB) list of addresses, inc. multiple meters – Gemserv, and also the DECC’s data collection. <p>What might ONS offer?</p> <p>See the various ideas.</p> <p>Postscript:</p> <p>If Gemserv does not provide a comprehensive address file, the regional Distribution Network Operator companies may be an alternative.</p>

5.7 Water – summary profile

Sources for this profile include information from interviews with two companies.

Industry summary	Comments for ONS
<p>The companies' products and services:</p> <ul style="list-style-type: none"> • Supply. Each water company has its own territory for water supply, and aims to have every property and sub-property (NB), domestic and business, on its database. An increasing % of domestic properties are metered, but many are still billed on the basis of rateable value. Meters are now fitted whenever a property is sold / bought. All businesses are metered. • Properties / addresses c.f. customers. The focus is on compiling a definitive list of properties, rather than customers. 	<p>Note that, unlike all the other examples, water companies:</p> <ul style="list-style-type: none"> • Are local monopolies, with 100% coverage of the market. There are 25 water companies in England and Wales. Thames Water covers c.75% of the population of Greater London, an area of great interest to ONS. • Often discuss industry issues and share knowledge with other water companies. • Focus on properties (& who should pay the bill), rather than details about customers.
<p>Address information held:</p> <ul style="list-style-type: none"> • Postcodes & PAF? > 99% postcoded & validated against PAF. Each company builds and maintains its own list of properties. Valuation Lists created by the VOA provide the basis (i.e. not PAF). Liaison with larger builders to capture new addresses. Addresses are checked, & sometimes subdivided, when meters are fitted; new information from Land Registry when properties are sold & bought (but no such record of change of tenant). Use of a large-scale GIS to map properties. • Extra address detail, e.g. multi-occupied? Yes, sometimes added. Some properties are split into separate flats / addresses. It's in the water companies' interests to get separate rateable values for each subdivision – increased revenue, e.g. 4 meters. Other big old multi-occupied houses can remain listed as a single property, & usually billed on their rateable value. However, local clusters of multi-let properties are apparent when mapped. • Each property has: <ul style="list-style-type: none"> • A Unique Property Reference Number. • Unmetered / metered flag. • Name of person to be billed (if possible). • Adding new & deleting old addresses? New – yes, the water company makes the connection (NB). "Our database on new homes is as good as anyone's". Demolitions – recorded, & remain on the database. Also record void (empty) addresses – e.g. c.3 or 4% in one particular area at any one time. • Additional customer addresses, e.g. landlord, or 2nd home owner? Aim that each property should have a person named for billing. A landlord may receive bills for several properties. 	<p>Addresses? (ONS interests)</p> <p>Water companies aim to have high quality information about addresses / properties (which may be of value to both tracking population and migration, and to the Census):</p> <ul style="list-style-type: none"> • Multi-occupied: sub-divided into flats or bed-sits. • Rented. • New – new connection of supply (NB). • Empty /void (NB).
<p>Person / Household information collected:</p> <ul style="list-style-type: none"> • Applying to become a customer? The occupier is 	<p>People / households? (ONS interests)</p> <p>Water companies collect very little</p>

Industry summary	Comments for ONS
<p>required to inform the water company, but the onus is on the water company to identify a named person for billing.</p> <ul style="list-style-type: none"> • Person / customer detail? Hope to get surname & first name if possible (no other demographic details). • Household detail? Some household information is collected when meters are fitted, but not good quality. • Date that customer was added to the database? (To spot recent arrivals). Yes (would need to look back to see when billing name changed). • Validation / accuracy? The focus is on getting someone at the property to pay the bill, rather than checking the person. Rather than fraud, the main problem is non-payment (esp. by private tenants). 	<p>information about people or households.</p>
<p>Customers' transactions (aka consumption, billing & payments):</p> <ul style="list-style-type: none"> • Billing. Billed according to rateable value, or amount used (metered). Bills sent twice a year. Consumption varies by season & neighbourhood type. • Payment. Encourage Direct Debit – easy, cheap. Also cheques by post, standing orders, internet banking, payment cards. • Non-payment? There's no compulsion to pay – can't cut them off (legislation in 2000). Citizen's Advice Bureaux advise people in debt to give water bills low priority. The landlord is not obliged to tell the company the name of the tenant, nor obliged to pay. Debt is concentrated in rented accommodation – esp. short lets & students; "this is common to all water companies". • Payment – non standard? Some customers (e.g. landlords, 2nd home owners) pay the bills for more than 1 address. ALSO some areas may be part of a "Schedule", where a third party does all the billing (e.g. a council may bill thousands of tenants). 	<ul style="list-style-type: none"> • Non-payment as an indicator of "difficult" people and properties: high turnover / migration, young, students, single, and also multi-occupied, rented, and difficult to contact. • However, note that some of these patterns will be masked by non-standard payment mechanisms.
<p>Broader issues:</p> <ul style="list-style-type: none"> • Any other useful information? (All included above). • Own customer analyses / segmentations? Analysis of consumption using geodeographics. • Any sharing / pooling of records with other organisations? Some sharing with electricity suppliers, but problems of differing geographical areas. Need to be aware of DPA requirements. • What might ONS offer in return for help? Better population stats! Also, mention of company corporate responsibility. 	

5.8 Telecoms – summary profile

Sources for this profile include information from interviews with two of the big four mobile telecoms suppliers.

Industry summary	Comments for ONS
<p>The companies' products and services:</p> <ul style="list-style-type: none"> • Products? The mobile telephone & broadband market has 4 main players, each with >15m customers. <p>Mobile phone customers split into 2 broad types:</p> <p>Post Pay (monthly contracts) – people signing up give their name & address, & are subject to a credit search.</p> <p>Pre Pay (no contract). They either buy a SIM card, & use a phone they have got from elsewhere, or buy a SIM + phone. They then top up, e.g. at an ATM. Names & addresses are not known unless the customer decides to register. Pre Pay customers outnumber Post Pay customers.</p> <p>Telecoms companies also offer broadband (mobile, and also sometimes fixed). These are more likely to be on contract, rather than Pre Pay.</p> <ul style="list-style-type: none"> • Payment channels? A wide range. • Single view of each customer? Close to it. • Coverage across the UK? Yes, and also overseas operations. 	<p>Note:</p> <ul style="list-style-type: none"> • The increasing ubiquity of mobile phones – owned by 78% of households in 2007⁹. • c.50% of these are used by customers with a recorded name & address. • Little else is known about customers' personal or household demographics. • However, data collected on transactions / use is instant, and immensely detailed.
<p>Person / Household information collected:</p> <ul style="list-style-type: none"> • Application / purchase? See above – Contract customers provide their name and address, & sometimes DoB and sex, (but not usually any other personal or household information). Some Pre Pay customers register. • Updating? Only address, for billing. • Validation / accuracy? Yes, Contact customers are ID & credit checked. 	<p>People / households? (ONS interests)</p> <p>Telecoms application forms & registrations harvest names & addresses for c.50% of customers, but often little more detail. They may, however, perhaps capture a higher % of younger people than Retail or Financial Services, and therefore help to identify:</p> <ul style="list-style-type: none"> • New customers (inc. those with foreign names, who might be in-migrants). • Overseas origin (using names). • Young males / females. • Students. <p>Disadvantages:</p> <ul style="list-style-type: none"> • Not truly representative samples / subsets. <p>Advantages:</p> <ul style="list-style-type: none"> • Validation of data at time of application. • Flows: timeliness, esp. large volumes of recent applicants (hundreds of

Industry summary	Comments for ONS
	<p>thousands p.a.), which could give pointers to recent in-migrants with foreign names.</p> <ul style="list-style-type: none"> • The stock: huge volumes / coverage.
<p>Address information held:</p> <ul style="list-style-type: none"> • Postcodes & PAF? Yes, for the c. 50% of customers with name & address, usually cleaned to PAF. • Extra address detail? No. • Single view of the customer's address? Yes. • Additional addresses? No. <p>• ALSO, could infer the address for the other customers by triangulating position to < 100 metres, e.g. if used on successive evenings, presumed home. (But still wouldn't have the name).</p>	<p>Addresses? (ONS interests)</p> <p>Telecoms companies seek only basic address information, primarily for credit checking, but also for billing (although only a minority pay by post) and promotions, etc.</p> <p>None of ONS's "Address" topics of interest are therefore likely to benefit.</p>
<p>Customers' transactions:</p> <ul style="list-style-type: none"> • Customers' transactions are recorded in full detail. Date, time, voice or text, duration, whether to another mobile, etc. • ALSO, location of each call, inc. those from home address, and also overseas: <p>OFCOM requirement to hold ALL call detail, in case needed by the police.</p> <p>Bad debtors are simply cut off (not like water or fuel!)</p>	<p>Although demographic & address information is limited, Telecoms companies gather immense volumes of detailed call information, and this includes geographical position (c.f. home address) – see Broader Issues below.</p>
<p>Broader issues:</p> <ul style="list-style-type: none"> • Any other useful information? Lots of segmentation of usage. <ul style="list-style-type: none"> • Any sharing / pooling of records with other organisations? No, not within the industry. • However, the BT OSIS file does collate all changes to UK telephone numbers. <ul style="list-style-type: none"> • What might ONS offer in return for help? NOT more statistics – "we have enough of our own!" Any government land use data – natural features, buildings (for the engineers when planning coverage). However, colleagues in other parts of the business may well have wider interests in relationships with government. 	<p>The transaction / usage data may help create new information, e.g.:</p> <ul style="list-style-type: none"> • Recent in-migrants & out-migrants. • Local geographical concentrations of calls to particular overseas countries., • People using more than one residence. • Young males. Identified by application form – or perhaps by pattern of behaviour / use? <p>There's no established culture of pooling data within the industry.</p> <p>What might ONS offer? Companies emphasise good relationships with government, rather than seeking more data from ONS.</p>

6. ONS's interests, and potential sources of useful information

6.1 A preliminary assessment

Bringing together ONS's topics of interest (from Section 2), and the information collected by companies (from Section 5), the following table identifies those sources which it might initially be considered for further investigation.

Key	
	Might offer some value
	Doubtful value
	Not collected, or no value

Table 6.1

Topic	Comments	Possible sources	
People / Households			
Recent Migrant Flows. In-migrants and out-migrants, + current and previous addresses. Ideally by sex and age.	Large volumes of new customers: no previous address, but often record DoB and sex; scope for inferring country of origin from names. Some recent recording of ethnicity & religion. Scope for analysing mobile phone usage (longitudinal; overseas destinations; young males). Also, for existing customers, their new and old address.	Financial services, retailers, telecoms	
Population stock. Young males + very elderly (& ideally the whole customer population by age & sex)	Often 10m plus customers - DoB, sex, name. Will be biased according to the particular market, but might be useful context when analysing recent migrant flows.	Financial services, retailers, telecoms	
Students	Sometimes specifically targeted by financial services firms	Financial services	
People using more than one residence	Not readily available, but some possible scope	Telecoms - a long shot	
Non-white ethnic groups	Some recent recording of ethnicity & religion	Financial services	
First language not English	(Not recorded)		
Unemployed	(Not recorded)		
Single person households	Even if recorded, not usually updated		
Households with >6 people	Unlikely to be reliably recorded or updated		
Disabled	(Not recorded)		

Topic	Comments	Possible sources	
Addresses			
New (Newbuild or conversion)	Some pointers: clusters of new customers, clusters of new mortgages. Also water connections, + electricity & gas connections.	Water; then electricity. Also financial services. Possibly retail too.	Green
Empty	Some recording by utilities	Water; then electricity.	Green
Second residences	Not readily available, but some possible scope	Telecoms - a long shot	Orange
Multi-occupied	Several possible pointers	Water & electricity have the best information	Green
Rented	(Implicit in multi-occupied); also buy-to let	Financial services	Green
Difficult to contact: gates, phones	Possible scope from retail delivery operations, and utility visits	Retail, utility	Orange
Also - "Difficult" / Unlikely to respond	Addresses where residents have carried out fraud, or have been in debt, may well be a good proxy for anticipating difficulties in Census coverage	Financial services, inc. pooled databases; electricity; water; also government databases	Green
Communal establishments	Possible scope from utility lists, but there are probably better trade lists (e.g. hotels)		Red
Census fieldwork			
Recruitment difficulties - local	No information uncovered		Red
Unsafe areas	Mentioned by retail delivery, & also water industry	Retail, water	Orange
Time best for making contact	No information uncovered		Red
Internet take-up in local areas	Scope for analysing internet customers' postcodes using geodemographics; also telecom broadband customers. Then focus on the areas with lowest usage	Retail, financial services; also telecoms	Green
Tel Nos (+ email addresses)	Collected by most, but variable quality, and use of individual details may be prevented by the DP Act		Red
Census Processing			
Data on Income (for modelling)	Accurate data collected by financial services	Financial services	Green

6.2 Narrowing the focus on topics and sectors – recommended priorities

Taking just those topics of interest for which there may be some data of value (flagged green above) the following table focuses on those sectors which might be approached for further investigation, both to explore their own data, and also any pooled national databases.

Table 6.2

Possible sources	Topic
Retail	Recent Migrant Flows. In-migrants and out-migrants, + current and previous addresses. Ideally by age & sex
	Population stock. Ideally by age & sex
	Internet take-up in local areas
Financial Services	Recent Migrant Flows. In-migrants and out-migrants, + current and previous addresses. Ideally by age & sex
	Population stock. Ideally by age & sex
	New addresses (Newbuild or conversion)
	"Difficult" addresses - debt or fraud
	Rented addresses
	Internet take-up in local areas
	Data on Income (for modelling)
Electricity	New addresses (Newbuild or conversion)
	Empty addresses
	Multi-occupied addresses
	"Difficult" addresses - debt or fraud
Water	New addresses (Newbuild or conversion)
	Empty addresses
	Multi-occupied addresses
	"Difficult" addresses - debt or fraud
Telecoms	Recent Migrant Flows. Scope for analysing mobile phone usage (longitudinal; overseas destinations; young males).
	Population stock.
	Internet take-up in local areas

Seeking to narrow the focus further, this personal view starts from the premise that both population estimates and migration, and the Census, are vital to ONS, and is based on the following factors:

- Population estimates and migration. The data sources currently used are not ideal. Future potentially comprehensive government administrative sources such as e-Borders, and data from the Department for Work & Pensions, offer great potential, but it may be wise to seek additional sources of populations and their movements for

triangulation, and, ideally, confirmation. Identifying new housing developments can also be locally important.

- Census. Much detailed information is already available for planning coverage, and anticipating low response, but additional sources might improve the accuracy and currency of the address register yet further, and also help to target precisely those areas and addresses which are least likely to respond.

Recommended priorities:

1) Recent migrant flows, and population stock

- Explore further with a major retailer, such as Tesco.

2) New, empty, and multi-occupied addresses

- If not done already by ONS, research the electricity industry's national meter address file (Gemserv), with the help of E.ON, and also investigate further the comprehensive address files compiled by selected regional water companies, with Thames Water's area as of greatest interest to ONS.

3) "Difficult" addresses, subject to bad debt or fraud

- Explore with financial services their pooled national files on debt and fraud, seeking the advice of Nationwide.

6.3 Encouraging companies to help

It was striking that, when asked what might ONS do to in return for their help, none of the companies immediately mentioned payment.

The following factors might encourage companies' cooperation and involvement:

- The prospect of better population estimates and Census counts, which are valuable for their business decisions.
- A major motive for some companies is to have good relations with government at a corporate level: as one of them put it, "We want to be seen to be on the side of the angels."
- Start with a short and simple wish list of the data sought.
- Minimise the work involved for the company – what do they do already?
- Emphasise that ONS is as concerned about data confidentiality and security as they are.

- Persons / households: rather than seeking identifiable records, it may be best to seek anonymised records, or simple counts aggregated at postcode level.
- Addresses: these should be less sensitive; ideally selected lists of properties; otherwise, aggregated to postcode level.
- Float the possibility of initially working in their offices, rather than transferring sensitive files to ONS.

6.4 Timescales, and potential next steps

Improving population estimates is already a very high priority for ONS, and Census planning is now also in top gear. If commercial sources are to be investigated, this needs to be done in the lead up to the Census in March 2011, and, if findings are positive, commercial data files also captured at the time of the Census for comparative purposes.

7. Conclusions and recommendations – and a postscript

7.1 Conclusions

In seeking to identify new information which may be useful to ONS, fourteen companies across the Retail, Financial Services, Leisure, Energy, Water and Telecoms sectors have explained in some detail the data that they collect about their millions of customers. Each sector has different characteristics in the ranges of products offered, sales channels used, and the detail collected about customers and their addresses. However, as one would expect, all collect comprehensive information about purchases and transactions (and often debt and fraud).

Such company customer databases usually have the disadvantages that they are not representative samples or subsets of the national population as a whole – there may be biases by region and status, and the impact of marketing campaigns, for example; also the updating of records can be patchy. However, on the plus side are the large stocks of customers (often > 10 million), and the large flows of new customers, and the timeliness of the data. If viewed as potential sources of insight – rather than precise statistics – about recent trends, these may offer real value. The local incidence of debt and fraud may also be a good social indicator. Again, the Energy and Water industries put great store in maintaining accurate and timely address files, and these too may be of value.

When the topic was raised of possibly making data available to ONS, it was striking that the possibility of payment was not mentioned. Rather, the emphasis was on seeking better population statistics, and, more generally, good corporate relations with government. However, issues of data security were recognised as very important, and any decision to release information might be need to be referred to senior management.

Section 6 of this report has brought together ONS's list of topics of interest for population estimates and migration, and the Census, and identified data from commercial companies which may be of value. This has been narrowed down to three proposed priorities, which are included in the recommendations below.

7.2 Recommendations

- 1) Recent migrant flows, and population stock.** Explore further with a major retailer, such as Tesco, the scope for compiling regular small area statistics.
- 2) New, empty, and multi-occupied addresses.** Check the content and availability of the electricity industry's national meter address file, with the help of E.ON; also investigate the comprehensive address files compiled by selected regional water companies, starting

with Thames Water as the area of greatest interest to ONS.

- 3) **“Difficult” addresses, subject to bad debt or fraud.** Explore financial services’ pooled national files on debt and fraud, seeking the advice of Nationwide.
- 4) **Assess the potential value of customer data for additional population topics.** ONS’s target topics are: internet take-up; rented addresses; income; and mobile phone usage (especially to track international movements).
- 5) **Assess the potential value of customer data for other topics.** Companies have very rich data on spending on particular products and services, and this could be a valuable supplement to the Expenditure and Food Survey.
- 6) **Consider encouraging more sharing and pooling of customer records.** This is already well developed in some sectors, especially Financial Services, and ONS should consider encouraging its extension (perhaps in alliance with other government departments) to new topics of interest, to produce more statistics “for the public good”.

7.3 A postscript

We, as members of the public, are leaving many electronic trails as we go about our daily lives. In addition to our numerous contacts with commercial companies illustrated in this report, we also appear on many other databases held by government, and our internet click streams also capture a record of our interests and behaviour. Whilst the prospect of compiling a detailed profile of every individual may be unnerving, the potential for deriving anonymous aggregate population statistics is enormous.

Annex A. ONS's Specification of requirement

1. Background and Introduction

The AUTHORITY requires the CONTRACTOR to provide services relating to the investigation and analysis of the information that is collected by commercial companies as possible sources of statistics, or at least valuable recent intelligence.

2. Objectives

The deliverables of the work carried out by the CONTRACTOR will be, but are not limited to the following;

2.1 The CONTRACTOR will produce a statement of ONS's priorities for creating new statistics. These will include but are not limited to the following;

- New topics
- Improvements to current subject detail
- Geographical detail
- Accuracy of data produced
- Currency
- Other areas to be agreed between the CONTRACTOR and the AUTHORITY.

2.2 The CONTRACTOR will provide an overview of relevant information that is collected by commercial companies in different sectors, some of which are as follows:

- Retail,
- Utilities,
- Telecoms,
- Financial services
- Other sectors to be agreed between the CONTRACTOR and the AUTHORITY.

The CONTRACTOR will focus on existing customer databases and transactions but will be expected to make a professional judgement on any other sources that should be included too and obtain confirmation from the AUTHORITY.

2.3 The CONTRACTOR will make observations about ONS's requirements from potential suppliers. The CONTRACTOR will do so through a consultation of the Demographic User Group (DUG) members seeking their observations and opinions. Due to their individual positions in the market, they will be able to provide an ideal perspective on a range of sectors. The current membership comprises Abbey, Argos, Barclays, Co-operative Group, E.ON, John Lewis, Marks & Spencer, M&S Money, Nationwide, O2, Sainsbury's, Tesco, The Children's Mutual, and Whitbread. This is list is not comprehensive, but provides a minimum indication of the vast range of sectors covered.

2.4 The CONTRACTOR will make conclusions using the information they have gathered and researched via the various sources, findings will be presented in a report. This will include:

- What information exists that it is of interest to ONS
- Which targets are worth pursuing
- Identify areas which are sometimes mentioned, but have no realistic prospect of success.
- How companies should be pursued. Such as who to approach and how, minimising the work involved for the company, and creating benefits for the company.

Annex B. Interview Topic Guides

Retail

Customer Databases
Scope: Focus on the main retail business (c.f. any financial services, etc)
Names & addresses of customers who have purchased in last 12 months? (i.e. only active customers who are identifiable)
Their method(s) of purchase? (Visit shop, order online, catalogue, telephone, mail order, other?)
Contacts held on one or several databases?
Use of a loyalty card? (Are its users a subset of the above?)
Do you have single view of all customers?
Approximate total number active in last 12 months?
Is that in England / Wales / GB / UK? Particular local areas?
Person / Household data collected
Name – Forename as well as surname?
DoB (or Age, & the year it was recorded?)
Sex?
Other personal demographics? Coverage? Are they updated?
Any household details? Coverage? Are they updated?
Date that customer was added to the database? (to spot recent arrivals)
Customer Address(es)
>99% postcoded?
As given by customer, or corrected to PAF?
Extra info, e.g. Flat C on 3 rd floor?
Single view of address? (Or alternative versions?)
Additional address, e.g. 2 nd home? (Or regarded as another customer?)
Delivery (contact) – any difficulties?
Purchases
Method & Place?
Date & Time?
Products?
Prices?
Comments on coverage & quality?
Any other useful data or information collected? Own analyses / segmentations?
Any sharing of records / summaries with other organisations?
If any sharing with ONS, what might they offer you in return?

Financial Services

Product / Customer Databases
Scope: Names & addresses of customers who have been active in last 12 months?
Their various products: current a/c, savings, mortgage, insurance, credit card, etc?
Their various channels: branch, ATM, post, phone, internet, IFAs, etc?
Customers held on one or several databases?
Do you have single view of all customers?
Approximate total number active in last 12 months?
Is that in England / Wales / GB / UK? Particular local areas?
Person / Household data collected
Forename & surname
DoB (or Age, & the year it was recorded?)
Sex?
Other personal demographics? Coverage? Are they updated?
Any household details? Coverage? Are they updated?
Date that customer was added to the database? (To spot recent arrivals)
Validation / accuracy? (E.g. to prevent fraud)
Customer Address(es)
>99% postcoded?
As given by customer, or corrected to PAF? (Presume postal address is vital?)
Extra info, e.g. Flat C on 3 rd floor?
Single view of address? (No alternative versions?)
Additional address, e.g. 2 nd home? Workplace?
Contact customer by post, telephone, email? (c.f. retailers' home delivery)
Products: Transactions, (& Statements as a record)
Channel & Place? (e.g. ATM at Victoria, or Branch at MK)
Date & Time?
Product & Amount?
Receipts & Debits. Sources identified?
Comments on coverage & quality?
Any other useful data or information collected? Own analyses / segmentations?
Any sharing of records/summaries with other organisations? (e.g. CACI, CIFAS)
If any sharing with ONS, what might they offer you in return?

Leisure

Customer Databases
Scope: Consumer / domestic users of various leisure brands (your list.....)
Names & addresses recorded of customers who have purchased in last 12 months? (i.e. only active customers who are identifiable)
Their method(s) of purchase? (Visit the restaurant / bar / sports centre, etc. Any other services, e.g. request home delivery?)
Contacts held on one or several databases?
Use of membership cards, or loyalty cards? (Are users a subset of the above?)
Do you have single view of all customers?
Approximate total number active (& identified) in last 12 months?
Is that in England / Wales / GB / UK? Particular local areas?
Person / Household data collected
Name – Forename as well as surname?
DoB (or Age, & the year it was recorded?)
Sex?
Other personal demographics? Coverage? Are they updated?
Any household details? Coverage? Are they updated?
Date that customer was added to the database? (to spot recent arrivals)
Customer Address(es)
>99% postcoded?
As given by customer, or corrected to PAF?
Extra info, e.g. Flat C on 3 rd floor?
Single view of address? (Or alternative versions?)
Additional address, e.g. 2 nd home? (Or regarded as another customer?)
Delivery (contact) – any difficulties?
Purchases
Place?
Date & Time?
Products / services?
Prices?
Comments on coverage & quality?
Any other useful data or information collected? Own analyses / segmentations?
Any sharing of records / summaries with other organisations?
If any sharing with ONS, what might they offer you in return?

Electricity & Gas

Customer / Product Databases
Scope: Focus on the main domestic retail business (c.f. business use, generation, any financial services, etc)
Names & addresses of customers who have purchased in last 12 months? (i.e. only active customers who are identifiable)
Their products: electricity, gas, other?
Their various payment channels: post, direct debit, standing order, internet banking, payment meter, etc?
Customers held on one, two, or several databases?
Do you have single view of all customers?
Approximate total number active in last 12 months? Electricity, Gas
Is that in England / Wales / GB / UK? Which local areas?
Person / Household data collected
Applying to become a customer – the process
Forename as well as Surname?
DoB (or Age, & the year it was recorded?)
Sex?
Other personal demographics? Coverage? Are they updated?
Any household details? Coverage? Are they updated?
Date that customer was added to the database? (To spot recent arrivals)
Validation / accuracy? (E.g. to prevent fraud)
Customer Address(es)
>99% postcoded?
As given by customer, or corrected to PAF? (Presume postal address is vital?)
Extra info, e.g. Flat C on 3 rd floor?
Single view of address? (Or alternative versions?)
Additional address, e.g. 2 nd home? Property owner's address, etc?
Contact customer by post, telephone, email? Also visit to read meter(s)?
Purchases: Billing & Payments
Amount used & cost. Estimated consumption? Erratic amounts by quarter?
Date / Time
Comments on coverage & quality?
Any other useful data or information collected? Own analyses / segmentations?
Any sharing of records/summaries with other organisations? (e.g. TRANSCO)
If any sharing with ONS, what might they offer you in return?

Water

Customer / Product Database
Scope: Focussing on domestic (c.f. business or institutional) addresses which have been supplied water in the last 12 months – and the people living at those addresses
Database of addresses, with customer names attached? (Rather than vice versa?)
Do some customers (e.g. landlords, 2 nd homes) pay the bills for more than 1 address?
Approximate total numbers of addresses / households / customers active in the Water Supply area? (Is a simple map of the area available?)
Person / Household data collected
Applying to become a customer – the process
Forename as well as Surname?
DoB (or Age, & the year it was recorded?)
Sex?
Other personal demographics? Coverage? Are they updated?
Any household details? Coverage? Are they updated?
Date that customer was added to the database? (To spot recent arrivals)
Validation / accuracy? (E.g. to prevent fraud)
Customer Address(es)
>99% postcoded?
As given by customer, or corrected to PAF?
Extra info, e.g. Flat C on 3 rd floor?
Multi-occupied / subdivided addresses?
Single view of address? (Or alternative versions, e.g. PAF + Own + NLPG?)
Adding new & deleting old addresses?
Additional customer address, e.g. Landlord's or 2 nd home owner's?
Contact customer by post, telephone, email? Also visit address to read meter(s)?
Purchases: Billing & Payments
Metered & non-metered consumption, & cost?
Quarterly billing? Seasonal variations in consumption?
Customers' various payment channels: post, direct debit, standing order, internet banking, payment meter, etc?
Addresses / customers who don't pay?
Comments on coverage & quality?
Any other useful data or information collected? Own analyses / segmentations?
Any sharing of records/summaries with other organisations?
If any sharing with ONS, what might they offer you in return?

Telecoms

Customer / Product Databases
Scope: Focus on domestic consumers (mobiles, broadband, other?), c.f. business Customers with a name, and address (NB?), active in the last 12 months
Products / purchases: monthly agreements; pay-as-you-go; sales of handsets; others?
Various payment channels: post & cheque, direct debit, standing order, internet banking, ATM top-ups, etc?
Customers held on one, two, or several databases?
Do you have single view of all customers?
Approximate total number active in last 12 months?
Is that in England / Wales / GB / UK? Particular local areas?
Person / Household data collected
Applying to become a customer – the process
Forename as well as Surname?
DoB (or Age, & the year it was recorded?)
Sex?
Other personal demographics? Coverage? Are they updated?
Any household details? Coverage? Are they updated?
Date that customer was added to the database? (To spot recent arrivals)
Validation / accuracy? (E.g. to prevent fraud)
Customer Address(es)
For those customers on the database, >99% postcoded?
As given by customer, or corrected to PAF?
Extra info, e.g. Flat C on 3 rd floor?
Single view of address? (Or alternative versions?)
Additional address, e.g. 2 nd home? Property owner's address, etc?
Contact customer by mobile (text, voice), email, & postal address (how important?)
Usage / Purchases: Billing & Payments
Itemised calls & their cost
Date / Time
Location where each call made? Scope for analysis?
Comments on coverage & quality?
Any other useful data or information collected? Own analyses / segmentations?
Any sharing of records/summaries with other organisations? E.g. OFCOM?
If any sharing with ONS, what might they offer you in return?

Annex C. Acknowledgements – interviewees, and additional contacts

Keith Dugmore, of Demographic Decisions, would like to express his thanks to all the people in commercial companies who agreed to be interviewed (some of whom wish to remain anonymous), and also the following people at ONS who have made time to provide information and to offer their views.

Person	Division / Branch
Roma Chappell	Centre for Demography
Steve Smallwood	Centre for Demography
Ben Humberstone	Centre for Demography
Richard Pereira	Centre for Demography
Frank Nolan	Methodology
Neil Hopper	Methodology
Louisa Blackwell	Methodology
Owen Abbott	Methodology
Garnett Compton	Census