



Department
for Business
Innovation & Skills

**MARKET ASSESSMENT OF
PUBLIC SECTOR INFORMATION**

Written by Deloitte

MAY 2013

A report by Deloitte for the Department for Business, Innovation and Skills

Deloitte

Stonecutter Court

1 Stonecutter Street

London

EC4A 4TR

Web: [UUwww.deloitte.co.uk](http://www.deloitte.co.uk)

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Business, Innovation and Skills.

Department for Business, Innovation and Skills

1 Victoria Street

London SW1H 0ET

www.gov.uk/bis

URN BIS/13/743

May 2013

Contents

Acknowledgements	9
Executive Summary	10
Key findings.....	10
Policy context.....	12
Defining public sector information and the market	13
Public sector information dimensions.....	13
Public sector information holders and customers	14
The value of public sector information today.....	16
The link between public sector information and growth	16
Generating value from public sector information.....	16
Value from different public sector information datasets	17
Understanding how public sector information creates value in the UK.....	19
Estimates of the value of public sector information and value to PSIHS	21
Barriers to realising the full potential of public sector information in the UK	24
A taxonomy of barriers	24
Legislative barriers	25
Economic barriers	27
Access barriers	32
Summary of key research questions answered	35
1. Introduction and approach.....	42
Project scope and objectives	42
Approach.....	43
Data sources	43

Structure of this report.....	44
2. Public sector information definitions and market definition	46
Definition of public sector information	46
Definition of the public sector	47
The nature of public sector information.....	48
The availability of public sector information	49
The complexity of public sector information	50
The speed at which public sector information is updated	50
Public sector information as a by-product or purpose-collected	50
The content of public sector information	51
The format of public sector information.....	52
The distribution channels of public sector information	54
The cost of public sector information	55
Market definition.....	55
Chapter summary.....	57
3. The supply and demand sides of the public sector information market	59
Supply side of the market.....	59
Supply of datasets nationally	59
Supply of datasets locally.....	64
Trading Funds	66
Attitudes and policies of public sector information holders	69
Demand side of the market.....	70
Public sector information use and re-use	71
Consumer perceptions of public sector information and unmet demand	75
Public sector information customer types	80

Different business models of public sector information intermediaries	86
Data scientists.....	91
Chapter summary.....	93
4. Public sector information regulatory landscape.....	95
Policy and regulatory stakeholders	95
The legal framework for publishing and sharing public sector information	96
Public sector information licensing and the regulatory framework	97
The Data Protection Act.....	101
The European dimension	103
The European Directive on the Re-Use of Public Sector Information	103
The proposed European Data Protection Regulation	103
The INSPIRE Directive.....	104
Chapter summary.....	105
5. The current value of the public sector information market	106
Understanding the value of public sector information	107
Modelling approach and assumptions.....	108
Stage 1: consumer and producer surplus approach – the welfare approach.....	108
Stage 2: indirect and induced value of public sector information	109
Narrow economic value estimates	110
Wider societal value estimates.....	110
‘Blue sky’ opportunities	116
Detection and Identification of Infectious Diseases (2006)	116
Tackling Obesity (2007)	116
Global Food and Farming (2011).....	117
Computer Based Financial Trading (2012)	117

Reducing the Risks of Future Disasters (2012)	117
The Future of Identity (2013).....	117
Chapter summary.....	118
6. Barriers and policy implications.....	120
Generating value.....	120
A taxonomy of barriers	121
Legislative barriers	122
Economic barriers	128
Access barriers	133
Chapter summary.....	140
7. Conclusion.....	141
Closing thoughts	141
Appendix 1: Glossary	143
Appendix 2: Acronyms used in this report	154
Appendix 3: Literature Review	156
Overview	156
The economic impact.....	157
The economic contribution of public sector information.....	158
The costs of providing public sector information	161
Providing value for money from public sector information	161
Summary.....	162
The broader impact.....	163
Maximising the impact of public sector information.....	164
Pricing models.....	164
The impact of zero or marginal cost pricing on innovation.....	166

Case studies	168
Public sector efficiencies.....	168
Business innovation	169
Country case studies.....	170
Appendix 4: Further statistics	174
Companies House.....	174
Land Registry	174
The Met Office.....	175
Ordnance Survey	179
Appendix 5: Empirical methodology.....	181
General approach and limitations	181
Definition of value.....	182
Methodologies considered	183
The quantification approach for the current value of public sector information.....	186
Stage 1: consumer and producer surplus approach – the welfare approach.....	187
Stage 2: indirect and induced value of public sector information	193
Appendix 6: Transport sector case study	194
Overview	194
Information released by Transport for London	194
Accessing information.....	195
Usage levels.....	196
Benefits	197
Quantification of benefits.....	198
TfL ‘lost customer hours’ data.....	198
Apps using TfL data	198

The model	200
Findings.....	200
Summary.....	202
Information relating to transport beyond London	204
Appendix 7: Further case studies	215
The healthcare and life science sector.....	215
The public sector.....	221
Opportunities in the local public sector	222
Appendix 8: Government Officials informal consultation.....	226
Consultation questions.....	226
Background	226
Consultation responses.....	228
Appendix 9: Bibliography	230
Academic Papers & Reports	230
Departmental Open Data Strategies	233
Other Documents	234

Acknowledgements

This report has been produced by Deloitte between October 2012 and March 2013. The lead authors are Andrew Tong, Haris Irshad and Daniel Revell Ward.

The Deloitte team acknowledges the support of Stephan Shakespeare and the Data Strategy Board, the Cabinet Office, the Open Data Institute, other public sector bodies and the wider open data community who have contributed their time, data and comments to this report. In particular, the team would like to express its thanks to the different start-up companies and established companies who have contributed to the case study analyses.

Executive Summary

Key findings

This is the first UK-wide market assessment of public sector information¹. It spans the use and re-use of public sector information at the UK-level, regionally and locally by a wide range of businesses, civil society groups, government and members of the general public. The aim of this market assessment is to establish a robust evidence base on its value and to highlight the policy implications flowing from an examination of how public sector information could be utilised further.

The research has covered three broad thematic areas:

- definitions of public sector information and its characteristics;
- how public sector information is used and re-used inside and outside of government; and
- barriers to fully exploiting the value of public sector information, including issues around competitiveness, funding and regulation.

The research has used a combination of literature reviews, analysis of secondary sources, stakeholder interviews and case studies. The key research findings are summarised below.

Key research findings

- This report has estimated that the **value of public sector information to consumers, businesses and the public sector** in 2011/12 was approximately **£1.8 billion** (2011 prices).
 - This is a mid-point estimate, with the sensitivity analyses giving a range between £1.2 billion and £2.2 billion
- However, the use and re-use of public sector information has **much larger downstream impacts affecting all areas of society beyond the direct customer**.
- The **UK is a global leader in releasing public sector information** for use by its citizens, businesses and policymakers. Through initiatives such as the data.gov.uk, the Open Data User Group, Departmental Data Strategies, as well as the establishment of the Data Strategy Board and Open Data Institute, **the UK has taken significant steps to creating a world-class public data infrastructure**.
 - While there is no central figure on the number of public sector information datasets currently being made available, a review of selected data portals suggests the number could exceed 37,500 from over 750 different publishers with over 2.5 million downloads by October 2012
- There is a **link between the provision and use/re-use of public sector information and economic growth**. Public sector information is used by businesses, individuals and the public sector to:
 - **stimulate innovation and develop new products and services;**
 - **hold public service providers to account, promote democratic engagement and foster greater transparency and better policymaking;**

¹ Public Sector Information covers the wide range of information that public sector bodies collect, produce, reproduce and disseminate in many areas of activity while accomplishing their public tasks.

The terms data, information and knowledge are frequently used for overlapping concepts. The main difference is in the level of abstraction being considered. Data is a broad term, embracing others, but is often the lowest level of abstraction, information is the next level and, finally, knowledge is the highest level. Unless otherwise stated, this report uses the term 'data' to include 'information' and vice versa. Accordingly, a 'dataset' in this context will be a collection of 'data' and/or 'information'.

- **reduce barriers to entry** into markets and **address information asymmetries**; and
- **generate network effects** that drive disruptive change by connecting increasing numbers of consumers and businesses
- The most popular, and potentially most valuable, datasets include **geo-spatial, environmental, transport, health** and **economic data**, with the construction, real estate, finance and insurance, public sector and arts, entertainment and recreation sectors being some of the largest users and re-users of public sector information and open data.
- Case studies examining the downstream impacts created by different organisations using and re-using public sector information have shown the monetary value of some of these benefits can be in the order of millions, if not billions. For example:
 - **publishing data on adult cardiac surgery** is estimated to have reduced mortality rates, which in turn has an **economic value in excess of £400 million p.a.**
 - **using live data from Transport for London in apps** can save users time to the **economic value of between £15 million and £58 million p.a.**
- While an aggregate figure on the social value of public sector information is difficult to reach without more information on the way data is used and how it then permeates society, on the basis of conservative assumptions, it is estimated this figure **could be in excess of £5 billion** for 2011/12 (2011 prices).
 - This estimate is likely to increase as public sector information is used more widely and in more impactful ways.
- Adding this social value estimate to the calculated value of public sector information to consumers, businesses and the public sector, gives an **aggregate estimate of between £6.2 billion and £7.2 billion** in 2011/12 (2011 prices). Future uses of public sector information that have the potential to generate much more value include **greater combining of public and private sector information**, exploiting the benefits of **linked data**, **embedding geospatial and location data across more and more products and services**, and **more informed policymaking based on better utilised public sector information**.
- Using a taxonomy originally developed by the Data Strategy Board, this report has considered a number of **barriers** that may be preventing the UK from fully realising the benefits of public sector information:
 - **Legislative barriers**: while current legislation, guidance and regulations are not hindering the public sector information market themselves, there is often a perception (right or wrong) that some legislation, guidance and regulations prevent datasets from being released and shared thereby reducing the availability of datasets. The Open Government Licence is widely seen as an effective means of improving the availability of public sector information;
 - **Economic barriers**: the issue of charging for public sector information datasets is complex, with a lack of data from both providers and users/re-users making it hard to conduct cost-benefit analyses on data release. While significant progress has been made in making more datasets available, there remains a perception (actual and perceived) that charging creates barriers to using and re-using public sector information. In the short-term there are a number of opportunities for improving the availability of certain datasets; and
 - **Access barriers**: improvements continue to be made to reduce fragmentation and improve the consistency of public sector information; however there remains scope for improvements in the accessibility of public sector information datasets to the general public. Equally, in some parts of the private and public sector, there is a skills shortage preventing the effective extraction of value from public sector information.

Policy context

As highlighted by Rt. Hon. Francis Maude MP in his foreword to the Open Data White Paper (June 2012), data is the twenty-first century's raw material: *"its value is in holding governments to account; in driving choice and improvements in public services; and in inspiring innovation and enterprise that spurs social and economic growth"*². Beginning with transparency statements originally set in the Coalition Agreement³, the UK has continued to push forward with the release of more public sector information, much of which meets the criteria of being 'open data'.⁴ Recent developments include (these are not exhaustive):

- the establishment of the **Data Strategy Board (DSB)**⁵ and the **Open Data User Group**⁶ (ODUG) to create the maximum value for businesses and people across the UK from data held by the **Public Data Group**⁷ (PDG) members and beyond;
- the establishment of the **Transparency Board**⁸ to drive forward the Government's transparency agenda, set open data standards across the whole public sector and facilitate the release of more public sector information;
- the creation of the **Open Data Institute (ODI)**, an independent, non-profit company funded by the Government and the private sector to incubate, nurture and mentor new ideas using open data (and the first organisation of its kind in the world);
- the publication of new **Open Standards Principles**⁹ for all government bodies to comply with making IT more open, cheaper and better connected;
- discussions at a European level on the revision of the **Public Sector Information Directive** and reforms to data protection rules;
- the publication of **departmental open data commitments** in their business plans;
- the on-going implementation of the Government's **Transparency and Open Data** agenda led by **Cabinet Office**. This includes the publication of updates on the **Government Digital Strategy**¹⁰ providing details on how services are being moved online to achieve efficiencies and how Cabinet Office is supporting improved digital capabilities across departments;
- the completion of a competition awarding investment to Glasgow to demonstrate how a city of the future could work by hosting the Technology Strategy Board's **Future City Demonstrator**¹¹; and
- specific policy announcements such as:

² Available at: www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential

³ Available at: www.direct.gov.uk/prod_consum_dg/groups/dg_digitalassets/@dg/@en/documents/digitalasset/dg_187876.pdf

⁴ Open government data is a subset of public sector information. It is data which can be used, re-used and re-distributed freely by anyone - subject only at most to the requirement to attribute and share-alike.

⁵ See www.gov.uk/data-strategy-board and the Terms of Reference at www.gov.uk/government/uploads/system/uploads/attachment_data/file/32384/12-673-terms-reference-data-strategy-board-and-public-data-group.pdf

⁶ See <http://data.gov.uk/odug>

⁷ The members of the PDG are Ordnance Survey, the Met Office, Land Registry and Companies House.

⁸ See www.gov.uk/government/policy-advisory-groups/134

⁹ See www.gov.uk/government/news/government-bodies-must-comply-with-open-standards-principles

¹⁰ See www.gov.uk/government/news/update-on-government-digital-strategy-aligning-it-with-digital

¹¹ See www.theodi.org/news/glasgow-wins-%C2%A324m-boost-after-recognising-%E2%80%9Crevolutionary-impact%E2%80%9D-open-data

- the **Open Data Immersion Programme**¹² to accelerate innovative open data ideas and develop market-ready businesses;
- the expansion of the Technology Strategy Board's **Innovation Voucher** programme available to SMEs to use open data to help commercialise their ideas and develop new products and prototypes; and
- other funding for a DSB **Breakthrough Fund** and upgrades to **Ordnance Survey OpenData products**¹³.

Together these developments are helping place the UK at the forefront of public sector information and open data initiatives globally. Research by Deloitte suggests that the UK has the most page views of any open data portal in the world¹⁴.

However, while the UK holds a number of strengths in the availability, analysis and value extraction of public sector information, there are challenges in fully exploiting the potential value from public sector information. There are logistical challenges around what data should be made available, when, in which format and the choice of funding model to use. There also exist challenges around ensuring businesses, civil society and the public sector itself are able to effectively use and re-use public sector information so as to fully exploit its value.

The Government has recognised that these challenges may be becoming barriers and on 22 October 2012 launched the **Shakespeare Review** led by Stephan Shakespeare, Chair of the Data Strategy Board. This review covers the entire breadth of the public sector information market (current and future) in order to make recommendations to Ministers on how to widen access to public sector information and consider new and innovative opportunities for open data.¹⁵

As part of the Shakespeare Review, **Deloitte** has been commissioned by the Department of Business, Innovation and Skills (BIS) on behalf of Stephan Shakespeare, to undertake an independent assessment of the market for public sector information, in order to establish a robust evidence base on its value and to highlight the policy implications flowing from an examination of how public sector information could be utilised further.

Defining public sector information and the market

Public sector information dimensions

Broadly speaking the term 'public sector information' refers to **data**¹⁶ and **information**¹⁷ that the public sector collects, produces, reproduces, publishes and disseminates in many areas of activity while accomplishing their public task or other duties¹⁸. In some limited cases for particular public sector information datasets¹⁹, there will also be private sector suppliers of data and information that can be considered substitutes for public sector information.

¹² See www.theodi.org/news/odi-launches-%C2%A3850k-scheme-create-businesses-open-data

¹³ See www.gov.uk/data-strategy-board#recent-news

¹⁴ Although it should be noted that page views do not always translate to use and re-use of data.

¹⁵ The full Terms of Reference can be found at: www.bis.gov.uk/assets/biscore/innovation/docs/i/12-1233-independent-public-sector-information-review-draft-terms.pdf

¹⁶ Defined as qualitative or quantitative statements or numbers that are assumed to be factual and not the product of analysis or interpretation.

¹⁷ Defined as outputs of processes that summarise, interpret or otherwise represent data to convey meaning. The term data typically includes information and this document follows this convention.

¹⁸ The collection, production, reproduction, publication and dissemination of public sector information by the public sector can be done through official channels or through third parties in the private and third sectors.

¹⁹ Defined as collections of data and information.

Figure 1: Public sector information dimensions

Public sector information datasets differ across a number of dimensions:

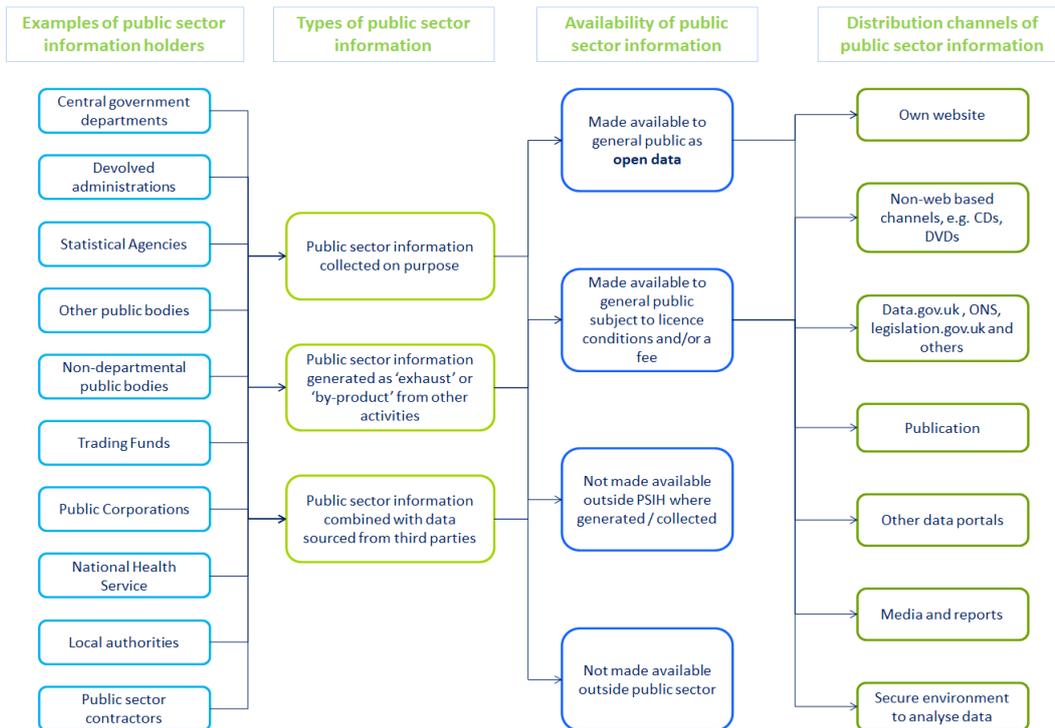
- whether the public sector information is made available to the general public and, if so, under what conditions (if any);
- the complexity of the dataset in terms of the number of records and variables, whether it is anonymised or aggregated or is quantitative or qualitative;
- how often the dataset is updated or replaced;
- whether public sector information is generated or collected as part of a public body’s public task or whether its generation is the result of other activities not related to public sector information;
- the content of the public sector information dataset;
- the electronic or non-electronic format of the dataset;
- the ways in which the public sector information dataset is distributed; and
- the cost of generating/collecting/maintaining/updating the public sector information dataset.

Source: Deloitte analysis

Public sector information holders and customers

Public sector bodies that generate, collect and disseminate public sector information (hereafter referred to as Public Sector Information Holders or ‘PSIHs’) will themselves differ in size, role and purpose and the extent to which, and how, they make the information and data available. Figure 2 below summarises.

Figure 2: Public sector information holders and routes to market



Source: Deloitte analysis. Note this illustrative is not exhaustive. For example, under the ‘20 year rule’, historical material previously unavailable is made available to the public²⁰.

As Figure 2 illustrates, public sector information is generated by a plethora of different PSIHs. It can be **generated specifically** (e.g. the Met Office collecting weather data), as a **by-product of**

²⁰ In 2013, the government began its move towards releasing records when they were 20 years old instead of 30. See www.nationalarchives.gov.uk/about/20-year-rule.htm#text for more details.

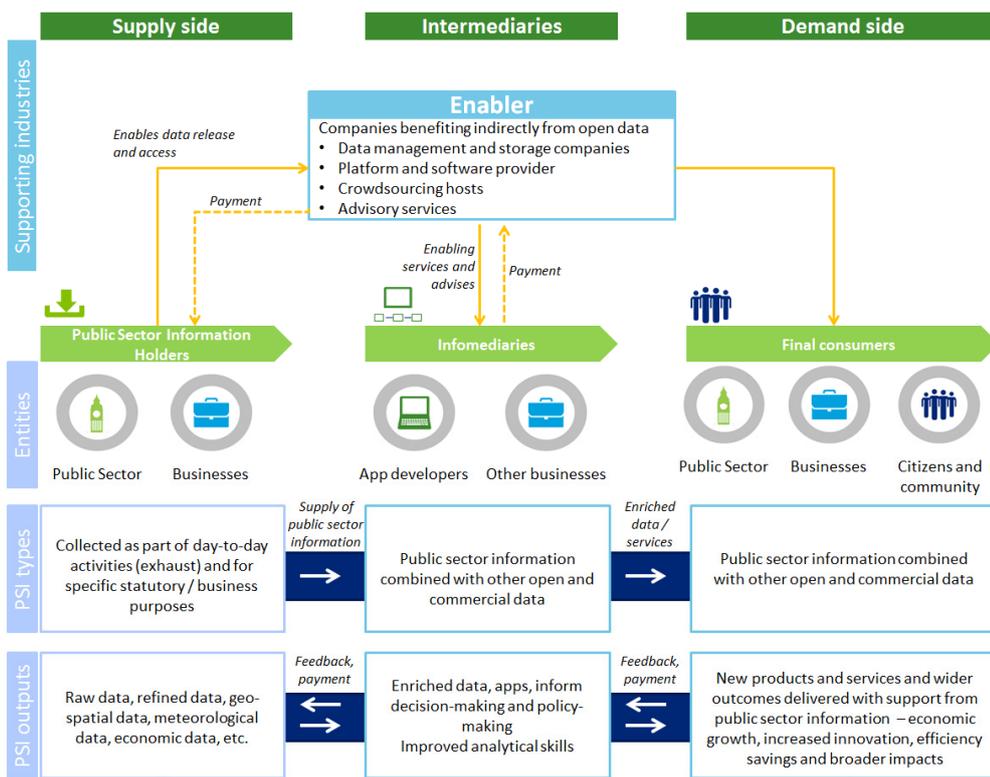
other activities (e.g. salary data - so called **exhaust data** that is generated through the performance of regular activities that are not data collection specific) or **combined with third party data** (e.g. financial data analysed and combined with other statistics by the Bank of England that is then used to inform economic forecasts).

While data on the total number of public sector information datasets is not recorded, a review of selected data portals reveals there to be over 37,500 datasets currently being made available as open data by PSIHS (this is likely to be an underestimate given a wealth of public sector information is also available through other routes). Focusing on data.gov.uk, analysis shows that the Office for National Statistics, Department for Communities and Local Government and the Health and Social Care Information Centre – with the analysis showing that of a total of 781 different publishers (as of 27th February 2013), the top ten suppliers of public sector information supplied over half of all datasets published on the website.

On the demand side of the market, customers come from the private sector, civil society, individuals and government itself. Some customers will use the data as direct inputs into products, services and research, while others will be repackaging the data for others to use – so-called infomediaries.

The market for public sector information can therefore be summarised as shown in Figure 3 below.

Figure 3: Public sector information market



Source: Deloitte analysis

As Figure 3 shows, on the supply side, PSIHS consist of the public sector and private business – though in practice the overwhelming majority will be supplied by the public sector. Intermediaries take this information and data and host and repackage it for a wider audience – in some cases this means augmenting the public sector information with other data elements. On the demand side, final consumers (which can include PSIHS themselves) can use and re-use the public sector information to develop new products, inform decision-making, improve research and make efficiency savings. As public sector information is increasingly used and re-used, actors across the supply, intermediary and demand sides will improve their data analysis skills, which can raise competitiveness and drive economic growth.

The value of public sector information today

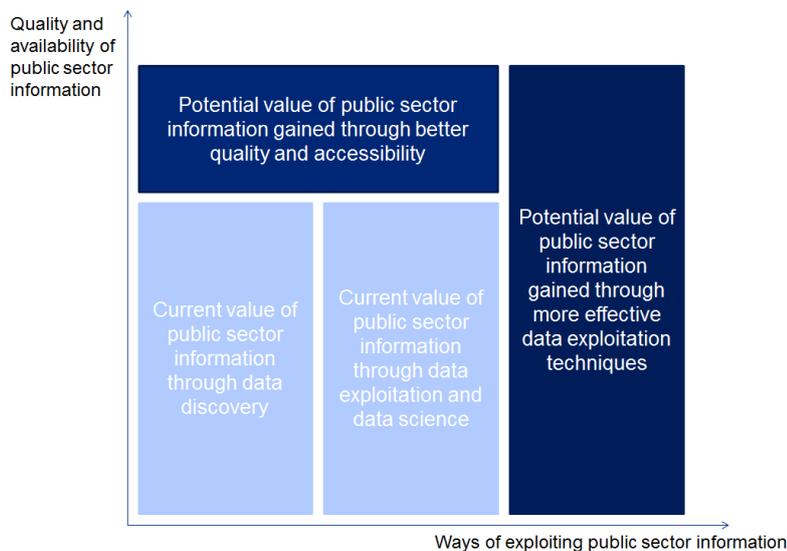
The link between public sector information and growth

The economic importance of public sector information has increased radically with the spread of new communication technologies, most notably the Internet, and the development of a ‘knowledge economy’ in which value is generated through innovation in information and services. There is now a considerable body of academic and other literature supporting the view that the greater **availability** and **accessibility**²¹ of public sector information can boost innovation and facilitate economic growth. Vickery (2011) and others have concluded that “knowledge is a source of competitive advantage in the ‘information economy’, and for this reason alone it is economically important that public information is widely diffused”²².

Generating value from public sector information

Following a review of the evidence it is apparent that the two main ways in which value is being generated currently, which will likely remain the case in the future, is through **data discovery** and **data exploitation**. The former relates to making better and, in some cases more, public sector information available and making it more accessible – what might be loosely term **supply-side** considerations. The second dimension relates to using and re-using public sector information better – **demand-side** considerations. This might be through reducing consumer risk aversion to using public sector information, improving data exploitation techniques, changing cultures and improving data analysis skills (which, in turn can improve competitiveness).

Figure 4: Generating value from public sector information



Source: Deloitte analysis. Boxes not to scale.

Value is thus generated through exploiting existing datasets to identify insights through statistical analysis, ‘data-mashing’ and visualisations. The value of public sector information in the UK can therefore be increased both by increasing the quality (and quantity) of public sector information (increasing the potential for data discovery) or by better using existing and new datasets.

While, in some cases, more public sector information being available can also increase value, there is not a linear relationship between quantity of public sector information and its value. What is

²¹ In this case, accessibility refers to the conditions attached to the use and re-use of public sector information. These conditions can take the form of a fee, limitations on who has access the data and limitations over how the data can be used.

²² Quoted in Vickery, G (2011) *Review of recent studies on PSI re-use and related market developments*.

key is the quality of this data and its amenability to data analytics. Simply releasing more datasets, irrespective of their quality, is likely to only have a minimal value impact.

Value from different public sector information datasets

Clearly the value of public sector information will vary according to the identity of the final beneficiary, the dataset in question and how and when the information is being exploited. Figure 5 highlights how different public sector information datasets can generate different levels of value.

Figure 5: Value of different public sector information datasets

It should be noted that in and of themselves, public sector information datasets do not carry any intrinsic value. Value is a function of customers being able to use the datasets to generate revenue and jobs, assist in decision making, and promote transparency and accountability – these impacts are discussed in the following sections.

The evidence points to the value of any given public sector information dataset to society as being positively correlated with:

- **the content of the dataset** – there are certain content themes that have well-established uses and re-uses or may be fundamental to the provision of services, products and types of research. Where this is the case, certain content themes (e.g. geo-spatial data or transport data) will positively influence the value of the public sector information dataset to society;
- **the flexibility of the dataset** – where datasets can be used in multiple ways to generate insights (e.g. house price data that can be used as proxies for a range of factors such as environmental conditions or school performance), the relative value of the dataset to society may be higher than a single-use dataset;
- **the accuracy, comprehensiveness and speed of refresh of the dataset** – as one would expect, the value of public sector information dataset will increase with its accuracy, comprehensiveness and speed of which it is updated (e.g. economic statistics); and
- **the ability to link the data** – the easier it is to link a given dataset with other datasets and other forms of information will increase its flexibility and comprehensiveness, and again can increase its value to society.

Of course, it should be noted that the value of a given public sector information dataset to society will vary over time, with some datasets becoming more or less valuable as new uses are discovered or rival datasets used – and accordingly predicting the future value of any given dataset is difficult.

Building on existing analyses, it is possible to create a data intensity matrix to compare how different types of datasets are used by different economic sectors in the UK currently (see Chapter 3 of the main report). The analysis shows that the construction, real estate, financial and insurance, public sector and arts, entertainment and recreation sectors are some of the largest users and re-users of open data²³. Dataset types most commonly being used include demographics, economic, environmental and geo-spatial and housing data.

From an analysis of data.gov.uk dataset requests, the most popular dataset category requested is location (or geospatial) data. This reinforces the impression that this category of data generates some of the highest levels of value. Environment, transportation, health and society are also data that attract particularly high levels of requests. Requests for data used to scrutinise and hold government accountable, such as government, finance, policy, administration and spending data are significantly lower, suggesting that this attracts less interest. However, even in these categories the numbers of requests are not insignificant, and the threat of accountability arising from the data, rather than the specific use of the data may be the primary driver of value.

Source: Deloitte analysis

Nonetheless, at a broad level, from the literature surveyed for this evidence review, it is clear that the release of public sector information has the potential to generate significant economic value through **stimulating innovation, addressing market failures** (such as barriers to entry and information asymmetries), **facilitating new ways of working** and **creating network effects** arising from more and more users of public sector information generating new insights and cross-fertilisation of ideas, helping with the creation of **new markets**.

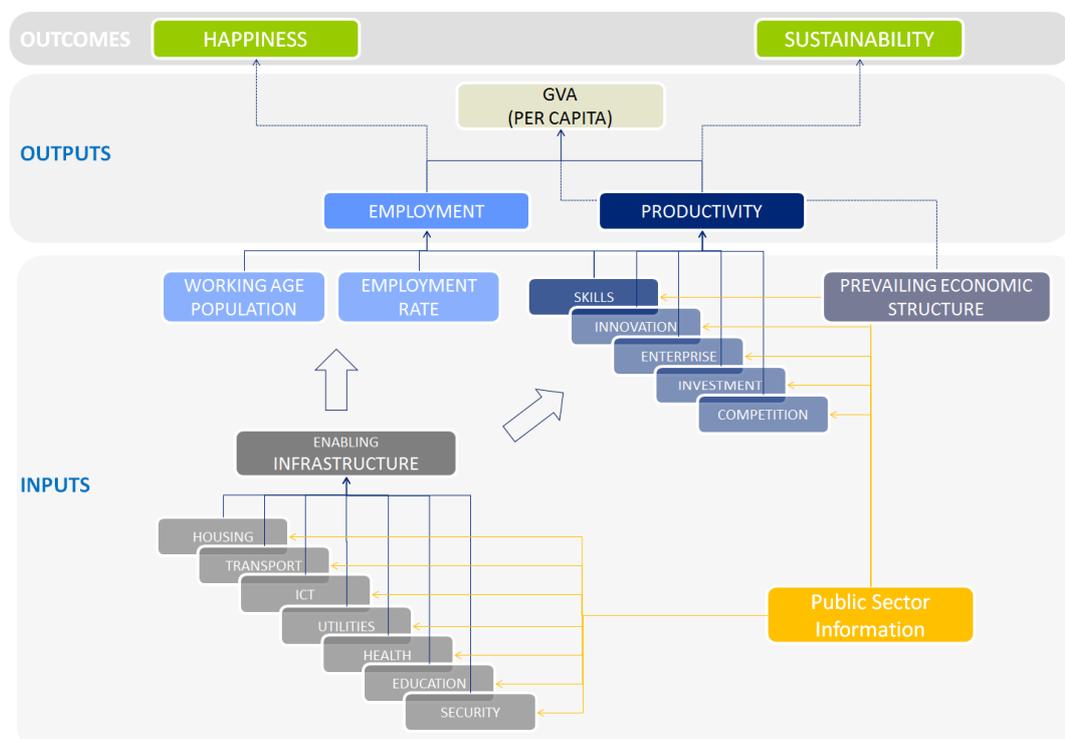
Conceptually, one can see how public sector information can drive economic growth and wider prosperity in the form of happiness and sustainability. The simplified framework, shown below in Figure 6, illustrates the long-term drivers of the UK's **economic growth** and **prosperity**.

²³ While this analysis has been carried out on the basis of open data, one might hypothesise that a similar picture holds for public sector information.

Measurable outputs (such as Gross Value Added²⁴, employment and productivity) as well as less easily measured outcomes (such as happiness and sustainability) are determined by two key drivers in the form of **productivity** and **employment levels**. In other words, in the long-term and considering the supply side of the economy only, the economic output of the UK is a function of only two things: the number of people engaged in gainful employment and the amount each person in employment is capable of producing.

In turn, these two key drivers of economic growth and prosperity are determined by **seven necessary enablers**. These are related to the infrastructure required to facilitate long-term growth: a deficit in these enablers will equate to a supply-side constraint on economic growth in the long-run. This could be caused either by limiting the growth in the working population (through insufficient affordable housing capacity limiting labour mobility) or by acting as a drag on productivity growth (through below-par ICT connectivity or a sclerotic transport system). The amount each worker produces is determined by skill levels; the extent of innovation in products and processes; the degree of investment in capital; entrepreneurial activity; and, lastly, levels of competition. Expanding and enhancing these enablers can therefore positively impact employment and productivity, which in turn can generate economic growth and greater prosperity.

Figure 6: Long-term UK economic growth framework



Source: Deloitte analysis based on HMT and BIS analysis. See www.hm-treasury.gov.uk/d/ACF1FBD.pdf for more details.

As is shown in Figure 6, the introduction of more public sector information can have a positive impact on infrastructure and other enablers. For example, it can stimulate innovation, help enhance skills, promote competition and enterprise and attract investment to new products and services. In terms of infrastructure, public sector information can help generate efficiency savings across public services and improve business decision-making.

²⁴ Which can be thought of as analogous to GDP.

Understanding how public sector information creates value in the UK

Having reviewed the literature, it is useful to disaggregate the different types of value generated by public sector information according to the beneficiary:

- **the direct value of public sector information to producers and suppliers (the PSIHs):** these are the benefits accruing to producers and suppliers of public sector information through the sale of public sector information or related value-added services;
- **the indirect value of public sector information arising from its production and supply:** the benefits accruing up the supply chain to those organisations interacting with and supplying PSIHs (but not directly using or re-using public sector information), and the benefits accruing to those organisations where employees of PSIHs and supply chain organisations spend their wages;
- **the direct use value of public sector information to consumers of public sector information:** the benefits accruing to businesses, civil society, individuals and the public sector from directly using and re-using public sector information for a variety of purposes; and
- **the wider societal value arising from the use and re-use of public sector information:** the benefits to society of public sector information being exploited, which are not readily captured elsewhere.

Using the above disaggregation, Figure 7 overleaf shows how value can be created across the public sector information market, from the perspective of different participants along the supply chain. Three (hypothetical) broad examples are shown:

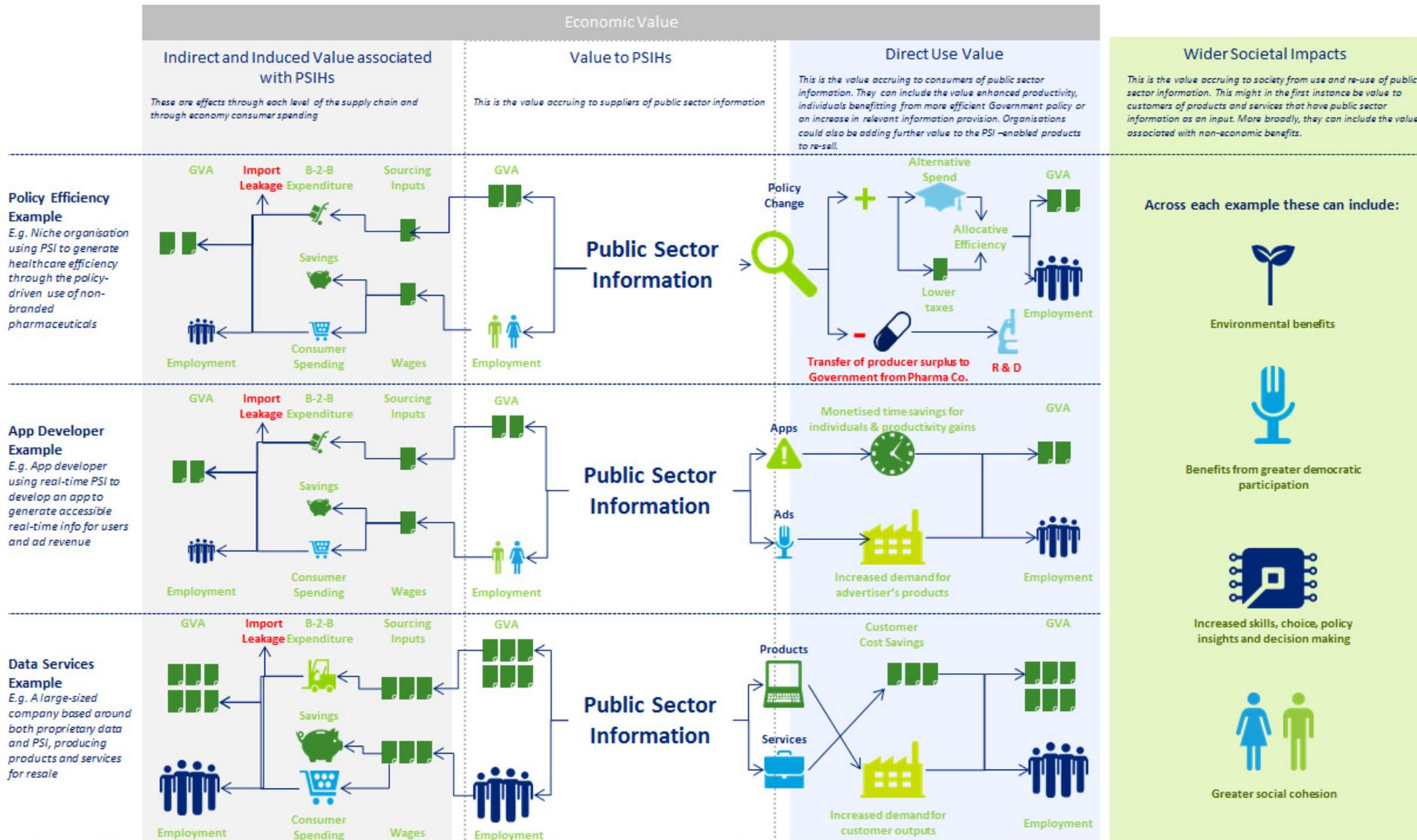
- a **policy efficiency example** where value is generated by the re-use of health data;
- an **app developer example** where value is generated by the use of data as a key input into transport apps that seek to save time; and
- a **data analytics example** where value is generated by the use and re-use of data to generate customer and business insights.

Value is generated across each element of the supply chain. PSIHs who produce and disseminate the public sector information can create value through employing staff to collect and organise the public sector information. If they also sell public sector information or value-added products, value will be created through the revenue attributable to public sector information. Or, if the PSI is made available for free under the Open Government License (OGL) this has the potential to maximise its use by third parties to generate added-value in the supply chain (see downstream impact).

Upstream along the supply chain, value will be created as a variety of third party organisations doing business with PSIHs (such as IT suppliers, operations and maintenance firms, caterers, recruitment agencies, etc.) will receive orders from the PSIHs and earn revenue. Similarly, businesses such as retailers where employees from PSIHs and other organisations in the supply chain spend their wages will also earn revenue and generate jobs.

Downstream along the supply chain, value is created by consumers of public sector information directly and indirectly using the information and data. In the policy efficiency example, this is through using the public sector information to identify efficiency savings; in the app developer case value is created through the sales of the app and the direct financial benefits to users in the form of time saved; and in the data analytics example, value comes from efficiency savings and better targeted products for consumers – raising revenue.

Figure 7: Economic and social value from public sector information



Source: Deloitte analysis. In the Figure 7, the term Gross Value Added (GVA) is used to capture economic value (from profits, wages and rents).

Even for these three archetypal examples, the wider societal value from public sector information is more challenging to measure as it captures wider benefits arising from the use and re-use of public sector information that are typically not measured in monetary terms. The literature discusses a number of ways in which public sector information can have broader value impacts through:

- **increasing democratic participation:** giving citizens and businesses access to public sector information allows them to perform their own analyses of salient issues, make more informed choices about public service providers and interact with policymakers to challenge their assumptions and improve the policymaking process;
- **promoting greater accountability:** for example through the scrutiny of costs of public service provision and benchmarking comparable services;
- **greater social cohesion:** for example, by providing more information on the provision and distribution of services, public sector information can be used to dispel myths on who receives certain public services;
- **generating environmental benefits:** such as reducing congestion and pollution through the release of better traffic and transport data which helps drivers to better plan journeys; and
- **identifying previously unknown links between different policy areas:** through data-mash ups it may be possible to develop system-wide solutions that holistically seek to address the root of policy challenges.

This value is potentially significant and is likely to have a major influence on overall societal wellbeing.

Estimates of the value of public sector information and value to PSIHs

Economic value

Figure 7 reflects the complex nature of the public sector information market and its participants for just three very specific examples²⁵. Very little evidence exists to accurately identify the ways in which public sector information acts as an input in the productive process and the importance it has in generating value. There is no central database tracking data and information collected and stored by the public sector and many businesses are reluctant to disclose how they use and re-use public sector information. As part of this research, steps have been taken to remedy these gaps through stakeholder consultations, but the fact remains that it is necessary, in many cases, to make a number of simplifying assumptions to arrive at indicative quantitative value estimates. Whilst Deloitte is content that these estimates make best-use of the information provided within the scope and parameters of the study, the estimates quoted in this report should be considered with regard to these caveats.

Using an adapted bottom-up methodology outlined by the Office of Fair Trading in its 2006 *Commercial Use of Public Information report*²⁶ that is consistent with HMT Green Book guidance²⁷, this report has estimated that the **value of public sector information to consumers, businesses and the public sector is currently approximately £1.8 billion** in 2011/12 in 2011 prices (what is termed the economic value). This figure includes the direct value of public sector information to PSIHs, the indirect value of public sector information arising from its production and supply, and the direct use value of public sector information to consumers of public sector information – for shorthand, this is referred to as the **narrow economic value of public sector information**.

²⁵ Figure 7 is also limited to the economic and social value within the UK. Clearly, there will be wider international network and other effects that spillover outside the UK. These effects, which potentially could be significant, are outside the scope of this report.

²⁶ Available at: www.offt.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861. In particular, see Annex G for the modelling framework. See Appendix 5 for a more detailed description of the methodology used in this study, including assumptions around additionality and the counterfactual.

²⁷ See www.hm-treasury.gov.uk/data_greenbook_index.htm

Figure 8: Estimates of the current value of public sector information in the UK in 2011/12 (2011 prices)

Value category	Central scenario	High scenario	Low scenario
Direct consumer surplus	£1.6 billion	£2.00 billion	£1.00 billion
Producer surplus	£0.1 billion	£0.1 billion	£0.1 billion
Indirect and induced value (supply chain and consumer spending)	£0.1 billion	£0.1 billion	£0.1 billion
Total	£1.8 billion	£2.2 billion	£1.2 billion

Source: Deloitte analysis

Subjecting this modelling approach to sensitivity tests²⁸ generates a lower bound value of £1.2 billion, and an upper bound value of £2.2 billion for the value of public sector information in 2011/12.

Wider societal value

The use and re-use of public sector information has **much larger downstream impacts affecting all areas of society**. Public sector information can act as a catalyst for positive creative destruction – the process of generating innovation, identifying and making efficiency savings, helping officials, business leaders and ordinary citizens make better policy and business decisions and promoting accountability and transparency.

Through the use of case studies, it is possible to derive some indications of the financial scale of different components of social value.

Figure 9: Case study examples of public sector information generating value

Sector	Example
Healthcare and life sciences	<ul style="list-style-type: none"> Using NHS prescribing data to identify efficiency savings of up to £1.4 billion per annum from switching from branded to generic drugs Reducing mortality rates following cardiac surgery leading to a positive impact on standards
Transport	<ul style="list-style-type: none"> Using traffic data provided by the Highways Agency to build a tool to better plan road trips – potentially saving motorists up to £6.5 million per annum (value of time saved) Using public sector information on roadworks to better co-ordinate utilities work and journey planning – leading to estimated benefits of around £25 million per annum for local authorities and road users in terms of efficiency savings and reduced congestion Embedding public sector information on real-time transport data in apps in London can save users between £15 million and £58 million in terms of time saved, each year

See main report for further details of report authors and details

Whilst these individual estimates of social value from the different case studies cannot be simply summed (some are cost-savings, some are time-savings some are economic value and some are non-quantifiable), they suggest that the **total social value of public sector information is likely to be significantly greater than the narrow economic value presented here**. A conservative

²⁸ Through altering assumptions on demand curve shapes, elasticities and the usage of public sector information datasets. Alternating these assumptions does not materially affect producer surplus and indirect and induced impacts.

estimate of the wider social value of public sector information suggests a value at least three times the size of the £1.8 billion economic value may be appropriate²⁹, i.e. over £5 billion p.a.

However, without further data on the relationship between public sector information use and social outcomes, it is not possible to say definitively what the value might be.

Total value

Aggregating the calculated economic value and the conservative estimate of wider societal value generates a total current value of public sector information of between £6.2 billion and £7.2 billion in 2011/12 (2011 prices).

Future uses

Given the unpredictable nature of innovation, with the most valuable new products often being among the least anticipated, it is risky to offer firm predictions of the future uses and value of public sector information. Nonetheless, given the current trajectory of increasing volumes of public sector information being made available in ever more accessible formats, it seems reasonable to hypothesise that the larger part of the value to be generated from public sector information lies in the future. Much of this value may come from combining public sector information with information from other sources, such as details of consumer spending habits held by supermarkets and other retail firms, or details of domestic and commercial properties.

Many of these future uses may be extensions of the ways in which public sector information is already being used. For example, a recent study calculated that the benefit to the local government sector from the use of geospatial data was £232m over five years, through a variety of measures such as more efficient routes for waste collection.³⁰ It is to be expected that such uses will become increasingly widespread throughout the public sector at both a local and central level, as more information becomes available, methods of harnessing this information are improved, and best practice becomes more widely adopted, thus increasing the value derived from public sector information.

Further benefits are likely to emerge for private individuals, business and other organisations as methods of exploiting public sector information are developed and improved. More effective ways of data exploitation can include:

- greater system-wide analysis through more sharing and combining datasets together to consider policy and business issues from a range of non-traditional perspectives;
- unlocking the potential from linked data: and better integrating; and
- greater data-mashing with private sector and individual datasets.

The case studies explored in the main report illustrate some of the ways in which public sector information is generating value. Such examples are likely to proliferate as available data proliferates and the expertise to exploit it is developed. An example is the case of Honest Buildings, described in Chapter 6: this SME is exploring ways to help businesses reduce their energy bills by combining Energy Performance Certificate data with privately held data, potentially delivering significant savings and reducing the UK's energy usage. Such examples may be expected to proliferate as access to information increases, and innovative uses are developed.

In addition, there are likely to be benefits from wholly new uses of public sector information. Numerous additional possibilities for achieving such savings and improving policy may become apparent as more information is made available and expertise in its use increases. In addition, there are likely to be more 'blue sky' opportunities, such as those identified by the Government

²⁹ This is based on insights from individual case studies by other authors which have variously estimated the wider value from public sector information use in the health care sector, geo-spatial, meteorological and other sectors in the order of billions.

³⁰ The Value of Geospatial Information to Local Public Service Delivery in England and Wales, 2010

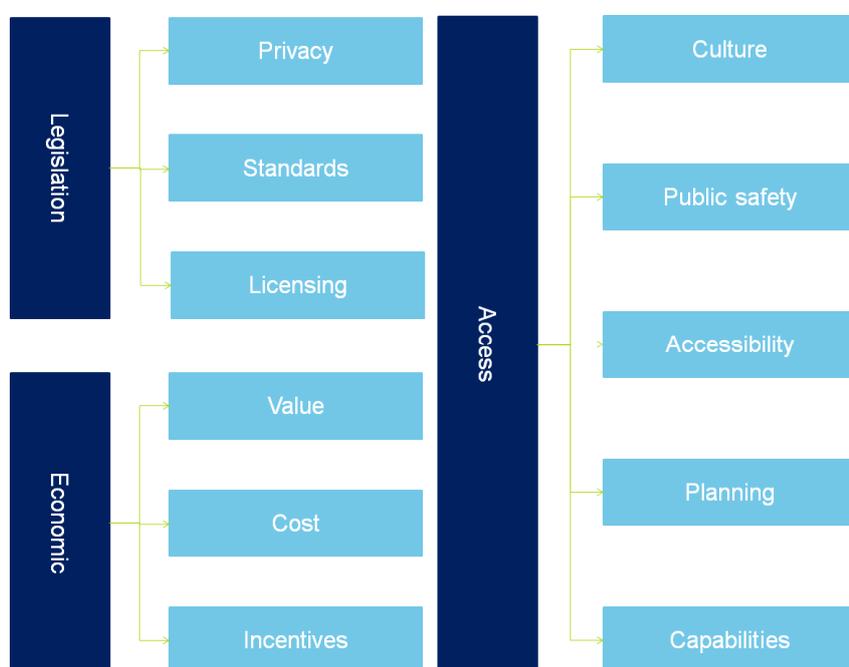
Office for Science. These are explored further in Chapter 5, but include improving management and resilience of food supplies, tackling obesity, and detecting and identifying infectious diseases. These require combining public sector information with other sources of information rapidly and making the resulting insights available to decision makers in a timely fashion to allow prompt responses to crises and informed policy decisions.

Barriers to realising the full potential of public sector information in the UK

A taxonomy of barriers

Adapting a taxonomy originally developed by the Data Strategy Board, the research has examined the evidence to identify the areas where barriers may currently exist. Note, some of the themes under each barrier overlap, e.g. privacy could be captured under legislation or access.

Figure 10: Taxonomy of barriers



Source: adapted from DSB

Legislative barriers include whether certain acts of legislation and other regulations reduce the usability of public sector information datasets by consumers. There are also further questions over whether current assurance/accreditation standards for public sector information remain effective and whether there is scope for existing licensing arrangements to be improved.

Economic barriers include questions of which datasets to release (which datasets yield the highest value) and how their costs (if any) should be covered.

Access barriers to maximising the value from public sector information can include:

- a reluctance to publish or share datasets;
- a bias against using non-traditional datasets;
- a reluctance to use public sector datasets because of concerns around quality, reliability and on-going support; and
- a lack of skills to fully exploit the value of public sector information.

It should be noted that the extent of these barriers may differ between public sector information datasets and also between the different types of users.

Each of these barriers is discussed in more detail below.

Legislative barriers

Privacy and the impact of current regulations and legislation

The review of available evidence suggests that by and large, the current legislative and regulatory environment around public sector information is not acting as a barrier to generating value and market development. While a full legal analysis has been beyond the scope of this study, particular acts and regulations such as those covering **Data Protection and Human Rights legislation have not emerged as preventing the development of new products or services using public sector information**³¹. The majority of stakeholders consulted have not reported that these generic legislative acts are currently preventing them from using and re-using public sector information, but, in some cases specific legislation controlling a particular data collection exercise do.

However, the evidence received does suggest there are some current **challenges around perceptions and attitudes towards data release**. In particular:

- regulations such as **Data Protection are sometimes used as a shorthand justification for not sharing public sector information within the public sector**, with PSIHs not always able to translate their awareness of their rights and duties into scenarios where public sector information is released or shared, causing a barrier; and
- **when a policy decision is taken not to release public sector information datasets to the general public, the reasons are often not well articulated or the conditions attached to access are overly restrictive.**

Stakeholders have cited examples where the Data Protection Act (and other legislation) has been used as a reason for a PSIH to withhold information from the general public. However, as noted above, in reality, the Act should rarely be a barrier to sharing information. Where public sector information datasets do contain personal details, these may need to be aggregated or anonymised. Where this is done effectively the Data Protection Act, in most cases, no longer applies to the information and it can be made available.

This is not to downplay data protection issues. The impact of breaches leading to the release of personal information can be extremely serious. Further, even if data is anonymised, people may be reluctant to report it if they believe it can have wider consequences, e.g. reporting crime data may adversely affect house prices³². These issues notwithstanding, PSIHs are beginning to use innovative methods to test different ways of anonymising datasets – for example the Ministry of Justice worked with statisticians, the private sector and the academic community to avoid the ‘jigsaw effect’³³ occurring from the release of offender data.

Within the public sector, there are also questions surrounding the ability of different public bodies to share datasets with one another. These have been explored in depth in the recent report by the **Administrative Data Taskforce**³⁴. It notes that as regards the legal gateways established to allow departments and other public sector bodies to share information without obtaining the consent of the data subjects, “recent experience demonstrates that link-specific gateway legislation is both cumbersome and inefficient.” As a solution to this, the Taskforce recommends the creation of a generic legal gateway to reduce this barrier to the exploitation of public sector information and clarify the legal position around sharing data. Note that this applies to sharing of data within the

³¹ This observation should be caveated with the note that, at the time of writing, there are a number of initiatives to revise various regulatory frameworks at the European level.

³² See for example: www.guardian.co.uk/money/2011/feb/01/police-crime-website-house-prices

³³ The process of combining anonymised data with auxiliary data in order to reconstruct identifiers linking data to the individual it relates to.

³⁴ Source: Administrative Data Taskforce (2012), www.esrc.ac.uk/funding-and-guidance/collaboration/collaborative-initiatives/Administrative-Data-Taskforce.aspx

public sector and with certain accredited third parties such as researchers, rather than publication for use by the general public.

Insights

- **Existing regulations and legislation governing public sector information do not appear to be acting as actual barriers to realising the full potential value of public sector information. However, the manner in which some of these regulations and legislation are interpreted can lead to overly risk averse behaviour and can create barriers.**
- **Increasingly effective anonymisation techniques and an approach across the public sector that emphasises granting access to as much data that is compatible with privacy and security has the potential to improve access to public sector information.**

Licences

Licensing conditions play an important role in facilitating (or preventing) the full exploitation of the value of public sector information. The ideal standard widely acknowledged by stakeholders was licensing public sector information under the Open Government Licence (OGL), and increasing amounts of public sector information are being made available as open data under it, even by organisations that have traditionally charged for data³⁵. Some groups of stakeholders have argued that if the OGL were used for all public sector information, this could substantially increase the openness of the UK's public sector information and could remove many of the barriers to use and re-use that currently inhibit the realisation of the full value of public sector information.

However, some argue that the public sector information landscape is characterised by its diversity, and for this reason a 'one size fits all' solution is unlikely to be practical or necessarily desirable. For cases where the release of information under the OGL is not considered appropriate and cost recovery is justified, the introduction of a generic charging licence could, in principle, address the current complexity of charging arrangements for public sector information, completing the UK Government Licensing Framework and simplifying licensing arrangements for users. Some stakeholders pointed out that by having a charge, users and re-users can be re-assured of the quality of data and be able to expect a certain level of service (this is discussed in more detail below). However, one concern raised was that such a licence could encourage some PSIHs who are not covered by the OGL, to charge for certain datasets.

Insights:

- **Making public sector information available under the Open Government Licence is seen by a large number of stakeholders as an effective means of removing barriers to exploiting the value of public sector information.**
- **If there are instances where a generic Government Licence for charging for public sector information are applicable, such a Licence would need to be drafted in a way to avoid incentivising charging when there is no strong justification or leading it to become the default alternative to the Open Government Licence.**

Eligibility restrictions on accessing particular datasets

With respect to access restrictions to public sector information datasets (such as the datasets only being available to researchers or in secure environments), in many cases the rationale for these restrictions are clear. The rationale may cover national security reasons or genuine data protection concerns. However, some stakeholders have reported that in some cases, **the rationale for restrictive access to certain datasets is not always clear, appears overly restrictive or may**

³⁵ For example, Ordnance Survey has released 11 datasets, and the Met Office, Land Registry and Companies House have each opened up some of their data.

no longer be relevant, although it should be noted that just because a stakeholder feels the rationale is overly restrictive, that does not necessarily make it so.

For example, some commercial start-up companies have reported difficulties in accessing different versions of the National Pupil Database as non-research organisations. While restrictions on access are in this case clearly needed to protect sensitive and personally identifiable information; there is a valid question as to whether there are ways that some or all of this information could be made available to commercial organisations to develop new products and services to help parents, teachers and other stakeholders to improve decision-making, increase accountability and identify good practice. Without this, opportunities for innovations using this and other datasets are likely to be missed, meaning loss of the value that they could potentially create.

Insights:

- **The eligibility restrictions imposed around certain datasets are typically due to reasons of national security, data protection and other sensitivities.**
- **The rationale behind these restrictions is not always clearly articulated and may, in some cases, no longer apply. In some cases, this may be preventing opportunities for innovation to take place.**

Economic barriers

The evidence reviewed suggests a number of key economic challenges around maximising the value from public sector information in the UK. Primarily these relate to the value and cost of datasets, which can be grouped into two key themes:

- **which datasets are made available to the general public;** and
- **which datasets consumers have to pay to access for (funding models).**

Data gaps

There is currently no national or local information asset register covering the amount and type of public sector information datasets held³⁶ (irrespective of whether it is being released to the public) or the detailed costs of collection and dissemination. Similarly, there is little evidence on how consumers themselves are using and re-using particular datasets.

While this has meant this report has been forced to make a number of assumptions in order to reach quantitative estimates, there is a more significant impact on PSIHs themselves. By not being able to accurately ascribe value to different datasets, PSIHs are generally unable to reach evidence-based decisions as to which datasets to publish, how to publish them and what support to provide. One consequence of this is that the costs (which are readily measurable *a priori*) of making particular public sector information datasets available may appear to be much larger than the benefits – as the benefits case cannot be clearly articulated due to a lack of evidence.

Indeed, as part of this report, a sample of government officials have been consulted over the reasons why public sector information datasets are not released: over 50 per cent of those responding identified resources as an issue preventing the release of more public sector information, with the following response being typical: *“in a time of diminishing resources and the need to make best use of the resources we have; the time and the cost of ensuring data validity before release is an issue.”* This was one of the most common reasons given for why data is not released, but not the only one – others include data protection and legislation.

Stakeholders have identified particular datasets (such as an aggregate Energy Performance Certificate database) that are currently not available to them (either for free or for a fee) that they could use to generate new products and services. By not making these datasets more widely

³⁶ Though the UK is not alone in this respect.

available (or articulating the reasons they are not available), PSIHs may, whether knowingly or not, be acting as a barrier to realising the full potential of public sector information.

There are routes to address these data gaps. One route might involve conducting a public sector-wide audit of public sector information (which also covers the customers using and re-using it). This would involve both primary research to determine the stock of public sector information held across all PSIHs and also identifying how this data is being used commercially across different economic sectors and actors. This would help match costs and benefits to individual datasets and building this report to identify those datasets that can generate the most value for the UK. The costs of doing such an audit would not be insignificant and for it to be effective, it would need to be repeated in future years (though the on-going costs may be lower than the one-off start-up costs). However, the benefits of doing such an audit could be outweighed by PSIHs having a much better sense of which datasets are creating the most value, which can guide future policy decisions around releasing data.

Other routes to addressing data gaps might involve better tracking of usages (perhaps through measuring how often data is re-used in other sources), requiring all new public sector information datasets to be logged centrally or crowdsourcing an audit across the public sector. Another potential route for the long-term, at least conceptually, might be a true single conduit for the access of all public sector information.

The Open Data User Group (ODUG) has been set up as an advisory group to Government and provides a channel through which the potential users of data, from all areas of the community, can identify their need for data and set out the benefits they expect this data to deliver. This will help PSIHs understand in more detail the potential benefits which the release of individual datasets can be expected to deliver.

Through the completion of data gaps, it would be possible for government to articulate what is meant by the term **core reference dataset** (beyond the definition contained in the Open Data White Paper). The issue of core reference datasets is also raised in recent work by APPSI on the national information framework for public sector information and open data.³⁷ Consideration could include:

- the features of a particular dataset that make it a core reference dataset;
- the funding arrangements of core reference datasets (see below);
- the obligations of suppliers of core reference datasets; and
- the rights of citizens to core reference datasets.

Insights:

- **There are significant data gaps when it comes to public sector information. This lack of data can, in some cases, lead to inertia with certain public sector information datasets not be released or conversely, undue attention being given to datasets that are unlikely to generate significant value but have a low cost of dissemination.**
- **There are a number of routes to addressing these data gaps. These range from a detailed, regular audit of public sector information to improved tracking of current usage and a single conduit for all public sector information.**

³⁷ The issue of core reference datasets is also raised in recent work by APPSI on the national information framework for public sector information and open data, www.nationalarchives.gov.uk/documents/nif-and-open-data.pdf.

Funding models

Related to the question of making the business case for more public sector information being released is the issue of how the cost of generation, collection, retention and dissemination of the datasets should be funded. This also leads to questions around pricing for access to public sector information.

This is a complex issue and data on costs (and benefits) is not fully available. Further, one must take account of different costs being incurred across the different stages of the data lifecycle. For example, in the case of so-called ‘exhaust’ or ‘by-product’ data that is generated as a result of PSIHs conducting their day-to-day and other activities and duties, the marginal cost of its generation will, by definition, be negligible or very low compared to the activity that caused the data to come about; but the marginal cost of its dissemination to the wider public may be higher and vice versa for some ‘purposely collected data’. There will also be additional costs in formatting the data for use and further costs if support is provided to the public in re-using the data.

The issue of charging and funding models for public sector information is highly contentious and extremely complex. For example, there remains an element of subjectivity as to what constitutes a dataset and what constitutes a ‘value-add’ service – with some PSIHs arguing that what is being charged for is not the public sector information itself, rather its interpretation and analysis.

As part of this report, a wide range of opinions have been expressed as to whether there is any economic or other justification for charging for public sector information datasets. On the one hand, these arguments include:

- **charges for datasets create barriers to entry and expansion** for SMEs and individuals to develop new products and services;
- the **charges prevent SMEs and individuals from ‘experimenting’ with the datasets** before they purchase to see if they are able to derive value from them, thereby making it hard to develop business cases; and
- any **lost revenues to PSIHs from releasing datasets for no cost will be recovered by the Exchequer** in the long-run through increased tax revenues and more jobs being created.

In contrast, arguments have been put forward to support current pricing arrangements include:

- **aligning a revenue stream with a particular dataset will ‘protect’ it from any reductions in funding**, allowing PSIHs to continue to supply this even if they themselves must make other savings;
- **a price can be interpreted as a signal of consumers’ willingness to pay for a particular dataset’s quality** and a commitment by the PSIH to maintain this and offer support; and
- **charging for certain datasets is necessary given they include elements of commercial or international datasets.**

As the most visible PSIHs that charge for certain public sector information, a great deal of attention has been focused on the four Trading Funds that make up the Public Data Group (PDG): Ordnance Survey, the Met Office, Land Registry and Companies House. The view has been expressed by some stakeholders that these PSIHs hold some of the most valuable datasets and there are strong arguments that these should be treated as **core reference datasets** available to all at no direct cost to the general public.

It is not helpful to treat these PSIHs as a single group – indeed, there are a number of other public bodies that charge for access to datasets³⁸ and there are a number of other Trading Funds outside the PDG. The four Trading Funds in the PDG differ in their sources of revenue, their public service duties and whether the public sector information they hold can be classed as ‘**exhaust**’ data or

³⁸ For example, the Office of National Statistics has a subscription charge for certain datasets.

‘purpose-collected’ data. Further, it should be noted that substantial progress has been made by the four Trading Funds in recent years to make increasing volumes of data available as open data³⁹ and there continue to be moves in this direction.

However, despite these positive steps, **there remains a perception among many consumers and commentators⁴⁰ that they are unable to access certain datasets for reasons of cost** and this is creating a barrier to business growth. A number of studies (e.g. Pollock, 2011 and others – see Appendix 3) have argued that releasing these datasets as open data will have significant welfare benefits.

The impact of a cost recovery model for public sector information, compared to the Open Government Licence model, is summarised below.

Figure 11: Potential impact of the cost recovery model on the public sector information market

Taxpayer funded public sector information	Public sector information supplied on a cost recovery basis
<ul style="list-style-type: none"> Public sector information typically made available at no cost to users under the Open Government Licence 	<ul style="list-style-type: none"> Access to public sector information involves a cost to the user and may come with restrictions
<ul style="list-style-type: none"> There may be strong public policy reasons for having these datasets available at no cost 	<ul style="list-style-type: none"> The costs of collection (rather than dissemination) are significant and cannot be borne by the public purse
<ul style="list-style-type: none"> Will not include other commercial or international datasets 	<ul style="list-style-type: none"> May include other data sourced under licence
<ul style="list-style-type: none"> The availability of public sector information at no cost to the user can contribute to the transparency agenda by increasing access to the widest possible customer base, irrespective of the ability to pay 	<ul style="list-style-type: none"> Access to Public sector information is restricted on the basis of cost
<ul style="list-style-type: none"> Will not typically include other services or any other guarantees 	<ul style="list-style-type: none"> May also include bespoke value-add services and guarantees of data quality and continuity

Source: Deloitte analysis

It is very difficult to perform a robust cost-benefit analysis of different funding model options for PSIHs, not least because PSIHs differ greatly from one to the next. While some data is available as to the costs incurred from collection and dissemination, this is not typically openly available or apportioned by dataset, nor is data readily available on how their customers are using the datasets and generating revenue and value.

More fundamentally, very little data is available on what the benefits might be if charging models were to radically change – as it is very difficult to predict how businesses and individuals might use datasets in the future to generate new products and services and by implication impact economic growth. It is also important to note that, as per the HMT Green Book⁴¹ guidance, any benefits from a change in charging structures should include not just increased tax receipts but wider social benefits and costs in terms of organisational impacts.

Even in the case of the four Trading Funds that make up the PDG, estimating the effects on Exchequer revenue of releasing all their public sector information as open data is a difficult task, even in spite of the information made available by the Trading Funds as part of the study. There

³⁹ For example, see www.ordnancesurvey.co.uk/oswebsite/products/os-opendata.html

⁴⁰ See for example www.freeourdata.org.uk/ and other references in Appendix 3.

⁴¹ Available at www.hm-treasury.gov.uk/data_greenbook_index.htm

are differing views on precisely which revenues should be considered relevant to PSI, considering factors such as specific revenues from the data, the cost of collecting the data, the extent to which Government 'buys' data from itself and so forth.

That notwithstanding, on the basis of the information made available to this study, it is possible to indicatively calculate that the cost effects on Exchequer revenue of continuing to collect and disseminate Trading Funds' public sector information in its current guise without charging for it. This cost is estimated to be in the order of £395 million on an annual basis⁴². However this figure is without regard to the extent that government pays for public sector information from the four Trading Funds in the PDG (this varies significantly between Trading Funds). On the basis of information provided by sales channel, it can be estimated that the annual loss to the Exchequer would be lower, as government would no longer need to purchase these public sector information datasets – it could use them at no cost. **In this scenario, the direct loss to the Exchequer on an annual basis might be of the order of £143 million.** This figure may be lower still if there are efficiency savings to be made if fewer dedicated sales and marketing resources are required by Trading Funds.

Following the HMT Green Book approach to account for the wider social and economic benefits, it is important to note that in a world without charging, private sector entities (consumers and businesses) that currently pay for access to public sector information provided by the Trading Funds would benefit by this amount – £143 million – and some of this may be recouped in the medium-term as a result of additional economic activity generating tax revenues.

An additional group that is currently deterred by having to pay for the data would also benefit as they are able to access the data. Estimating the size of this latter group is difficult but directionally it is clear that removing charging would mean more people and businesses, not fewer, would be able to access and benefit from this data⁴³. Conversely, organisations who are at an advantage in using their own proprietary information for commercial advantage, might find their competitive advantage eroded if more public sector information is released to act as a publically available substitute to that data.

The situation is thus complex, and while the above example is stylised, it does suggest the quantum necessary for the associated benefits to outweigh the costs. **Without more detailed and accurate data on both costs and benefits (including wider social benefits) it is not possible to reach a clear conclusion on this issue.**

However, this is not to say there is no room for improvement today in the provision of public sector information that carries a charge and a number of steps are being taken in this respect. These include⁴⁴:

- much more **communication** about what existing licences allow consumers to access and use / re-use the public sector information for, building on existing efforts by the trading funds and other PSIHs to build awareness among the user community;
- offering substantial **discounts** to SMEs and individuals;
- implementation of a '**royalties**' model for consumers to exploit the value of public sector information up to a certain value before a charge is applied;
- greater provision of '**sandbox**' or secure environments in specialised locations across the UK to allow consumers to explore datasets;
- greater **provision of out-of-date public sector information** at no cost to the general public to allow consumers to experiment; and

⁴² This figure comprises of Trading Funds' operating surpluses and the cost of data collection.

⁴³ There would accordingly be second-round effects for people consuming products provided by these businesses, etc.

⁴⁴ These options have not been costed.

- greater use of ‘**hack days**’ to demonstrate the value of particular public sector information datasets.

Of course, not all of these initiatives will be applicable to all PSIHs that charge for public sector information – there is no ‘one-size-fits-all’ model – but they should work with the grain of the market and build on existing initiatives to release more data. In some cases there may be significant logistical or legislative challenges to overcome before the above suggestions are implemented.

Insights:

- **The issue of charging for public sector information datasets and their funding models is complex. While significant progress has been made by a number of major PSIHs in this area in simplifying charging structures and making more public sector information available as open data, there remains a perception that barriers exist and there is scope for improvement.**
- **Ways to improve access could include greater communication on licence conditions, discounts to certain consumers, a royalties model, use of a ‘sandbox’ model, greater provision of out-of-date information and more hack days.**
- **There is currently a lack of data to definitively conduct a cost-benefit analysis across different funding models across the range of PSIHs currently charging. This is an area for further analysis through primary research with the direct customers of public sector information and a detailed cost apportionment exercise to assign costs to individual datasets.**

Access barriers

The evidence reviewed suggests a number of access barriers to fully maximising the value of public sector information:

- **difficulties around finding where public sector information is located;**
- **a lack of skills and understanding to fully exploit public sector information;**
- **the format and reliability of public sector information; and**
- **a reluctance to use and share public sector information.**

Public sector information fragmentation

Fragmentation in the supply of public sector information continues to be a problem for many consumers, even following the establishment of a number of data portals. A sample review of websites done as part of this study has found that too often it is difficult to locate a clear point of contact, or establish who has ownership of and responsibility for a particular dataset. Even on data.gov.uk, where contact details are provided, these are often generic enquires email addresses rather than named contact email addresses.

In contrast, this report has received feedback that when consumers have found the relevant point of contact in a PSIH their experience has been very positive with a productive dialogue being established. Indeed, some individual PSIHs have established clear procedures for consumers seeking to raise queries or challenge restrictions, although this is not yet happening in all areas of the public sector. Dialogue can also take place between users: data portals such as data.gov.uk and the London Datastore have busy forums and request pages which encourage an active dialogue between information users and re-users and information holders.

Public sector information fragmentation can, at best, raise transaction costs from dealing with public sector information and, at worst, deter users and re-users from using public sector information altogether, or make this impossible to achieve. **Reducing fragmentation across the**

public sector could save up over £50 million per annum in terms of reduced transaction costs and time saved for data specialists⁴⁵.

However, it should be noted that the existing fragmentation and opacity of some public sector information datasets has **created market opportunities for some intermediaries to develop new products** that aggregate datasets and present it in innovative ways. Thus, PSIHs need to consider who is using their data and how, and whether there is a risk of crowding out innovation if they intervene to reduce fragmentation in such a way that duplicates what the private sector can provide.

Insights:

- **Improvements are ongoing to reduce the time taken to locate public sector information datasets. Reducing this can help reduce transaction costs and the overall cost of doing business.**

Data scientists and the skills gap

A number of studies have recently been published contending that advanced economies face a skills gaps in so-called '**data scientists**'. Although statisticians and experts on quantitative analysis have long existed, data scientists differ from these existing professions in a number of important ways. As well as being able to work with large volumes of structured and unstructured data, they are able to translate these analyses into policy and commercial-ready insights and effectively communicate them to a range of stakeholders, often using innovative tools and visualisations. Key to this is the ability to "identify rich data sources, join them with other, potentially incomplete data sources, and clean the resulting set." Data scientists will also be conversant with the vocabulary of public sector information and open data and have the skills to create and manipulate large datasets and linked data. In many ways they therefore resemble scientists more closely than traditional data analysts.

There is a fear that a lack of data scientists will reduce the UK's competitive advantage. The evidence received suggests that in the UK:

- there is increasing demand for individuals with a portfolio of skills able to manipulate quantitative data, present it in innovative ways and generate commercial and policy insights from it;
- many of the individuals performing these roles have no specialised training, but rather have learned on the job and / or have a science/computation/mathematics background;
- businesses rarely designate specific 'data scientist' roles; rather, such analyses are done across a combination of professions such as statisticians, economists, researchers, analysts, policy and commercial managers – a dedicated data scientist would embody elements of all these roles;
- certain industries such as pharmaceuticals, financial services, professional services and retail are increasingly dependent on these skills sets and a shortage of them would reduce the UK's international competitive advantage.

Based on ONS figures for 2011, **around 1.5 million workers in the UK, representing around 5 per cent of the active workforce, are employed in job categories that are likely to involve elements of the role of the data scientist**, but which individually may not be termed data scientists. The average annual median wage of these workers was over £36,000 in 2011 which is higher to the national annual median wage of £26,000.

⁴⁵ Based on an assigned average value of time spent by individuals in occupations that regularly use public sector information. See main report and appendices for details of calculations.

Economic theory suggests that a skills shortage will manifest itself in the form of large wage differentials between ‘data scientists’ and other comparable professionals. A recent report found an observable pay premium for ‘big data’ staff in 2012, with salaries around 20 per cent higher than those for IT staff as a whole.⁴⁶ This may persist due to lags between training and entering or re-entering the workplace, but economic theory suggests it will eventually dissipate in the long-run as supply increases to meet demand and is able to exploit the full value of public sector information. While evidence⁴⁷ suggests there are a number of sector initiatives to reduce the talent learning curve and establish precedence, these initiatives will take time to work through the system with **the interim consequence being that some value of public sector information may remain locked up**. The general scarcity and increasing competition for these skilled workers from the private sector makes it harder to construct the infrastructure for world class public sector information. A shortage of data scientists also hinders efforts to scale-up public sector information data analytics.

Within the public sector, concerns have been highlighted over **a lack of skills and familiarity to work effectively with data**. These concerns should not be overstated as public sector officials have a long history of using public sector information to inform policymaking without having dedicated data scientists. What the concerns appear to be directed at are cultural biases against using public sector information from outside home PSIHS, as well as having the necessary skills to combine and manipulate Big Data and Linked Data. One stakeholder has observed that *“data [owners] do not fully appreciate the power and potential of open data. There may also be cultural resistance to change: for instance, through not trusting or being able to exploit new, untried and untested third-party datasets in analysis and policy advice.”*

Where public sector employees do not have a numerate background, or are not accustomed to working with large datasets, they could benefit from increased training and support in this area. However, this research reveals understandable concerns over resources at a time when budgets are under considerable pressure. The main report outlines a number of cost-effective routes to build the public sector skills base, which include taking massively open online courses (MOOCS), creating incentives to use public sector information as part of day-to-day business, and convening the public and private sectors to work together to explore public sector information.

Insights:

- **While there may be gaps in the supply of ‘data scientists’, economic theory suggests that in the medium- to long-term, the number of data scientists will increase, filling the supply gap and reducing the current wage premium.**
- **However, in the short-term, this may mean public sector information is left under-exploited and associated value remains locked out. The general scarcity and increasing competition for these skilled workers can make it harder to construct the infrastructure for world class public sector information and scale up efforts to exploit its value.**
- **There are some low-cost solutions that can be explored in order to quickly improve the skills base to be able to effectively manipulate and extract value from public sector information.**

Format and reliability

The release of data in an unfinished form raises concerns for the public sector too because of concerns that the public may be provided with data that is inaccurate and potentially misleading, which can have negative consequences.

⁴⁶See e-skills UK, ‘*Big Data Analytics: an assessment of demand for labour and skills, 2012-2017*’ (January 2013)

⁴⁷ See Deloitte Tech Trends 2013, available at www.deloitte.com/view/en_GB/uk/services/consulting/technology/technology-trends/abbffbfdad4ac310VgnVCM3000003456f70aRCRD.htm

With respect to the format of public sector information, its importance may vary between customers. For casual consumers, it is clearly of the utmost importance, and they may most value format. In contrast, professional users may be more concerned with consistency of service and data and can accommodate changes to format. Equally, intermediaries may positively value low quality data as they can provide a service to improve the quality and format of public sector information for wider consumption.

Evidence as to how significant a barrier this currently forms to the use and re-use of public sector information is mixed. In general, there seems to be steady improvement in all the areas of format, although there remains work to be done – especially with regard to ensuring consistency in format and upgrading the star rating of datasets. Progress is also being made with data being updated more consistently. Examples of this include data released by the Met Office, and transport data released directly by Transport for London as well as through data.gov.uk. The Open Standards Principles⁴⁸, the Standards Hub⁴⁹, the Open Standards Board and the public sector will be crowdsourcing, researching and implementing data standards for Government IT systems. The user challenges it will focus on are likely to cover standards relating to formats and meta-data that should help to provide data on reliability. Some of these open data standards may be made compulsory for central government use.

However, a minority of stakeholders have complained that datasets released by different PSIHS are not always easy to combine and work across, because of variations either in the content or the format. This is a particular issue with data produced by local authorities, with each local authority often adopting their own standards and procedures. This report recognises that there are on-going efforts, such as e-PIMS, to tackle this issue and secure a greater degree of standardisation across PSIHS.

Further, there appears to be scope for improvement is greater certainty and clarity over the publication schedule of public sector information datasets and what users can expect from PSIHS. In particular, stakeholders have noted the value in having a cover sheet setting out the limitations of the data, its release schedule, explaining outliers and providing links to previous analyses (i.e. greater use of **metadata**⁵⁰).

Insights:

- **Improvements continue to be made on the quality, format and consistency of public sector information.**
- **There is scope for improvement, especially around the greater provision of metadata.**

Summary of key research questions answered

Further details behind each of the key findings, addressing the Data Strategy Board's overall research questions are summarised in Figure 12 below.

⁴⁸ See www.gov.uk/government/news/government-bodies-must-comply-with-open-standards-principles

⁴⁹ See <http://standards.data.gov.uk/>

⁵⁰ That is, broadly speaking, data about data.

Figure 12: Key research questions answered

Theme	Findings
Definitions of public sector information and its characteristics	
<ul style="list-style-type: none"> • What are the main market segments for public sector information both existing and emerging? • What is the composition of these markets in terms of companies, their governance and their outputs 	<ul style="list-style-type: none"> • While there is no central figure on the total number of public sector information datasets currently being made available across all parts of the public sector, a review of selected data portals suggests the number could exceed 37,500 from over 750 different publishers with over 2.5 million downloads • The key suppliers of public sector information are public sector bodies. An initial assessment shows that of a total of 781 publishers (as of 27th February 2013 on selected data portals), the top ten suppliers of PSI supplied over half of all datasets published on the website, with the Office of National Statistics, the Department for Communities and Local Government and Health and Social Care Information Centre the three largest suppliers by number of datasets. • Customers of public sector information including the private sector, civil society, individuals and government itself. Business models include developing apps that use/re-use public sector information, data enrichers, data enablers and data marketplaces. • Analysis suggests the construction; real estate; financial and insurance; public sector; and arts, entertainment and recreation sectors are some of the largest users of public sector information (and public sector open data). • As well as being used for commercial and policy purposes, public sector information can be used for scientific research, data journalism and holding public service providers to account. • Based on ONS figures for 2011, it is estimated that around 1.5 million workers in the UK (5 per cent of the active workforce) are employed in jobs that are likely to involve direct exposure to public sector information and big data. Analysis of ONS data of jobs that involve 'data science' suggests an average annual median wage of over £36,000 compared to the national annual median wage of £26,000 in 2011. • The extent to which public sector information is a key input into companies' products and services will depend on the company in question, with smaller companies likely to be more reliant on it than larger companies, with their products and services more likely to use public sector information as the critical input. • Public sector information acts as an input as either the main or supplementary data point for complex algorithms and analyses used in products and services, as a source of insights (perhaps summarised as data dashboards or visualisations) for business or policy decisions, as API feeds for apps and so forth.
<ul style="list-style-type: none"> • What are the characteristics of the markets for public sector information, including secondary markets? 	<ul style="list-style-type: none"> • The supply side of the market is mostly populated with public bodies supplying public sector information – there are few private sector substitutes.

Theme	Findings
	<ul style="list-style-type: none"> • Broadly speaking, public sector information customers on the demand side of the market can be split into seven archetypes: larger data companies, SMEs creating apps, SMEs creating efficiency solutions, not-for-profit organisations, individuals, the public sector and other non-data specialist companies. These companies vary in size (from sole traders to larger multi-national companies), turnover (those having established, profitable businesses and those who have yet to report a profit), the dependency on public sector information and the types of public sector information used. • The demand side of the market for public sector information includes a range of different business models including: data marketplaces/infomediaries, apps-based re-use, data enrichment and data enablers. • Downstream, insights and inputs from public sector information can be found in a range of products and services. Products and services range from credit scoring, economic forecasts, weather apps, transport apps, navigation systems and other mapping tools, research reports, tools to assist choice (e.g. in education, health or housing) etc. Downstream, public sector information can also be used to improve transparency and improve public services and hold providers to account.
<ul style="list-style-type: none"> • What do past studies tell us about the markets and how they function? 	<ul style="list-style-type: none"> • There is very little previous literature on the functioning of public sector information markets, possibly due to data limitations around the supply of datasets. The literature has instead focussed on delineating links between public sector information and growth and welfare.
<ul style="list-style-type: none"> • What does the current evidence tell us about the likely evolution of these markets? 	<ul style="list-style-type: none"> • The available evidence suggests that as better quality public sector information becomes available and accessible, customers will be able to use it in increasingly sophisticated ways to exploit its value and drive growth outcomes. • The literature suggests the provision of more and better public sector information can raise sales (a 5.7 per cent increase in sales in restaurants displaying 'good' scores), improvements in public health, a productivity boost (around 0.6 per cent of GDP in the case of New Zealand geo-spatial data) and faster growth (firms growing 15 per cent faster in countries where certain data was either free or priced at marginal cost) – see Appendix 3 for more details. These figures are based on particular examples and are not directly transferable to the UK. • Adapting the McKinsey analysis used by Policy Exchange, improvements in efficiency between 1 and 5 per cent caused by better use of public sector information can lead to annual savings of between £1 and 8 billion nationally and around £70 million in local government (on a smaller savings ratio). • Benefits are likely to come from further release of more and better data, but the release of swathes of data in and of itself will not guarantee value, it is how society as a whole innovates with it that counts.
How public sector information is used and re-used inside and outside of government	
<ul style="list-style-type: none"> • What are the different types of public sector information and can we estimate the value that they 	<ul style="list-style-type: none"> • Public sector information differs by its availability to the public, its format, its complexity and speed of replacement, its content, its channel of distribution and the costs of collection, generation,

Theme	Findings
<p>hold for government and the economy?</p>	<p>maintenance and dissemination.</p> <ul style="list-style-type: none"> • This report has estimated that the narrow value of public sector information to consumers, businesses and the public sector in 2011/12 was approximately £1.8 billion (2011 prices). This is a mid-point estimate, with the sensitivity analyses giving a range between £1.2 billion and £2.2 billion • However, the use and re-use of public sector information has much larger downstream impacts affecting all areas of society beyond its direct customers. Case studies examining the downstream impacts created by different organisations using and re-using public sector information have shown the monetary value of some of these benefits can be in the order of millions, if not billions. • For example, publishing data on adult cardiac surgery has estimated to reduce mortality rates, which in turn has an economic value in excess of £400 million per annum and using live data from Transport for London in apps can save users time to the economic value of between £15 million and £58 million per annum. • While an aggregate figure on the social value of public sector information is difficult to reach, on the basis of conservative assumptions, the figure could be in excess of £5 billion in 2011/12 (2011 prices). This figure is likely to increase as public sector information is used more widely and in more impactful ways. • This yields a total economic and social value of the market for public sector information of between £6.2 billion and £7.2 billion in 2011/12.
<ul style="list-style-type: none"> • What does the evidence tell us about the size and potential market for public sector information? Is there reliable data available? 	<ul style="list-style-type: none"> • While a complete dataset on the number of public sector information datasets is currently not collected, a preliminary estimated (based on a sample of data portals) suggests the volume of public sector information datasets exceeds 37,500 – but this is likely to be an underestimate. • As discussed above, the narrow economic value of the market can be estimated to be between £1.2 billion and £2.2 billion in 2011/12. • The total economic and social value of the market can be estimated to be between £6.2 billion and £7.2 billion in 2011/12. • There is currently insufficient reliable data to be able to make estimates on the size of the potential market for public sector information in the future. However, the increased availability of better and more public sector information, more linked data and better tools and techniques to exploit it mean the value of the market is likely to grow.
<ul style="list-style-type: none"> • What are the gaps in the evidence base and what can we do in the short term to address them? 	<ul style="list-style-type: none"> • The key gaps in the evidence base which, if filled, could lead to more precise estimates on the size, include: (i) the total number of public sector information datasets collected across the public sector; (ii) the individual cost elements behind each dataset and (iii) how public sector information datasets are being used and re-used by customers and others downstream. • In the short term, these gaps could be addressed by (i) a more detailed survey of public bodies to understand what datasets are being held and disseminated; (ii) a survey and consultation of

Theme	Findings
	<p>customers of pay-for public sector information; (iii) a wider survey of users of open government data or some form of limited tracking of how this information is re-used.</p> <ul style="list-style-type: none"> One route to increasing the amount of data on the use and re-use of public sector information would be to foster a climate of greater openness, taking into account commercial sensitivities around how personal sector information is used/re-used. For example, one way of collaborating and fostering more openness in future, providing it can be carried out in a non-too-onerous manner, might be providing data free to third parties on the pre-condition that the information flow becomes a 'two-way street' for Government and policymakers, e.g. <i>"you can have our data, but we'd like to know how it is being used and re-used to give us insight, benefit us and in turn UK society"</i>.
<ul style="list-style-type: none"> Can we assess the current and future for the; UK market, EU market, Global market? 	<ul style="list-style-type: none"> This report has assessed the current market for public sector information in the UK and considered how it might evolve. Many of these insights will be transferable to other markets and, subject to overcoming any data limitations, it would be possible to assess the current and future potential of public sector information markets overseas. However, we note the potential issues arising from a straight read-across of outcomes from one jurisdiction to another. As well publicised examples, public sector information value creation in Denmark and New Zealand are widely cited as a reason for releasing more data. The UK is a larger and more complex economy and data landscape, and as such the impacts of similar policy-changes might be quite different.
<ul style="list-style-type: none"> What indicators should we use to measure our success in widening access to public sector information? 	<ul style="list-style-type: none"> Some organisations are already developing Key Performance Indicators around the quality and format of data being made available. However, success is ultimately determined by the ability of public sector information customers to be able to effectively and efficiently access and exploit datasets as much as would be reasonably expected. Metrics could be created to measure progress against the barriers identified in this report.
<p>Challenges to fully exploiting the value of public sector information, including issues around competitiveness, funding and regulation</p>	
<ul style="list-style-type: none"> What are the strengths and weaknesses of the UK market? 	<ul style="list-style-type: none"> The strengths of the UK market relate to the commitment by government to release more and more public sector information and an emerging eco-system of companies and individuals able to exploit it. The report has identified a series of barriers around legislation, economics and access which may hinder the growth of the UK market.
<ul style="list-style-type: none"> What are the key issues affecting the data market? 	<ul style="list-style-type: none"> The report has identified an issue around consumers' and suppliers' perceptions on the availability of certain public sector information datasets. The report has also noted on-going complexities around charging for certain datasets – the paucity of data on the total stock of public sector information being held/collected, the different cost components behind individual datasets and how datasets are being used and re-used by customers and the wider public, makes it difficult to make decisions around the

Theme	Findings
	<p>provision of public sector information.</p> <ul style="list-style-type: none"> This lack of data (see above) is preventing accurate cost-benefit analyses and identifying where scarce resources should be allocated to maximise the benefit from public sector information.
<ul style="list-style-type: none"> What are the medium and longer term trends in the data market? 	<ul style="list-style-type: none"> It seems reasonable to hypothesise that the larger part of the value to be generated from public sector information lies in the future. Much of this value may come from combining public sector information with information from other sources (e.g. held by the private sector), such as details of consumer spending habits held by supermarkets and other retail firms, or details of domestic and commercial properties. Further benefits are likely to emerge for private individuals, business and other organisations as methods of exploiting public sector information are developed and improved. More effective ways of data exploitation can include: <ul style="list-style-type: none"> greater system-wide analysis through more sharing and combining datasets together to consider policy and business issues from a range of non-traditional perspectives; unlocking the potential from linked data: and better integrating; and greater data-mashing with private sector and individual datasets. In addition, there are likely to be benefits from wholly new uses of public sector information. Numerous additional possibilities for achieving such savings and improving policy may become apparent as more information is made available and expertise in its use increases. The main report summarises some examples of 'blue skies' thinking around public sector information.
<ul style="list-style-type: none"> How far does regulatory regime enable the market and how far does it constrain it and in what ways? 	<ul style="list-style-type: none"> Existing regulations and legislation governing public sector information do not appear to be acting as actual barriers to realising the full potential value of public sector information. However, the manner in which some of these regulations and legislation are interpreted can lead to overly risk averse behaviour that can create barriers.
<ul style="list-style-type: none"> What licensing arrangements are currently in place for PSI and how do they enable open data? 	<ul style="list-style-type: none"> The Open Government Licence is widely regarded by stakeholders as an effective means of removing barriers to exploiting the value of public sector information. However, this is not always used and does not cover all public sector bodies.
<ul style="list-style-type: none"> How does the cost of public sector information affect the data market? 	<ul style="list-style-type: none"> At the broadest level, charging reduces the use of public sector information, but it can also act as a signal of quality for (prospective) users. The issue of charging for public sector information datasets and their funding models is complex. While significant progress has been made by a number of public sector bodies in this area in simplifying charging structures and making more public sector information available as open data, there remains a perception that barriers exist, with the cost of datasets deterring use. There is a lack of data (see above) to definitively conduct a cost-benefit analysis across different funding models. This is an area for further analysis via detailed analysis of customer bases, costs and

Theme	Findings
	how data is used.
<ul style="list-style-type: none"> How clear are the standards that apply to public sector information and how do they affect the market? 	<ul style="list-style-type: none"> Improvements continue to be made on public sector information datasets' format and consistency, with initiatives such as the Open Standards Principles and APPSI definition likely to contribute to this – leading to common understanding and expectations around public sector information.
<ul style="list-style-type: none"> How can we categorise the innovative potential of the market and how can we unlock that potential? 	<ul style="list-style-type: none"> Given the current trajectory of increasing volumes of public sector information being made available in ever more accessible formats, it seems reasonable to hypothesise that the larger part of the value to be generated from public sector information lies in the future. Much of this value may come from combining public sector information with information from other sources, such as details of consumer spending habits held by supermarkets and other retail firms, or details of domestic and commercial properties. Many of these future uses may be extensions of the ways in which public sector information is already being used. For example, a recent study calculated that the benefit to the local government sector from the use of geospatial data was £232m over five years, through a variety of measures such as more efficient routes for waste collection. It is to be expected that such uses will become increasingly widespread throughout the public sector at both a local and central level, as more information becomes available, methods of harnessing this information are improved, and best practice becomes more widely adopted, thus increasing the value derived from public sector information. Further benefits are likely to emerge for private individuals, business and other organisations as methods of exploiting public sector information are developed and improved. In addition, there are likely to be benefits from wholly new uses of public sector information. Numerous additional possibilities for achieving such savings and improve policy may become apparent as more information is made available and expertise in its use increases. In addition, there are likely to be more 'blue sky' opportunities, such as those identified by the Government Office for Science. These are explored further in Chapter 5, but include improving management and resilience of food supplies, tackling obesity, and detecting and identifying infectious diseases. These require combining public sector information with other sources of information rapidly and making the resulting insights available to decision makers in a timely fashion to allow prompt responses to crises and informed policy decisions. Addressing the barriers identified in this study will certainly work to unlock this potential, but it is hard to foresee specifically where innovation might take place in the UK. Often innovation takes place in areas which are hard to predict, with first-movers benefitting accordingly. Quite possibly, the next data innovators are already en-route to innovation on the back of public sector information.

1. Introduction and approach

This introductory chapter sets out the scope of this report, sets out its approach and provides an outline of its contents.

Project scope and objectives

The Department for Business, Innovation and Skills on behalf of Stephan Shakespeare has commissioned Deloitte⁵¹ to undertake an independent assessment of the market for public sector information, in order to establish a robust evidence base on its value and to highlight the policy implications flowing from an examination of how PSI could be utilised further. The Deloitte report forms the evidence base for the Shakespeare Review on public sector information.

When published, the Shakespeare Review will consider the full breadth of the public sector information market, both current and future. It will consider how the private sector, civil society and general public uses and re-uses public sector information, as well as the potential benefits for how the public sector can use and re-use its own data.⁵²

The key elements of the Deloitte market assessment include:

- a **rapid evidence review** to inform the definition of public sector information, users and re-users and what the market looks like and could look like in the future;
- a **desk-based data collection exercise** setting out what the current and latent market for public sector information looks like;
- the construction of an **analytical framework tracing** based on the literature charting the routes of impact publication sector information can have;
- an **'as is' market analysis of public sector information** in the UK, taking into account the suppliers and competitors, consumers, nature of competition, value chain, regulatory landscape, uses and re-uses of public sector information and identifying instances of market failure;
- **three case studies** examining how public sector information is used and re-used in the private and public sectors. These include the health sector, use and re-use of public sector information within government, and a 'deep dive' into the use of public sector information in the transport sector;
- **high level quantitative analyses**, where the data exist, assigning monetary estimates to the current and potential value of public sector information; and
- identifying **policy implications** on the different challenges to maximising the full potential of public sector information.

⁵¹ Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited ("DTTL"), a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.co.uk/about for a detailed description of the legal structure of DTTL and its member firms.

Deloitte MCS Limited is a subsidiary of Deloitte LLP, the United Kingdom member firm of DTTL.

This publication has been written in general terms and therefore cannot be relied on to cover specific situations; application of the principles set out will depend upon the particular circumstances involved and Deloitte recommends that you obtain professional advice before acting or refraining from acting on any of the contents of this publication. Deloitte MCS Limited would be pleased to advise readers on how to apply the principles set out in this publication to their specific circumstances. Deloitte MCS Limited accepts no duty of care or liability for any loss occasioned to any person acting or refraining from action as a result of any material in this publication.

⁵² BIS, Independent Public Sector Information Review: Terms of Reference

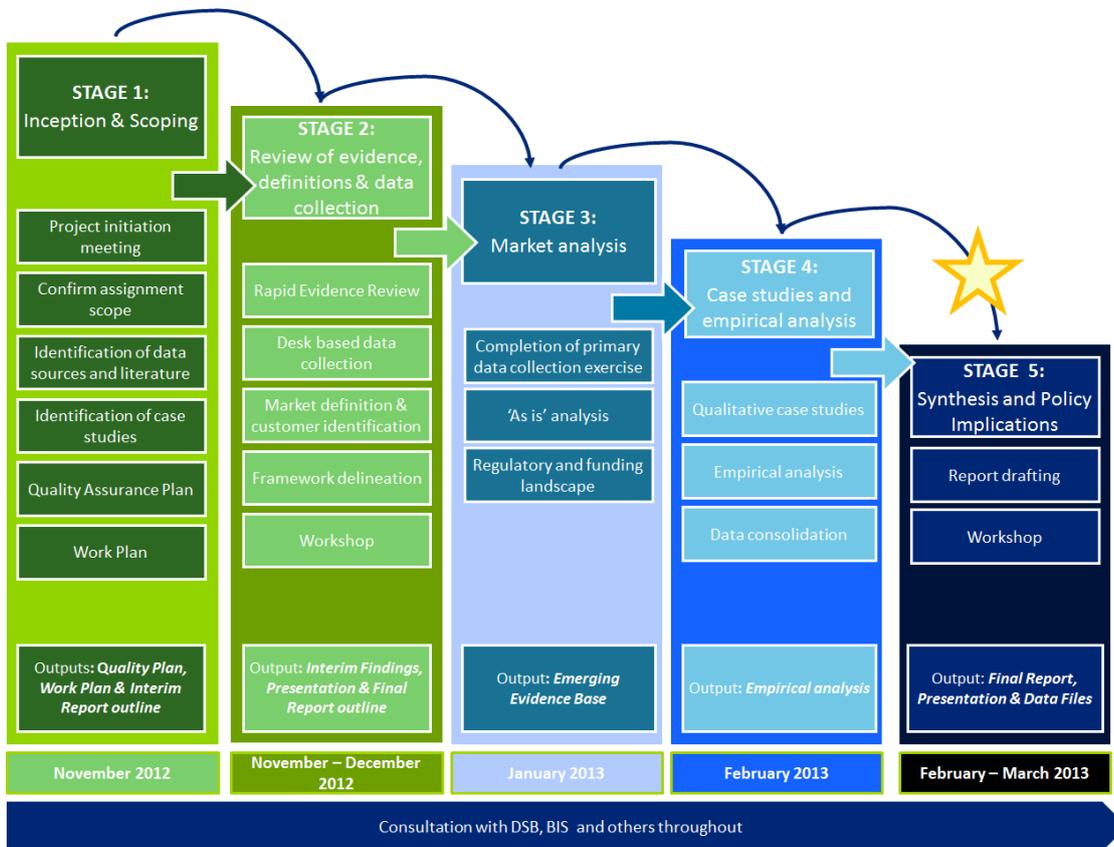
This report takes as its starting point that the public sector is equivalent to Category O of the Standard Industrial Classification (SIC 2007) nomenclature – Public Administration and Defence; Compulsory Social Security. However, when appropriate, public bodies outside this category (e.g. health, transport and education) are considered and included.

The report covers all levels of government: UK, devolved, regional and local. Where appropriate, it also considers the international dimension.

Approach

This report has adopted Deloitte’s standard five-stage methodological approach to develop the evidence base.

Figure 1.1: Deloitte methodology summary



Source: Deloitte

This final report concludes the research.

Data sources

There is a paucity of data around public sector information in the UK making it difficult to generate accurate figures on the size, value and potential of public sector information. This notwithstanding, data on public sector information and other proxies has been collected from the following sources:

- **web-based PSI portals** such as data.gov.uk and local government data stores and other statistical agencies such as National Statistics;
- **existing literature** on public sector information including the Open Data White Paper, Deloitte research, studies by academics and other public bodies such as the Office of Fair Trading (OFT);

- **conversations with public sector information stakeholders** such as the Open Data User Group, the Trading Funds, regulatory bodies, government departments, business users and re-users and consumers;
- **annual reports** of public sector information suppliers such as Trading Funds and departmental open data strategies;
- **internal Deloitte research** that draws on client experience and the use and re-use of public sector information;
- **international comparisons** such as from the EU, Australia and the USA.

As set out in Figure 1.1 this data was primarily collected between November and December 2012⁵³. The approach to data collection has been to focus on gaining an overview of the public sector information market at the broadest level, covering the widest possible range of issues and underlying trends, noting that primary research was excluded from the project's terms of reference.

Accordingly, all model results and data analysis should be read in conjunction with the relevant caveats.

Structure of this report

This evidence base contains all of Deloitte's analysis for this research project and is structured as follows:

- **Chapter 1 – Introduction and approach:** outlines the scope of this project and Deloitte's broad approach to the research;
- **Chapter 2 – Public sector information definitions and market definition:** considers the different definitions used for public sector information and goes on to discuss how public sector information varies across a number of dimensions;
- **Chapter 3 – The supply and demand sides of the public sector information market:** provides an overview of PSIHs and some statistics on the supply of datasets, before moving to examine how the datasets are used and re-used by businesses, individuals, community groups and the public sector;
- **Chapter 4 – The regulatory landscape for public sector information:** contains an overview of the different regulations, guidelines and pieces of legislation that govern the operation of the public sector information market;
- **Chapter 5 – The current value of the public sector information market:** presents the report's calculations of the economic value of the market currently, as well as a series of case studies that highlight the wider social value of public sector information;
- **Chapter 6 – Barriers to maximising the value of public sector information and their policy implications:** discusses the identified challenges to maximising the value of public sector information in the UK; and
- **Chapter 7 – Conclusion:** contains some closing thoughts on the subject.

The report's appendices provide further methodological and analytical background. These are:

- **Appendix 1 – glossary:** setting out the terms used in this research
- **Appendix 2 – acronyms** used in this research
- **Appendix 3 – literature review:** considers previous research in this area;

⁵³ Although data continued to be received by the project team until March 2013.

- **Appendix 4 – further statistics:** in particular on the Trading Funds that make up the Public Data Group;
- **Appendix 5 – empirical methodology:** provides a full description of the assumptions and approaches underpinning quantitative estimates;
- **Appendix 6 – transport detailed case study:** contains additional details on the transport centre case study methodology;
- **Appendix 7 – other case studies:** additional details on various case studies conducted for this research;
- **Appendix 8 – results from an informal Government Officials survey;** and
- **Appendix 9 – bibliography.**

2. Public sector information definitions and market definition

A key foundation stage in any market analysis is to establish which products and services constitute the market under discussion. An effective definition of public sector information and the parameters of the market provide a strong framework for further analysis.

Definition of public sector information

Since their establishment, public bodies across the world have generated and retained a wealth of data and information. The scale and range of this information can be overwhelming, as the following examples demonstrate⁵⁴:

- HM Revenue and Customs interacts with over 40 million customers;
- UK departments, agencies and other public bodies procure over £243 billion worth of goods and services;
- around six million people work in the public sector, each of whom has a record on performance, sickness absences, skills and years of service; and
- each year nearly 700,000 students make 2.7 million applications to university.

This data and information is typically collectively referred to as public sector information. However, definitions of public sector information can, and do, vary. A review of the literature suggests that there are a number of definitions in current use globally.

Figure 2.1: selected definitions of public sector information

Definition	Source
Information and data collected, reproduced and disseminated by the public sector covering many areas of activity, such as social, economic, geographical, weather, tourist, business, patent and educational information.	<i>Re-use of Public Sector Information Regulations 2005 (SI 2005/1515), Reg.4(2)</i> ⁵⁵
Information, including information products and services, generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for the Government or public institution, taking into account the relevant legal requirements and restrictions.	OECD, <i>Recommendation of the Council for Enhanced Access and more Effective Use of Public Sector Information [C(2008)36]</i> ⁵⁶
Information, data or content collected by and/or held by a public body. The information may or may not be	OFT, <i>The Commercial Use of Public Information (CUPI), [OFT861]</i> ⁵⁷

⁵⁴ Taken from *Deloitte Analytics – Insight on tap: improving public services through analytics*. Available at: www.deloitte.com/view/en_GB/uk/industries/government-public-sector/67360f23824e0310VgnVCM2000001b56f00aRCRD.htm

⁵⁵ Available at www.legislation.gov.uk/uk/si/2005/1515/made

⁵⁶ Available at www.oecd.org/internet/interneteconomy/40826024.pdf

⁵⁷ Available at www.ofc.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861

Definition	Source
Crown copyright information.	
Material with the essential purpose of providing Government information to the public. Material with the essential purpose of artistic expression is unlikely to be treated as PSI for the purpose of the policy.	Office of the Australian Information Commission, <i>Principles on open public sector information: Report on review and development of principles</i> , (May 2011) ⁵⁸
Information held by a public sector organisation, for example a government department or, more generally, any entity which is majority owned and/or controlled by government.	Pollock, R, <i>The Economics of Public Sector Information</i> (2008) ⁵⁹

Note: in some cases the original definition has been summarised for purposes of brevity

As Figure 2.1 demonstrates, definitions of public sector information fall across a continuum ranging from all information and data generated, held and disseminated by the public sector (content agnostic) to the definition of public sector information explicitly covering certain types of data and information (content specific). Importantly, all definitions are neutral with respect to the format of public sector, whether it is qualitative or quantitative, structured or unstructured and whether or not it is subject to user fees⁶⁰.

For the purposes of this report, public sector information is defined as covering the wide range of information that public sector bodies collect, produce, reproduce and disseminate in many areas of activity while accomplishing their public tasks. This is consistent with the recently published glossary of terms by the Advisory Panel on Public Sector Information (APPSI)⁶¹.

This report acknowledges that this definition is not entirely uncontested. In particular, there is debate over when particular public sector information datasets become value-added services rather than information and data. For example, Deloitte has heard views that the above definition of public sector information is too narrow and fails to capture intangible advice given out to customers that draws upon public sector information, but of which no record is kept⁶². However, these reservations notwithstanding, the above remains a reasonable working definition for the purposes of this evidence base and allows some parameters to be set around the analysis.

It is important to note that for the purposes of this report, as is conventional practice, the term information (output of such processes that summarise, interpret or otherwise represent data to convey meaning) is taken to include data unless otherwise specified.

Definition of the public sector

A corollary of defining public sector information is the need to define what is meant by the public sector itself. At the most simplistic level, the public sector can be defined in reference to the Standard Industrial Classification (SIC) codes; where the public sector is Category O of SIC(2007) covering 'public administration and defence; compulsory social security'⁶³. While this is a useful starting point it does not capture other activities that might reasonably be considered as part of the public sector such as education, health and transport services as well the public funding of other activities such as the arts and other cultural bodies.

⁵⁸ Available at:

www.oaic.gov.au/publications/reports/Principles_open_public_sector_info_report_may2011.html#_Toc293927686

⁵⁹ Available at: www.rufuspollock.org/economics/papers/economics_of_psi.pdf

⁶⁰ The definitions also capture case data collected as required by legislation, e.g. highways data

⁶¹ Available at: www.nationalarchives.gov.uk/appsi/appsi-glossary-a-z.htm#apps-p

⁶² For example, interpretation of datasets given over the phone.

⁶³ See www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/standard-industrial-classification/index.html for more details on SIC codes

There is some debate as to whether data and information produced by cultural centres, such as museums and libraries, should be considered public sector information, especially in cases where these receive public subsidy. For the purposes of this report cultural institutions are considered to lie within the public sector, unless otherwise specified⁶⁴. For similar reasons, while not part of the public sector, private and third sector contractors providing public services may be considered as producers of public sector information. Universities and schools although not within the scope of the PSI Directive are significant in public sector terms and are included in the definition of the public sector for the purposes of this report.

Thus, for the purposes of this report, the public sector is taken to refer to include (but not necessarily be limited to):

- national and devolved government (Ministerial and non-Ministerial) departments and their executive agencies;
- Non-Departmental Public Bodies;
- the National Health Service;
- the Judicial Service;
- the Armed Forces and Police Service;
- Public Corporations and Trading Funds;
- Independent Panels and Inquiries; and
- Local Authorities.

In reality it is difficult to clearly define the boundaries of the public sector and the above list is not exhaustive, but is deliberately broad to capture the wide range of public sector actors, ranging from the departments of state, to the BBC and the NHS, through to local borough councils.

The nature of public sector information

As defined above, the term public sector information captures a vast range of information and data gathered from diverse sources, for a wide range of purposes, and stored in many different formats, with the single unifying feature that it is collected by public sector bodies. It includes data gathered intentionally, the collection of which may be one of the organisation's main tasks; and information gathered incidentally while performing other functions. Public sector information thus has a number of dimensions, as set out below

Figure 2.2: public sector information dimensions

Public sector information can be categorised along a number of non-competing dimensions:

- whether the public sector information is made available to the general public and, if so, under what conditions (if any);
- the complexity of the dataset in terms of the number of records and variables, whether it is anonymised or aggregated or is quantitative or qualitative (its verbosity);
- how often the dataset is updated or replaced (its velocity);
- whether the public sector information is generated or collected as part of a public body's public task or whether its generation is the result of other activities not related to public sector information;
- the content of the public sector information dataset;
- the electronic or non-electronic format of the dataset;
- the ways in which the public sector information dataset is distributed; and

⁶⁴ The revised EC PSI Directive (see chapter 4) also adds the cultural sector (archives, libraries and museums) within scope.

- the cost of generating/collecting/maintaining/updating the public sector information dataset.

Source: Deloitte analysis

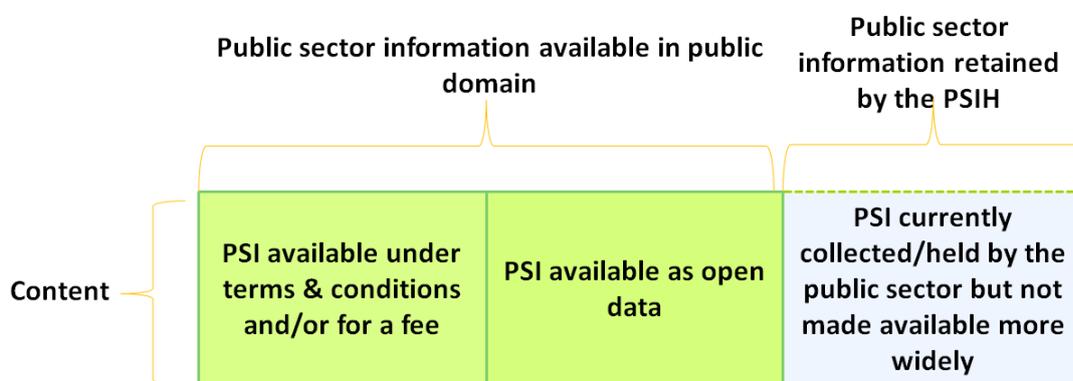
The following section defines these dimensions and explores each in more detail.

The availability of public sector information

It does not automatically follow that public sector information will always be made available to the public or outside the public sector body that originally collected/generated it. Equally, it may not be made immediately following its collection. For example, some data releases are not made public immediately to avoid risking public safety, e.g. some Met Office data, which must be validated to ensure it is correct; this is contrast to certain economic statistics which can be revised over time.

Figure 2.3 below summarises how public sector information can be retained and disseminated.

Figure 2.3: public sector information in the UK



Source: Deloitte analysis

It is important to highlight that public sector information released as open data is a sub-set of the total public sector information market. To recall, public sector open data is data that has been made available free of charge for anyone to use as they wish and meets the criteria of being accessible, in a digital, machine readable format and free of restrictions on use or redistribution. The 'open definition' is "a piece of data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike"⁶⁵. The data portal data.gov.uk contains a wealth of public sector open data⁶⁶ covering a range of topics ranging from data on tariffs⁶⁷ to homelessness statistics⁶⁸. Examples of data only available to the general public under terms and conditions or for a fee include the National Pupil Database and subscriptions to specialist statistics by the Office of National Statistics⁶⁹. Data currently collected by the public sector, but not made widely available outside the public sector, include many national security information datasets.

The Open Government Licence⁷⁰ sets out the conditions under which individuals and users can use and re-use PSI – though not all data and information available under the OGL this meets the open data criteria (i.e. it may not be in a digital machine readable format). This report discusses the regulations around public sector information in more detail in Chapter 4.

⁶⁵ See <http://opendefinition.org/>

⁶⁶ Much public sector information is now classified as open data, but the term open data itself is much broader, covering any data – for example, data provided by private companies, individuals, and not for profit organisations – which is made available free of charge to the public.

⁶⁷ See <http://data.gov.uk/dataset/uk-tariff-codes-2009-2010>

⁶⁸ See http://data.gov.uk/dataset/statutory_homelessness_statistics_england

⁶⁹ See www.nomisweb.co.uk/

⁷⁰ See www.nationalarchives.gov.uk/doc/open-government-licence/

The complexity of public sector information

The level of detail of a public sector information dataset is referred to as its verbosity, with more complex, larger datasets being more verbose. Verbosity may be a function of the number of variables contained in the dataset, the number of records it holds and the time period it covers. Public sector information datasets can either focus on aggregated data or be at an individual unit level – the unit can vary between individuals, social groups, businesses, industries, economic sectors and so forth – the more disaggregated, the more verbose.

The size of public sector information datasets can also vary considerably, with only a small proportion being truly classed as Big Data, although this may not always be reported in the statistics. For example, the Companies House dataset is substantially larger than, say statistics on football banning orders, but both are counted as a single dataset by the official statistics. Equally, the Met Office supplies around 200Mb data per day to data.gov.uk, but this is contained in just three large datasets. The size of individual datasets also has implications as to how it is distributed (see below).

The speed at which public sector information is updated

It is important to distinguish between public sector information that exists in a relatively static form, and public sector information which is continually updated. The frequency with which a public sector information dataset is updated is known as its velocity.

In some cases, public sector information datasets will only be updated infrequently, perhaps annually or, in the case of the Census, every ten years. Typically, updates are made when an additional cohort of data is available and the original dataset is extended to include this new data. Examples of such datasets include information on departmental expenditure, educational attainment of GCSE students, or numbers of welfare claimants – all of which are updated with supplemental data at certain intervals (the time series expands). Other public sector information datasets may also be updated periodically, but these updates replace existing datasets (there is no time series of data) – such as the case of economic forecasts where the latest forecast supersedes the previous one.

Other types of public sector information, however, are updated on a daily, hourly or on an even more frequent basis and the usefulness of the public sector information is directly related to how up-to-date it is. For example, Transport for London provides live travel information meaning that updates on disruptions and delays, as well as bus and tube departure times, are updated in real time. In these instances the need for the data to be regularly refreshed becomes of the utmost importance, as historic data may have relatively little value. The updated dataset further needs to be disseminated to users rapidly and in an appropriate format. This is commonly achieved through an application programming interface, or API, which allows updates to be embedded in a webpage, programme or mobile application. This information can be characterised as having a high velocity.

The distinction between the velocity of different public sector information datasets is significant because the speed at which datasets are updated can impact their end use, methods of dissemination and the level of commitment required from the PSIH. High velocity public sector information is often collected by an organisation dedicated to that purpose, as in the case of the Met Office and weather information. Given the time-sensitive nature of high velocity public sector information, users are likely to demand a higher quality of service and assurance that the public sector information will be up-to-date.

Public sector information as a by-product or purpose-collected

In some cases public sector information is generated in the course of a public sector body's other activities, rather than the information being generated or collected as its core activity. This is known as 'by-product' or 'exhaust' data or information. This applies, for example, to the data contained in the National Pupil Database; Home Office statistics on crime; salary information for employees of public sector bodies; and details of public expenditure. The Department for Education collects data on students in the course of its regular activities but the collection of this data is not its primary

purpose – it arises from, and facilitates, other activities. The same may be said for salary information published by government departments: the data is generated in the course of employing people to undertake the range of functions required of the department.

The fact that this public sector information is a by-product of an organisation's public task does not necessarily imply that its value to users is in any way reduced. Nor does it mean that its collection and dissemination is cost-free. There may be costs incurred in processing, storing and disseminating the public sector information⁷¹. However, the cost of collection is generally thought to be lower than purpose-collected data, since the public sector information is generated by activities which would have taken place regardless of whether or not the public sector information was collected⁷².

In other cases, the collection of public sector information is the purpose of a public sector organisation's activities and even its existence (part of its public service remit). This is the case, for example, with Ordnance Survey. In these cases the public sector information has been identified as desirable in its own right. It is also likely that this public sector information will be highly valued by users, since it has been actively identified as worth the cost of collection.

The distinction between public sector information as a by-product of other activities and public sector information that is purpose-collected is therefore significant, since it suggests differing costs of collection and perhaps also different user profiles and, potentially, value.

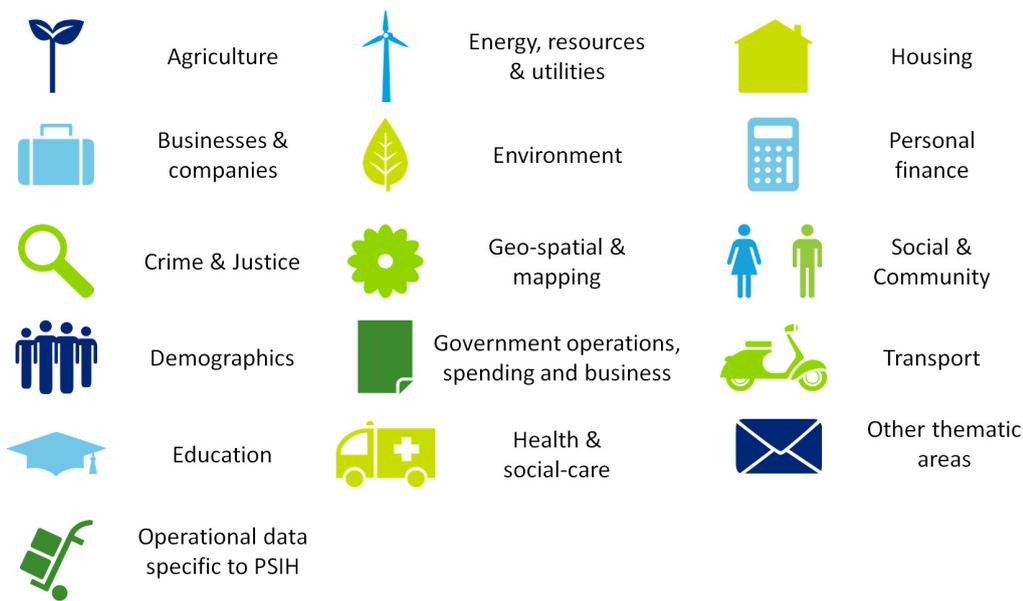
The content of public sector information

Public sector information can also be categorised on the basis of content. These content categories broadly follow the activities of the PSIH in question and cover all aspects of the UK's economic, social and cultural activities. In addition to specific datasets relating to the activities of their staff and agencies, all PSIHs will also be generating public sector information in the form of operational data and information covering the administrative and logistical activities of public sector actors such as budgets, organisational charts, HR statistics and pay scales.

Some examples of public sector information content categories are shown in Figure 2.4.

⁷¹ In some cases, this may be significant if the data needs to be cleansed and re-structured to allow consumers to effectively use and re-use it.

⁷² Costs will also vary according to how often the data is collected.

Figure 2.4: selected public sector information content categories

Source: Deloitte analysis

Where public sector information datasets contain personal information they may be anonymised or non-anonymised – if the latter, they are subject to restrictions on their release.

The format of public sector information

Public sector information can be downloaded in a wide variety of electronic file formats. At the time of writing, there were nearly 50 different formats of electronic files available for download at data.gov.uk (a reasonable proxy for public sector information currently available to the general public, although certainly not a comprehensive source) – a number of which are proprietary formats such as .pdf and .xls (even if the software readers are readily available).

The ten most popular file formats (constituting around 96 per cent of all datasets) were:

Figure 2.5: most popular public sector information file types on data.gov.uk

File type	Number of datasets (27 th February 2013)	Percentage of total (27 th February 2013)
.csv	2423	40%
.xls	1624	27%
.pdf	588	10%
.html	461	8%
.rdf	191	3%
.xml	189	3%
.zip*	123	2%
.wms	69	1%
.doc	59	1%
.txt	52	1%

Source: Deloitte analysis

*Once unzipped, the file could be any of the above other files.

Figure 2.4 refers to the format of public sector information currently available online at the time of writing on data.gov.uk⁷³. However, it is likely that there is public sector information currently retained (or even collected) by the public sector which has yet to be digitised and is not available electronically (such as archive and historical material⁷⁴). This public sector information may or may not be accessible to the general public, but can only be provided in hard copy or cannot be removed from site.

In addition, much public sector information is released in the form of an API⁷⁵. This allows developers to embed a data source in a webpage, programme or mobile application, allowing for updates to the data. Prominent examples of public sector information APIs include Transport for London, which provides live travel updates delivered via a webpage or mobile applications, and the Met Office weather information, which is periodically updated with the latest data.

Star rating of public sector information

Related to the issue of format is the usability of the dataset. There are different ways in which usability can be measured, and they are, to some extent, subjective. One of the more common methods is the Open Data five-star scoring mechanism suggested by Sir Tim Berners-Lee, which is used by data.gov.uk.

Figure 2.6: open data scoring

Score	Description
★	Make your stuff available of the Web (whatever format) under an open licence
★★	Make it available as structured data (e.g. Excel instead of image scan of a table)
★★★	Use non-proprietary formats (e.g. CSV instead of Excel)
★★★★	Use URLs to identify things, so that people can point at your stuff
★★★★★	Link your data to other data to provide content

Source: <http://5stardata.info/>

At the time of writing data.gov.uk has recently begun to provide data on the proportion of datasets attaining each star rating (currently in beta form). These figures indicate that, at the time of writing, nearly a quarter of datasets have attained a three-star rating. Only one per cent of datasets have been given a five-star rating, indicating they are linked data. Over 50 per cent of datasets have attained no stars – though this is due to many datasets not yet being rated⁷⁶ or the link referring to data such as live mapping services (for which a rating has not yet been agreed as they are not strictly data).

⁷³ It does not include other data portals such as legislation.gov.uk.

⁷⁴ Though some archive material, such as that held by The National Archives, can be digitised on demand and electronic copies supplied if a user wishes to pay for the service.

⁷⁵ A specification intended to be used as an interface by software components to communicate with each other. An API may include specifications for routines, data structures, object classes, and variables.

⁷⁶ This may be because the algorithm to award the star rating to datasets is unable to categorise certain datasets effectively.

Figure 2.7: star ratings of datasets on data.gov.uk

Score	Description	Number of datasets (27 th February 2013)	Proportion
TBC	N/A	842	9%
No stars	Unavailable	5117	56%
★	Unstructured data (e.g. PDF)	163	2%
★★	Structured data but proprietary format (e.g. Excel)	826	9%
★★★	Structured data in open format (e.g. CSV)	2004	22%
★★★★	N/A	N/A ⁷⁷	N/A
★★★★★	Linked data - data URIs and linked to other data (e.g. RDF)	107	1%

Source: www.data.gov.uk/data/search

The distribution channels of public sector information

The analysis undertaken for this report suggests that the key distribution route of public sector information by suppliers (or originators) is through websites⁷⁸. The next chapter considers in more detail the different websites from which public sector information is assessed, but broadly speaking the ways in which public sector information is accessed and disseminated in the first instance are:

- through **viewing** the public sector information online on a webpage;
- through **downloading** the public sector information from a webpage;
- through an **API** embedded in an app or webpage;
- via **secure terminal facilities** or **virtual laboratories** to view the data⁷⁹;
- other **published versions** of the work; and
- through **requesting** the public sector information via a webpage, telephone or other written request. The public sector information may be sent via email, an encrypted link, CD or DVD or hard copy.

Often the latter case is imposed for public sector information datasets that are too large to efficiently host online, that contain sensitive personal data or are subject to copyright conditions and which can only be released upon the receipt signing licensing terms and conditions. For example, given the nature of the data and information contained within it, the National Pupil Database can only currently be shared with researchers following the signing of specific terms and conditions. It should be noted that in any case this data can only be shared where there is a statutory gateway to permit this (as is the case with the National Pupil Database).

⁷⁷ Not included in the data

⁷⁸ Though, as discussed above, there is likely to be a wealth of historic public sector information that is not readily available online.

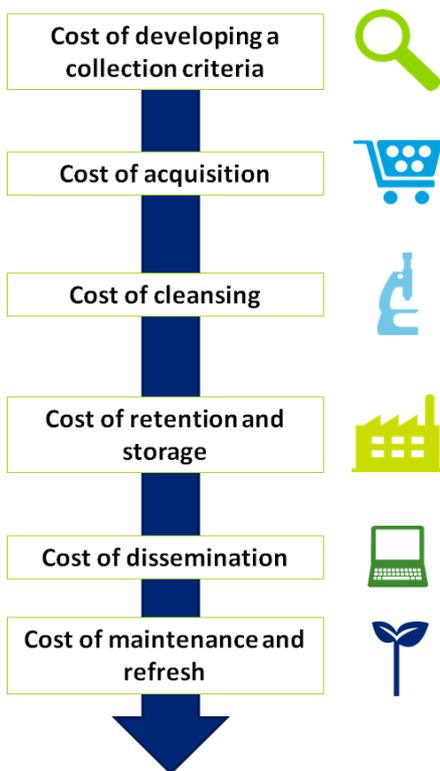
⁷⁹ For example, the ONS allows researchers access to micro-data in secure facilities.

The cost of public sector information

Public sector information differs in terms of cost. For certain types of public sector information the marginal cost of generation/collection may be relatively high (compared to dissemination) if this data is being collected specifically. Examples of the former include the geospatial data collected by Ordnance Survey, which requires use of aircraft and personnel on the ground. On the other hand the data collected by Transport for London arises from its day to day operations, and producing this data requires relatively little investment of resources beyond the infrastructure already in place to support operations, meaning that there is a low marginal cost of generation – although the costs of dissemination may be higher.

The public sector information cost lifecycle components are shown below.

Figure 2.8: public sector information cost lifecycle



Source: Deloitte analysis. Costs of acquisition may also include purchasing third party data that is combined with public sector information

Public sector information can be released to the general public at different parts of the cost lifecycle – PSIHs do not necessarily have to wait until the public sector information has been cleansed (data assured) and formatted into refined data before they more widely disseminate it. The marginal cost at each point of the lifecycle will differ, e.g. the marginal cost of dissemination may be very different to the marginal cost of collection. PSHs can release public sector information prior to cleaning as unrefined or raw data.

How public sector information is funded is discussed in more detail in subsequent chapters. Where public sector information datasets do carry charges these can be in form of one-off fees, ongoing subscription fees or a royalty model.

Having defined public sector information, the next section considers its market definition.

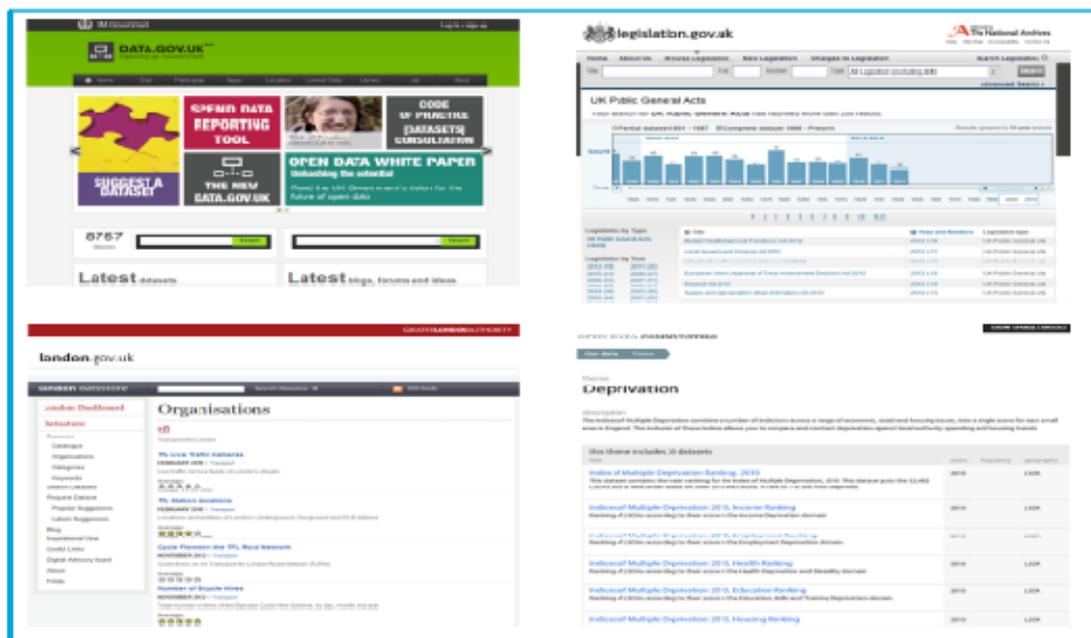
Market definition

Before considering the actors on the supply, intermediary and demand sides of the market, a preliminary analytical step is to define what is meant by the market for public sector information. As

set out by the OFT in its guidance for market definition⁸⁰, the process of defining a market typically begins by establishing the closest substitutes to the product or service in question. These substitute products are the most immediate competitive constraints on the behaviour of the organisations supplying the product or service in question⁸¹. As highlighted in an earlier chapter, in the case of public sector information the market definition exercise is complicated by the fact that different public sector information datasets vary substantially by content, quality, size and format.

As discussed in previously, at a first glance, the overwhelming majority of public sector information is provided by the public sector through PSIHs.

Figure 2.9: examples of public sector information web portals



However, it is not enough to stop at the PSIHs when considering who the suppliers of public sector information are – there may be substitute suppliers outside the public sector. Figure 2.10 highlights a selection of public sector information datasets, their suppliers and potential substitutes.

Figure 2.10: public sector information datasets and substitutes⁸²

Public sector information dataset	Public sector supplier	Potential substitute supplier (if any)
UK mapping data	Ordnance Survey	Google Maps, Open Street Map
Meteorological data such as weather forecasts	Met Office	Various private sector suppliers or overseas suppliers and academic institutes
Economic statistics	National Office of Statistics, Bank of England	None currently

Source: Deloitte analysis

⁸⁰ See Market Definition: understanding competition law (OFT403) available at www.of.gov.uk/shared_of/business_leaflets/ca98_guidelines/of403.pdf

⁸¹ Although some of these substitutes may come with greater restrictions over the use and re-use of public sector information.

⁸² The focus here is on substitutes for the underlying data in the public sector information dataset not the value-added service that may be associated with it (including intermediary services).

What this short comparison shows is that the presence of alternative suppliers varies with content, with some types of content being more readily substitutable: this could include mapping data or meteorological data. Of course, whether Google Maps is an effective substitute for Ordnance Survey will depend on whether consumers regard them of substitutable quality. This can be tested through the so-called hypothetical monopolist test⁸³.

In this case, taking the Met Office as an illustrative example, the question would be whether the Met Office could profitably raise prices for meteorological data by 5-10 per cent above competitive levels. If the answer is yes, then the test is complete and the market for meteorological data solely consists of the Met Office. If the answer is no, then this is due to consumers switching to other substitute products such as rival suppliers of meteorological forecasts (assuming they exist). In this case, the hypothetical monopolist test is repeated, assuming the Met Office also controls overseas suppliers. If it can profitably raise prices by 5-10 per cent then the market is defined; if not, the process is repeated with more substitute suppliers until the market is defined.

The ability of this report to carry out an empirical hypothetical monopolist test has been constrained by a lack of elasticity data on potential public sector information substitutes. Further it is not practical to carry out the test for the thousands of different public sector suppliers of public sector information. Accordingly, as a working market definition, this report broadly defines the market for public sector information as including only public sector suppliers, unless there are strong reasons to include other non-public sector suppliers (this may be the case for certain types of geo-spatial and mapping data).

The geographic market definition is restricted to the UK.

Chapter summary

- For the purposes of this report, public sector information is defined as covering the wide range of information that public sector bodies collect, produce, reproduce and disseminate in many areas of activity while accomplishing their public tasks.
- For the purposes of this report, the public sector is taken to refer to (but not be limited to):
 - national and devolved government (Ministerial and non-Ministerial) departments and their executive agencies;
 - Non-Departmental Public Bodies;
 - the National Health Service;
 - the Judicial Service;
 - the Armed Forces and Police Service;
 - Public Corporations and Trading Funds;
 - Independent Panels and Inquiries;
 - Local Authorities.
- Public sector information varies in terms of:
 - whether it is made available to the general public and, if so, under what conditions (if any);
 - its complexity in terms of the number of records and variables, whether it is anonymised or aggregated or is quantitative or qualitative (its verbosity);
 - how often it is updated or replaced (its velocity);

⁸³ The OFT defines this as a “test that seeks to establish the smallest product group (and geographical area) such that a hypothetical monopolist controlling that product group (in that area) could profitably sustain 'supra competitive' prices, i.e. prices that are at least a small but significant amount above competitive levels. That product group (and area) is usually the relevant market. If, for example, a hypothetical monopolist over a candidate product group could not profitably sustain supra competitive prices, then the candidate product group would be too narrow to be a relevant market. If, on the other hand, a hypothetical monopolist over a subset of a candidate product group could profitably sustain supra competitive prices, then the relevant market would usually be narrower than the candidate product group.” Available at http://www.of.gov.uk/shared_of/business_leaflets/ca98_guidelines/oft403.pdf

- whether it is generated or collected as part of a public body's public task or whether its generation is the result of other activities not related to public sector information;
 - by its content;
 - its format;
 - the ways in which it is distributed; and
 - the cost of generating/collecting/maintaining/updating it.
- This report broadly defines the market for public sector information as including only public sector suppliers unless there are strong reasons to include other non-public sector suppliers. The scope of the analysis is restricted to the UK.

3. The supply and demand sides of the public sector information market

This chapter provides an overview of the supply and demand sides of the public sector information market. It provides statistics around the supply of datasets and profiles the different types of public sector information holders (PSIHs).

The chapter then moves to consider intermediaries and the final consumers on the demand side of the market. It looks at the public sector information value chain, different business models in operation and the ways in which public sector information is currently used and re-used.

Supply side of the market

As highlighted in the preceding chapter, public sector information is collected and/or generated by thousands of organisations across the public sector – so-called Public Sector Information Holders (PSIHs). These PSIHs can be found in UK, devolved, regional and local government, as well as other public sector bodies such as the armed forces, police force, NHS, universities and others named in Chapter 2.

Supply of datasets nationally

When considering the identity of the main suppliers of public sector information it is important to distinguish between PSIHs who hold the most information and PSIHs who make the most information available to the general public. As an example, the Ministry of Defence and the Security Services may potentially hold the largest amounts of public sector information of any PSIH, but they may be one of the smallest suppliers of information to the market (either as open data or under restrictions). Given the lack of robust data as to the overall size of the public sector information market (i.e. public sector open data + public sector information available under restrictions or for a fee + public sector information collected/generated not disseminated to the general public), this report can only seek to make an approximation as to who the largest suppliers of public sector information are in terms of information made available to the public.

With respect to open public sector information⁸⁴, at a national level, much of this public sector information is made available through data portals, specifically data.gov.uk (the single largest source of public sector open datasets). An analysis of this provides a useful, illustrative tool for assessing the most prolific suppliers of public sector information amongst PSIH. An initial assessment shows that of a total of 781 publishers (as of 27th February 2013), the top ten suppliers of PSI supplied over half of all datasets published on the website.

The top ten data.gov.uk publishers are as follows:

⁸⁴ There is less data available on the number of datasets that are not available as open data or under the Open Government Licence (see Chapter 4). However, particular instances of datasets available under licence or for a fee are discussed in this and subsequent chapters and appendices.

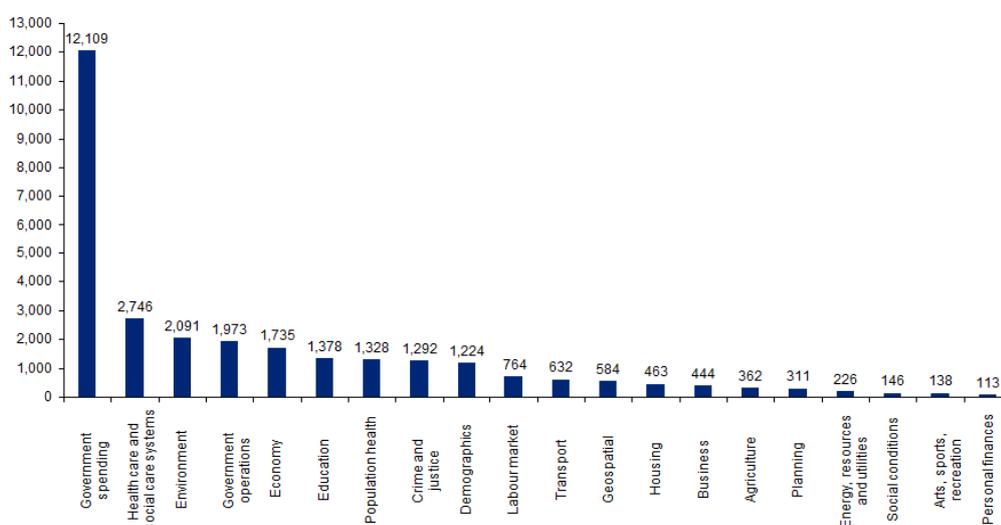
Figure 3.1: PSIHs with the largest number of datasets published on data.gov.uk, February 2013

Supplier	Number of open datasets (as of 27/02/13)	Proportion of all datasets on data.gov.uk
Office for National Statistics	847 ⁸⁵	12%
Department for Communities and Local Government	744	11%
Health and Social Care Information Centre	589	9%
British Geological Survey	379	6%
Department for Environment, Food and Rural Affairs	331	5%
Centre for Ecology & Hydrology	307	5%
Welsh Government	241	4%
Department of Health	240	4%
Department for Children, Schools and Families	227	3%
Home Office	213	3%
Total	4,118	60%

Source: Deloitte analysis of data.gov.uk

The content of the datasets available on data.gov.uk is also summarised below.

Figure 3.2: data.gov.uk dataset links, October 2012



Source: Deloitte analysis of data.gov.uk

The numbers in Figures 3.1 and 3.2 should be read carefully as they refer to the number of datasets only and by download links⁸⁶. Furthermore, they do not take account of the velocity and verbosity of the information.

As an example, as part of the evidence gathering process for this report, the Met Office has indicated that their daily submission to data.gov.uk amounts to around 200Mb of data, which undoubtedly makes them one of the largest suppliers of public sector information to data.gov.uk.

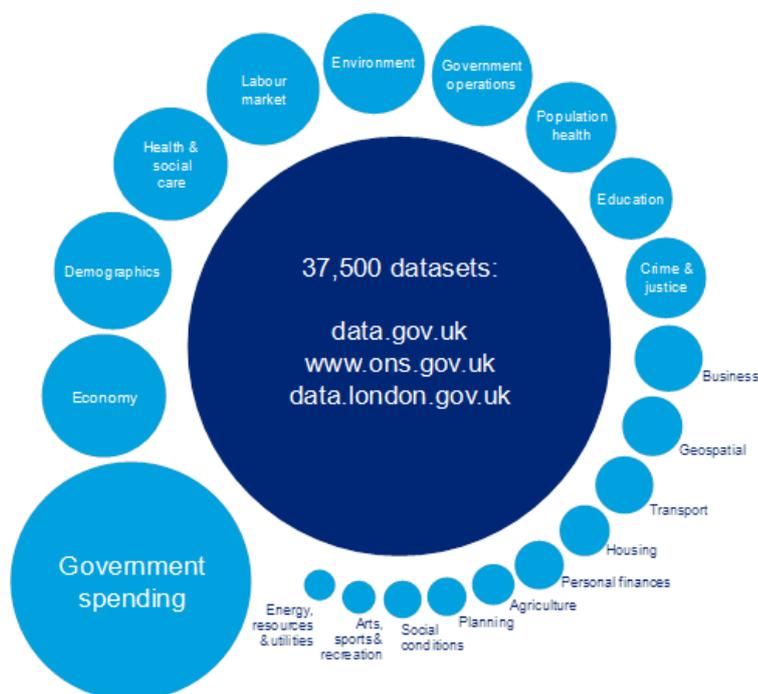
⁸⁵ This figure appears to refer to the main datasets (e.g. Census and Labour Force Survey) – as clearly, the ONS publishes significantly more individual pieces of data.

⁸⁶ I.e. there may be more than one dataset per download link so the numbers are likely to be an underestimate.

However this importance is disguised by the fact that the data/information is contained in just three datasets⁸⁷. It is therefore important to treat information on the number of datasets with a degree of caution: it is a reliable guide to neither volume nor value of public sector information. Nonetheless, as a first approximation, the analysis is useful to illustrate the long tail of PSIHS: whereas around half of all PSI datasets come from 10 PSIHS, the remaining datasets come from over 750 PSIHS.

Moving away from data.gov.uk, open public sector information is also available from a number of other websites. Deloitte's recent research with the ODI has examined the number of open datasets types (rather than links) available across three major data portals: data.gov.uk, the ONS and data.london.gov.uk – this is shown below.

Figure 3.3: number of open datasets across data.gov.uk, the ONS and data.london.gov.uk⁸⁸



Source: Deloitte / ODI analysis of various public sector information data portals

The number of links / datasets in both Figure 3.2 and 3.3 is dominated by government spending information and data. In some senses this is unsurprising as this type of public sector information is likely to have relatively low cost of collection and dissemination for PSIHS, and is relatively easy for them to make available online on data.gov.uk, subject to any data redactions.⁸⁹ PSIHS have also been incentivised to publish this type of public sector information under the Government's transparency agenda⁹⁰.

Other sources of public sector information available (at the time of writing) as open data include⁹¹:

- Companies House – with one major free open dataset;

⁸⁷ Indeed, the Met Office's daily files include modelling and observational data for over 5000 sites across the UK ingested by data.gov.uk which are then converted, per location, into a database row with observational and modelling data. These three files generate 250 million lines of data each hour.

⁸⁸ Another example of a website counting the number of datasets is the TWC LOGD Linking Open Government Data website: http://logd.tw.rpi.edu/iogds_data_analytics

⁸⁹ This is because it is largely 'by-product' data: for example, information on salaries, spending by departments, and organograms

⁹⁰ See for example: www.cabinetoffice.gov.uk/news/government-spending-data-published

⁹¹ This list is not exhaustive.

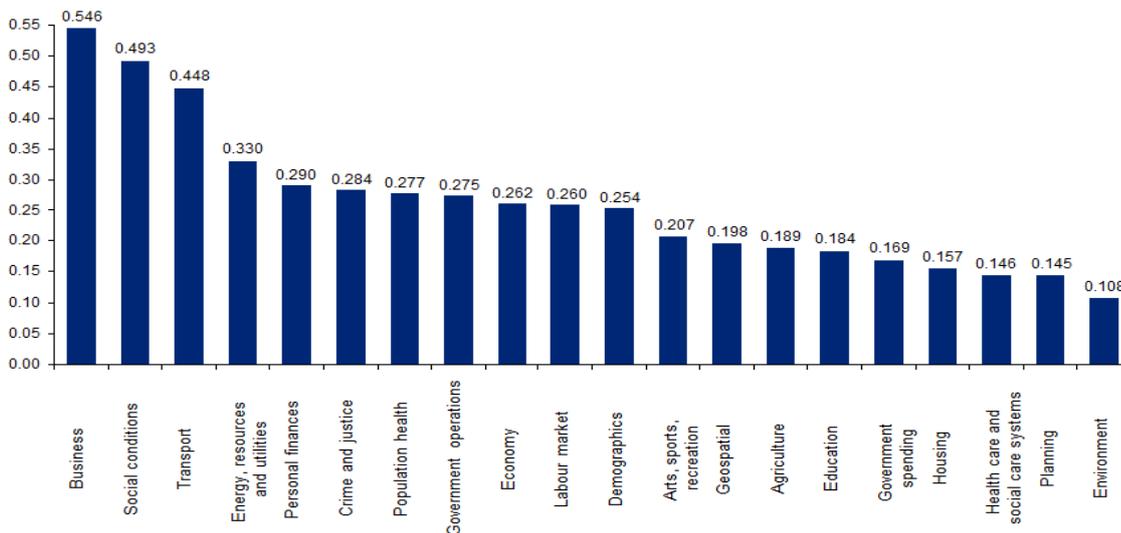
- The Met Office – with 118 open data sets (listed on data.gov.uk);
- Land Registry – with over 30 open data sets (available through data.gov.uk);
- Data.london.gov.uk – with over 550 open data sets;
- OS OpenData – with 12 mapping products that can be ordered online;
- Open Data Communities – with over 70 open data sets; and
- Legislation.gov.uk – with over 78,700 open data sets⁹².

A separate statistics appendix⁹³ provides further detail on datasets held by the four Trading Funds that make up the Public Data Group.

Downloads and popularity of datasets

Using data.gov.uk statistics again, it is also possible to examine the popularity of different public sector information open datasets in terms of page views.

Figure 3.4: data.gov.uk dataset normalised page views to show relativity, October 2012

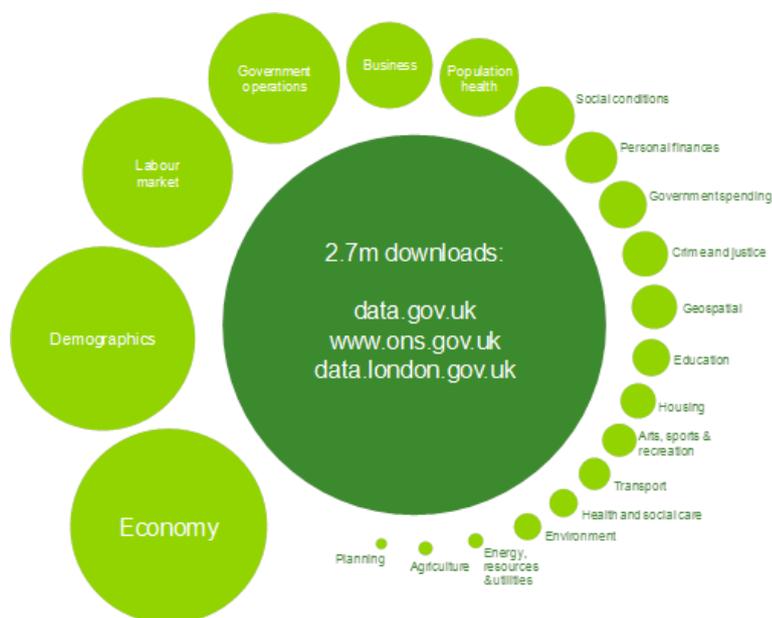


Source: Deloitte analysis

The equivalent chart for dataset downloads from a broader group of websites is shown below.

⁹² Individual acts etc. This figure is approximate and may not include all PSI available of legislation.gov.uk. There is an argument that this should be considered a single dataset, but given that the files are downloaded separately and not as one database they are here treated as separate datasets.

⁹³ See Appendix 4.

Figure 3.5: open data downloads

Source: Deloitte / ODI analysis of various public sector information data portals

As one might expect, business and economy datasets dominate, although this may simply be a function of the supply of these datasets exceeding those of other public sector information types. However, data on social conditions and transport also report high levels of relative popularity.

Between July 2012 and January 2013 there were over a million total visits to data.gov.uk and over 3.7 million page views⁹⁴. While this is a high level of traffic, many of these users would not have downloaded data (see below). In some cases this may simply have been due the visitor deciding they no longer wished to see the data or automated hits from web-trawlers. However, in other cases, this may be because individuals were unable to find the particular dataset they were looking for; or because they encountered broken links; or that they were transferred to the original PSIH's website to download the data rather than download it from data.gov.uk. Without a survey of users it is difficult to know the reasons why a visit ended with no dataset download. However, the National Audit Office found that 82 per cent of users left data.gov.uk directly from the homepage or data page, indicating that they may have not found the information they were seeking.⁹⁵ Note, this was a snapshot figure in early 2012 and data.gov.uk continues to evolve.

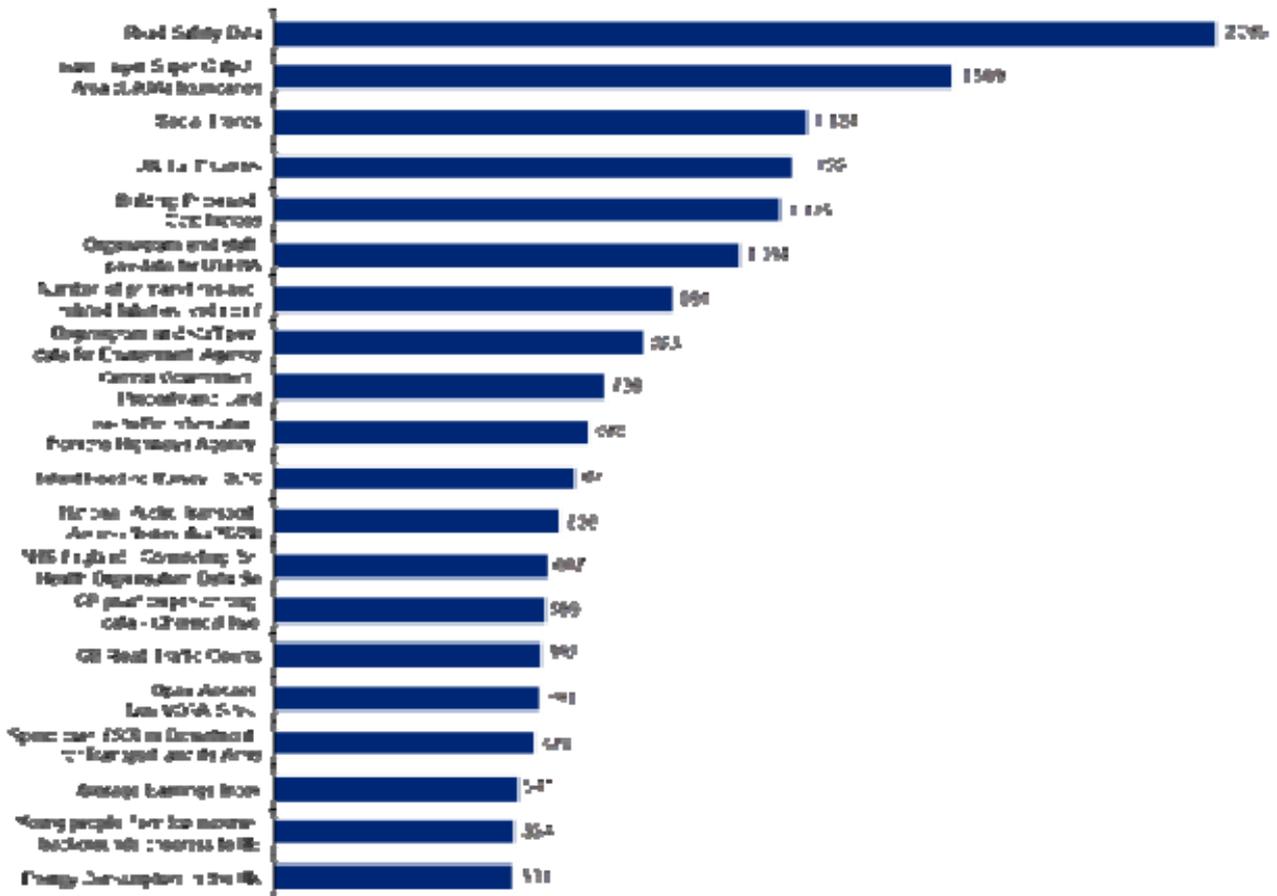
Figure 3.6 below indicates the most downloaded datasets from data.gov.uk between July 2012 and January 2013. In total there were over 80,000 downloads of datasets from data.gov.uk during this period.⁹⁶ While this is a significant number, it means that at most eight per cent of visitors downloaded a dataset – and fewer if some users downloaded more than one dataset, as seems likely.

⁹⁴ Source: www.data.gov.uk/data/site-usage#totals

⁹⁵ NAO, *Implementing Transparency*, April 2012. Available at www.nao.org.uk/wp-content/uploads/2012/04/10121833.pdf

⁹⁶ Source: www.data.gov.uk/data/site-usage/dataset

Figure 3.6: data.gov.uk most downloaded datasets (July 2012 – January 2013)⁹⁷



Source: www.data.gov.uk/data/site-usage/dataset

Of course, these statistics do not take into account views and downloads from other PSIHS websites and non-public sector PSIHS. Further, these statistics do not include public sector information sent to data.gov.uk but hosted on other platforms. For example, the Met Office data supplied to data.gov.uk is hosted by Microsoft Windows Azure Datamarket⁹⁸ - approximately 250 million rows of data per hour is hosted on this service and it is estimated that there are currently over half a million transactions per month, on average. Nearly two thirds of these transactions come from outside the UK suggesting UK public sector information can also help generate growth overseas.

Taking these into account, the numbers of total downloads of public sector information may be much higher.

Supply of datasets locally

As well as open public sector information being available at the national level, consumers can download and access local open public sector information. The Local Government Association carried out a survey of local authorities in late 2012 (*Local Government Transparency Survey 2012*)⁹⁹ to examine the provision of data at a local level. 113 local authorities responded out of

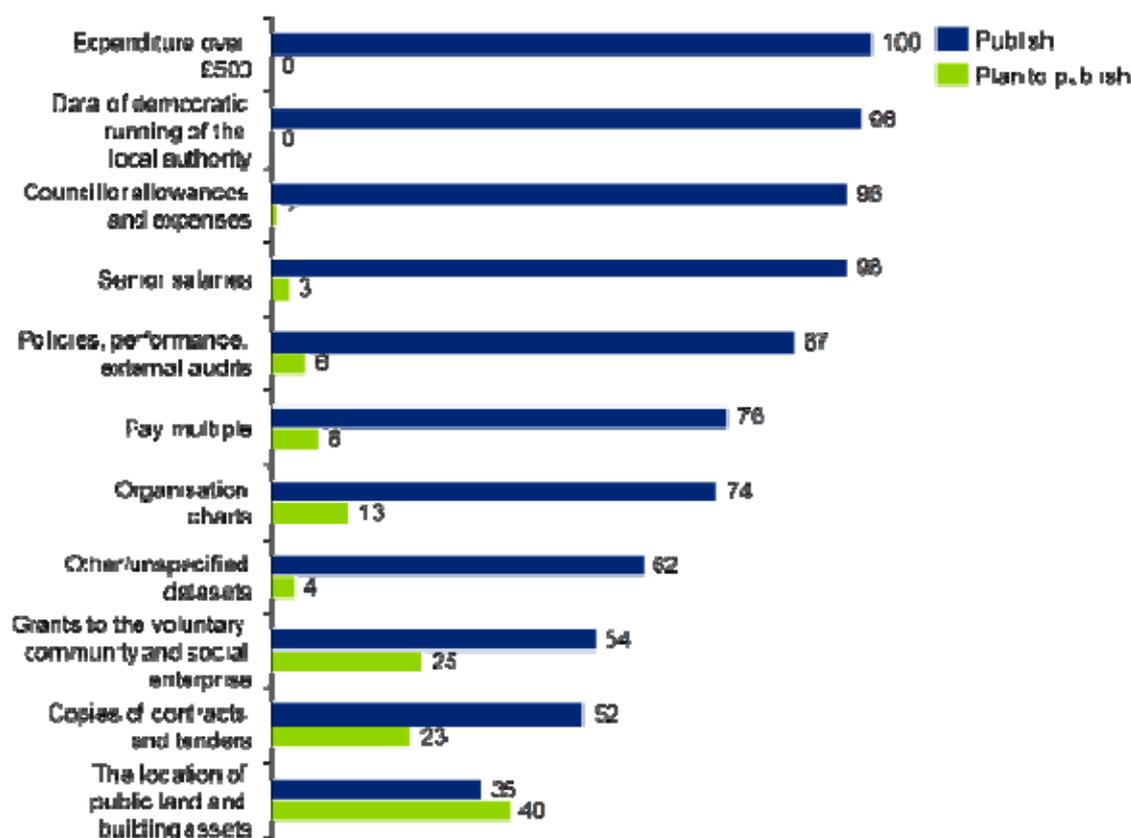
⁹⁷ "Downloads" is the number of times a user has clicked to download either an original or cached resource for a particular dataset. Download information is only available from 2nd December 2012 (<http://data.gov.uk/data/site-usage/dataset>)

⁹⁸ See <https://datamarket.azure.com/dataset/0f2cba12-e5cf-4c6d-83c9-83114d44387a>

⁹⁹ See www.local.gov.uk/web/guest/local-transparency/-/journal_content/56/10171/3825698/ARTICLE-TEMPLATE

128 respondents to the survey (the total surveyed was 346). The chart below shows the types of information that respondents reported they currently publish or plan to publish in the future (per cent).

Figure 3.7: Local Government Transparency Survey 2012 results – public sector information published and plans to publish (%)



Source: LGA, 'Local Government Transparency Survey 2012'

The data being made available by local government is often not published through a national data portal such as data.gov.uk. The survey revealed that of 104 respondents to the question, 14 per cent made information available through data.gov.uk and nine per cent through direct.gov. The table below shows the full responses to the question 'Where on the internet do you publish your open data/transparency data?'

Figure 3.8: Local Government Transparency Survey 2012 results – public sector information channels

Publication channel	Number	Per cent responding
Dedicated open/transparency data page	71	65
Directly within topic section pages	53	48
Data.gov.uk	15	14
Direct gov.uk (now disbanded)	10	9
Publication schema	23	21
Information asset inventory	0	0

Publication channel	Number	Per cent responding
Other	10	9

Source: LGA, 'Local Government Transparency Survey 2012'

Because local authority information is published through many different channels, it is difficult to get a sense of the number and types of users of the information that has been released.

Nonetheless, the LGA survey offers some insight in this regard. There were 104 responses to the question 'Do others use your open data?' The answers reveal that 16 per cent of local authorities were certain that others used their open data, while nearly half believed they did not. 36 per cent were unsure, indicating that it is often difficult to track how many users of open data there are, especially where this does not require registration by users.

14 local authorities responded that they knew who used their open data. Their answers are shown in the table below, showing that local community groups are particularly significant users of local authority open data. However, given the small sample size, these results should be treated with some caution.

Figure 3.9: Local Government Transparency Survey 2012 results – public sector information consumers

Users of open data	Number
Local community groups	7
Other councils/public services	3
Individuals	3
Local educational establishments	2
Local charities	1
Local businesses	1

Source: LGA, 'Local Government Transparency Survey 2012'

Trading Funds

Trading Funds are among the most visible PSIHS in the wider public sector information landscape, and are often the focus of attention in debates over both the value of public sector information and access for users¹⁰⁰. While this report focuses on the wider public sector landscape, of which the Trading Funds are only one component, given the nature of the public sector information that the Trading Funds collect, as well as the debate over issues around access, it would be remiss not to separately consider their role in the supply of public sector information.

While no Trading Fund is identical, collectively they occupy a special position in the public sector information landscape as bodies with a statutory requirement to fund their operations and meet their financial objectives from trading income. As such they are subject to a set of guidance documents, legislation and regulations and guidance that differ from most other public sector bodies. The key elements of this framework are described below (this list is not exhaustive and these do not exclusively apply to Trading Funds).

¹⁰⁰ These debates are discussed in depth in Chapters 5 and 6.

Figure 3.10: regulations and guidance covering Trading Funds¹⁰¹

Regulation / guidance	Content
Managing Public Money (MPM)	<p>The PDG Trading Funds are required to comply with the guidance on fees, charges and levies in Managing Public Money. The standard approach to setting charges for public services (including services supplied by one public sector organisation to another) is full cost recovery. This normally means recovering a 3.5 per cent real charge for the cost of capital. However, for services supplied into competitive markets, charges should be set at a commercial rate.</p> <p>Managing Public Money explains that the norm is to charge full cost for publicly provided goods and services. For commercial services, charges should be set at a commercial rate. Managing Public Money also explains that much information about public services should be made available either free or at low cost in the public interest. However, there are circumstances where charges are made. Public sector organisations can also charge for information which recipients intend to re-use. Managing Public Money explains that where data is supplied for re-use, the norm is to charge marginal cost. For value added data, and for all information supplied by Trading Funds, the norm is to charge at full cost plus an appropriate rate of return.</p>
European Union Directive on Re-use of Public Sector Information	<p>Under the 2005 legislation, charges for the re-use of information made by public sector bodies cannot exceed the cost of collection, production, reproduction and dissemination of documents; and a reasonable return on investment.</p> <p>The revision of this directive is proposing to move to marginal cost but allows for some exceptions.</p>
Information Fair Trading Scheme (IFTS)	<p>This sets and assesses standards for public sector bodies. It requires them to encourage the re-use of information and reach a standard of fairness and transparency. The National Archives use the IFTS to monitor the activities of the trading funds to ensure that they trade fairly, openly and transparently in information. The IFTS also covers other public bodies.</p>

Source: Public Data Group, 'Approach to Charging' (December 2012)

For clarity, it should be noted that Trading Funds have their own Trading Fund Order – statutory duties to charge are only in place for the Land Registry and Companies House, but the Trading Fund Order and Trading Fund Act set out the framework for all Trading Funds' operations.

It is important to be clear that these guidance documents and regulations, and especially the requirement to recover costs plus 3.5 per cent, constrain the ability of the Trading Funds to change their underlying business models beyond a certain point.

There are a number of Trading Funds currently operating in the UK, each arising out of distinctive historical circumstances and each with a very different purpose, remit and business model. Below, an overview of the four Trading Funds that constitute the Public Data Group (PDG) is provided. This draws on the work of the Shareholder Executive, December 2012¹⁰².

¹⁰¹ Note: the below is adapted from a Public Data Group paper, 'Approach to Charging', December 2012. This is available at www.gov.uk/government/publications/approach-to-charging. The regulations here discussed do not necessarily apply exclusively to trading funds.

¹⁰² Ibid.

Figure 3.11: Public Data Group Trading Funds

	Companies House	Land Registry	Met Office	Ordnance Survey
Status	Trading Fund Statutory Body	Trading Fund Statutory Body	Trading Fund	Trading Fund
Competitive Landscape	100% non-competed (statutory monopoly)	98% non-competed (statutory monopoly)	Mix of: <ul style="list-style-type: none"> • Non-competed and competed HMG contracts; and • Widely competed commercial (non-HMG) contracts. 	All products and services are traded in competed markets (apart from OS OpenData).
Sources of revenue	Registration and search customers pay statutory fees. Data users fund dissemination costs.	Register users fund data collection through statutory fees. Value added service users fund product development and dissemination.	HMG funds the core underpinning national infrastructure, research and development required for the national weather, core climate services and services to defence. Wholesale data users fund their own data dissemination costs. Users of value added competed services pay for the bespoke development of weather services and the transfer price of the data, which is on the same terms for all parties.	Data users fund data collection, product development and dissemination. Users of value-added services pay for the bespoke development of services and the transfer price of the data, which is on the same terms for all parties.
Public and private sector users	99% non-HMG 1% HMG	100% non-HMG	79% Core HMG 4% Competed HMG 17% non-HMG	43% PSMA, OSMA ⁽¹⁾ 43% other Chargeable incl. Private Sector 14% OpenData™ (for private and public sector innovation)

(1) PSMA: Public Sector Mapping Agreements; OSMA: One Scotland Mapping Agreement
Source: Shareholder Executive

As described above, Trading Funds are required to generate their own revenue, which is to say that they do not in general receive centrally allocated funds. They are required to recover the full cost of service provision plus the cost of capital, set at 3.5 per cent. The exception is where Trading Funds operate in competitive markets, in which case they are required to price their services at market rates to avoid distorting competition. Even where Trading Funds provide information or services to public sector organisations, they 'sell' these at agreed rates. This is an important distinction, making the Trading Fund business models qualitatively different to that of other publicly funded PSIHs which provide 'free' services to public sector users (though certain services may be charged for). The Trading Fund model is intended to drive efficiency by forcing Trading Funds to hold down costs in the same way as a commercial business would do.

Further detailed statistics on the amount of data being made available and under which charging models is provided in Appendix 4.

Attitudes and policies of public sector information holders

As part of wide-ranging stakeholder consultations, Deloitte conducted an informal consultation of Government officials across national departments to seek to develop indicative estimates of the amounts of public sector information currently held and currently made available and understand some of the challenges currently faced. Full details of the consultation and its results can be found in Appendix 8¹⁰³.

The most important findings from the survey included:

- departments currently vary significantly in the amount of public sector information they make available to the public;
 - around a third of respondents to the informal survey indicated that up to a quarter of public sector information held by the department (and in some cases its executive bodies) was currently being made available to the public. Other respondents were unable to give estimates
- of the public sector information made available to the public, all respondents said it was available either entirely free of charge or at least 75 per cent was available at no cost; and
- 33 per cent of respondents report that less than five per cent of their staff were dedicated to data collection, processing and dissemination (though a much higher proportion would be involved as part of their wider duties), while 50 per cent were unable to estimate the proportion. The exceptions were dedicated statistical bodies, which owing to their nature employ many more staff dedicated to data-related functions.

All respondents indicated that their PSIH planned to make more public sector information available to the public over the next 12 months, as set out in departmental open data strategies.

The officials consulted in this informal consultation work with public sector information issues on a regular basis and are likely to have had a high degree of awareness of public sector information issues. A recent survey by Listpoint¹⁰⁴ surveyed a wider number of civil and public servants to examine levels of awareness of the benefits of open data and public sector information initiatives. The key highlights are shown below:

- 78 per cent of those surveyed did not know about government plans for open data and the benefits that follow;
- 57 per cent of those surveyed did not know how to access datasets, how to interpret them or how to best apply data standards;
- Over 75 per cent of those surveyed did not know their role in delivering the open data agenda; and
- 72 per cent recognised that knowing how to access, share and use data will be increasingly important over the next three years.

Again, it should be noted that there may be an element of self-selection in the survey results.

¹⁰³ This informal survey should not be treated as statistically robust as there is no general public sector asset register to accurately capture the size of the market, but it does provide a useful indication of some of the challenges faced. While responses were received from a number of departments, responses were not received from some major departments which, a priori, are expected to hold significant amounts of public sector information.

¹⁰⁴ See: www.intellectuk.org/media-centre/member-press-releases/8967-first-benchmark-survey-of-open-data-understanding. The research was conducted in late December 2012 and more data is available on Listpoint's website: www.listpoint.co.uk/media. More than 1000 responses were received across central and local government, non-departmental bodies, the NHS and police forces.

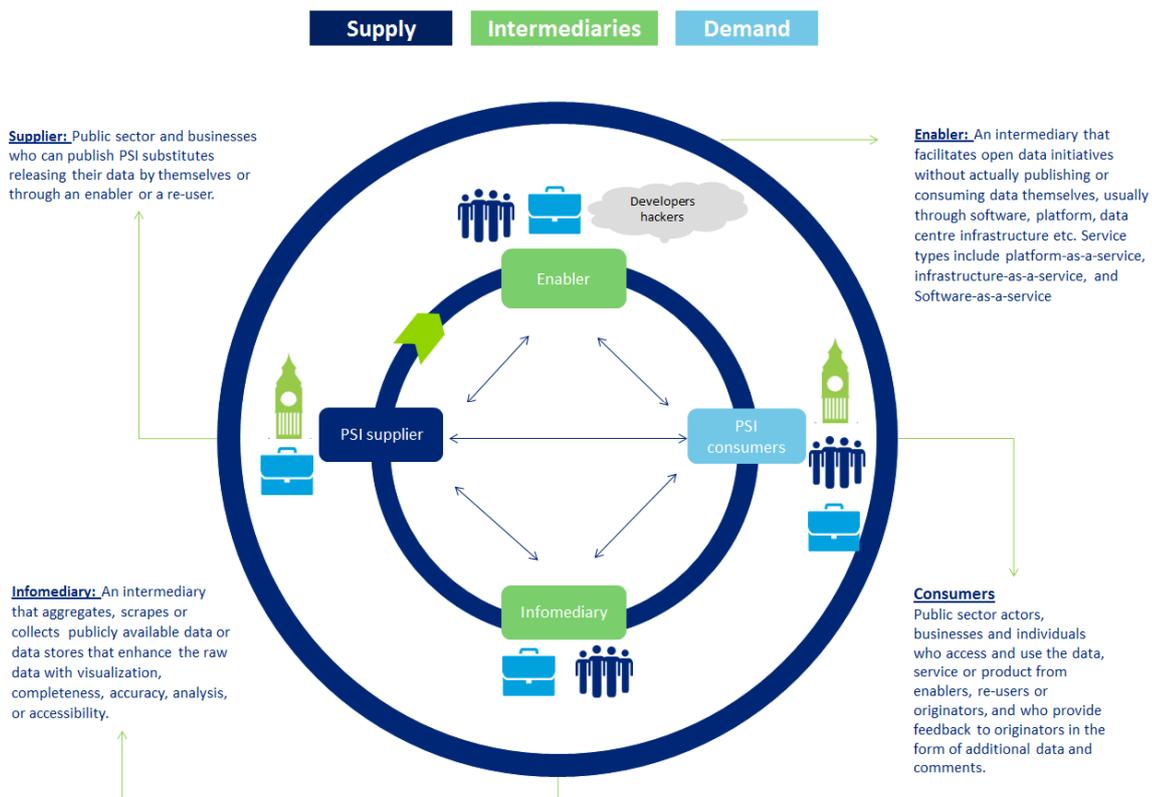
Demand side of the market

Consumers on the demand side of the public sector information market can be divided into two broad categories based on whether they **use** or **re-use** public sector information. Use of public sector information refers to exploiting the data and information in a way that is commensurate with the original purpose within which the public sector information was produced. An example might be using economic statistics produced by the ONS to inform economic policymaking and investment decisions. In contrast, re-use of public sector information refers to the use of the dataset in a way other than the initial purpose for which it was collected/generated. An example might be correlating data on environmental and atmospheric conditions to the academic performance of school children. Clearly, the distinction between use and re-use is not hard and fast, as some aspects of use could be interpreted as re-use.

Ultimately, when considering final consumers, almost every inhabitant of the UK is, at some point, a downstream user of public sector information in some form or other – often indirectly where the data is an input into another value-added service or product. This might be through a transport information or weather forecasting mobile app, an Ordnance Survey map, or another product that relies on public sector information to function.

The different types of consumers of public sector information can be charted to form a circular value-chain. As Figure 3.12 highlights, the role of intermediaries in making public sector information accessible to a wider audience is currently crucial. One might expect that as general awareness of public sector information increases and the capacity to perform analysis directly increases, the role of intermediaries may diminish.

Figure 3.12: public sector information eco-system



Source: Deloitte analysis

The supply of public sector information out of PSIHs is not a one-way street - consumers and infomediaries can request more datasets to be released via the data.gov.uk request mechanism¹⁰⁵.

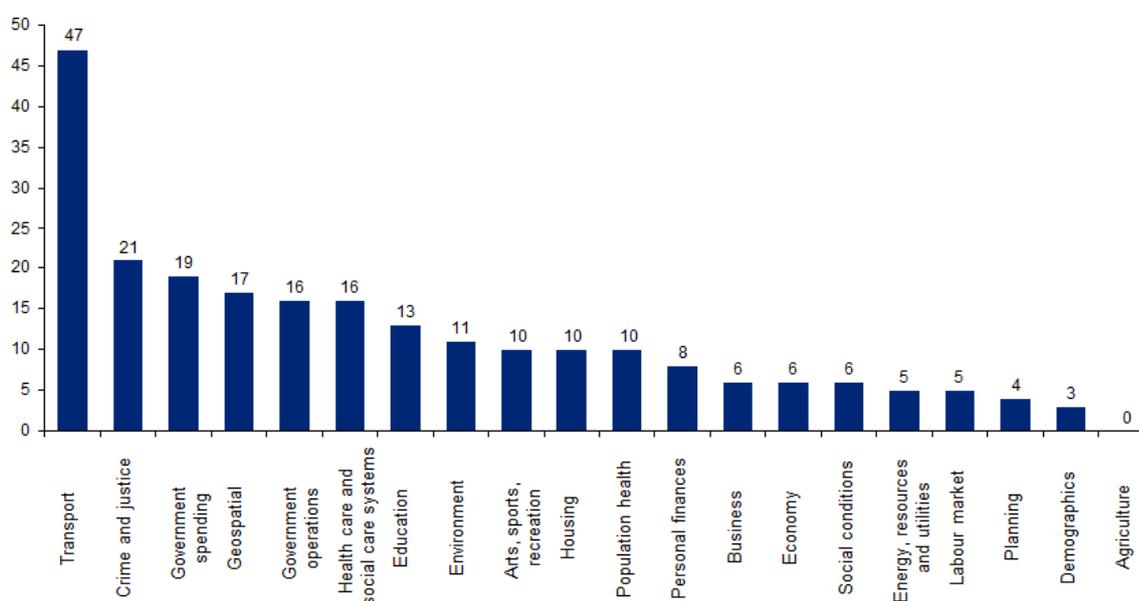
As discussed above, there is a paucity of statistics on public sector information users and re-users. The following sub-sections present the available statistics on use and re-use in terms of download statistics, consumers perceptions of the market, different business models in operation and a profile of the emerging data scientist role.

Public sector information use and re-use

Download and usage statistics

One approach to measuring the extent to which public sector information is re-used is to consider the number of apps that take currently available public sector information as an input. Figure 3.13 below charts the number of mobile apps that use public sector information as a direct input from data.gov.uk by content category. One might infer that these apps could not have been created in the absence of these public sector information datasets.

Figure 3.13: data.gov.uk dataset mobile apps, October 2012¹⁰⁶



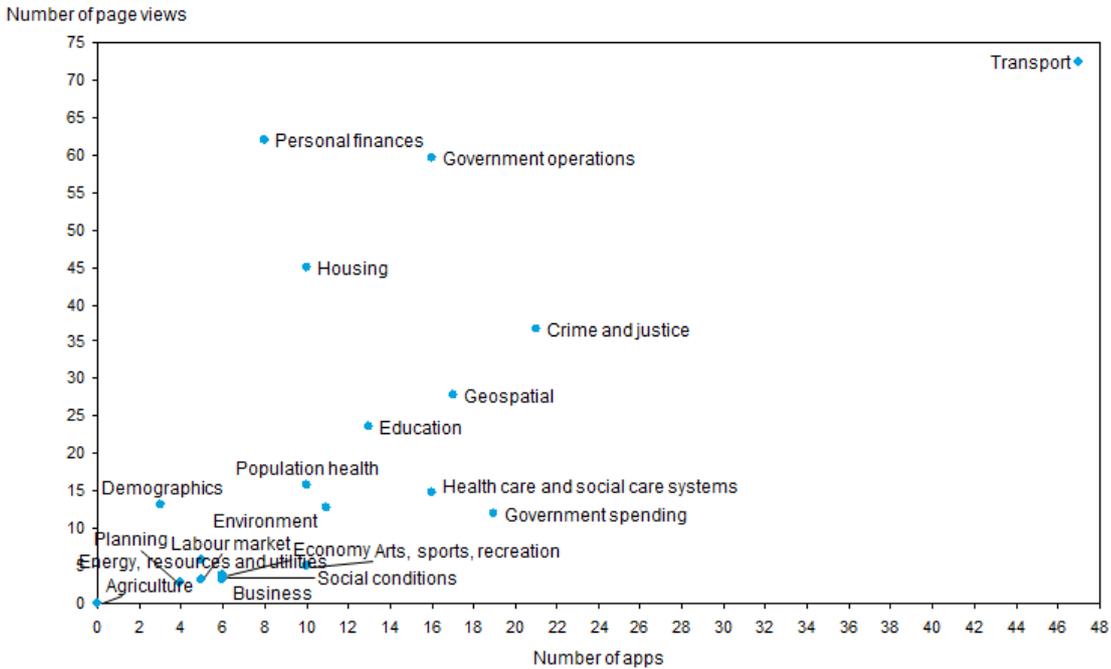
Source: Deloitte analysis of data.gov.uk data using Google Analytics

What is particularly interesting in Figure 3.14 is that the most viewed datasets on data.gov.uk (as shown in Figure 3.2) do not entirely correspond with those that are most used to develop apps. Figure 3.14 compares page views to dataset apps.

¹⁰⁵ See <http://data.gov.uk/node/add/data-request>

¹⁰⁶ This analysis was done in October 2012. The number apps that use data.gov.uk datasets is likely to have risen since then.

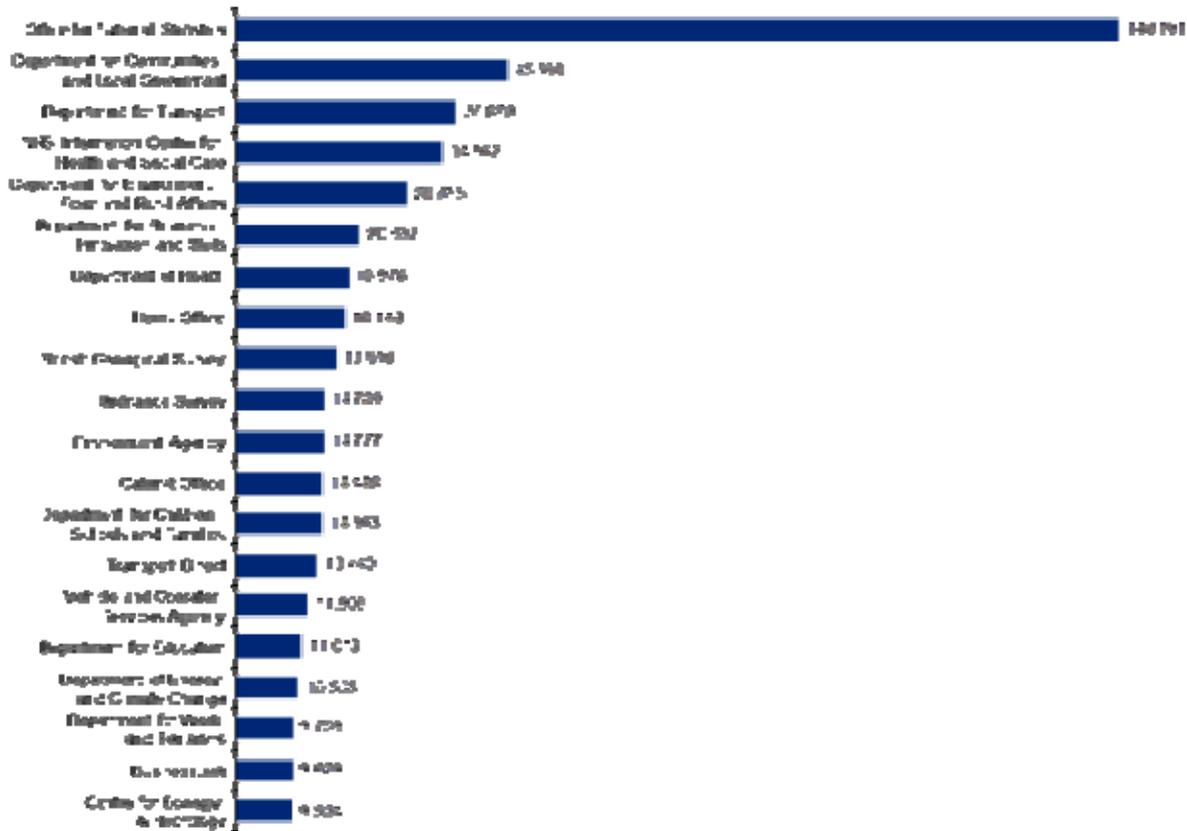
Figure 3.14: data.gov.uk dataset apps versus page views, October 2012



Source: Deloitte analysis of data.gov.uk data using Google Analytics

Public sector information covering transport is clearly the most popular both in terms of page views and number of apps developed, with government operations data, personal finance information, housing, crime and justice, geospatial and education also attracting high levels of attention.

Statistics from data.gov.uk on the page views received by publishers on the site indicate that the Office for National Statistics has received by far the most attention from users, with over three times the views of the next most viewed publisher, DCLG. This is only partially a consequence of the fact that the ONS is the most prolific publisher on data.gov.uk, and suggests that ONS datasets are particularly interesting to users.

Figure 3.15: data.gov.uk most viewed publishers (July 2012 – January 2013)¹⁰⁷

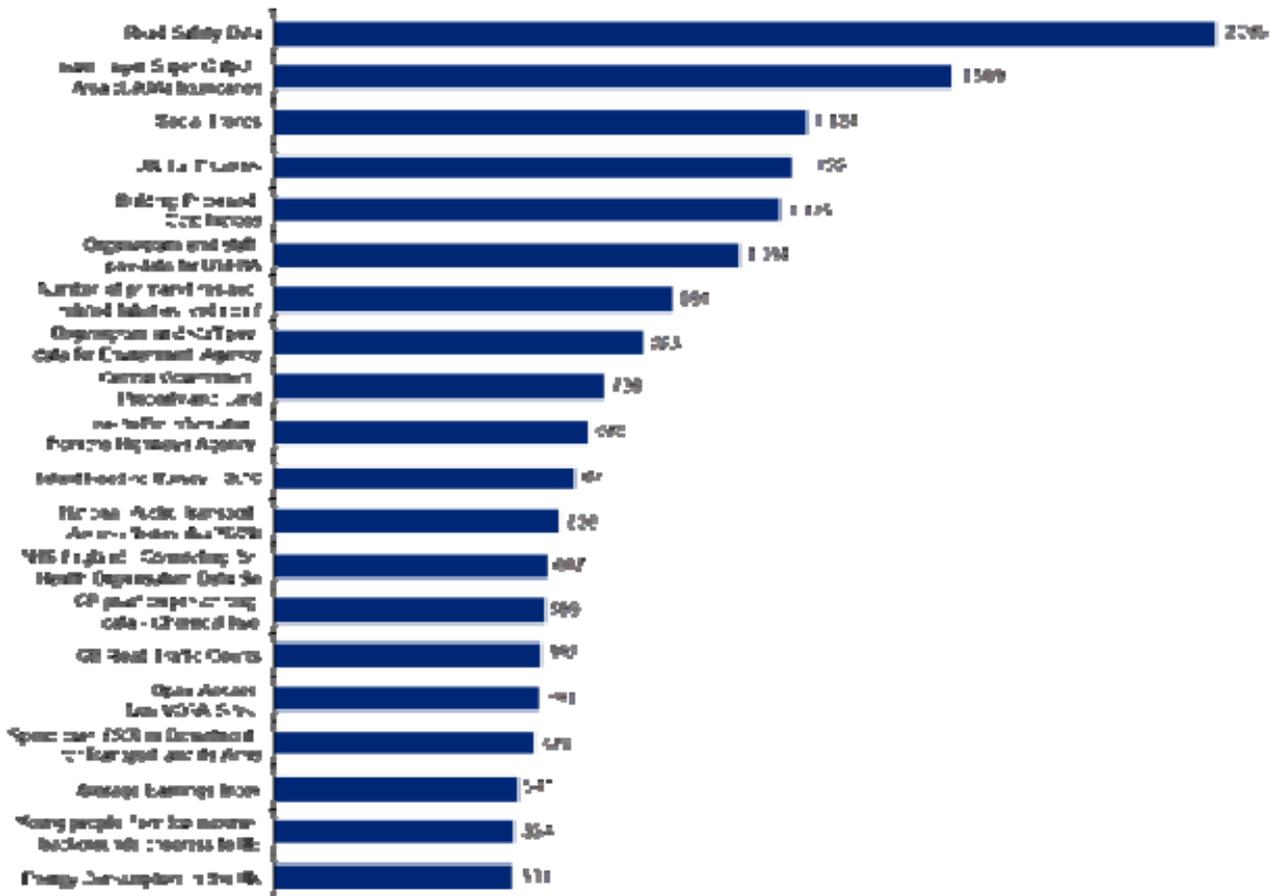
Source: Deloitte analysis of <http://data.gov.uk/data/site-usage/publisher>

Data.gov.uk statistics on the most downloaded datasets give some indication as to which information users believe holds the most value and may also indicate where PSIHS should focus their efforts in improving quality and access. Interestingly, many of the datasets in the list of the top 20 most downloaded fall into one of the areas that this report identifies as areas where public sector information offers particularly promising potential for generating value (see Chapter 5), including:

- transport, especially for roads – live traffic data, traffic counts, road safety;
- health – the GP prescribing data has obviously attracted a high level of interest following the Mastodon C work;
- property – prices and energy efficiency; and
- social research – low income students in higher education, social trends, average earnings.

¹⁰⁷ “Views” is the number of times a page was loaded in users’ browsers. (<http://data.gov.uk/data/site-usage/publisher>)

Figure 3.16: data.gov.uk most downloaded datasets¹⁰⁸



Source: www.data.gov.uk/data/site-usage/dataset

Again, it should be noted that the analysis refers only to open data available on data.gov.uk and therefore misses public sector information from PSIHS such as the Met Office or Land Registry as well as non-public sector PSIHS. The review of the Trading Funds in the Public Data Group above suggests the different ways their data is used and re-used.

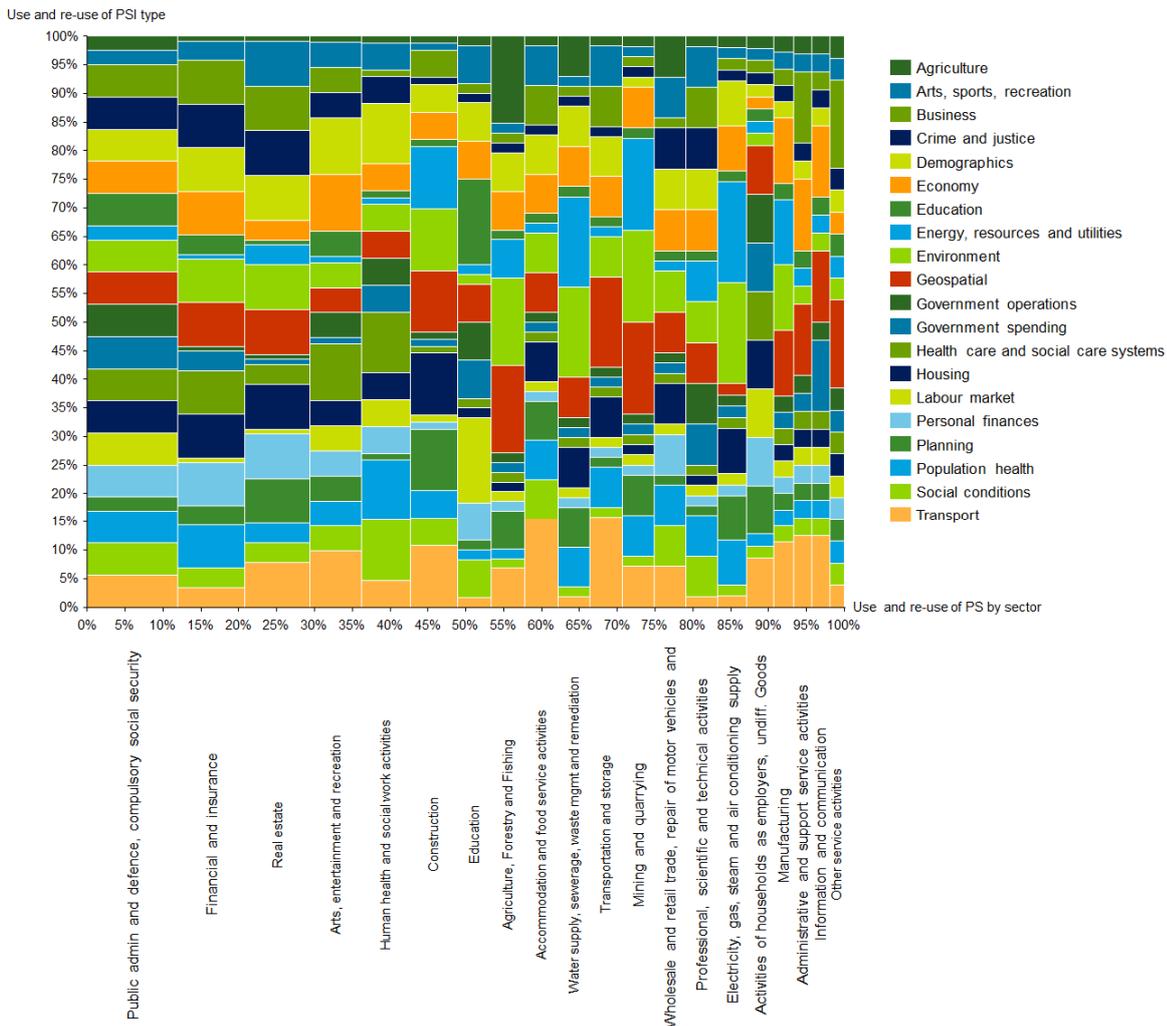
Use of public sector information across different sectors of the economy

Another way to consider the public sector information demand side is to explore the use of all types of open data (public sector information and private sector and individual open data) across different economic sectors. Building on existing analysis by Deloitte, it is possible to create a data intensity matrix which compares the use of different types of data by industries in the UK¹⁰⁹. Figure 3.18 shows this report’s estimates as to which economic sectors most consume different types of open data. It shows, as might be expected, that businesses in the agricultural, forestry and fishing sectors are major users and re-users of agricultural open data, but they also use/re-use geo-spatial, environmental and energy, resources and utilities data.

¹⁰⁸ "Downloads" is the number of times a user has clicked to download either an original or cached resource for a particular dataset. Download information is only available from 2nd December 2012 (<http://data.gov.uk/data/site-usage/dataset>).

¹⁰⁹ This has been developed on the basis of balance sheet analysis of major UK firms.

Figure 3.17: Use of data across economic sectors



Source: Deloitte analysis

The x-axis charts the different sectors of the economic and the y-axis different types of public sector information. The width of each economic sector columns compares the relative ‘consumption’ of open data of each sector. For example, ‘public admin and defence, compulsory social security’ (i.e. the public sector) consumes over 10 per cent of all data and of this figure, 5 per cent of its consumption is on transport data, 5 per cent on social conditions data, etc. As Figure 3.17 suggests, the construction, real estate, financial and insurance, public sector and arts, entertainment and recreation sectors are some of the largest users and re-users of open data. While this analysis has been carried out on the basis of open data, one might hypothesise that a similar picture holds for public sector information.

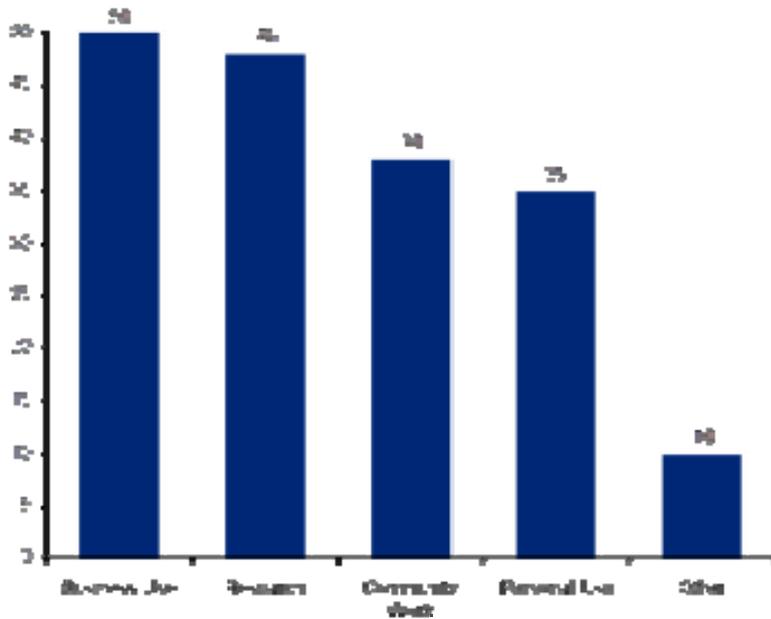
Consumer perceptions of public sector information and unmet demand

In order to gauge consumer perceptions and the unmet consumer demand for public sector information, Deloitte has undertaken an analysis of publically available requests for information and data on two data portals: data.gov.uk and the London Datastore. Both these portals allow users to lodge requests for additional types of public sector information. For a number of reasons, these results should be treated as indicative only: the sample size is small and inevitably self-selecting. It is likely that there are many more potential users who have not engaged with the data portals. Nonetheless, this analysis provides an indicative insight into the types of public sector information that users are interesting in gaining access to, as well as the uses to which this public sector information might be put.

Data.gov.uk requests

This analysis is based on 100 requests posted publically on data.gov.uk between 26 September 2012 and 18 November 2012. As of the end of November 2012 there were 472 requests posted publically on data.gov.uk; however, the detail contained in the requests was much reduced after the first 100. For this reason the 100 most recent requests are used as the sample for this analysis.

Figure 3.18: usage categories cited in data requests (%)¹¹⁰



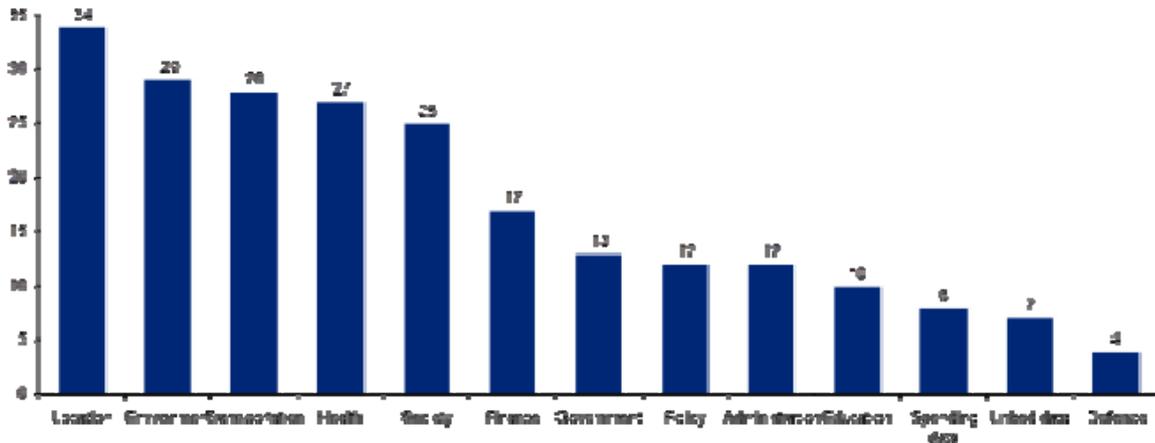
Source: Deloitte analysis of data.gov.uk data requests: based on 100 requests

The requests in the sample most commonly cite business use and research as the intended use for the public sector information, at around 50 per cent for each of these categories.

In terms of the data categories requested, it is location (or geospatial) data that tops the list of requests. This reinforces the impression that this category of data generates some of the highest levels of value – or at least that this is where users perceive value to lie. Environment, transportation, health and society are also data that attract particularly high levels of requests. Requests for data used to scrutinise and hold government accountable, such as government, finance, policy, administration and spending data are significantly lower, suggesting that this attracts less interest. However, even in these categories the numbers of requests are not insignificant.

¹¹⁰ Note that where the sum of responses shown in the charts is greater than 100, this is because more than one category is selected in some data requests. For example, a request may cite both business and personal use.

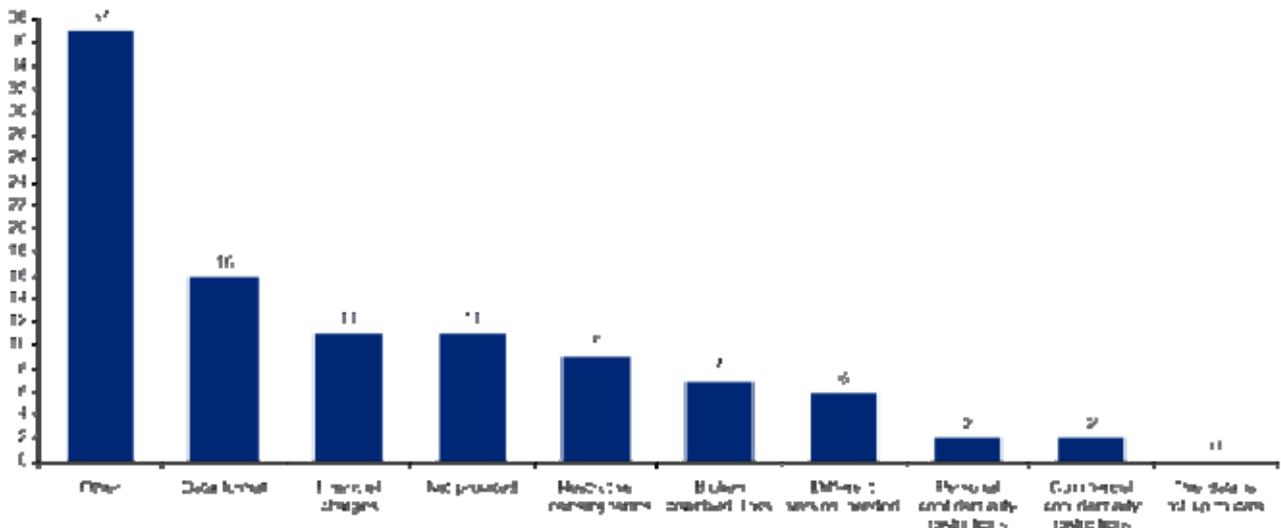
Figure 3.19: data categories cited in data requests (%)



Source: Deloitte analysis of data.gov.uk data requests: based on 100 requests

Of the specific barriers cited in accessing public section information, data format and financial charges are the most commonly mentioned – though these are both behind the ‘other’ category.

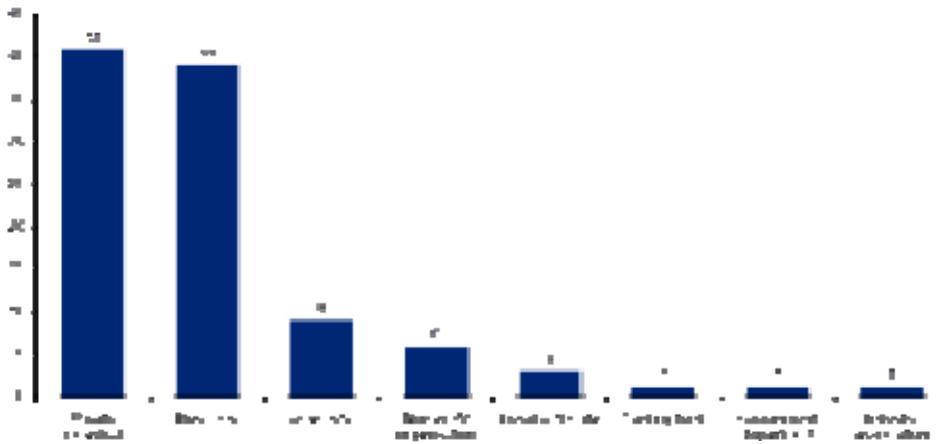
Figure 3.20: barriers cited in data requests (%)



Source: Deloitte analysis of data.gov.uk data requests: based on 100 requests

The large majority of requests are generated by private individuals and business, at 41 per cent and 39 per cent of the sample respectively.

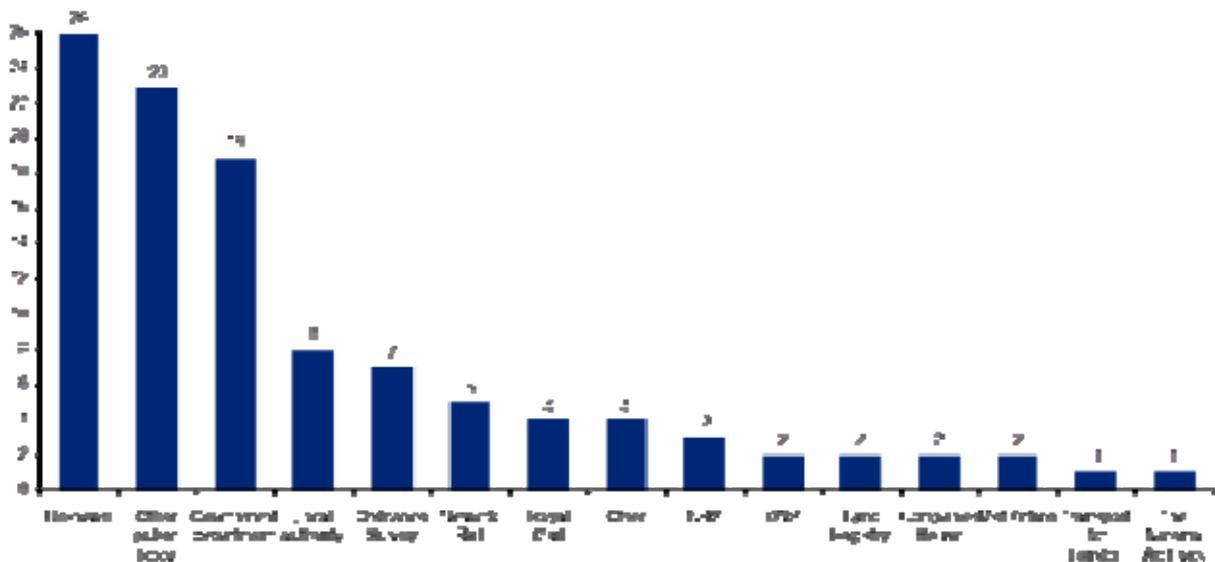
Figure 3.21: affiliation of those requesting data (%)



Source: Deloitte analysis of data.gov.uk data requests: based on 100 requests

The most commonly cited data holders are ‘unknown’ and ‘other public body’, suggesting that users may be unclear where the data they seek is held. Government departments are cited in 19 per cent of requests, although this category covers a large range of PSIHs.

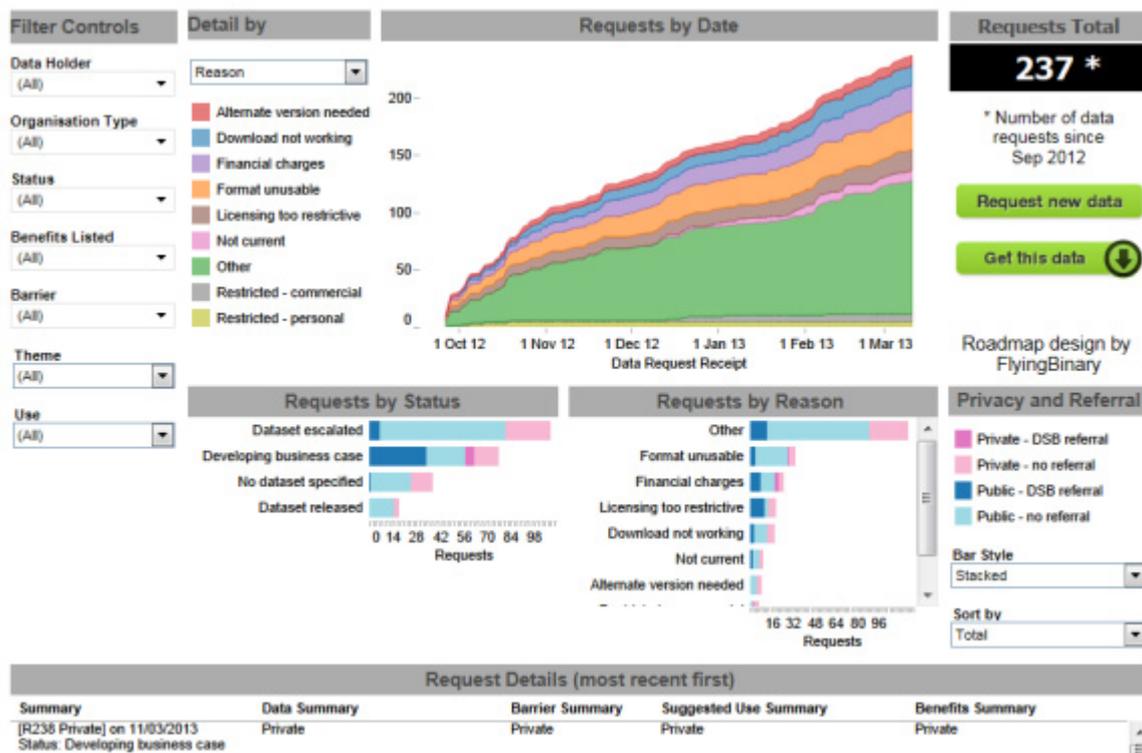
Figure 3.22: data holders cited in requests (%)



Source: Deloitte analysis of data.gov.uk data requests: based on 100 requests

The Open Data User Group has recently launched a new dashboard on data.gov.uk, providing a ‘roadmap’ of data requests. This promises to make the volume and nature of requests more transparent and easier to analyse. As the volume of data on requests increases, this should offer additional insights into the types of data that users and potential users would like to access, and the barriers they face.

Figure 3.23: data.gov.uk request dashboard



Source: www.data.gov.uk/odug-roadmap

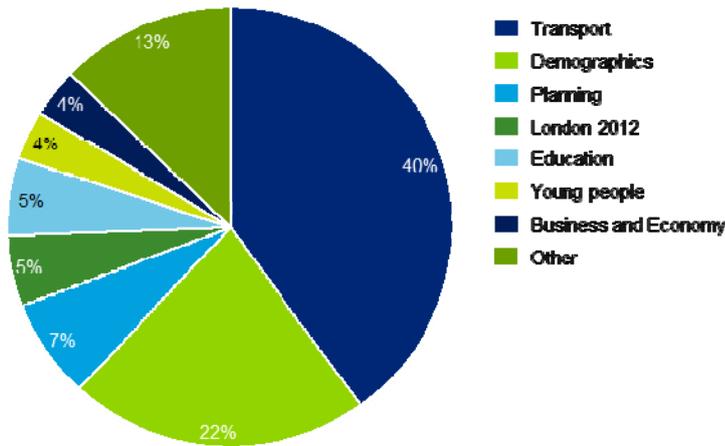
London Datastore requests

A similar analysis is present below using data requests from the London Datastore. The sample size for this analysis is 50 of the data requests with the most 'wants' on london.data.gov.uk, out of a total of at least 870 published requests (as of 14 December 2012). The website allows users to vote for requests by adding 'wants', which function in a similar manner to 'likes' on Facebook. The dates of the requests included in the analysis run from February 2010 to August 2012. The option to submit a request is currently closed: the website notes: "please note we have temporarily suspended the dataset suggestion function as a result of recent spamming activity on the site." Indeed, the majority of the most recent submissions appear to be spam.

Many of the datasets requested are now available, suggesting that the London Datastore has been responsive to requests (although it is possible that these requests overlooked already published data, or that its publication was already planned). Although the sample is small and some of the requests are over two years old, these requests provide an indicative insight into the sorts of data to which users and potential users would like to gain access. Note that some of the requests cite more than one usage category and data category, meaning that in these cases the sum of the responses is greater than the sample size.

As with data.gov.uk (and reflecting particularities in London), transport is the most commonly requested type of data, at 40 per cent of requests, with demographic data the second most common at 22 per cent. This is likely to reflect the high perceived value of these types of data, particularly on a local level, as witnessed by the rapid increase in the number of mobile apps available that use Transport for London data.

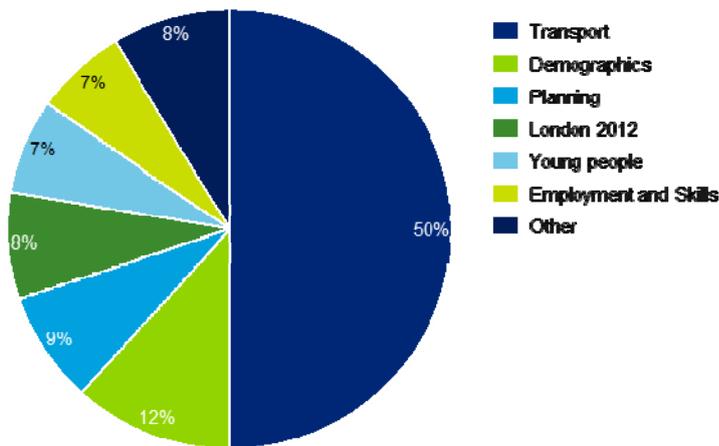
Figure 3.24: most requested data categories



Source: Deloitte analysis of London.data.gov.uk data requests

When the number of ‘wants’ received by request is analysed, transport data is even more popular, with 50 per cent of ‘wants’ in the sample. This is testament to the perceived value of transport data to users – although it may also reflect lack of awareness as to the possibilities offered by other, less obviously valuable types of data.

Figure 3.25: most popular data categories by number of ‘wants’



Source: Deloitte analysis of London.data.gov.uk data requests

Public sector information customer types

While this report has not conducted in-depth primary research to explore the details of how different customers use and re-use public sector information, it is possible to create seven archetypes of customers based on the available evidence¹¹¹. These groups are summarised in Figure 3.26 below.

¹¹¹ It has not been possible to specify, in most cases, precisely what types of public sector information different types of customers use and re-use, nor the revenue they generate from this data. Nor, in most cases, is it possible to discern exactly how businesses use public sector information, i.e. its role as an input into products and services as part of the value chain. This is largely because users, for reasons of commercial sensitivity, have been unwilling to share such details.

Figure 3.26: public sector information customer archetypes

	 Large data services company	 Independent SME app developer using public sector information	 Independent SME focused on efficiency solutions	 Individual user	 Community Groups (incl. not-for profits)	 Other non-data companies	 Public Sector
Number of employees	<ul style="list-style-type: none"> Large – likely to be over 1,000 	<ul style="list-style-type: none"> Small – likely to be up to 10 staff 	<ul style="list-style-type: none"> Small – likely to be up to 10 staff 	<ul style="list-style-type: none"> Individual members of the public 	<ul style="list-style-type: none"> Volunteers or paid staff 	<ul style="list-style-type: none"> Varies 	<ul style="list-style-type: none"> Varies
How public sector information is used / re-used	<ul style="list-style-type: none"> As a contributory input into larger data analyses 	<ul style="list-style-type: none"> As the key input into products and services 	<ul style="list-style-type: none"> As the key input into products and services 	<ul style="list-style-type: none"> As an input into data discovery and own analyses 	<ul style="list-style-type: none"> As an input into data discovery and own analyses As the key input into products and services 	<ul style="list-style-type: none"> As a minor input into day to day activities (extent varies) 	<ul style="list-style-type: none"> As a minor input into day to day activities (extent varies)
Economic value generated	<ul style="list-style-type: none"> Contributes to profits generated New products and services 	<ul style="list-style-type: none"> Contributes to profits generated New products and services 	<ul style="list-style-type: none"> Contributes to profits generated New products and services Efficiency savings across public and private sector 	<ul style="list-style-type: none"> Contributes to improved choice and decision making, leading to savings 	<ul style="list-style-type: none"> Contributes to improved choice and decision making, leading to savings Efficiency savings across public and private sector 	<ul style="list-style-type: none"> Contributes to profits generated New products and services Efficiency savings across public and private sector 	<ul style="list-style-type: none"> New products and services Efficiency savings across public sector

Source: Deloitte analysis

As noted earlier, almost every aspect of the economy will use public sector information in some respect (either directly as an input into a production process or indirectly as a final consumer using products and services built on or around public sector information); it is therefore probably true to say that these archetypes are not absolutely comprehensive in their coverage of all aspects of public sector information use and re-use. Nonetheless, based on this report's survey of information users, they capture the most important areas where public sector information is used as an input and generates value.

Large data services company

Companies that offer data services use data from a variety of sources, including public sector information, to deliver value-added products to clients. They are likely to offer some or all of the following services:

- **Credit Services:** helping organisations to evaluate the risks and rewards associated with providing credit to consumers and businesses, enabling clients to make better informed lending decisions.
- **Decision Analytics:** providing the analytical skills and specialist software products that enable organisations to increase the speed and quality of their decision-making, helping clients to optimise their lending strategies and to implement changes quickly.
- **Marketing Solutions:** helping organisations to find new customers and to take advantage of opportunities for expanding existing relationships, enabling clients to communicate with prospective customers in the most effective way and with the most appropriate offer.
- **Interactive Enabling:** allowing consumers to view their credit report online and to monitor changes to their credit records.

Depending on the service offered, a range of types of public sector information is likely to be used as an input. This includes information on companies (such as supplied by Companies House), demographic information, and economic forecasts. This information is likely to provide one input among many, and so only a proportion of revenue could be directly traced back to public sector information. In many cases, the products being sold by the large data services company could still be produced, albeit not as effectively. In isolated cases, substitutes may also be found to the public sector information, albeit potentially at greater cost or reduced quality.

Independent or SME app developer

App developers are predominately individuals or small companies¹¹² with the technical skills to create smartphone applications presenting publically available data, including public sector information, to a wider base of non-technical users. They may release these apps free of charge, or for payment; or they may release multiple versions, i.e. a free version to download but containing advertising, and the other requiring payment but no advertisements. An example is the developer Routemaster, responsible for releasing the London Transport Live and London Transport Pro apps using Transport for London data.¹¹³

The apps market as a whole is fragmented among many individuals and small enterprises and individual profitability is not easily known. Nonetheless they are an important conduit to a wider range of users and re-users for public sector information.

The type of public sector information used varies depending on the nature of the app. For example, transport apps will use road and rail data; weather forecasting apps will use data from the Met Office; property price apps will use data from the Land Registry; etc. In many cases there will be no substitute for public sector information as an input, or any substitute may be prohibitively priced or of lower quality. High quality, reliable and competitively priced or free public sector information

¹¹² Though of course, larger companies will also produce apps, but not as their sole activity.

¹¹³ See https://play.google.com/store/apps/developer?id=RouteMaster&hl=en_GB

therefore underpins many app products, and is likely to play an increasing role as more public sector information become available and the app market matures.

SME focussed on policy efficiency solutions

SMEs using public sector information and other information sources to develop products that aid policy efficiency can usefully be grouped into a third archetype.

An example is ELGIN, which via its service roadworks.org publishes streetworks and other highway information on the web. It is designed to facilitate coordination of activities between neighbouring authorities and statutory service providers, such as utility companies, enabling them to reduce road space occupation and to meet their statutory obligations under the Traffic Management Act.¹¹⁴

The ELGIN business model is a good example of a shared geospatial 'cloud-based' service, where, for a relatively small upfront subscription, joining authorities get access to shared services and obviate developing and supporting their own computer system. The approach also supports interoperability and data sharing, with all information being presented in a common format on a single website.

In this instance public sector information (transport information, in this case) is critical to the business model of roadworks.org, without which it could not exist. This is therefore an example of a business that has been enabled through the release of public sector information, and which acts as an 'enabler' to facilitate improved use of this information (see below).

In many cases, business models may rely on combining public sector information with other sources of information. For example, Honest Buildings intends to combine public sector information such as the Energy Performance Certificates database (environmental/energy information) with other sources of information on properties and real estate services, helping property owners to realise energy savings through improvements to their properties.¹¹⁵ In these instances the business could often exist without the public sector information, but the value of its offering is greatly enhanced by access to this information. This is therefore an example of an area where government could add value and aid the development of new markets through targeted release of key public sector information datasets (the Energy Performance Certificates database being a key current example).

Individual users

Individuals with the technical skills and interest to work with data may decide to use their skills to derive insights from data (research or other purposes) or present it in a form that is more accessible to non-technical users. For example, upon the release of road construction project data by the Department of Transport in Edmonton, Canada, a local application developer decided to create a mobile app for smart phones and similar devices to access the map interface. He saw the data as useful for the population and decided to make a contribution to his new home city, not as a commercial venture. He reported that creating the app was a relatively straightforward task for him, due in part to his extensive development experience and the high quality of the data set and metadata provided by the city, indicating that these individuals often lack the resources to exploit data if it is not presented in a suitable format.¹¹⁶

In these cases, the monetary value generated by the use of public sector information may be small or non-existent, but the product may produce wider social and economic benefits. In general, availability of public sector information appears to be a precondition for the emergence of many of these-user generated products, as alternative sources of information may not exist or may be prohibitively priced. However, this varies on a case by case basis.

¹¹⁴ See <http://roadworks.org/>

¹¹⁵ See <http://www.honestbuildings.com/>

¹¹⁶ Centre for Technology in Government, 2012. *The Dynamics of Opening Government Data: A White Paper*. http://www.ctg.albany.edu/news/press_ogsap2_20121204

Not for profit organisations

Not for profit organisations may use public sector information for a wide range of purposes, including holding elected officials to account, strengthening civil society, and providing means for citizens to monitor and influence areas of interest to them such as the environment and local services. As such public sector information use or re-use by these organisations does not necessarily have a revenue impact, but is likely to have wider social, economic or public interest benefits. Their task would in many cases be impossible to perform without access to public sector information. The benefits arising from their actions can therefore in many cases be largely attributed to the availability of public sector information.

General data user (non-specialist company)

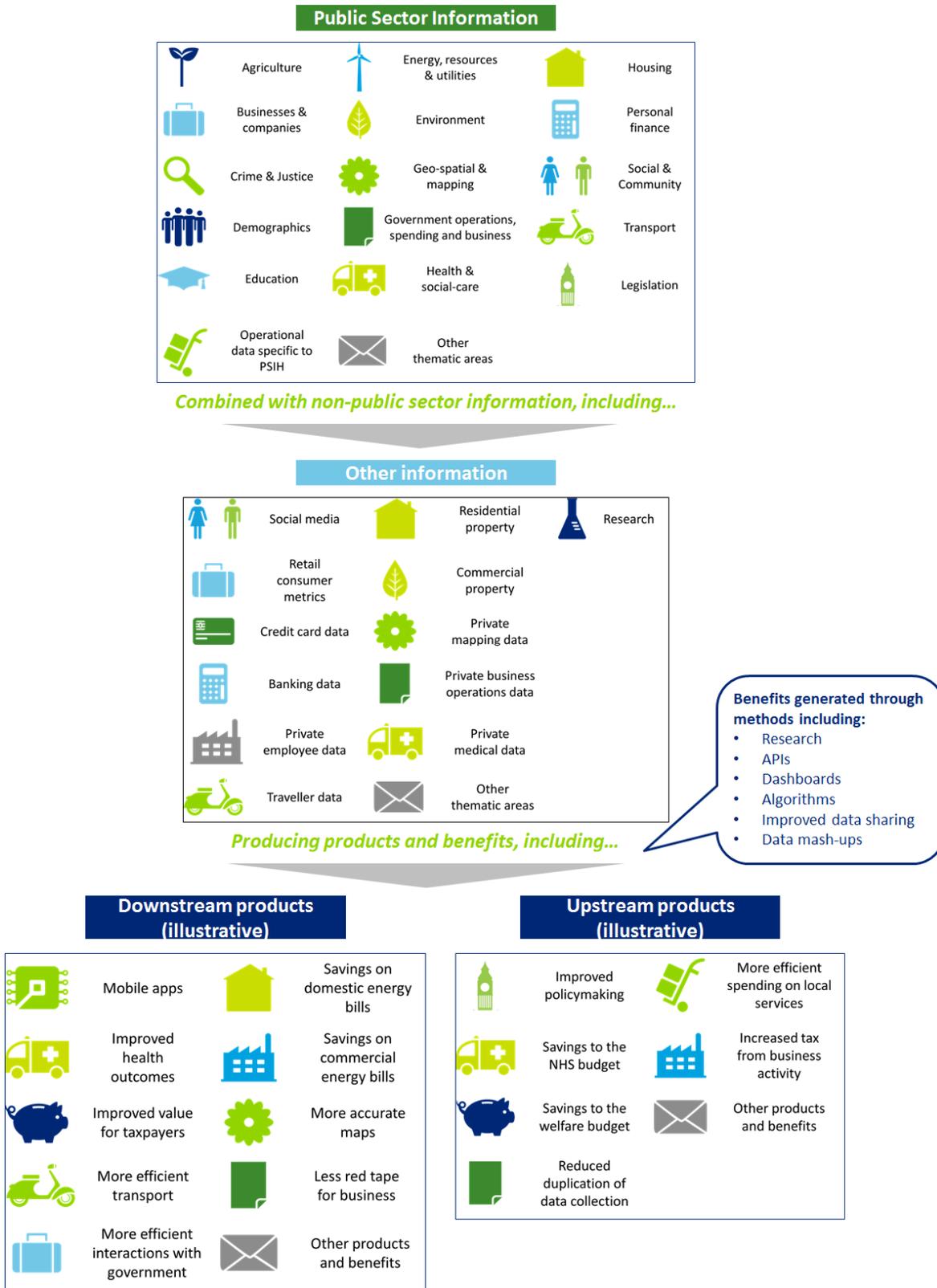
Many companies, notably professional services firms, use public sector information as a matter of course as a key source to inform their work. An example of public sector information commonly used by such companies is ONS data on the economy. The conduct of their operations would become more difficult and more expensive were public sector information unavailable. However, while these firms work with public sector information and other data on a day to day business, they differ from data services companies in that they do not explicitly sell data services. The proportion of their revenue attributable to public sector information is therefore likely to be lower than for the data service companies, and it is likely that they could turn to alternative sources of information were the public sector information unavailable. In these cases public sector information generally improves the product or service offered, rather than being integral to its existence.

Public sector bodies

Public sector bodies not only produce but also use public sector information in their day to day operations. Uses vary widely, ranging from use of address data by local authorities to use of economic data by central government departments. Public sector information use by public sector bodies does not necessarily generate revenue but may assist with achieving more efficient and effective operations, leading to cost savings and improved outcomes.

The ways in which these different archetypes generate value from public sector information is shown below. Value can be generated through its independent use or re-use or by combining public sector information with other data sources to become important inputs into products, services, decision-making and other outcomes. As noted above, public sector information acts as an input as a data point for complex algorithms and analyses, as a source of insights (perhaps summarised as data dashboards or visualisations), as API feeds for apps and so forth.

Figure 3.27: public sector information value chain



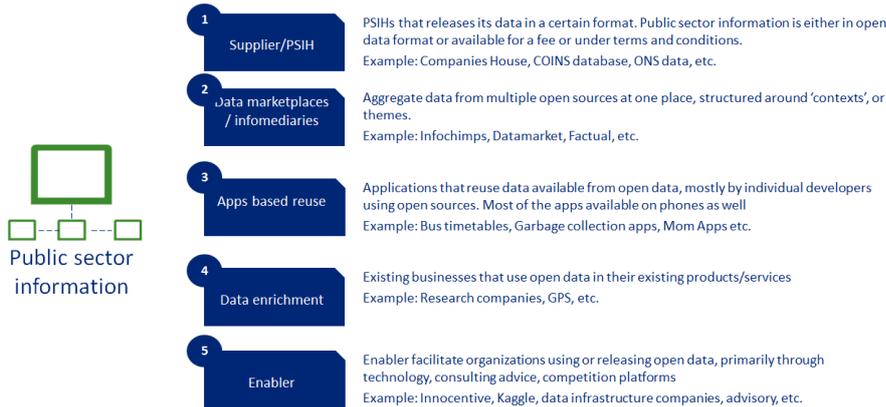
Source: Deloitte analysis

The following section focused on a particular group of business models used by public sector information intermediaries, who can be found in the SME and large data service companies archetypes.

Different business models of public sector information intermediaries

The ways in which different intermediaries use and re-use public sector information varies by their business models. To recap, the public sector information value chain includes suppliers (PSIHs), intermediaries (enablers and infomediaries) and final consumers. Figure 3.28 summarises the different business models currently in operation across the public sector information market.

Figure 3.28: public sector information and open data business models



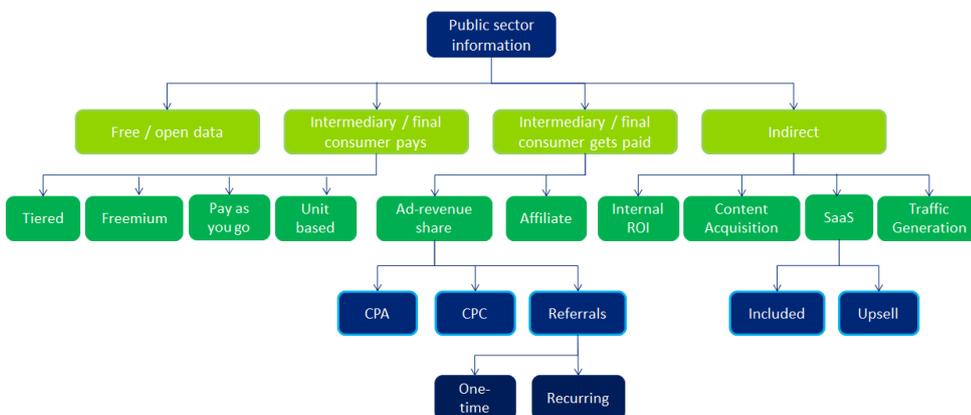
Source: Deloitte analysis

As Figure 3.28 shows, there is clearly overlap in business models that rely upon open data more generally with those that rely specifically on public sector information (either as open data or otherwise). The business models here are a way of deriving economic benefit out of public sector information and open data. Many of these might be interrelated, such as, in some cases, Open API is the data feed for Apps based reuse, and hence an Enabler. The following sections briefly outline each business model, as well as the economic impact arising from each (the relationship between economic growth and public sector information and open data is explored in much more detail in the next chapter).

PSIHs as public sector information suppliers – business models

The collection, retention and distribution of public sector information is not without cost – though the level of cost will vary by dataset and PSIH. In some cases, PSIHs will seek to recover these costs from intermediates and final consumers – indeed, they are often incentivised to do so in terms of legislative requirements or other business objectives. The different ways in which public sector information can be monetised are shown below.

Figure 3.29: public sector information monetisation models



Source: Adapted from Programmerweb, <http://programmersweb.blogspot.co.uk/>

The choice of which monetisation model to use will depend on each PSIH's strategic objectives and the type of public sector information in question. For example, Trading Funds are explicitly required to make a return on data collected, retained and disseminated.

There are a number of benefits to releasing public sector information for PSIHs regardless of the monetisation model. Through open application programming interfaces (APIs) PSIHs can improve their brand visibility, explore new distribution channels, extend innovative external product development, and tap into new communities.

Intermediaries as data marketplaces – business models

Intermediaries acting as data marketplaces or infomediaries have very successfully monetised public sector information and other open data as a third-party reuse, mostly through freemium model or advertisement revenues. Figure 3.30 provides some examples of different business models.

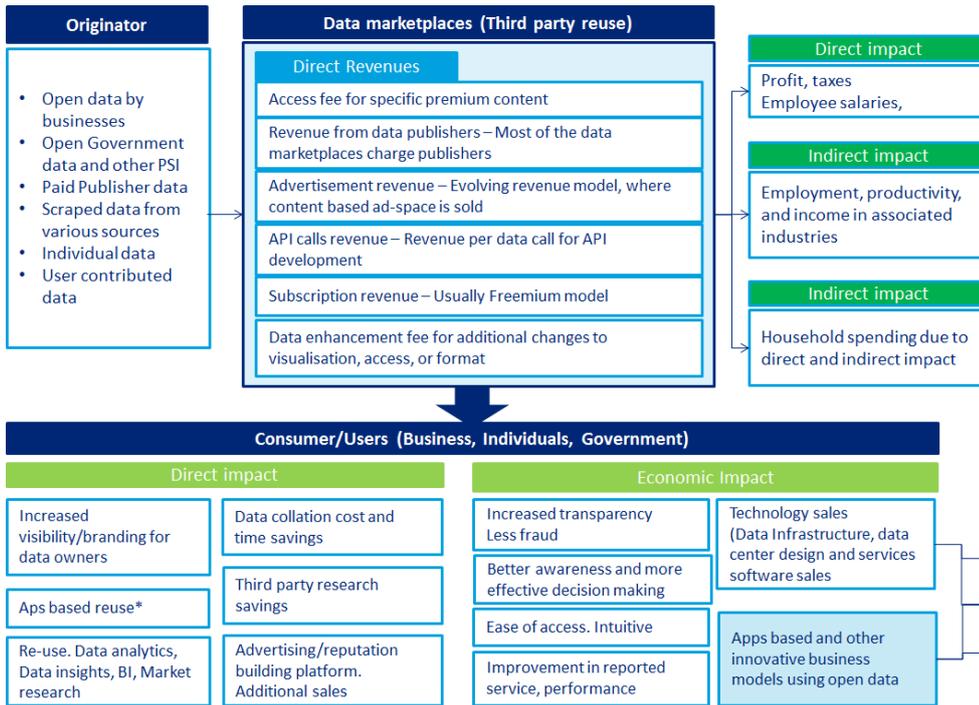
Figure 3.30: data marketplaces business models

Website	Data Domain	Monetisation	Sources
Factual	High quality local data Travel, finance, sports, autos, movies, music, TV, books, health, food, politics, education, science, arts	Currently free Pay-per-use pricing for every API call with subscription Integrate third party deals in apps	Open data sources, community contribution, web crawling
Infochimps	General public sector information	Charge data sellers Charge data buyers	User submitted data Data from public sources
Datamarket	Public sector information Statistical data	Charge data sellers Charge data buyers	Datasets from public sources User submitted datasets
Freebase	General public sector information	Paid for higher volumes of data calls	Community curated data Public datasets imports
Kasabi	All-purpose public sector information, including BBC linked data, Geonames	Charge data consumers	Public datasets User submitted datasets
Timetric	Public sector information economic data	Free public datasets Paid exclusive datasets	Public sources of economic data User uploaded datasets
Xignite	Financial data	Paid subscription	Aggregate data from leading financial data sources, public datasets, and user uploaded data

Source: Deloitte analysis

Chapter Five discusses the ways in which the use and re-use of public sector information can generate impact, but focusing just on data marketplaces and infomediaries, one can see how other businesses, individuals and the public sector benefit from their products and services - not just the data marketplaces and infomediaries themselves, through increased profits.

Figure 3.31: impact of data marketplaces



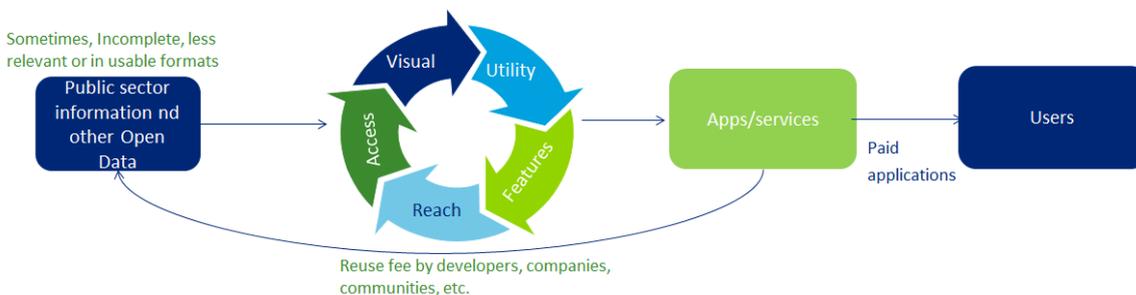
Source: Deloitte analysis

Apps based re-use – business models

App developers take advantage of refined public sector information and other open data processed by infomediaries and enablers to produce apps for end-users – either for their smart phones, tablets, laptops or other devices. The barriers to entry into the app development market are very low and there are a range of established distribution channels available.

Some apps are made available for free, while others are available upon a fee or on a freemium basis. Consumers’ willingness to pay for apps will ultimately depend on their expected utility from using these apps. For example, though people can access *The Guardian* as a mobile site on their iPhone, The Guardian iPhone app is among the top paid application. Similarly, citizens in San Francisco pay for more reliable and more interactive iBARTLive (Bay Area Rapid Transit System), although it is available as a mobile site, and is integrated with Google Maps. It may also be that willingness to pay for apps exceeds willingness to pay for websites, owing to the greater convenience.

Figure 3.32: app value-chain

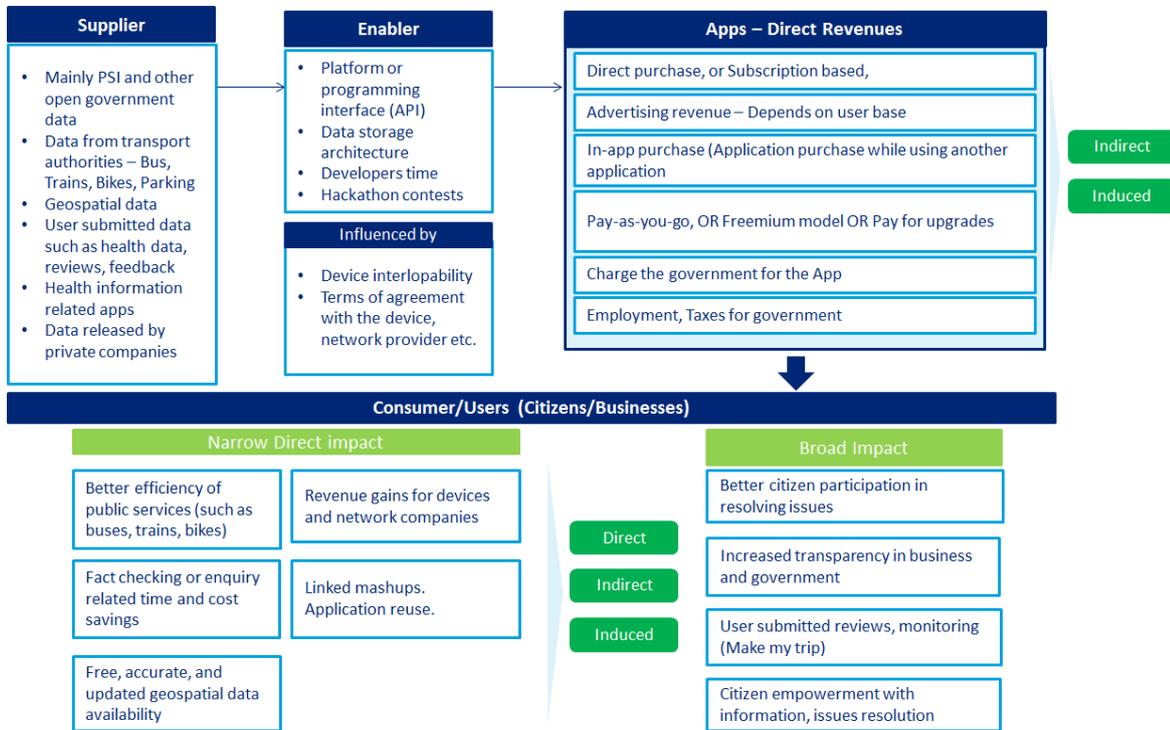


Source: Deloitte analysis

Figure 3.33 above shows the value-chain for app development from public sector information and other open data. Clearly, there are significant benefits to be had from app-based public sector information and open data re-use. Beginning with narrow effects, businesses creating apps can create new jobs and generate wealth. This in turn can generate supply-chain or indirect impacts: changes in the number of jobs and income in associated industries that supply inputs to

organisations developing apps. Finally, there will be induced or consumer impacts through spending by households associated with businesses creating apps and the supply chain that results in changes to jobs and income. Additionally, there may be broad effects such as apps facilitating greater participation in policymaking or more informed choices.

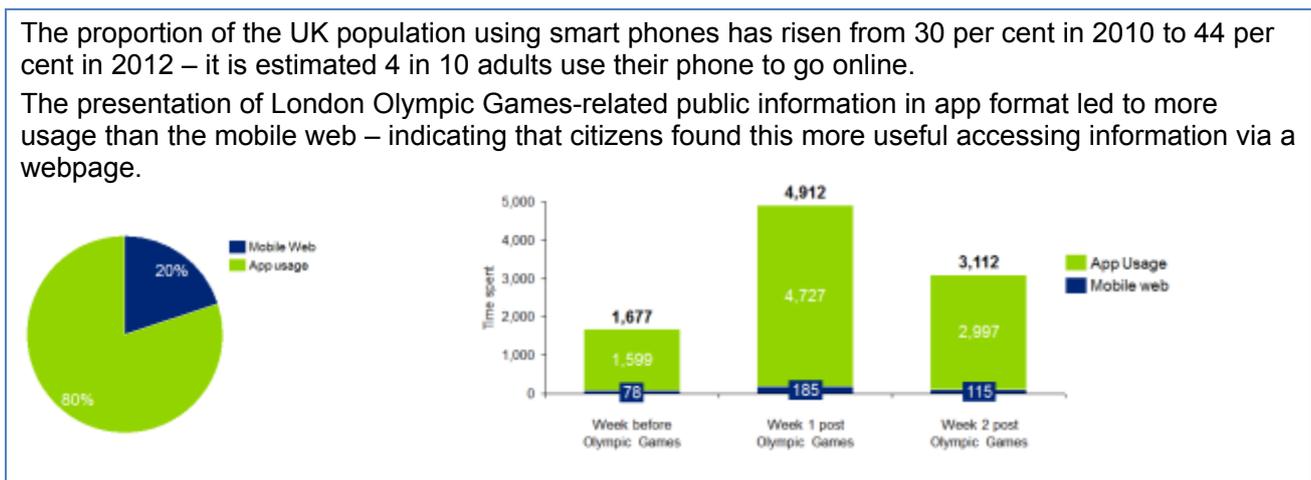
Figure 3.33: impact of app re-use



Source: Deloitte analysis

Public sector information and open data backed 'Apps' have already reduced fact-checking costs/time considerably, and led to increase in citizen awareness and partnership, with the case study below providing an example of this.

Figure 3.34: London Olympic Games – public sector information app use



Source: Adult Media Use and Attitude Report, 2012, Ofcom; Neilson

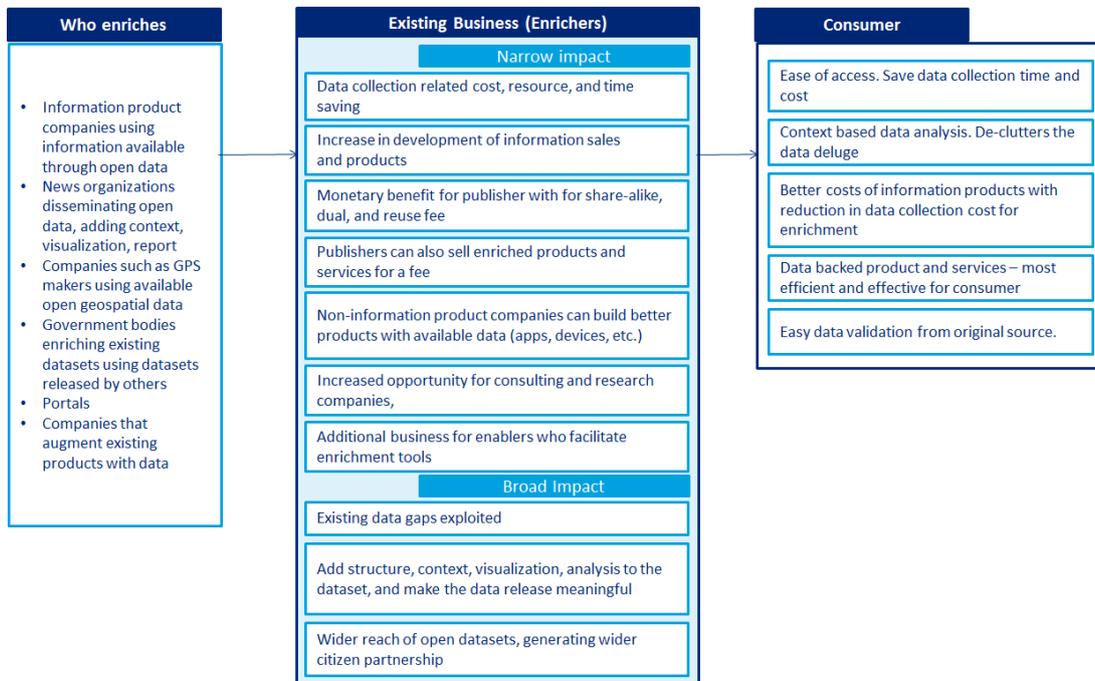
Data enrichment – business models

Data enrichment refers to those consumers who take advantage of public sector information to improve and expand their existing products and services. In its broadest sense this could be improved decision-making with reference to the latest economic and demographic statistics published by the ONS, or more specifically, it could be the embedding of newly available

educational datasets to refine existing educational products focusing on those features that truly improve outcomes.

The types of business which ‘enrich’ data and their end consumers are summarised below in Figure 3.35.

Figure 3.35: data enrichers



Source: Deloitte analysis

As can be seen from Figure 3.35, data enrichers can be found across all aspects of the economy.

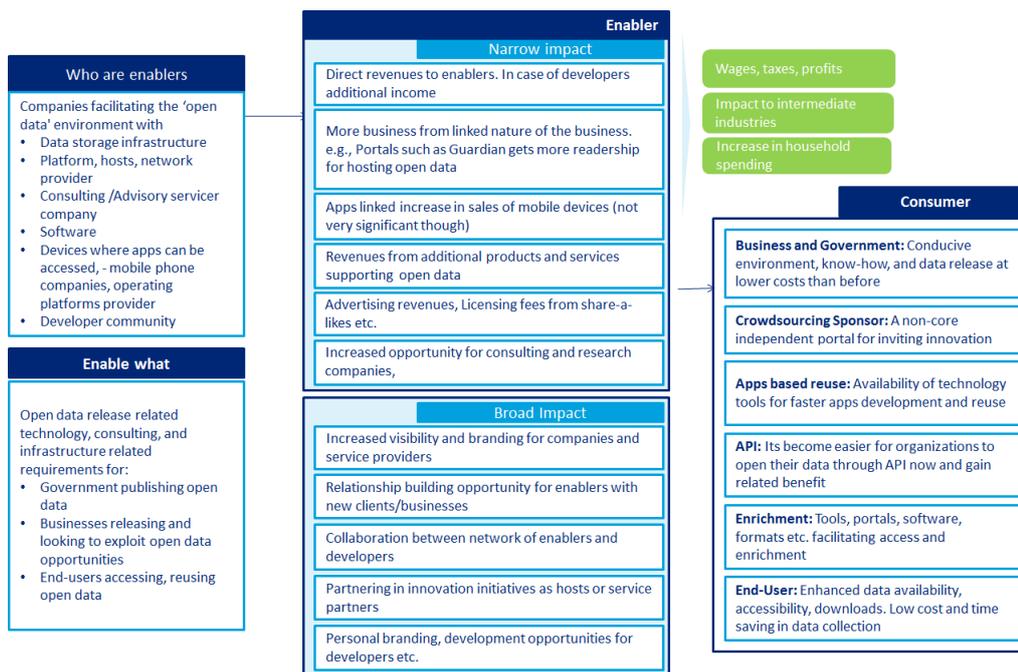
Data enablers – business models

The final category of business model this report has identified is data enablers, who are another group of intermediaries. Data enablers are businesses that facilitate the public sector information and open data environment through:

- data storage infrastructure;
- platform, hosts, network provider;
- consulting /advisory services;
- software provision;
- providing devices where apps can be accessed, - mobile phone companies, operating platforms provider; and
- the developer community.

Through their work, enablers allow other businesses, government and individuals to more readily exploit public sector information.

Figure 3.36: data enablers



Source: Deloitte analysis

Data scientists

Individuals using and re-using public sector information on behalf of their organisation or for personal use are likely to have a range of analytical skills and abilities. The types of roles that are developing around information and data exploitation are well demonstrated by the example of the emerging role of the 'data scientist'.

Figure 3.37: profession profile: data scientists

A key profession in the world of open data, public sector information and data analytics is the 'data scientist', which was recently described by The Harvard Business Review as "the sexiest job of the 21st century". The data scientist is described as "a high-ranking professional with the training and curiosity to make discoveries in the world of big data."

Although statisticians and experts on quantitative analysis have long existed, data scientists differ from these existing professions in a number of important ways. As well as being able to work with large volumes of structured and unstructured data, they are able to translate these analyses into policy and commercial-ready insights and effectively communicate them to a range of stakeholders, often using innovative tools and visualisations. Key to this is the ability to "identify rich data sources, join them with other, potentially incomplete data sources, and clean the resulting set." In many ways they therefore resemble scientists more closely than traditional data analysts.

Data scientists are generally drawn from the ranks of graduates in numerate and scientific fields, including mathematics, economics, and computer science. Scientific qualifications with a less immediately obvious relevance to data science, such as ecology and astrophysics, have also generated successful data scientists, as many of the core traits of practitioners in these fields are similar: an ability to generate and test hypotheses and the curiosity to penetrate to the heart of a problem.

The data scientist is a best thought of as a "hybrid of data hacker, analyst, communicator, and trusted adviser." Keys skills required for the role of data scientist include:

- The ability to write computer code
- A foundation in maths, statistics and probability
- Skill in communicating the 'story' the data tells to a non-technical audience
- Ability to display complex information visually

The key traits of a typical data scientist may include:

- Curiosity
- Empathy and a feel for the key issues
- Creativity and imagination

Source: Harvard Business Review, 'Data Scientist: The Sexiest Job of the 21st Century' (October 2012);

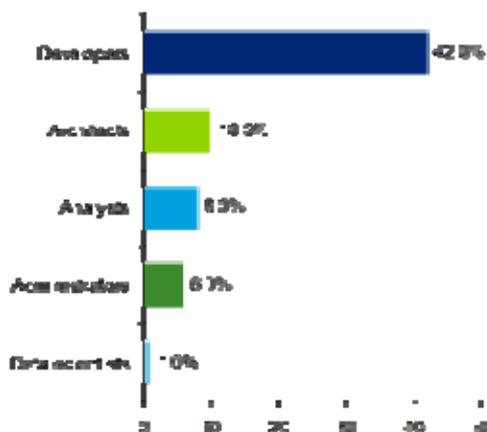
Examples of businesses directly employing data scientists include:

- **GE** uses data science to optimize the service contracts and maintenance intervals for industrial products;
- **Google** uses data scientists to refine its core search and ad-serving algorithms;
- **Zynga**, a games company, uses data scientists to optimize the game experience for both long-term engagement and revenue; and
- **Kaplan**, a test preparation company, uses its data scientists to uncover effective learning strategies.

Based on ONS figures for 2011, around 1.5 million workers in the UK, representing around five per cent of the active workforce, are employed in job categories that are likely to involve elements of the role of the data scientist¹¹⁷. These figures are indicative only, owing to limitations with the ONS data and the use of SOC occupational classifications, but indicate how widespread the requirement for data analysis skills has become. The importance of these workers to the UK economy is likely to be greater than the numbers indicate as they are usually employed in high-skilled, high-productivity, and high-paid job categories.

A recent report¹¹⁸ has estimated current and future demand for staff with data skills. It found that data scientists, while important, accounted for less than 1 per cent of all advertised positions for big data staff in the third quarter of 2012. However, when the analysis expanded to consider other roles not labelled data scientists, but likely to perform related roles (i.e. developers, IT architects and analysts), these were among some of the most commonly advertised roles.

Figure 3.38: most commonly advertised big data roles, Q3 2012 (% of total)



Source: 'Big Data Analytics' (January 2013)

The report found an observable pay premium for big data staff, with salaries around 20 per cent higher than those for IT staff as a whole. This reflects the rapid increase in demand for big data

¹¹⁷ These categories include: science, engineering and technology professionals, IT professionals, statisticians, economists, management consultants, actuaries, research managers and architects and systems designers.

¹¹⁸ e-skills UK, 'Big Data Analytics: an assessment of demand for labour and skills, 2012-2017' (January 2013)

staff and a potential shortage of skills: the report found that demand has risen by 912 per cent over the last five years. This includes a 350 per cent rise in demand for data scientists. This growth is projected to continue: depending on the growth scenario adopted, the report anticipates growth in demand for big data staff of between 13 per and 23 per cent per annum. Taking a midpoint of 18 per cent would mean the generation of 28,000 gross job opportunities per annum by 2017, with 132,000 gross job opportunities created between 2012 and 2017.

Analysis of ONS SOC occupations that involve an element of data science for this report found that the average annual median wage of these 'data science' workers was over £36,000 in 2011 which compared favourably to the national annual median wage of £26,000 in 2011.

Chapter summary

- While data on the total number of public sector information datasets is not recorded, a review of selected data portals suggests the number could exceed 37,000.
- Some of the key suppliers of public sector information, in terms of number of datasets, include the Office for National Statistics, the Department for Communities and Local Government and the Health and Social Care Information Centre. Collectively, the top ten publishers of public sector information make up 60 per cent of all datasets on data.gov.uk.
- Datasets on government spending are among the most available datasets.
- Levels of awareness of public sector information vary across staff in PSIHs. Over 50 per cent of civil and public servants, according to one survey, did not know how to access datasets, how to interpret them or best apply data standards.
- The demand side of the market includes infomediaries and final consumers. Business models include developing apps that use/re-use public sector information, data enrichers, data enablers and data marketplaces.
- Building on existing analyses, it is possible to create a data intensity matrix to compare how different types of datasets are used by different economic sectors in the UK currently. The analysis also shows that the construction, real estate, financial and insurance, public sector and arts, entertainment and recreation sectors are some of the largest users and re-users of open data.
- Dataset types most commonly being used include demographics, economic, environmental and geo-spatial and housing data.
- From an analysis of data.gov.uk dataset requests, the most popular dataset category requested is location (or geospatial) data. This reinforces the impression that this category of data generates some of the highest levels of value. Environment, transportation, health and society are also data that attract particularly high levels of requests.

- As well as being used for commercial and policy purposes, public sector information can be used for scientific research, data journalism and holding public service providers to account.
- Some of the most viewed and requested datasets relate to geospatial and transport information. Of the named barriers identified as preventing access to public sector information, data format and financial charges were among the most commonly mentioned.

4. Public sector information regulatory landscape

This chapter provides an overview of the key stakeholders involved in the governance and policy direction of public sector information, as well as the main regulations and legislation influencing its delivery. It is not intended to be an exhaustive survey of the regulatory landscape but provides a high level overview of the salient points.

Policy and regulatory stakeholders

The public sector information market is complex and evolving. The market also continues to be the subject of a number of policy and other regulatory and legislative initiatives. Many of the policy questions around public sector information link to wider agendas on transparency, accountability and growth - within which there are a range of initiatives. Some of these currently include:

- a number of consultations covering a range of issues from **Freedom of Information, data release guidelines**, the **right to data** and **intellectual property**;
- a number of forums considering public sector information and transparency issues including the **Data Strategy Board**, **Transparency Board**, the **Public Data Group**, the **Open Data User Group**, the **Administrative Data Taskforce**, the **Advisory Panel on Public Sector Information** and others; and
- updating existing policies on data and transparency for bodies such as the Research Councils.

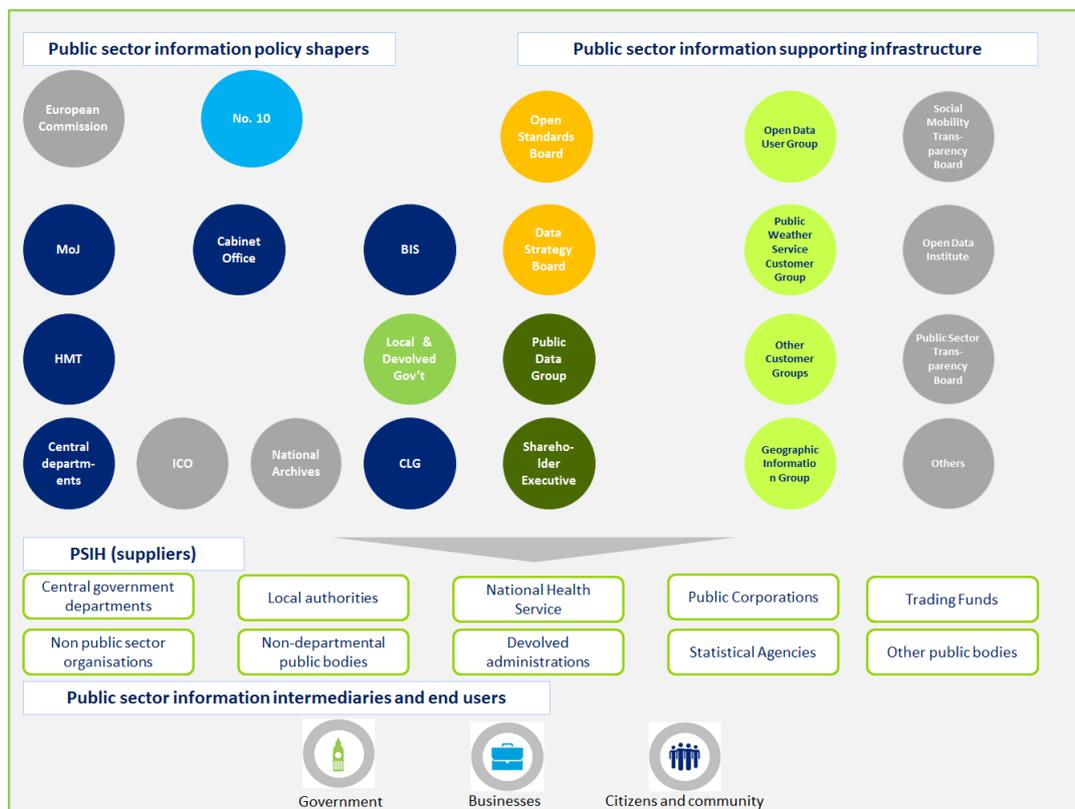
A full list of contemporary initiatives in this area can be found on the Cabinet Office website¹¹⁹.

These initiatives notwithstanding, it is possible to sketch out a provisional schematic for the market that captures the different players and their roles and responsibilities. As well as PSIHs, there are a range of actors involved in the governance of public sector information and open data policy and regulations – all of which have a bearing on how the market develops.

The highly simplified schematic in Figure 4.1 illustrates the number of different policymakers and others involved in shaping the public sector information landscape. Within each PSIH there will be staff directly involved in public sector information policymaking and others more focused on delivery. In a number of cases, policy shapers will also be part of the public sector information supporting infrastructure and may also be public sector information suppliers and consumers.

¹¹⁹ See www.gov.uk/government/policies/improving-the-transparency-and-accountability-of-government-and-its-services

Figure 4.1: simplified public sector information market schematic (not definitive) (March 2013)



Source: Deloitte analysis. * Some policy shapers will also have operational duties, e.g. Cabinet Office is responsible for data.gov.uk

As the market evolves, one might expect increasing numbers of supporting public sector information agencies to emerge, as well as organisations in the private and voluntary sectors, as in the case of the Open Data Institute.

The legal framework for publishing and sharing public sector information¹²⁰

The ability or inability of government departments and other PSIHs to publish and share public sector information they hold is dependent on their possessing one or more of the following:

- **express statutory powers:** these are powers specifically given through legislation;
- **implied statutory powers:** these are powers not specifically given in legislation but which can be implied from it because without these powers the public body would be unable to carry out its functions as specified in legislation; or
- **common law powers:** these are powers that exist outside of legislation. The Ram Doctrine (after Sir Greville Ram, 1945) states that a Crown government department has all the powers of a natural person and does not need to demonstrate any statutory power or authority to carry out an action, unless that action is expressly or implicitly precluded by statute¹²¹.

¹²⁰ Note, this section relies heavily on the report by the Administrative Data Taskforce in December 2012: *Improving Access for Research and Policy*. This is available at: www.esrc.ac.uk/images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf.

¹²¹ Administrative Data Taskforce, 'The UK Administrative Data Network: Improving Access for Research and Policy' (December 2012) (ADT)

There are notable cases where statutory powers restrict or supersede common law powers. For example, the Department of Education shares individual pupil data under the Education Act 1996, which sets the limits for sharing this data.¹²² However, it continues to share some of its aggregated data under common law powers. Consequently there are statutory restrictions on the sharing of individual pupil data which do not necessarily apply to the sharing of the same data in aggregate form. This level of complexity means both that it is not always clear whether existing legal powers (statutory and common law) are sufficient to allow a particular instance of data sharing; and also that even where there is a legal basis to permit sharing or release of data, there can be uncertainty over the legal position leading to public sector bodies ‘playing it safe’ and restricting data sharing. Given that unlawfully sharing information relating to an identifiable person or legal entity is an offence that can carry a prison sentence of up to two years,¹²³ this caution is understandable. Nonetheless, it can create additional barriers to sharing of information even where it is permitted under statute or common law.

Where two or more public sector bodies attempt to share information, the level of complexity is even greater. It may be that one body is legally permitted to share its information but the other is not. In this case the second body may be able to carry out the linking work, but this still raises the question of whether the resulting analysis can legally be shared with the first body. The level of complexity only increases as the number of public bodies involved rises – and many of the most interesting opportunities for linked data may involve information held across the full spectrum of the public sector.

In specific cases a ‘legal gateway’ may be created through legislation in order to explicitly allow the sharing of information. An example is data collected under the Statistic of Trade Act 1947, which may be provided to local authorities owing to subsequently created legal gateways, the Employment and Training Act 1973, as amended by the Employment Act 1988. These allow the sharing of relevant information with the persons listed under section 4(3) of the 1973 Act.

However, as is indicated by the above example, these legal gateways are complex and time-consuming to set up – it has been reported it can take as long as two years, by which time the window of opportunity to enhance a policy through sharing information may well have passed. Having been created, the legal gateway only applies to a very specific instance of information sharing, while the proliferation of such gateways continually increases the legal and regulatory complexity of the public sector information landscape. In order to be clear as to whether a particular instance of information sharing is permitted, an employee of a public sector body requires a substantial knowledge and understanding of legally complex gateway arrangements.

In its recent report the Administrative Data Taskforce recommended the creation of “a generic legal gateway for research access to and linkage between administrative data.”¹²⁴ The rationale is to create a more consistent, efficient and timely legal framework for sharing information, so as to enhance policymaking with evidence-based insights and to ensure that information holders have clear guidance as to when it is appropriate to share information.

Public sector information licensing and the regulatory framework

The UK Government Licensing Framework

Public sector information provision is governed by the UK Government Licensing Framework (UKGLF), overseen by The National Archives. It provides “a policy and legal overview of the arrangements for licensing the use and re-use of public sector information both in central

¹²² At the time of writing, the DfE was consulting on new access arrangements to the National Pupil Database.

¹²³ ADT (2012), p. 16

¹²⁴ ADT (2012)

government and the wider public sector.”¹²⁵ The key licences comprising the UKGLF are summarised below.

Open Government Licence

The Open Government Licence (OGL) was introduced in 2010 as a licensing mechanism to facilitate the use and re-use of public sector information. It is designed to impose as few conditions as possible on users, permitting them to:

- Copy, publish, distribute and transmit the information;
- Adapt the information; and
- Exploit the information commercially, including by combining it with other information and including it in a product or service.

Notably, information published under the OGL does not require users and re-users to register or pay any charge.

The main condition of use for information published under the OGL is that users must acknowledge the source of the information by providing any attribution statement specified by the information provider.

Other conditions and limitations are that:

- Users may not use the information in a way that suggests official status, or that they have the endorsement of the information provider;
- Users may not mislead others, nor misrepresent the information or the information provider; and
- Users must ensure that their use of the information does not breach the Data Protection Act 1998 or the Privacy and Electronic Communications Regulations 2003 (EC Directive).

The OGL also limits the liability of the information provider by specifying that the provider is not liable for any errors or omissions in the information, and does not guarantee the continued supply of the information.

The OGL replaces the previous Click-Use Licence as the default licence for a wide range of information owned by the Crown, including information previously made available under the Click-Use Licence and source code and software originated by the Crown under Framework 1 of the NESTA agreements. It is designed to be interoperable with other widely used models including the Creative Commons Attribution Licence and the Open Data Commons Attribution Licence. It is increasingly widely used across the entire public sector, and its use is not confined to Crown bodies. For example, over 270 local authorities currently use the OGL.

Source: The National Archives

Non-Commercial Government Licence

Although the default position is that information should be published under the OGL, there are circumstances under which it is not appropriate to allow commercial use of information. In these circumstances information providers can publish under the Non-Commercial Government Licence.

This licence allows the use and re-use of information free of charge, permitting users to:

- Copy, publish, distribute and transmit the information;
- Adapt the information; and
- Combine the information with other information.

The main restriction on the use of information published under this licence is that it may not be used for purposes intended to confer commercial advantage or private monetary compensation. These restrictions also apply to any onward licensing of the information, for example when combined with other information. As with the OGL, users are required to acknowledge the source of the information.

Source: The National Archives

¹²⁵ See www.nationalarchives.gov.uk/information-management/government-licensing/the-framework.htm

In certain circumstances public sector bodies are permitted to charge for use and re-use of information on the basis of cost recovery and a reasonable rate of return. The most notable examples of this are the Trading Funds, although they are not the only examples.

In such cases, the specific terms of the licence vary depending on the requirements for the product or service in question. However, the National Archives is in the process of developing a standard charged licence under the UKGLF, currently in beta version and undergoing development at the time of writing. The aim is to “provide a straightforward set of terms which deliver an effective standard approach.”¹²⁶ This is intended to provide greater consistency and transparency around licensing conditions where information is charged for, both for the purchaser and the provider of the information, and thereby reduce costs for the provider and increase the accessibility of the information for users.

An additional licence is the Open Parliament Licence, which is designed to enable parliamentary material to be shared and used in a manner consistent with the principles of the OGL.¹²⁷

Crown copyright and the Information Fair Trader Scheme

The National Archives manages Crown copyright by licensing material, and by granting delegations of authority to certain public sector bodies for certain information to allow them to license the material they hold. Delegations are only offered to certain organisations, notably trading funds; other organisations must write a business case setting out why this is appropriate.

This is regulated through the Information Fair Trader Scheme (IFTS), which is designed to ensure that re-users of information are treated fairly by public sector information providers. IFTS accreditation is based on a full audit of information trading activities and is intended to show that “their processes and policies are compliant and consistent with government policy on information trading and that they meet the needs of existing or potential customers.” All public sector bodies that license information can apply to join the IFTS. Organisations currently accredited include Land Registry, Ordnance Survey, the Met Office and Companies House, among other trading funds.¹²⁸

The IFTS is based on the following principles:¹²⁹

- **Maximisation:** the default position should be that information can be re-used, in the absence of strong reasons to the contrary;
- **Simplicity:** regarding processes, policies and licences;
- **Innovation:** actively facilitate the development of new and innovative forms of re-use;
- **Transparency:** regarding terms of re-using, charging details, and what information is available;
- **Fairness:** re-users should be treated in a non-discriminatory way, and information holders should not use their position to compete unfairly; and
- **Challenge:** a robust complaints process.

Exceptions to marginal cost pricing

As discussed above, most public sector information is made available at marginal cost, which in practice usually means free of charge, especially where the information is provided online. However, there are exceptions to this. Where a PSIH wishes to charge above marginal cost for information, it must make a business case to the Office of Public Sector Information¹³⁰ under the

¹²⁶ See www.nationalarchives.gov.uk/information-management/government-licensing/charged-licence.htm

¹²⁷ See www.parliament.uk/site-information/copyright/

¹²⁸ See www.nationalarchives.gov.uk/information-management/ifts/full-accreditation.htm

¹²⁹ See www.nationalarchives.gov.uk/information-management/ifts/principles.htm

¹³⁰ The Office of Public Sector Information operates from within The National Archives and has a statutory role under the PSI Regulations for the investigation of complaints.

exceptions process. An application will generally be made if an information holder has added value to the information and therefore wishes to treat it as a commercial product.

The case for an exception is judged on the following criteria:¹³¹

- Is it essential to produce the information as part of government's core duties and therefore vital to the workings of government?
- Is the information directly funded by the taxpayer, either through it being collected for the purposes of government or produced with the purpose of informing the public?
- Is the Department or Agency the sole producer of the information or can the information be obtained from other sources?
- What effect would charging for the information have on the level of re-use?
- Is the Department or Agency able to provide a statement of commitment to Information Fair Trader principles signed by its Permanent Secretary or Chief Executive?

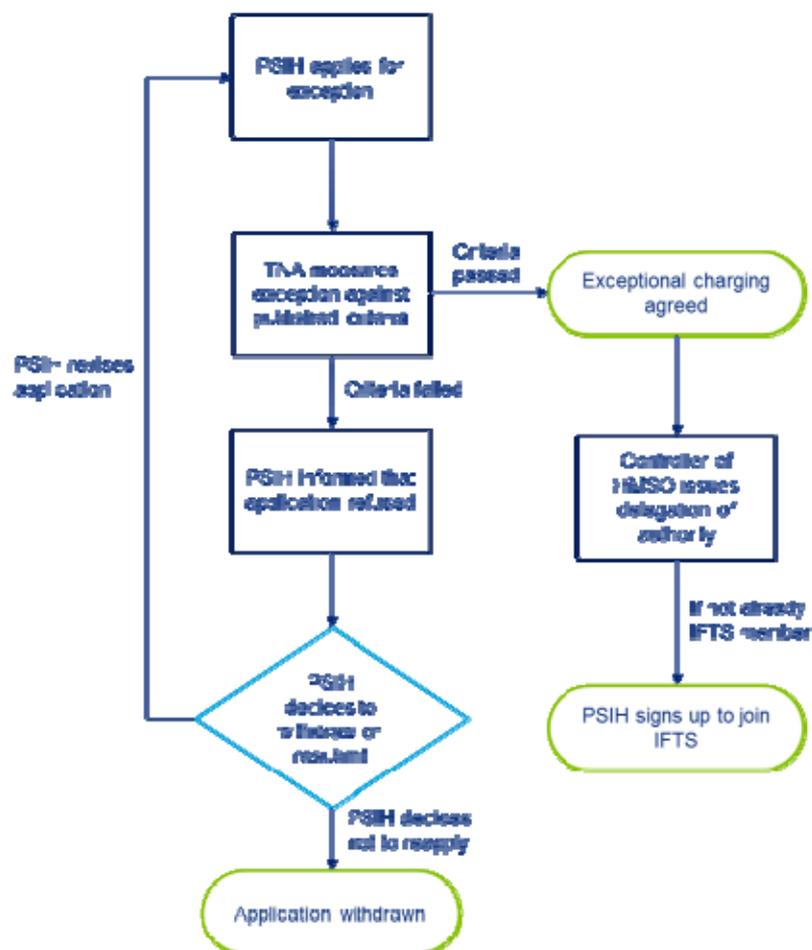
These criteria are designed to ensure that charging above marginal cost does not compromise the core principles of public sector information policy, especially:

- maximisation of access, re-use and innovation;
- openness and accountability; and
- competitive neutrality.

The below schematic illustrates the exceptions application process.

¹³¹ See www.nationalarchives.gov.uk/documents/information-management/criteria-exceptions-marginal-cost-pricing.pdf

Figure 4.2: the exceptions application process



Source: adapted from www.nationalarchives.gov.uk/documents/information-management/how-we-deal-with-exception-application.pdf

The Data Protection Act

The Data Protection Act 1998 (DPA) was enacted to bring UK law into line with the EU Data Protection Directive 1995, which required Member States to protect citizens' rights and freedoms, including their right to privacy, with regard to the processing of personal data. As such it plays a significant role in the public sector information landscape, particularly with regard to sharing and publishing data which may have personal privacy implications.

The DPA covers any data which concerns a living and identifiable individual. Anonymised or pseudonymised data is not considered personal data and is therefore not covered by the DPA – in this respect the UK differs from several other EU Member States. Where anonymisation is reversible, however, the data does fall within the scope of the DPA.

The individual who is the subject of the data that is held or processed is entitled to:

- view the data for a 'subject access fee';
- request that incorrect information be corrected;
- require that data is not used in any way that could cause damage or distress; and
- require that their data is not used for direct marketing purposes.

The subject of the data must consent to the collection and storage of their information. The data holder has an obligation to inform the data subject the purposes for which it is being used and to whom it has been disclosed, and the data subject may be able to withdraw consent. In any event,

consent is not assumed to last forever – in general it is assumed to last for as long as the data needs to be processed.

In some cases the DPA may be seen as a barrier to the release or sharing of information, or it may be that information holders are unsure of their obligations and therefore err on the side of caution and resist releasing information even where this is permitted. However, from conversations with stakeholders in the information and regulatory communities, notably The National Archives and the Information Commissioner's Office, it seems clear that while the DPA requires information holders to handle information with care: it should not be a barrier to its use and re-use. For example, if the information undergoes appropriate anonymisation its use and re-use is no longer constrained by the DPA, because anonymised information is not considered personal information.

It should be noted that the reform of the EU legal framework on the protection of personal data, proposed by the European Commission in 2012 and currently undergoing negotiation, may eventually replace the DPA with a new EU regulation. The impact on the UK regulatory landscape is currently unclear since the details of any new regulation are as yet unknown.

The case study below is an example of the anonymisation process to which data can be subjected, in order to ensure that it does not contravene the principles of personal privacy or fall within the remit of the DPA.

Case study: avoiding the 'jigsaw effect'

Anonymisation of datasets is a key plank of the transparency agenda. It is particularly relevant where datasets are of high value for researchers, policymakers and the wider public, but also contain personal and potentially confidential details of individual citizens. Publication of such personal details would be a breach of privacy. In order to make these datasets available without compromising privacy, key individual identifiers (names, dates of birth, addresses and so forth) are removed prior to release. High profile examples of these types of datasets include the National Pupil Database, court sentencing data, and data on offenders and reoffenders.

However, it is sometimes possible to 'de-anonymise' data through a process called 'jigsaw identification'. This involves assembling information from other sources which allows re-identification of an individual through a process of triangulation. This process becomes easier the more information is readily available, which is a matter of concern given the ever-expanding volumes of information available through the Internet, especially with the rise of social media and indeed the growth of government transparency programmes. In addition, the growing level of computational power at the disposal of individuals brings the resources required within the reach of an ever greater number of people.

The authors of a recent paper¹³² therefore describe 'best practice' for the anonymisation of sensitive datasets, based on the experience of the Ministry of Justice in anonymising its data on reoffending. As a first step, the data was anonymised by Ministry of Justice statisticians, who aggregated the data into ranges for each characteristic as follows: gender, age, offence, establishment/trust, previous offences, whether reoffended and number of re-offences.

This data was then passed to academics and postgraduate students with relevant experience at the LSE, Royal Holloway and Southampton University. They were asked to identify the anonymised subjects.

One disclosure resulted from this process, out of over 200,000 cases. This was possible thanks to the profile of an offender named on a local news site. In response to this, the MoJ statisticians further aggregated the data by removing information about the offence committed. The data was then passed to data security specialists Detica, where a team confirmed that the data was secure against the jigsaw effect.

By involving three diverse groups of testers, the effectiveness of the anonymisation process was tested more robustly and the MoJ was able to release the anonymised data with high level of confidence that individual privacy had been respected. The involvement of the students was seen as a particularly important element of the test, as their approach resembled that of hackers and they subjected the data to unorthodox approaches, highlighting any unforeseen risks or weaknesses in the anonymisation process.

As a consequence, the MoJ statisticians acquired more detailed knowledge of the properties of the data a greater understanding of the level of aggregation required to preserve the anonymity of the subjects.

¹³² Kieron O'Hara, Edgar Whitley and Philip Whittal, 'Avoiding the jigsaw effect: experience with Ministry of Justice reoffending data' (December 2011)

Case study: avoiding the 'jigsaw effect'

The authors suggest that this method could usefully be applied across government, as a way of ensuring that transparency is accompanied by robust methods and a respect for privacy in which the public can have confidence.

Source: Kieron O'Hara, Edgar Whitley and Philip Whittal, '*Avoiding the jigsaw effect: experience with Ministry of Justice reoffending data*' (December 2011)

The European dimension

In addition to UK statute and common law, the European Union (EU) also plays a significant role in the legal and regulatory framework for public sector information. The main relevant regulations are the European Union Directive on the Re-Use of Public Sector Information (the 'PSI Directive'), and the European Union Data Protection Directive (see box-outs below). It should be noted that both of these are undergoing revision, with the revised versions likely to be adopted during the course of 2013. These debates are ongoing, and whichever of these dynamics becomes the dominant trend at the European level may have a significant effect on the development of the public sector information landscape in the UK.

The European Directive on the Re-Use of Public Sector Information

The breakout box below summarises the current PSI Directive and the proposals for its revision.

The European Directive on the Re-use of Public Sector Information and the proposed changes to it

In 2003 the European Union introduced the Directive on Re-use of Public Sector Information (2003/98/EC) which introduced a common legislative framework regulating how public bodies should make their information available for re-use, including removing barriers and improving transparency. This required public bodies to:

- be transparent on conditions of re-use;
- avoid any form of discrimination between users, including a re-use by the public sector body itself;
- deal with applications for re-use within a set maximum time; and
- not enter into exclusive arrangements, except under exceptional circumstances.

In December 2011 the European Commission, as part of an Open Data package, presented proposals to revise the 2003 Directive.¹³³ These revisions are likely to create a presumption of openness on public sector information, with key changes, according to the EC, including:

- bringing new bodies under the scope of the Directive, including cultural bodies such as museums, libraries and archives;
- limiting the fees that can be charged by public authorities at marginal cost as a rule;
- introducing independent oversight over re-use rules in member states; and
- making machine-readable formats the norm for information held by public authorities

Negotiations are continuing at the time of writing.

The revised Directive is expected to be adopted in Europe in the course of 2013 and come into force through regulations in UK law in 2014-15.

The proposed European Data Protection Regulation

The breakout box below summarises the proposed regulation on data protection.

¹³³ See http://ec.europa.eu/information_society/policy/psi/revision_directive/index_en.htm

The proposed European Data Protection Regulation

In January 2012 the European Commission published a draft General Data Protection Regulation. If adopted, this would replace the current Data Protection Directive (95/46/EC) and any relevant member state legislation, including the UK's Data Protection Act. Unlike a Directive, a Regulation takes direct effect with no need to be transposed into law for each member state.

There are aspects of the Regulation, in its initial draft form as proposed by the European Commission, that have caused concern among holders and users of public sector information. In particular, the draft was seen to favour the right to privacy of individuals and one possible outcome of this is that the Regulation may impose restrictions and conditions on the linking of datasets and use of 'big data', where this involves personal data.

In its proposed form the Regulation introduces new procedures for the processing of personal data. These may increase the bureaucratic burden on organisations and could disincentivise the release of public sector information where this becomes too burdensome. This is significant since the use of anonymised personal data is among the most potentially valuable exploitations of public sector information, both for improving policy and for wider welfare gains.

The Regulation appears to provide derogations from particular requirements for the use of personal data for historical, statistical and scientific research purposes. To qualify for these derogations, personal data must be processed in accordance with Article 83: personal data should be used only if anonymous data is insufficient, and where possible identifying information should be kept separate from other information. This may make the use of public sector information for research and policy purposes easier, although other areas of the proposed Regulation may continue to pose challenges.

The Regulation is currently undergoing consideration and amendment by the European Parliament and Council of Ministers, with adoption likely either during the course of 2013 or possibly in 2014.

The INSPIRE Directive

The Inspire Directive (2007/2/EC), which came into force in May 2007, is intended to enhance availability, access and sharing of spatial information across EU member states. It is undergoing implementation in stages, with full implementation required by 2019. The essential objective is to enable sharing of geospatial information among public sector organisations, and improve public access to this information. The information to which the regulation applies falls with 34 'themes'.¹³⁴

INSPIRE is based on the following principles:

- data should be collected only once and kept where it can be maintained most effectively;
- it should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications;
- it should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes;
- geographic information needed for good governance at all levels should be readily and transparently available; and
- easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.¹³⁵

¹³⁴ Available to view at <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/list/7>

¹³⁵ See <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/48>

This is an example of data sharing best practice which could productively be replicated for other types of public sector information. However, it also indicates the difficulties involved in achieving such agreement. It took six years for the Regulation to be agreed upon; and full implementation was not scheduled until a further 12 years after the date of agreement, in 2019.

Chapter summary

- Despite the public sector information market being relatively young and still maturing, there are a number of key policymakers and policy 'shapers' supporting and influencing its evolution.
- The ability and inability of PSIAs to publish and share public sector information is influenced by express statutory powers, implied statutory powers and common law powers.
- The UK Government Licensing Framework provides the policy and legal overview of the arrangements behind the use and re-use of public sector information and includes an Open Data-style licence.
- The Information Fair Trader Scheme (IFTS) accreditation scheme, which is open to all public sector bodies, has been designed to ensure that re-users of public sector information are treated fairly by PSIAs, and contains principles around complaints handling, transparency and simplicity.
- Exceptions to marginal cost pricing for public sector information can be made in certain cases where specified criteria are fulfilled.
- The Data Protection Act 1998 does not apply to truly anonymised or pseudonymised data and there are a number of examples of how this can be done.
- The key European regulations affecting public sector information in the UK are currently being revised with final versions expected to be adopted later in 2013.

5. The current value of the public sector information market

There is no single recognised methodology for estimating the value of public sector information and previous estimates vary considerably. Further, the lack of data on the number of public sector information datasets, consumers' willingness to pay and usage statistics makes any estimate subjective due to the need to make a number of modelling assumptions.

Nonetheless, based on the available information and following a number of informed, conservative assumptions developed with stakeholders and on the basis of previous research, this report has applied a pragmatic approach to estimate the value of public sector information market in the financial year 2011/12. While it adapts a previously used methodology, it makes a number of changes to reflect the need to capture the entire public sector information market rather than just an individual segment. This means that the estimates presented here are not readily comparable with previous estimates of the value of public sector information in the UK¹³⁶.

The primary focus of the estimation exercise has been to calculate the so-called narrow economic value of public sector information, which captures the direct value to consumers and PSIHS as well as the supply-chain and consumer spending value impacts. However, given that public sector information is known to have significant broader, social value impacts, case study analyses have been used to illustrate this current value.

In some instances, these case studies are augmented with ad hoc calculations that provide an indication as to what the broader societal value of public sector information might be in financial terms.

¹³⁶ For example, the figures estimated here are not directly comparable with the £16 billion figure quoted in the Open Data White Paper.

Understanding the value of public sector information

At the outset, it is important to clearly define what is understood by the term value¹³⁷. Previous studies have considered value in terms of market value (defined as either total turnover or profits accruing to consumers of data and PSIHs), consumer¹³⁸ and producer¹³⁹ welfare and broader (non-monetary) value. Having reviewed the evidence, it appears appropriate to disaggregate value into four distinct components which can capture elements of market, consumer, producer and wider value:

- **the direct value of public sector information to producers and suppliers (the PSIHs):** these are the benefits accruing to producers and suppliers of public sector information through the sale of public sector information or related value-added services;
- **the indirect value of public sector information arising from its production and supply:** the benefits accruing up the supply chain to those organisations interacting with and supplying PSIHs (but not directly using or re-using public sector information), and the benefits accruing to those organisations where employees of PSIHs and supply chain organisations spend their wages;
- **the direct use value of public sector information to consumers of public sector information:** the benefits accruing to businesses, civil society, individuals and the public sector from directly using and re-using public sector information for a variety of purposes; and
- **the wider societal value arising from the use and re-use of public sector information:** the benefits to society of public sector information being exploited, which are not readily captured elsewhere.

The first three types of value can be termed economic value or narrow economic value and can be measured using standard economic methodologies to derive a monetary estimate for value and the associated employment figure (see below).

The final type of value is harder to measure as it captures wider benefits arising from the use of public sector information – these are typically not measured in monetary terms. The literature discusses a number of ways in which public sector information can have broader impacts:

- **increasing democratic participation:** giving citizens and businesses access to public sector information allows them to perform their own analyses of salient issues, make more informed choices about public service providers and interact with policymakers to challenge their assumptions and improve the policymaking process;
- **promoting greater accountability:** for example through the scrutiny of costs of public service provision and benchmarking comparable services;
- **greater social cohesion:** for example, by providing more information on the provision and distribution of services, public sector information can be used to dispel myths on who receives certain public services;
- **generating environmental benefits:** such as reducing congestion and pollution through the release of better traffic and transport data which helps drivers to better plan journeys; and

¹³⁷ This is distinct from 'value-add' services that might be supplied by PSIHs using public sector information as an input.

¹³⁸ This is the value or benefit consumers of public service information enjoy over and above the price they pay for it (including if the price is zero).

¹³⁹ This is the value accruing to PSIHs when public sector information is purchased by consumers. It is typically analogous to profit received.

- **identifying previously unknown links between different policy areas:** through data-mash ups it may be possible to develop system-wide solutions that holistically seek to address the root of policy challenges.

Thus, for the purposes of this report, the total value of public sector information is defined as the benefits accruing to the suppliers, users and re-users (i.e. consumers) of the information and data in terms of profits generated, jobs created and supported (narrow economic value) and the wider benefits to society arising from the exploitation of public sector information.

It is important to clarify that value defined in this way is not directly the same as the market value of public sector information. Market value typically refers to the volume of sales multiplied by the price of the product. In this way it includes labour costs, capital costs and the intermediate costs of production which are not included in the Deloitte definition. Conversely, market value will not capture indirect effects and wider societal benefits.

Modelling approach and assumptions

Appendix 5 contains full details of the modelling approach and assumptions used. This report adopts a three-stage approach to valuation of public sector information for the financial year 2011/2:

- **Stage 1:** estimating the value of public sector information to PSIHs and the value to direct consumers (users and re-users) of public sector information using a bottom up approach that quantifies consumer and producer surplus; and
- **Stage 2:** estimating the value of the associated indirect and induced impacts to PSIHs using Input-Output multipliers.
- **Stage 3:** estimating a 'ready-reckoner' value of wider value based on other available research.

The report does not seek to systematically quantify the value of the broader social impacts of public sector information. There is a lack of reliable evidence on the linkages (correlated and causal) between the consumption of public sector information and democratic, social, environmental and political impacts.

However, notwithstanding this caveat, a third stage of the approach includes a number of case study analyses to illustrate the types of impacts that public sector information can have – and in some cases these include ad hoc calculations to give an indication of the quantum of value that is being generated. This is then used in conjunction with 'uplift' ratios taken in other studies to allow us to generate a broad, order-of-magnitude estimate for wider, and thus the total value of public sector information in the UK.

Again, it should be stressed that the lack of reliable data on public sector information¹⁴⁰ makes the quantification of value difficult and heavily reliant on assumptions. To reflect this inherent uncertainty the estimates contain upper and lower bounds. Further research (and especially data collection) is required to develop more comprehensive estimates.

Stage 1: consumer and producer surplus approach – the welfare approach

The report uses the so-called bottom-up approach developed by the OFT/DotEcon (2006)¹⁴¹. As the OFT/DotEcon report discusses, a bottom-up approach is preferable to a top-down approach given the latter's tendency to over-attribute causality and generate biased estimates. This 'welfare' approach seeks to quantify the consumer surplus¹⁴² derived from the use and re-use of public

¹⁴⁰ Namely, the total number of public sector information datasets held by PSIHs, the number of users by dataset and how users are exploiting public sector information.

¹⁴¹ This report is publicly available and the methodology replicable once the necessary data has been sourced.

¹⁴² Defined as the value or benefit consumers of a product or service enjoy over and above the price they paid for it. It is the difference between the price consumers pay and the maximum price they are willing to pay

sector information and the producer surplus¹⁴³ accruing to PSIHS from the generation, collection and dissemination of public sector information in the financial year 2011/12.

Following the DotEcon approach, the welfare approach can be estimated from summing the current net surplus with the total producer surplus from the supply of public sector information.

$$\text{Net surplus} = \text{Consumer Surplus} + \text{Producer Surplus}$$

There are standard formulae to calculate consumer surplus and producer surplus (see Appendix 5). The calculations use data on revenues attributable to public sector information, number of known public sector information datasets and available download data. Assumptions are made relating to:

- price elasticity of demand for different dataset categories;
- the proportion of users going on to actively use and re-use downloaded/viewed datasets; and
- the shape of the demand curve.

The shape of the demand curve has an important bearing on the size of the consumer surplus. While it may be the case that there are linear demand curves for public sector information that carries a fee, the demand curves may exhibit non-linearities and discontinuities when the public sector information is available free of charge – the level of consumer surplus may be much larger or smaller when the demand curve is not linear.

This modelling exercise has sought to recognise these uncertainties and the upper and lower bounds alter our assumptions across the three above variables.

Further details of the modelling approach taken can be found in Appendix 3.

Stage 2: indirect and induced value of public sector information

The indirect value of public sector information refers to the value generated in associated industries that supply inputs into PSIHS. The induced value of public sector information refers to the value generated when households spend wages paid out by PSIHS and industries involved in the supply chain. The report estimates the size of these value impacts using the UK Domestic Use Matrix for 2005 (latest available) sourced from the ONS and the direct use estimate derived above.

Appendix 5 contains more details of the approach, but broadly speaking it has involved taking the direct use estimate of producer surplus from Stage 1 (analogous to operating profit) and converting this into gross output (GO) and gross value added (GVA) on the basis of available data. GO TYPE I and TYPE II multipliers are then used in an Input-Output setting to consider the upstream business-to-business purchasing effects (indirect) and consumer spending effects (induced). The GO multiplier used in this process was estimated to be 3.0.

To allow a comparable estimate of value to the original producer surplus (and thus allow aggregation), the results are converted back into GVA and then operating profit, by definition, providing a comparable surplus estimate. The ‘surplus multiplier’ – the ratio of indirect and induced surplus to the direct use surplus – was estimated to be 1:2.4 (3.4 in total). In other words, for each £1 generated as producer surplus in direct use, a further £2.40 is generated via indirect and induced effects.

Per worker productivity estimates are used in conjunction with estimates of GVA to provide an indicative level of employment supported in organisations supplying inputs to the public sector information supply chain and supplying goods and services to consumers.

¹⁴³ Defined as the value accruing to producers of goods and services when their output is purchased by consumers. In traditional supply and demand analyses, producer surplus is calculated as the difference between the lowest amount the producer would be willing to sell the good/service for and the price the producer actually sold it for. In many cases this is equal to profit.

Narrow economic value estimates

Taking account of the above caveats and limitations, the estimates of the current value of public sector information in the UK in 2011/12 are shown below.

Figure 5.1: estimates of the current value of public sector information in the UK in 2011/12 (annual value)

Value category	Central scenario	High scenario	Low scenario
Direct consumer surplus	£1.6 billion	£2.00 billion	£1.00 billion
Producer surplus	£0.1 billion	£0.1 billion	£0.1 billion
Indirect and induced value (supply chain and consumer spending)	£0.1 billion	£0.1 billion	£0.1 billion
Total	£1.8 billion	£2.2 billion	£1.2 billion

Source: Deloitte analysis. Note these figures are not comparable with other estimates of the value of public sector information (such as the previous DotEcon estimates) due to differences in the scope of analysis and methodology used.

Note the calculated values for producer surplus and indirect and induced impacts do not alter between scenarios as the data is more robust and fewer assumptions are made. Changes to elasticity, demand curves and data only affect direct consumer surplus estimates.

As noted in the introduction, these figures are not comparable with previous estimates of the value of public sector information as earlier attempts have focused on specific datasets or sought to include selected social impacts.

As the review of evidence has shown, public sector information can generate value across a much wider range of indicators – suggesting that the total current value of public sector information far exceeds £1.8 billion. The following section considers the wider social value of public sector information in the UK.

Wider societal value estimates

The use and re-use of public sector information can generate a range of benefits, which in turn can create value for consumers and producers alike. Some of the types of benefits that have demonstrable and, in some cases, calculable financial implications include:

- the value of **reduced carbon emissions, reduced fuel use and time saved** due to reduced congestion through using apps and other tools that rely on live public sector information APIs, for example to give live transport updates;
- the value of **efficiency savings in the public sector**¹⁴⁴ through identifying cost savings arising from analysis of different public sector information datasets;
- **improvements to decision making and choice** due to new insights from public sector information which, in turn, helps reduce barriers to entry for new entrants to different markets, helping generate economic growth;

¹⁴⁴ It is assumed, based on the evidence reviewed, that cost savings arising from the use and re-use of public sector information are likely to exceed value, as defined. Consider a hypothetical saving of £1 billion through using public sector information to reduce healthcare costs. This benefits the Exchequer and gives the Government an additional £1 billion to spend or an equivalent amount in tax cuts. However, what it can also do is reduce value as defined by producer surplus as healthcare companies may see face falls in profit and they may choose to reduce R&D. The exact balance will depend on international flows of value and how the £1 billion is redistributed.

While the evidence suggests the cost savings will be greater than any fall in value, it is for these reasons of uncertainty that this impact has been considered separately from other narrow economic impacts.

- promoting **greater accountability in public services and public life** through the availability of public sector information datasets that help individuals rate performance and pursue accountability; and
- **better policy making** using public sector information datasets that improve value for money and the efficacy of policy.

To illustrate the ways in which public sector information currently generates value, the report presents a number of case studies in industries already exhibiting some of the benefits of using and re-using public sector information:

- the healthcare and life sciences sectors;
- the transport sector¹⁴⁵; and
- the public sector.

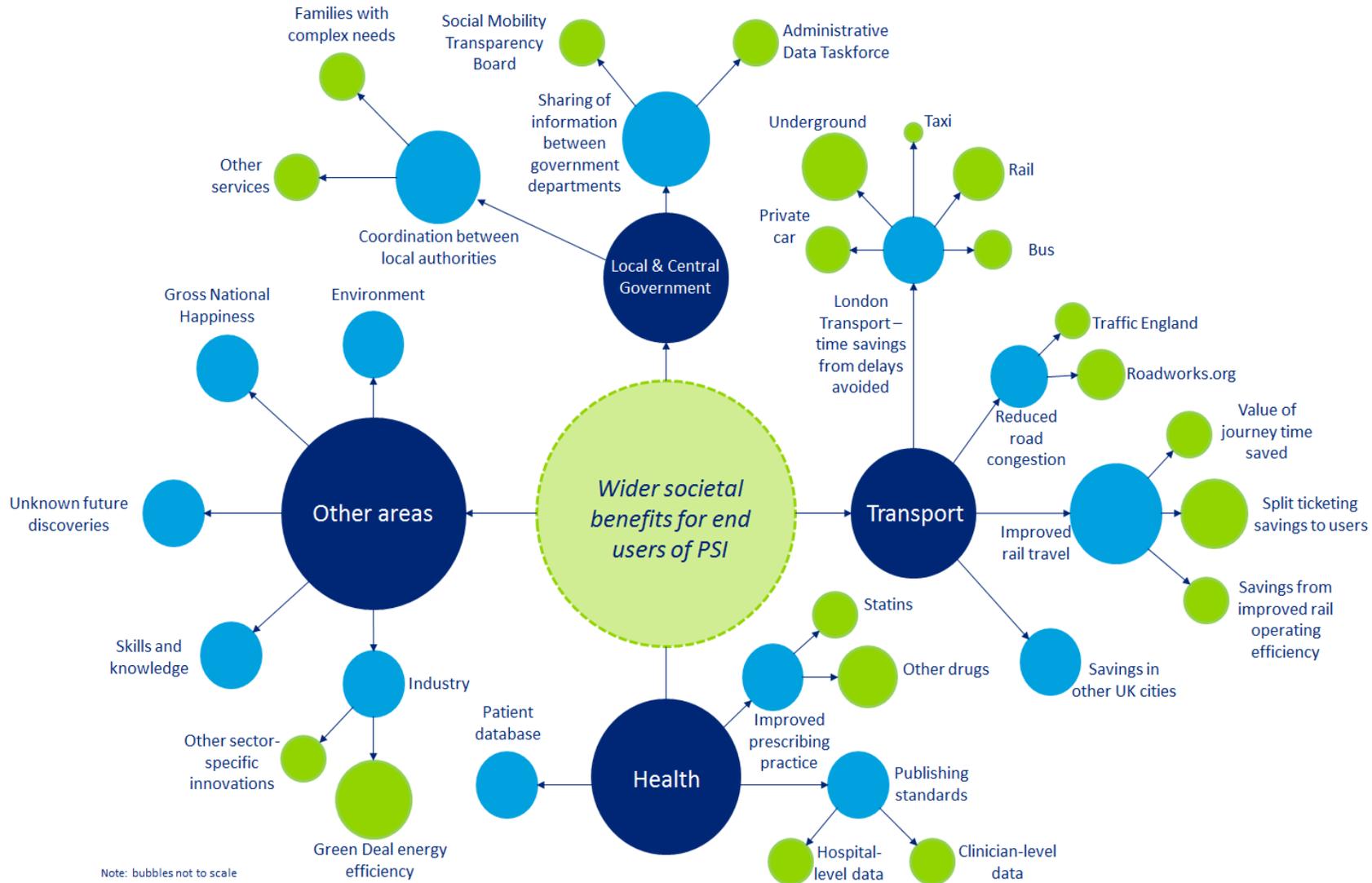
Where appropriate, indicative estimates of the financial value being generated by public sector information are provided. It is not, of course, always possible to measure benefits in financial terms. Figure 5.2 on the following page offers an indication of the types of benefits and impacts. Note that this includes only the benefits from the discussed case studies, and should be treated as indicative of the much larger total value of the wider societal value of public sector information. A summary of the case studies is also included on the following pages; see the appendices for further detail on the sources and any empirical methodologies used for these.

Whilst these individual estimates of social value from the different case studies cannot simply be summed (some are cost-savings, some are time-savings some are economic value and some are non-quantifiable), they suggest that the **total social value of public sector information is likely to be significantly greater than the narrow economic value presented here**. A conservative estimate of the wider social value of public sector information suggests a value at least three times the size of the £1.8 billion economic value may be appropriate¹⁴⁶, i.e. **over £5 billion p.a.**

¹⁴⁵ A 'deep dive' in this sector is placed in Appendix 6.

¹⁴⁶ This is based on insights from individual case studies by other authors which have variously estimated the wider value from public sector information use in the health care sector, geo-spatial, meteorological and other sectors in the order of billions.

Figure 5.2: wider societal benefits of public sector information



Source: Deloitte analysis

This chapter concludes with some 'blue skies' opportunities for the use and re-use of public sector information, which are not currently producing tangible benefits but which studies indicate may prove valuable opportunities in the future.

The health sector

As is well-known, the health sector collects a wealth of data on patients and patient outcomes. The case studies below highlight some of the opportunities that can arise out of analysing public sector information datasets, but also some of the challenges around making this data more widely available.

Publication of mortality rates following cardiac surgery

Case study summary:

Publishing data on mortality rates following adult cardiac surgery appears to be associated with a decline in mortality. There are various theories as to why this is the case, including competitiveness among surgeons which leads to a rise in performance, increased awareness among healthcare professionals, and public pressure for higher standards. However, there are also concerns that the apparent decline in mortality may reflect 'gaming' of the mortality data.

Size of the prize:

A decline in mortality rates is, in itself, desirable. The economic value depends on the value attributed to a statistical life and the costs of providing treatment under various outcomes/pathways. Taking a median value from a range of recent studies, the value of lives saved among those undergoing adult coronary artery surgeries in NHS centres in north-west England in 2005 was around **£55 million**. This suggests that the total value of lives saved per annum could **exceed £400 million** for England and Wales, if similar benefits were observed in all regions.

Mastodon C – identifying NHS prescription savings from big data

Case study summary

By using data on prescribing practice across England, variations in spending on different classes of drugs can be identified. It is then possible to calculate the potential savings to be achieved by moving from prescribing branded to generic drugs.

Size of the prize

For statins alone, the study found that the NHS could save around **£200 million per year** by reducing prescriptions of branded in favour of generic versions.

When extended to all classes of drugs, the total potential savings could amount to **£1.4 billion per year**, according to work by the British Medical Journal.

A patient database for the NHS – the challenges to extracting value from large datasets

Case study summary:

A central NHS patient database could offer significant savings to the NHS as well as improving standards of care and the patient experience. However, there are significant technical, privacy and cultural hurdles to overcome if this is to be made a reality. This case illustrates both how attractive the prize of harnessing the power of large public sector information datasets can be, but also the difficulties these can present.

Size of the prize:

Difficult to estimate, but if the system is delivered as planned the savings are likely to run to billions of pounds. An initial illustrative estimate identified **£4.4 billion of potential savings**, although these are not all directly related to the patient database.

The transport sector

The transport sector is an example of rich data with a high velocity. The case studies below show how this data is being used to generate both narrow economic value and wider societal value.

The break-out box below illustrates some of the areas of research and policy where sharing of data could generate new insights and lead to improved policy outcomes.¹⁴⁷

Transport for London data

Case study summary:

Transport for London has released considerable volumes of information on the London transport network, including live network updates. This case study assesses the value of time saved due to avoidance of disruption on the transport network by travellers, using data published by Transport for London and accessed via mobile apps.

Size of the prize:

The value of time saved due to avoidance of disruption in 2012 is estimated at **between £15 million and £58 million**, depending on the modelling assumptions used. This is likely to be a conservative estimate: for example, the figure would be higher if a higher value for working time were used instead of a valuation based on commuting. It also omits other aspects of value generated by Transport for London data, for example in routine (non-disrupted) journeys by allowing travellers to plan their journeys more efficiently.

Nonetheless, this order of magnitude represents a significant annual time saving. By way of comparison, the HS2 impact assessment calculates operational time savings equivalent to £440 million per year in 2012 values. Therefore releasing relatively low-cost data, which is in many cases a by-product of other activities, may generate time savings versus a nominal baseline valued in excess of 10 per cent of the time savings resulting from a major national infrastructure project such as HS2.

Traffic England

Case study summary:

Traffic England offers live traffic data to road users, enabling users to plan their journeys so as to avoid congestion, roadworks and other conditions likely to cause delays. They are thereby able to save time and reduce fuel waste.

Size of the prize:

As a high level estimate, if 50 per cent of monthly users save ten minutes each on their journeys over the course of a month, the value of time saved could equate to **£6.5 million per year**. This is likely to increase as more road users become aware of the service, and as increasing congestion on the UK road network increases the value of live traffic information.

Improving access to fragmented information – roadworks.org

Case study summary:

By providing a platform that combines information on planned and on-going roadworks into a single database, roadworks.org provides a resource for road users that allow them to easily see potential disruptions to their journey. It also reduces costs to local authorities from their own bespoke platforms, and improves communications between utilities companies and local authorities.

Size of the prize:

A recent ELGIN report estimated the total benefits at around **£25 million per annum**. This includes 'tangible savings' of £6.3 million to local authorities as a result of efficiency savings, and 'intangible benefits' of £19 million due to reduced congestion.

The public sector

The final set of case studies examine how the public sector itself is beginning to exploit public sector information to drive efficiencies, improve public services and enhance policymaking.

¹⁴⁷ *The UK Administrative Data Research Network: Improving Access for Research and Policy* (December 2012)

Area of research and policy	Description
Social mobility	Linking data on education, training, employment, unemployment, income and benefits
Causal pathways over the life course	Linking data on education, health, employment, income and wealth
Support for the elderly	Comparative analysis of access to and provision of social care support for the elderly
Poverty	Linking data on housing conditions, health, incomes and benefits
Social care for children	Linking indicators of parental employment, social background and childcare
Offence and re-offence	Linking data on offending and re-offending behaviour, income, benefits, health and mental health

An example of data sharing at a local level – families with complex needs

Examples of data sharing between local councils and other public bodies are proliferating across the UK. These are often responses to the twin pressures of deep funding cuts and intractable problems involving multiple agencies.

An example of this sort of problem is the case of families with complex needs, often referred to in the press as 'troubled families'. These families may combine issues such as mental health problems, children out of school, and long term worklessness and benefit dependency, meaning that they fall within the remit of multiple public sector bodies including social services, the Police, and the local and national welfare services. These services are estimated to cost around £75,000 per family per year.¹⁴⁸

Increased coordination between local councils and other local partners may prove both more cost efficient and more effective in resolving the problems faced by such families. The councils of Greater Manchester, Leicestershire and Bradford are working together to improve information sharing and management in this area.

The project aims to develop a single toolkit for information sharing, combining existing guidance and approaches. This is intended to be applicable to both the councils and the agencies working with families with complex needs.

In addition the project is intended to lay the foundations for a culture more conducive to information sharing, addressing issues such as different professional cultures of sharing, lack of training and expertise, and differing interpretations of legislation.

In practice, the steps needed to achieve this can appear prosaic but are nonetheless potentially powerful enablers of an environment in which information can more easily be shared within and between organisations.

As the project is still ongoing it is too early to judge its effects, whether in terms of reduced costs or improved outcomes. Nonetheless, this appears to be a positive example of the potential of greater sharing and exploitation of data between councils and other public sector bodies in response to a policy problem which has proved unresponsive to a siloed approach.

¹⁴⁸ Association of Greater Manchester Authorities, 'Improving Information Sharing and Management: A National Exemplar Project'

The Justice Data Lab

It is currently difficult for many providers of offender services, particularly in the voluntary and community sector, to access re-offending data relevant to the offenders they work with. As a consequence, organisations may encounter significant difficulties in measuring the effectiveness of their rehabilitation work with respect to a reduction in re-offending. The lack of access to high-quality re-offending information has also prevented some organisations learning from and improving the services they deliver; and has made it difficult or impossible for them to demonstrate their impact to commissioners.

In order to address this problem the MoJ has set up the Justice Data Lab, announced by Secretary of State, the Right Honourable Chris Grayling MP in December 2012. It is currently in a pilot phase running for a year from April 2013. The Justice Data Lab will provide organisations with aggregate re-offending data specific to the offenders they have been working with, and that of a matched control group. This is intended to allow them to understand their specific impact in reducing re-offending. The intention is that supporting organisations by providing easy access to high-quality re-offending information will allow them to focus only on what works, better demonstrate their effectiveness, and, it is hoped, ultimately cut crime in their area.

Participating organisations will supply the Justice Data Lab with details of the offenders they have worked with and information about the services they have provided. The Justice Data Lab will supply aggregate one-year proven re-offending rates for that group, and a matched control group of similar offenders. The re-offending rates for the organisation's group and the matched control group will be compared using statistical testing to assess the impact of the organisation's work on reducing re-offending. The results will then be returned to the organisation with explanations of the key metrics, and any caveats and limitations necessary for interpretation of the results.

This is an example of how data of a personal, sensitive nature may be used to inform those who have a specific need for this data. It indicates that these barriers can be overcome in a manner that does not compromise ethical considerations around data protection and the right to privacy, while still allowing the data to be used to improve policy and service delivery including by third parties. It also demonstrates the notion of data being a 'two-way street' between PSIH and user, noted elsewhere in the study, which can lead to improved outcomes for both parties.

However, this example also indicates that some investment into data analysis capability is likely to be required on the part of the public sector information holder (in this case the MoJ) – in other words, there is a cost involved. Depending on the success of this pilot over the course of 2013-14, it may be that this provides a model for granting controlled access to similar datasets held across the public sector.

Source: www.justice.gov.uk/downloads/justice-data-lab/justice-data-lab-user-journey.pdf

'Blue sky' opportunities

In addition to the specific examples already discussed, there are many areas in which public sector information may have a positive impact in the future. Foresight, part of the Government Office for Science,¹⁴⁹ has carried out a number of studies which highlight the role public sector information could play in improving our understanding of and ability to respond to the challenges our society confronts. These are summarised below.¹⁵⁰

Detection and Identification of Infectious Diseases (2006)

Novel Information Technology for the Early Detection of Infectious Disease events. This project identified a 'Grand Challenge' around using modern information and communication technology systems to gather and interpret timely and relevant data, and to deliver it to those managing an outbreak.

Tackling Obesity (2007)

On-going data collection and evaluation relating to the effect of different interventions was proposed as a core principle in tackling obesity. This was seen as a particularly important in view of the uncertainties surrounding the efficacy of certain interventions. The project stimulated the

¹⁴⁹ See www.bis.gov.uk/foresight

¹⁵⁰ With thanks to the Government Office for Science

creation of the National Obesity Observatory - but much more remains to be done. Mismatches between policy and research timescales remains a fundamental challenge.

Global Food and Farming (2011)

Policy makers are hampered by conflicting results from different models associated with the food system. The project argued the need to substantially improve datasets that provide the basis for model calibration and comparison. The project brought together a forum of modellers to help take this forward.

Computer Based Financial Trading (2012)

Improving collection and access of financial transaction data was a key recommendation: a trusted 'European Data Centre' was proposed. Such data would be a crucial new tool for regulators in identifying abuse spanning different markets. Also, if made easily available, it could unlock the resources of the scientific community to better understand evolving markets, and so help policy makers in the 'arms race' with algorithmic trading developers.

Reducing the Risks of Future Disasters (2012)

The report argues the need to improve the infrastructure for data collection for hazards: e.g. better coordinated effort on satellites and sensors. It also proposes establishing a much needed database of evidence on the costs and benefits of interventions to reduce disaster risk. Importantly, the data in such a repository would need to be quality assured, and easily accessible to decision makers at different levels in different countries.

The Future of Identity (2013)

The report considered how notions of personal and social identity in the UK might change over the next ten years. Policies will need to take into account the multiple nature of identities and how policies might affect groups differently, or individuals' different times and places. The report explicitly considers the commercial value of identity through the use of 'Big Data'. This is likely to become crucial to private sector organisations, but also has the potential for criminal exploitation, for example through opportunities for identity theft.

Chapter summary

- The value of public sector information can be disaggregated into four broad categories according to the beneficiary:
 - the direct value of public sector information to producers and suppliers (the PSIHs): these are the benefits accruing to producers and suppliers of public sector information through the sale of public sector information or related value-added services;
 - the indirect value of public sector information arising from its production and supply: the benefits accruing up the supply chain to those organisations interacting with and supplying PSIHs (but not directly using or re-using public sector information), and the benefits accruing to those organisations where employees of PSIHs and supply chain organisations spend their wages;
 - the direct use value of public sector information to consumers of public sector information: the benefits accruing to businesses, civil society, individuals and the public sector from directly using and re-using public sector information for a variety of purposes; and
 - the wider societal value arising from the use and re-use of public sector information: the benefits to society of public sector information being exploited, which are not readily captured elsewhere.
- The first three value categories can be grouped together as narrow economic value and have been estimated at £1.8 billion p.a. for the year 2011/12.
- The wider societal value of public sector information is harder to estimate. Public sector information can benefit society through a number of routes including:
 - increasing democratic participation: giving citizens and businesses access to public sector information allows them to perform their own analyses of salient issues, make more informed choices about public service providers and interact with policymakers to challenge their assumptions and improve the policymaking process;
 - promoting greater accountability: for example through the scrutiny of costs of public service provision and benchmarking comparable services;
 - greater social cohesion: for example, by providing more information on the provision and distribution of services, public sector information can be used to dispel myths on who receives certain public services;
 - generating environmental benefits: such as reducing congestion and pollution through the release of better traffic and transport data which helps drivers to better plan journeys; and
 - identifying previously unknown links between different policy areas: through data mash-ups it may be possible to develop system-wide solutions that holistically seek to address the root of policy challenges.
- This value is potentially significant and is likely to have a major influence on overall societal wellbeing. A conservative estimate of the wider social value of public sector information suggests a value at least three times the size of the £1.8 billion economic value may be appropriate, i.e. over £5 billion p.a.
- Adding this wider social value estimate to the calculated value of public sector information to consumers,

businesses and the public sector, gives an aggregate estimate of between £6.2 billion and £7.2 billion in 2011/12 (2011 prices).

6. Barriers and policy implications

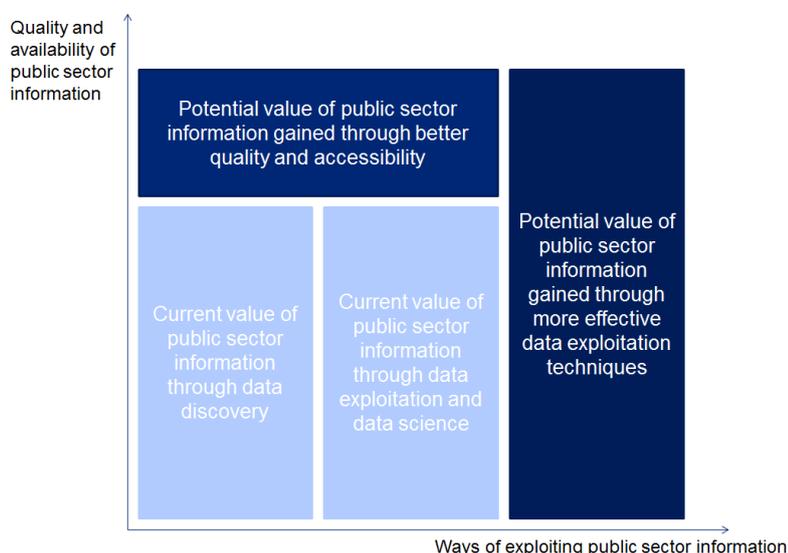
Using a taxonomy originally developed by the Data Strategy Board, this chapter considers the current barriers to the UK in fully realising the benefits and value of public sector information. The analysis is based on discussions with stakeholders and a review of the evidence.

The chapter also considers the policy implications of each barrier. The Shakespeare Review will be making its own recommendations on which barriers and associated challenges it believes are the most important and the steps that can be taken to address these.

Generating value

Having reviewed the literature and evidence, it is apparent that the two main ways in which value is being generated currently, which will likely remain the case in the future, is through data discovery and data exploitation. The former relates to making better and, in some cases more, public sector information available and making it more accessible – what might be loosely term supply-side considerations. The second dimension relates to using and re-using public sector information better – demand-side considerations. This might be through reducing consumer risk aversion to using public sector information, improving data exploitation techniques, changing cultures and improving data analysis skills (which, in turn can improve competitiveness).

Figure 6.1: exploiting the value of public sector information



Source: Deloitte analysis

The light blue boxes (which are not to scale) represent the value that is currently being generated from public sector information, with the first light blue box on the left showing the current value from data discovery and the second light blue box the current value through data exploitation and data science. The dark blue boxes illustrate the additional value that could be achieved if (i) better quality public sector information was released, (ii) more was made more easily available; and (iii) more public sector information datasets were released. While, in some cases, more public sector information being available can also increase value, it is important to note that there is not a linear relationship between quantity of public sector information and its value. What is key is the quality of

this data and its amenability to data analytics. Simply releasing more datasets, irrespective of their quality, is likely to only have a minimal value impact.

More effective ways of data exploitation can include:

- greater system-wide analysis through more sharing and combining datasets together to consider policy and business issues from a range of non-traditional perspectives;
- unlocking the potential from linked data: and better integrating; and
- greater data-mashing with private sector and individual datasets.

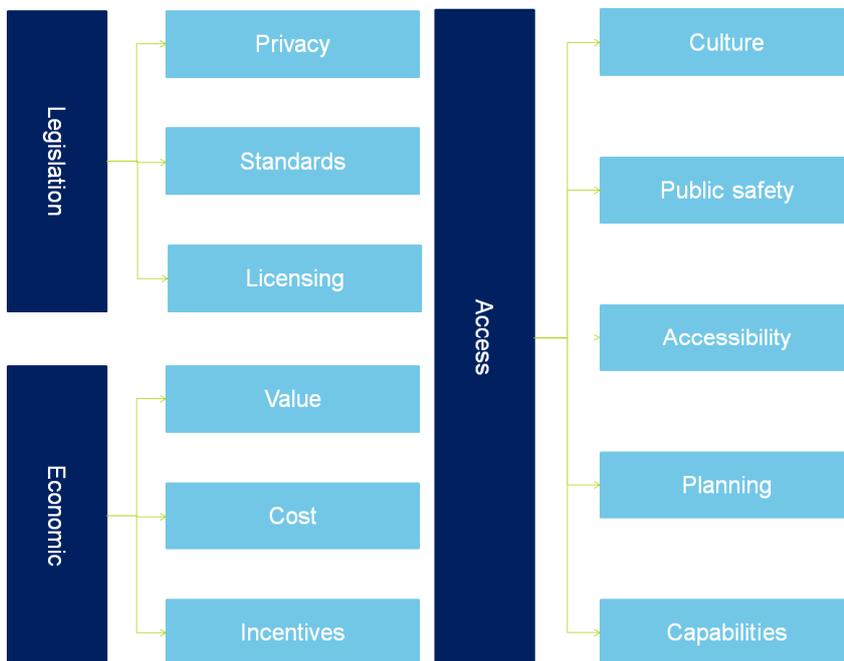
Value is thus generated through exploring existing datasets to identify insights through statistical analysis, 'data-mashing' and visualisations. The value of public sector information in the UK can therefore be increased both by increasing the quality, and in some cases quantity, of datasets available/accessible (increasing the potential for data discovery) or by better using existing and new datasets (increasing data exploitation).

This raises the question of what the barriers to the UK fully exploiting the value of public sector information are, and what the size of this potential value may be.

A taxonomy of barriers

This report adopts a taxonomy originally developed by the Data Strategy Board, which illustrates the four main areas where challenges or barriers may currently exist. This taxonomy of barriers is shown in Figure 6.2 below.

Figure 6.2: taxonomy of barriers



Source: adapted from DSB

Legislative barriers include whether certain acts of legislation and other regulations reduce the usability of public sector information datasets by consumers. There are also further questions over whether current assurance/accreditation standards for public sector information remain effective and whether there is scope for existing licensing arrangements to be improved.

Economic barriers include questions as to which datasets to release (which datasets yield the highest value) and how their costs (if any) should be covered.

Access barriers to maximising the value from public sector information can include:

- a reluctance to publish or share datasets;
- a bias against using non-traditional datasets;
- a reluctance to use public sector datasets because of concerns around quality, reliability and on-going support; and
- a lack of skills to fully exploit the value of public sector information.

It should be noted that the extent of these barriers may differ between public sector information datasets and also between the different types of users.

Each of these barriers is discussed in more detail below, alongside the relevant policy implications. Note, due to the overlap in some of the taxonomy areas, the narrative below does not always follow Figure 6.2 exactly.

Legislative barriers

Privacy and the impact of current regulations and legislation

The review of available evidence suggests that, by and large, the current legislative and regulatory environment around public sector information is not acting as a barrier to generating value and market development. While a full legal analysis has been beyond the scope of this study, particular acts and regulations such as those covering Data Protection and Human Rights legislation have not emerged as preventing the development of new products or services using public sector information¹⁵¹. The majority of stakeholders consulted have not reported that these generic legislative acts are currently preventing them from using and re-using public sector information, but, in some cases specific legislation controlling a particular data collection exercise do.

However, the evidence received does suggest there are some current challenges around perceptions and attitudes towards data release. In particular:

- regulations such as Data Protection are sometimes used as a shorthand justification for not sharing public sector information within the public sector, with PSIHs not always able to translate their awareness of their rights and duties into scenarios where public sector information is released or shared, causing a barrier; and
- when a policy decision is taken not to release public sector information datasets to the general public, the reasons are often not well articulated or the conditions attached to access are overly restrictive.

Some stakeholders have cited examples where the Data Protection Act (and other legislation) has been used as a reason for a PSIH to withhold information from the general public. However, as noted in previous chapters, in reality, the Act should rarely be a barrier to sharing information. Where public sector information datasets do contain personal details, these may need to be aggregated or anonymised. Where this is done effectively the Data Protection Act, in most cases, no longer applies to the information and it can be made available.

This is not to downplay data protection issues. The impact of breaches leading to release of personal information can be extremely serious. Further, even if data is anonymised, PSIHs may be reluctant to release it if they believe it can have wider consequences, e.g. reporting crime data may adversely affect house prices¹⁵². These issues notwithstanding, PSIHs are beginning to use innovative methods to test different ways of anonymising datasets – for example the Ministry of

¹⁵¹ This observation should be caveated with the note that, at the time of writing, there are a number of initiatives to revise various regulatory frameworks at the European level.

¹⁵² See for example: www.guardian.co.uk/money/2011/feb/01/police-crime-website-house-prices

Justice worked with statisticians, the private sector and the academic community to avoid the ‘jigsaw effect’¹⁵³ occurring from the release of offender data (see Chapter 5).

Access to the general public

With respect to other access restrictions to public sector information datasets (such as the datasets only being available to researchers or in secure environments) arising from specific pieces of legislation, in many cases the rationale for these restrictions are clear. The rationale may cover national security reasons or genuine data protection concerns. In many cases this data is released only to licensed researchers, and some stakeholders questioned the value of making it available to the wider public since its applicability is to a specialised area of research. However, other stakeholders have reported that in some cases, the rationale for restrictive access to certain datasets is not always clear, overly restrictive or no longer relevant.

For example, some start-up companies have reported difficulties in accessing different versions of the National Pupil Database as non-research organisations (see case study below). While restrictions on access are in this case clearly needed to protect sensitive and personally identifiable information; there is a valid question as to whether there are ways that some or all of this information could be made available to commercial organisations to develop new products and services to help parents, teachers and other stakeholders to improve decision-making, increase accountability and identify good practice. Without this, opportunities for innovations using this and other datasets are likely to be missed, meaning loss of the value that they could potentially create.

Case study: the National Pupil Database

The National Pupil Database (NPD) was created under the School Standards and Frameworks Act 1998, which instituted a legal gateway permitting the collection of personal data about students without their consent or the consent of their guardians. The information was to be drawn directly from school management systems. Until the late 2000s students and parents were not informed that this information was being collected and stored.

Since 1998, the National Pupil Database has expanded to contain ever more information about students. From initially being an annual census, information is now collected every term, and has expanded to include preschools, any provision of childcare that is funded by the state, exclusions of students and the reasons for exclusion, recipients of free school meals, and the mode of travel to school., among other fields. Because the initial gateway offered something approaching carte blanche in terms of the information that could be collected, it has been gradually expanded according to the needs of policymakers.

This means that the NPD is tremendously rich as a source of data. It also means that it contains a wide range of potentially sensitive and personally identifiable information. This information is stored permanently, often without the knowledge and in any case without the consent of the subject.

The NPD therefore creates a dilemma. On the one hand, the insights that could be derived from this information, especially when combined with other sources of information, are tantalising. For example, it offers researchers and policymakers a route into understanding the drivers of educational attainment, behavioural problems among children, and social mobility in later life. This information could also be used to generate a range of value-added services and products by private companies.

On the other hand, there are clear risks associated with the holding and sharing of such a quantity of sensitive personal information. For example, information about exclusions of pupils, including the reason, is stored indefinitely. If this information were to be released in an identifiable manner it could have a compromising effect on the individual in later life, for example when applying for work.

The Department for Education currently has a well-developed process in place for access to the information to safeguard against negative outcomes. Those who require access are required to demonstrate that this information is necessary research into the education achievements of pupils. The requests are processed by the DfE Data and Statistics Division (DSD), with requests for Tier 1 data (information that is both identifiable and highly sensitive) referred to the DfE Data Management and Advisory Panel. Where access is approved,

¹⁵³ The process of combining anonymised data with auxiliary data in order to reconstruct identifiers linking data to the individual it relates to.

Case study: the National Pupil Database

the DSD manages the supply of data.

In November 2012 the DfE launched a consultation on widening access to the NPD, with a view to granting access to: “persons (i) conducting research or (ii) providing information, advice and guidance or (iii) data based products and services for the purpose of promoting the education or well-being of children in England and who require individual pupil information for that purpose.”

However, these changes may still preclude some potentially valuable uses of the NPD. Restricting access along the lines children’s ‘well-being’ potentially forecloses innovative non-educational analyses. For example, a researcher exploring links between educational performance and culture or the environment would have to seek to justify their use of the NPD in ‘well-being’ terms – this may not be straightforward. Equally, always having to specify a reason for requesting the NPD may prevent speculative research or ‘randomised’ data-mash-ups.

Despite this, the personal nature of the information inevitably makes this a sensitive topic, as it may be felt that widening access increases the risk of accidental disclosure of the information, as well as allowing information to be used in ways the data subjects have not consented to. In its response to the consultation the ODI cited a number of uses to which aggregations of data from the NPD could be used as inputs to products and services developed by start-up business, but stated that “the aggregations these applications require could be generated by the Department for Education. None of these services require access to individual-level data.”¹⁵⁴

This case therefore illustrates the complexity of the barriers to access in cases of sensitive, identifiable information, and the tension that may arise between the right to personal privacy and the potentially rich uses of datasets containing personal information. At the time of writing, the DfE NPD consultation had closed with a response expected soon.

Source: partially based on an ODI talk, ‘Getting to grips with the National Pupil Database, 15th February 2013. Slides available at www.scribd.com/doc/125638490/Getting-to-grips-with-the-National-Pupil-Database-personal-data-in-an-open-data-world

See also www.education.gov.uk/researchandstatistics/national-pupil-database/b00212283/national-pupil-database

The case study below illustrates the case where the value of public sector information remains locked not for reasons of restricted access or cost, but simply because a given dataset is not being made available.

Honest Buildings – ‘LinkedIn for the real estate market’

Honest Buildings is a start-up that originated in the United States but is now expanding into the UK, and to this end is currently working with the ODI in London. It provides an example of how greater availability of public sector information can be combined with private sector and crowd-sourced information to generate new efficiencies in an industry sector.

The real estate sector – covering owners and occupiers of buildings, as well as providers of building services and improvements – currently depends on a relatively cumbersome approach for most interactions between parties. For example, businesses searching for new premises, or for improvements to their property are forced to enter into a potentially lengthy and resource-intensive process of inviting potential suppliers to tender for work.

Honest Buildings offers a platform similar to LinkedIn for the real estate industry, providing profiles for buildings, organisations, projects and people. The idea is to build up a transparent, searchable network of information on buildings, their owners and occupiers, suppliers of services and their project histories. This should make the process of interactions between these parties much more efficient, as potential buyers of services will be able to receive more relevant, informative tenders and suppliers will be able to adopt a more focussed approach to business development. Michael Adler, former CFO and executive vice president of the travel website Expedia.com, has said: “Honest Buildings is bringing the same type of efficiencies to the real estate market that

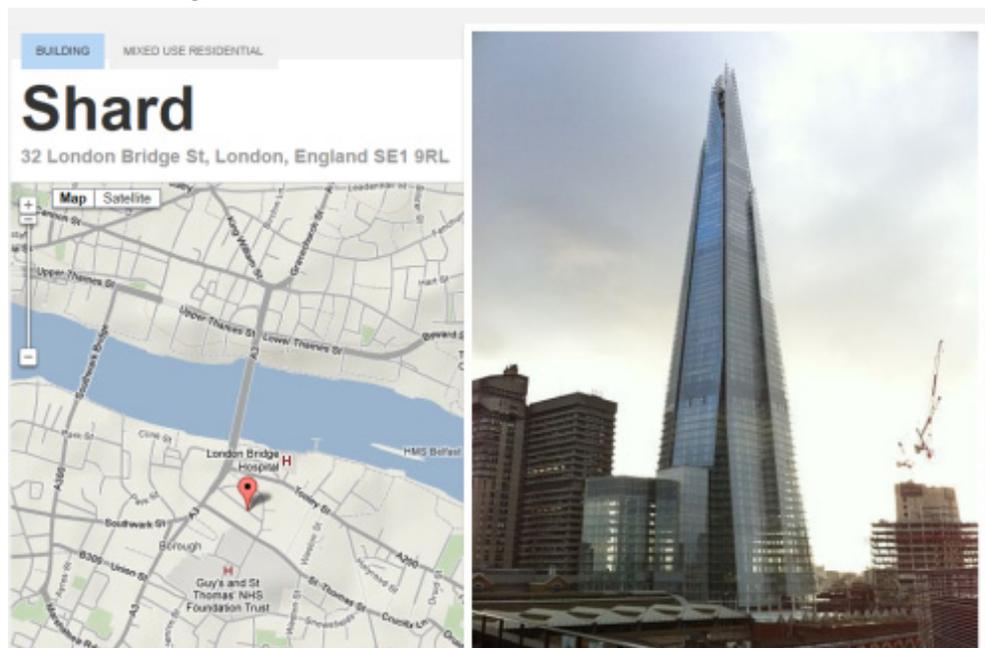
¹⁵⁴ See www.theodi.org/consultation-response/proposed-amendments-individual-pupil-information-prescribed-persons

Honest Buildings – ‘LinkedIn for the real estate market’

Expedia and Trip Advisor did to the travel industry.”¹⁵⁵

Much of the information on this platform will be crowd-sourced from building owners, occupiers and service providers, but Honest Buildings also supplements this with public sector information in order to add additional value. This includes property prices, rates, details of building specifications and energy efficiency details.

Honest Buildings



Source: <http://www.honestbuildings.com/>

Honest Buildings already have a presence in nine cities in the United States. In addition to their wider social networking platform they have also found success providing platforms for energy efficiency programmes in New York State and Connecticut. They are hoping to get involved with analogous initiatives in the UK like the Green Deal, GLA's RE:FIT programme, Cambridge Retrofit along with other major public/private sector led building initiatives. The Green Deal allows consumers to have a range of energy savings improvements made to their homes and then pay for the work through additions to their energy bill. This would allow the government and suppliers to more efficiently target those buildings with an Energy Performance Certificate (EPC) of F or G, the lowest ratings and therefore the highest priority for energy saving upgrades. RE:FIT and Cambridge Retrofit provide a framework focused for non-domestic buildings to be upgraded in the public sector in London and in the public and private sector in Cambridge respectively.

The potential economic benefits of widespread uptake of the Green Deal are significant. For example, this could generate large savings on domestic energy bills. The Energy Saving Trust calculate that loft insulation of 270mm could save the average three bedroom house up to £180 per year on their energy bill, and double glazing could save around £170 per year.¹⁵⁶ If 20 per cent of the UK's approximately 25 million households were able to achieve these combined savings of £350 per year, this would equate to an annual saving on domestic energy bills of £1.75bn.

In addition, and more directly relevant to the current focus of Honest Buildings, there is also the potential for savings on the energy bills of commercial properties. The scale of the potential savings is more difficult to

¹⁵⁵ See www.greenbiz.com/blog/2012/09/12/secret-life-buildings

¹⁵⁶ See www.energysavingtrust.org.uk/Energy-Saving-Trust/Our-calculations

¹⁵⁷ See www.gov.uk/government/uploads/system/uploads/attachment_data/file/47986/1002-energy-bill-2011-ia-green-deal.pdf

¹⁵⁸ Ernst & Young, 'Making energy efficiency your business', 2011, [www.ey.com/Publication/vwLUAssets/Making_energy_efficiency_your_business_-_Understanding_the_potential_of_the_non-domestic_Green_Deal/\\$FILE/EY_Making_energy_efficiency_your_business_-_Non_domestic_Green_Deal.pdf](http://www.ey.com/Publication/vwLUAssets/Making_energy_efficiency_your_business_-_Understanding_the_potential_of_the_non-domestic_Green_Deal/$FILE/EY_Making_energy_efficiency_your_business_-_Non_domestic_Green_Deal.pdf)

Honest Buildings – ‘LinkedIn for the real estate market’

estimate as there exists no comprehensive register of the UK’s commercial property stock. In addition commercial buildings are more heterogeneous in nature, meaning that it is more complex to assess the benefits of energy saving measures. As an indication, the 2010 DECC impact assessment for the Green Deal calculated potential energy savings of between £170 million and £330 million, based on an additional uptake of energy saving measures of 10 to 20 per cent above a business as usual scenario. However, these savings are set against capital costs of between £75 million and £140 million.¹⁵⁷

These are indicative estimates, but demonstrate the scale of the potential saving from widespread uptake of the Green Deal. They also omit the downstream impact, including improved energy security due to less reliance of imported energy, reduced carbon dioxide emissions and other negative environmental externalities, and greater consumer spending power due to reduced energy bills.

In addition, there is likely to be a wider economic uplift due to the generation of green jobs, support for the construction industry, and increased demand for raw materials, many of which can be sourced within the UK and would therefore not need to be imported. A recent study on the non-domestic Green Deal indicated a potential market size, for SME uptake of Green Deal measures, of between £470 million and £800 million by 2020, depending on the rate of uptake.¹⁵⁸ This would provide a substantial economic stimulus to this sector of the economy.

There is, in addition, a potential legislative imperative to upgrading building energy efficiency. It is proposed to ban the rental of buildings gaining an F or G EPC rating. According to Honest Buildings this applies to an estimated 20 per cent of the UK’s building stock, and at the current time such a ban would therefore throw the real estate market into crisis. Upgrading this stock is for this reason likely to become a priority over the next decade, and a task of this magnitude will require an efficient and transparent approach, with property owners fully aware of their supplier options and the relative costs and benefits of each.

Honest Buildings therefore appear to be an example of innovation using public sector information which could generate a range of tangible economic benefits, from energy savings and job creation to reducing transaction costs across an entire sector. However, there are barriers to accessing the information required. Most importantly, EPC register, administered by Landmark Information Group, is not currently accessible. Honest Buildings report that Landmark claim they are unable to give access to the database due to their contractual obligations to CLG.

The EPC register is an example of ‘by product’ data – although there might be a small cost associated with making it publically available (due to hosting requirements and user support), there would be no additional collection costs as this is data that is collected in the course of the normal operations of the EPC scheme. Release of this data therefore appears to be a win-win option which could unlock considerable potential value without leading to a loss of revenue or incurring significant costs to the information holder.

More generally, Honest Buildings has identified issues with fragmentation and disconnect across the public sector’s approach to information. It has frequently found it difficult to find the information needed, and that where this has been possible to locate, access has sometimes been stymied by confusion over permissions and ownership, meaning that information holders do not feel empowered to release the information. This reinforces the argument for greater coordination and clarity in the public sector’s approach to information, with an ‘open by default’ assumption for information and a clear framework defining the rules around information release. Support for providing information about what data is available, whether through centralised catalogues or decentralised search, would also help guide potential users to the information they need.

In addition to the barriers discussed above, there has been much discussion about core reference datasets in the context of the debates about releasing the Postcode Address File¹⁵⁹.

Access within the public sector

Within the public sector, there are also questions surrounding the legal ability of different public bodies to share datasets with one another. These have been explored in depth in the recent report by the **Administrative Data Taskforce**¹⁶⁰. It notes that as regards the legal gateways established to allow departments and other public sector bodies to share information without obtaining the consent of the data subjects, “recent experience demonstrates that link-specific gateway legislation

¹⁵⁹ See for example: <http://data.gov.uk/blog/odug-progress-on-a-national-address-dataset>

¹⁶⁰ Source: Administrative Data Taskforce (2012)

is both cumbersome and inefficient.” As a solution to this the Taskforce recommends the creation of a generic legal gateway to reduce this barrier to the exploitation of public sector information and clarify the legal position around sharing data. Note that this applies to sharing of data within the public sector and with certain accredited third parties such as researchers, rather than publication for use by the general public.

Insights

- **Existing regulations and legislation governing public sector information do not appear to be acting as actual barriers to realising the full potential value of public sector information. However, the manner in which some of these regulations and legislation are interpreted can lead to overly risk averse behaviour and can create barriers.**
- **Increasingly effective anonymisation techniques and an approach across the public sector that emphasises granting access to as much data that is compatible with privacy and security has the potential to improve access to public sector information.**
- **The eligibility restrictions imposed around certain datasets are typically due to reasons of national security, data protection and other sensitivities.**
- **The rationale behind these restrictions is not always clearly articulated and may, in some cases, no longer apply. In some cases, this may be preventing opportunities for innovation to take place.**

Licences

Licensing conditions play an important role in facilitating (or preventing) the full exploitation of the value of public sector information. The ideal standard widely acknowledged by stakeholders is licensing public sector information under the Open Government Licence (OGL), and increasing amounts of public sector information are being made available as open data under this licence, even by organisations that have traditionally charged for data. For example, Ordnance Survey has released 11 datasets as open data, there are over 6,000 downloads of Land Registry’s linked data each month and the Met Office and Companies House have also each opened up some of their data. There is therefore a visible trajectory towards a world in which increasing volumes of high quality data are available at low cost or for no charge to users.

Nonetheless, some groups of stakeholders have argued that if the OGL were used for all public sector information, this could substantially increase the openness of the UK’s public sector information and could remove many of the barriers to use and re-use that currently inhibit the realisation of the full value of public sector information.

The Open Data User Group, as a strong advocate of the OGL, has noted that “it needs to become widely adopted to make the right to data a reality. Conversely, if publishing under the OGL does not become the default action for public bodies, the right to data will remain an aspiration.”¹⁶¹ Indeed, it can be argued that should the OGL become the standard licence across all public sector bodies, this could substantially increase the openness of the UK’s public sector information and could remove many of the barriers (both in terms of charges levied and restrictive copyright and re-use conditions) that currently inhibit the realisation of the full value of public sector information.

However, others have argued that the public sector information landscape is characterised by its diversity, and for this reason a ‘one-size-fits-all’ solution is unlikely to be practical or necessarily desirable. For cases where the release of information under the OGL is not considered appropriate and cost recovery is justified, the introduction of a generic charging licence could, in principle, address the current complexity of charging arrangements for public sector information, completing the UK Government Licensing Framework and simplifying licensing arrangements for users. Some

¹⁶¹ Open Data User Group response to the consultation on the code of practice for datasets and beta charged licence (<http://data.gov.uk/sites/default/files/ODUG%20datasets%20CoP%20response.pdf>)

stakeholders pointed out that by having a charge, users and re-users can be re-assured of the quality of data and be able to expect a certain level of service (this is discussed in more detail below). However, one concern raised was that such a licence could encourage some PSIIs who are not covered by the OGL to charge for certain datasets.

Insights:

- **Making public sector information available under the Open Government Licence is seen by a large number of stakeholders as an effective means of removing barriers to exploiting the value of public sector information.**
- **If there are instances where a generic Government Licence for charging for public sector information are applicable, such a Licence would need to be drafted in a way to avoid incentivising charging when there is no strong justification or leading it to become the default alternative to the Open Government Licence.**

Economic barriers

The evidence reviewed suggests a number of key economic barriers around maximising the value from public sector information in the UK. Primarily these relate to the value and cost of datasets, which can be grouped into two key themes:

- **which datasets are made available to the general public;** and
- **which datasets consumers have to pay to access for (funding models).**

Data gaps

There is currently no national or local information asset register covering the amount and type of public sector information datasets held¹⁶² (irrespective of whether it is being released to the public) or the detailed costs of collection and dissemination. Similarly, there is little evidence as to how consumers themselves are using and re-using particular datasets.

While this has meant this report has been forced to make a number of assumptions in order to reach quantitative estimates, there is a more significant impact on PSIIs themselves. By not being able to accurately ascribe value to different datasets, PSIIs are generally unable to reach evidence-based decisions as to which datasets to publish, how to publish them and what support to provide. One consequence of this is that the costs (which are readily measurable *a priori*) of making particular public sector information datasets available may appear to be much larger than the benefits – as the benefits case cannot be clearly set out due to a lack of evidence.

Indeed, as part of this report, a sample of government officials have been consulted over the reasons why public sector information datasets are not released: over 50 per cent of those responding identified resources as an issue preventing the release of more public sector information, with the following response being typical: “*in a time of diminishing resources and the need to make best use of the resources we have, the time and the cost of ensuring data validity before release is an issue.*” This was one of the most common reasons given for why data is not released, but not the only one – others include data protection and legislation.

Stakeholders have identified particular datasets (such as an aggregate Energy Performance Certificate database) that are currently not available to them (either for free or for a fee) that they could use to generate new products and services. By not making these datasets more widely available (or articulating the reasons they are not available), PSIIs may, whether knowingly or not, be acting as a barrier to realising the full potential of public sector information.

There are various routes to address these data gaps. One route might involve conducting a public sector-wide audit of public sector information (which also covers the customers using and re-using

¹⁶² Though the UK is not alone in this respect.

it). This would involve both primary research to determine the stock of public sector information held across all PSIHs and also identifying how this data is being used commercially across different economic sectors and actors. This would help match costs and benefits to individual datasets and building this report to identify those datasets that can generate the most value for the UK. The costs of doing such an audit would not be insignificant and for it to be effective, it would need to be repeated in future years (though the on-going costs may be lower than the one-off start-up costs). However, the benefits of doing such an audit could be outweighed by the benefits of PSIHs having a much better sense of which datasets are creating the most value which can guide future policy decisions around releasing data.

Other routes to addressing data gaps might involve better tracking of usages (perhaps through measuring how often data is re-used in other sources), requiring all new public sector information datasets to be logged centrally, or crowdsourcing an audit across the public sector. Another potential route for the long-term, at least conceptually, might be a true single conduit for the access of all public sector information.

The Open Data User Group (ODUG) has been set up as an advisory group to Government and provides a channel through which the potential users of data, from all areas of the community, can identify their need for data and set out the benefits they expect this data to deliver. This will help PSIHs understand in more detail the potential benefits which the release of individual datasets can be expected to deliver.

Through the completion of data gaps, it would be possible for government to articulate what is meant by the term core reference dataset (beyond the definition contained in the Open Data White Paper). The issue of core reference datasets is also raised in recent work by APPSI on the national information framework for public sector information and open data.¹⁶³ Consideration could include:

- the features of a particular dataset that make it a core reference dataset;
- the funding arrangements of core reference datasets (see below);
- the obligations of suppliers of core reference datasets; and
- the rights of citizens to core reference datasets.

Insights:

- **There are significant data gaps when it comes to public sector information. This lack of data can, in some cases, lead to inertia with certain public sector information datasets not be released or conversely, undue attention being given to datasets that are unlikely to generate significant value but have a low cost of dissemination.**
- **There are a number of routes to addressing these data gaps. These range from a detailed, regular audit of public sector information to improved tracking of current usage.**

Funding models

Related to the question of making the business case for more public sector information being released is the issue of how the cost of generation, collection, retention and dissemination of the datasets should be funded. This also leads to questions around pricing for access to public sector information.

This is a complex issue and data on costs (and benefits) is not readily available. Further, one must take account of different costs being incurred across the different stages of the data lifecycle. For example, in the case of so-called 'exhaust' or 'by-product' data that is generated as a result of

¹⁶³ The issue of core reference datasets is also raised in recent work by APPSI on the national information framework for public sector information and open data, www.nationalarchives.gov.uk/documents/nif-and-open-data.pdf.

PSIHs conducting their day-to-day and other activities and duties, the marginal cost of its generation will, by definition, be negligible or very low compared to the activity that caused the data to come about; but the marginal cost of its dissemination to the wider public may be higher and vice versa for some 'purposefully collected data'. There will also be additional costs in formatting the data for use and further costs if support is provided to the public in re-using the data.

The issue of charging and funding models for public sector information is highly contentious and extremely complex. For example, there remains an element of subjectivity as to what constitutes a dataset and what constitutes a 'value-add' service – with some PSIHs arguing that what is being charged for is not the public sector information but its interpretation and analysis.

As part of this report, a wide range of opinions have been expressed as to whether there is any economic or other justification for charging for public sector information datasets. On the one hand, these arguments include:

- charges for datasets create barriers to entry and expansion for SMEs and individuals to develop new products and services;
- the charges prevent SMEs and individuals from 'experimenting' with the datasets before they purchase to see if they are able to derive value from them, thereby making it hard to develop business cases; and
- any lost revenues to PSIHs from releasing datasets for no cost will be recovered by the Exchequer in the long-run through increased tax revenues and more jobs being created.

In contrast, arguments have been put forward to support current pricing arrangements include:

- aligning a revenue stream with a particular dataset will 'protect' it from any reductions in funding, allowing PSIHs to continue to supply this even if they themselves must make other savings;
- a price can be interpreted as a signal of consumers' willingness to pay for a particular dataset's quality and a commitment by the PSIH to maintain this and offer support; and
- charging for certain datasets is necessary given they include elements of commercial or international datasets.

As the most visible PSIHs that charge for certain public sector information, a great deal of attention has been focused on the four Trading Funds that make up the Public Data Group (PDG): Ordnance Survey, the Met Office, Land Registry and Companies House. The view has been expressed by some start-up companies that these PSIHs hold some of the most valuable datasets and there are strong arguments that these should be treated as core reference datasets available to all at no direct cost to the general public.

It is not helpful to treat these PSIHs as a single group – indeed, there are a number of other public bodies that charge for access to datasets¹⁶⁴ and there are a number of other Trading Funds outside the PDG. The four Trading Funds in the PDG differ in their sources of revenue, their public service duties and whether the public sector information they hold can be classed as 'exhaust' data or 'purpose-collected' data. Further, it should be noted that substantial progress has been made by the four Trading Funds in recent years to make increasing volumes of data available as open data¹⁶⁵ and there continue to be moves in this direction.

However, despite these positive steps, there remains a perception among many consumers and commentators¹⁶⁶ that they are unable to access certain datasets for reasons of cost and this is creating a barrier to business growth. A number of studies (e.g. Pollock, 2011 and others – see

¹⁶⁴ For example, the Office of National Statistics has a subscription charge for certain datasets.

¹⁶⁵ For example, see www.ordnancesurvey.co.uk/oswebsite/products/os-opendata.html

¹⁶⁶ See for example www.freeourdata.org.uk/ and other references in Appendix 3.

Appendix 3) have argued that releasing these datasets as open data will have significant welfare benefits.

The impact of cost recovery model for public sector information, compared to the Open Government Licence model, is summarised below.

Figure 6.3: potential impact of the cost recovery model on the public sector information market

Taxpayer funded public sector information	Public sector information supplied on a cost recovery basis
<ul style="list-style-type: none"> Public sector information typically made available at no cost to users under the Open Government Licence 	<ul style="list-style-type: none"> Access to public sector information involves a cost to the user and may come with restrictions
<ul style="list-style-type: none"> There may be strong public policy reasons for having these datasets available at no cost 	<ul style="list-style-type: none"> The costs of collection (rather than dissemination) are significant and cannot be borne by the public purse
<ul style="list-style-type: none"> Will not include other commercial or international datasets 	<ul style="list-style-type: none"> May include other data sourced under licence
<ul style="list-style-type: none"> The availability of public sector information at no cost to the user can contribute to the transparency agenda by increasing access to the widest possible customer base, irrespective of the ability to pay 	<ul style="list-style-type: none"> Access to Public sector information is restricted on the basis of cost
<ul style="list-style-type: none"> Will not typically include other services or any other guarantees 	<ul style="list-style-type: none"> May also include bespoke value-add services and guarantees of data quality and continuity

Source: Deloitte analysis

It is very difficult to perform a robust cost-benefit analysis of different funding model options for PSIAs, not least because PSIAs differ greatly from one to the next. While some data is available as to the costs incurred from collection and dissemination, this is not typically openly available or apportioned by dataset, nor is data readily available on how their customers are using the datasets and generating revenue and value.

More fundamentally, very little data is available on what the benefits might be if charging models were to radically change – as it is very difficult to predict how businesses and individuals might use datasets in the future to generate new products and services, and by implication impact economic growth. It is also important to note that, as per the HMT Green Book¹⁶⁷ guidance, any benefits from a change in charging structures should include not just increased tax receipts but wider social benefits and costs in terms of organisational impacts.

Even in the case of the four Trading Funds that make up the PDG, estimating the effects on Exchequer revenue of releasing all their public sector information as open data is a difficult task, in spite of the information made available by the Trading Funds as part of the study. There are differing views on precisely which revenues should be considered relevant to PSI, considering factors such as specific revenues from the data, the cost of collecting the data, the extent to which Government ‘buys’ data from itself and so forth.

That notwithstanding, on the basis of the information made available to this study, it is possible to indicatively calculate that the cost effects on Exchequer revenue of continuing to collect and disseminate Trading Funds’ public sector information in its current guise without charging for it. This cost is estimated to be in the order of £395 million on an annual basis¹⁶⁸. However this figure

¹⁶⁷ Available at www.hm-treasury.gov.uk/data_greenbook_index.htm

¹⁶⁸ This figure comprises of Trading Funds’ operating surpluses and the cost of data collection.

is without regard to the extent that Government pays for public sector information from the four Trading Funds in the PDG (this varies significantly between Trading Funds). On the basis of information provided on sales channels, it can be estimated that the annual loss to the Exchequer would be lower, as government would no longer need to purchase these public sector information datasets – it could use them at no cost. In this scenario, the loss to the Exchequer on an annual basis might be of the order of £143 million. This figure may be lower still if there are efficiency savings to be made if fewer dedicated sales and marketing resources are required by Trading Funds.

Following the HMT Green Book approach to account for the wider social and economic benefits, it is important to note that in a world without charging, private sector entities (consumers and businesses) that currently pay for access to public sector information provided by the Trading Funds would benefit by this amount - around £143 million – and some of this may be recouped in the medium-term as a result of additional economic activity generating tax revenues.

An additional group that is currently deterred by having to pay for the data would also benefit as they are able to access the data. Estimating the size of this latter group is difficult but directionally it is clear that removing charging would mean more people and businesses, not fewer, would be able to access and benefit from this data¹⁶⁹. Conversely, organisations who are at an advantage in using their own proprietary information for commercial advantage, might find their competitive advantage eroded if more public sector information is released to act as a publically available substitute to that data.

The situation is thus complex, and while the above example is stylised, it does suggest the quantum necessary for the associated benefits to outweigh the costs. Without more detailed and accurate data on both costs and benefits (including wider social benefits) it is not possible to reach a clear conclusion on this issue.

However, this is not to say there is no room for improvement today in the provision of public sector information that carries a charge and a number of steps are being taken in this respect. These include:

- much more communication about what existing licences allow consumers to access and use / re-use the public sector information for, building on existing efforts by the trading funds and other PSIHS to build awareness among the user community;
- offering substantial discounts to SMEs and individuals;
- implementation of a ‘royalties’ model for consumers to exploit the value of public sector information up to a certain value before a charge is applied;
- greater provision of ‘sandbox’ or secure environments in specialised locations across the UK to allow consumers to explore datasets;
- greater provision of out-of-date public sector information at no cost to the general public to allow consumers to experiment; and
- greater use of ‘hack days’ to demonstrate the value of particular public sector information datasets.

These options have not been costed or their benefits estimated.

A number of these initiatives are already being pursued and there appears scope and benefit to their wider roll-out. Of course, not all of these initiatives will be applicable to all PSIHS that charge for public sector information – there is no ‘one-size-fits-all’ model – but they should work with the grain of the market and build on existing initiatives to release more data. In some cases there may be significant logistical or legislative challenges to overcome before the above suggestions are implemented. For example, under the IFTS differential charging is prohibited; and a single sandbox

¹⁶⁹ There would accordingly be second-round effects for people consuming products provided by these businesses, etc.

model may not effectively meet the needs of all developers. There are also questions whether PSIHs can effectively separate communication and marketing efforts promoting the use of their public sector information from marketing to promote their value-added services which may be in direct competition with the private sector.

Insights:

- **The issue of charging for public sector information datasets and their funding models is complex. While significant progress has been made by a number of major PSIHs in this area in simplifying charging structures and making more public sector information available as open data, there remains a perception that barriers exist and there is scope for improvement.**
- **Ways to improve access could include greater communication on licence conditions, discounts to certain consumers, a royalties model, use of a 'sandbox' model, greater provision of out-of-date information and more hack days.**
- **There is currently a lack of data to definitively conduct a cost-benefit analysis across different funding models across the range of PSIHs currently charging. This is an area for further analysis through primary research with the direct customers of public sector information and a detailed cost apportionment exercise to assign costs to individual datasets.**

Access barriers

The evidence reviewed suggests a number of access barriers to fully maximising the value of public sector information:

- **difficulties around finding where public sector information is located;**
- **a lack of skills and understanding to fully exploit public sector information;**
- **the format and reliability of public sector information; and**
- **a reluctance to use and share public sector information.**

Public sector information fragmentation

Fragmentation in the supply of public sector information continues to be a problem for many consumers, even following the establishment of a number of data portals. A sample review of websites done as part of this study has found that too often it is difficult to locate a clear point of contact, or establish who has ownership of and responsibility for a particular dataset. Even on data.gov.uk, where contact details are provided, these are often generic enquires email addresses rather than named contact email addresses.

In contrast, this report has received feedback that when consumers have found the relevant point of contact in a PSIH their experience has been very positive with a productive dialogue being established. Indeed, some individual PSIHs have established clear procedures for consumers seeking to raise queries or challenge restrictions, although this is not yet happening in all areas of the public sector. Dialogue can also take place between users: data portals such as data.gov.uk and the London Datastore have busy forums and request pages which encourage an active dialogue between information users and re-users and information holders.

Public sector information fragmentation can, at best, raise transaction costs from dealing with public sector information and, at worst, deter users and re-users from using public sector information altogether, or make this impossible to achieve. Reducing fragmentation across the

public sector could save up over £50 million per annum in terms of reduced transaction costs and time saved for data specialists¹⁷⁰.

However, it should be noted that the existing fragmentation and opacity of some public sector information datasets has created market opportunities for some intermediaries to develop new products that aggregate datasets and present it innovative ways. Thus, PSIHs need to consider who is using their data and how and whether there is a risk of crowding out innovation if they intervene to reduce fragmentation in such a way that duplicates what the private sector can provide.

Insights:

- **Improvements are on-going to reduce the time taken to locate public sector information datasets. Reducing this can help reduce transaction costs and the overall cost of doing business.**

Data scientists and the skills gap

A number of studies have recently been published contending that advanced economies face a skills gaps in so-called 'data scientists'. Although statisticians and experts on quantitative analysis have long existed, data scientists differ from these existing professions in a number of important ways. As well as being able to work with large volumes of structured and unstructured data, they are able to translate these analyses into policy and commercial-ready insights and effectively communicate them to a range of stakeholders, often using innovative tools and visualisations. Key to this is the ability to "identify rich data sources, join them with other, potentially incomplete data sources, and clean the resulting set." In many ways they therefore resemble scientists more closely than traditional data analysts.

There is a fear that a lack of data scientists will reduce the UK's competitive advantage. The evidence received suggests that in the UK:

- there is increasing demand for individuals with a portfolio of skills able to manipulate quantitative data, present it in innovative ways and generate commercial and policy insights from it;
- many of the individuals performing these roles have no specialised training, but rather have learned on the job and / or have a science/computation/mathematics background;
- businesses rarely designate specific 'data scientist' roles; rather, such analyses are done across a combination of professions such as statisticians, economists, researchers, analysts, policy and commercial managers – a dedicated data scientist would embody elements of all these roles;
- certain industries such as pharmaceuticals, financial services, professional services and retail are increasingly dependent on these skills sets and a shortage of them would reduce the UK's international competitive advantage.

Data scientists will also be conversant with the vocabulary of public sector information and open data and have the skills to create and manipulate large datasets and linked data.

Based on ONS figures for 2011, around 1.5 million workers in the UK, representing around 5 per cent of the active workforce, are employed in job categories that are likely to involve elements of the role of the data scientist, but which individually may not be termed data scientists. The average annual median wage of these workers was over £36,000 in 2011 which is higher to the national annual median wage of £26,000.

¹⁷⁰ Based on an assigned average value of time spent by individuals in occupations that regularly use public sector information.

Economic theory suggests that a skills shortage will manifest itself in the form of large wage differentials between ‘data scientists’ and other comparable professionals. A recent report found an observable pay premium for ‘big data’ staff in 2012, with salaries around 20 per cent higher than those for IT staff as a whole.¹⁷¹ This may persist due to lags between training and entering or re-entering the workplace, but economic theory suggests it will eventually dissipate in the long-run as supply increases to meet demand and is able to exploit the full value of public sector information.

Evidence suggests the market is beginning to address skills shortages¹⁷² through:

- sector specific leading practices, lowering complexity and the subsequent talent learning curve (including setting legal precedents for legal liability and compliance); and
- universities are having new certifications for big data disciplines.

However these initiatives will take time to work through the system with the interim consequence being that some value of public sector information may remain locked up. The general scarcity and increasing competition for these skilled workers from the private sector makes it harder to construct the infrastructure for world class public sector information. A shortage of data scientists also hinders efforts to scale-up public sector information data analytics.

Within the public sector, concerns have been highlighted over a lack of skills and familiarity to work effectively with data. These concerns should not be overstated as public sector officials have a long history of using public sector information to inform policymaking without having dedicated data scientists. What the concerns appear to be directed at are cultural biases against using public sector information from outside home PSIHS, as well as having the necessary skills to combine and manipulate Big Data and Linked Data¹⁷³. Comments from stakeholder conversations and the government official’s informal consultation reinforce this concern – the following being representative:

- “There is currently a lack of awareness and skills in blending and combining multiple datasets. This will need to be addressed if we are to fully exploit the power and potential of integrating disparate sources over the web.”
- “Data users do not fully appreciate the power and potential of open data. There may also be cultural resistance to change:, for instance, through not trusting or being able to exploit new, untried and untested third-party datasets in analysis and policy advice.”¹⁷⁴

Where public sector employees do not have a numerate or scientific higher education background, or are not accustomed to working with large datasets, they could benefit from increased training and support in this area. However, there are understandable concerns over resources at a time when departmental budgets are under considerable pressure. The danger, therefore, is the training in this area is viewed as a ‘nice to have’ which is not currently a budgetary priority and is therefore neglected. Were this to happen, much of the effort to open up public sector information and facilitate sharing across the public sector would be poorly spent, as public sector employees would lack the skills to exploit the available data – the PSIHS would only be able to operate on the supply of the market, lacking the skills to be innovative public sector information consumers.

¹⁷¹ See e-skills UK, ‘*Big Data Analytics: an assessment of demand for labour and skills, 2012-2017*’ (January 2013)

¹⁷² See Deloitte Tech Trends 2013, available at www.deloitte.com/view/en_GB/uk/services/consulting/technology/technology-trends/abfbfdad4ac310VgnVCM3000003456f70aRCRD.htm

¹⁷³ However, it should be noted that many cases the linking itself is carried out by specialist data companies, which may be an example of increased activity in this space creating or expanding a market for private companies. This minimises the internal skill base that public sector bodies are required to develop, reducing the burden of releasing data: they are able to release data in raw form for linking by third parties. Nonetheless, some public sector bodies which work extensively with data, such as HEFCE, are known to have developed in-house capabilities in this area. In addition some researchers in the field, recognising that they depend heavily on linked datasets, have developed the technical expertise to do this work themselves.

¹⁷⁴ Government officials consultation, November 2012

There are a number of routes around these budgetary concerns. One example is the use of massively open online courses (MOOCs) to enhance skills sets and increase exposure to data manipulation techniques. These courses are free of charge to the user and include offerings from a number of prestigious higher education institutions. For a limited time commitment over a period of weeks, generally between three and six hours, students receive online material from academics, as well as access to a range of study material, online forums, and in some cases certification following course completion and assessment. Using estimates of public service wage levels and assuming an eight week long course of six hours of study a week, the opportunity cost value of taking a MOOC during office hours is in the order of £500 per participant, although the benefits from improved skills are likely to be much higher.

In practical terms, measures could be taken via internal communications to raise awareness of the benefits of taking these courses. In addition it would be helpful if completion and accreditation of relevant courses was viewed favourably in performance assessments, which would act as an incentive. A commonly used approach to in transformation programmes is to identify 'champions' across all grades who can be encouraged to take the courses and then act as advocates among their colleagues; once a critical mass of understanding and enthusiasm is attained, this becomes self-reinforcing and employees should start taking the courses as part of their continuing training and development.

A selected list of relevant MOOCs, available at the time of writing, is included below for guidance.¹⁷⁵

Course name	MOOC provider	Higher Education Institution	Estimated time commitment	Web link
Passion driven statistics	Coursera	Wesleyan University	3-4 hours per week (6 weeks)	www.coursera.org/#course/pdstatistics
Web intelligence and big data	Coursera	IIITC	2-3 hours per week (10 weeks)	www.coursera.org/#course/bigdata
Introduction to data science	Coursera	University of Washington	10 weeks	www.coursera.org/#course/datasci
Statistics: making sense of data	Coursera	University of Toronto	6-8 hours per week (8 weeks)	www.coursera.org/#course/introstats
Data analysis	Coursera	John Hopkins University	3-5 hours per week (8 weeks)	www.coursera.org/#course/dataanalysis

¹⁷⁵ Note: the above is a selective list, and is not intended to be a comprehensive survey of all relevant MOOCs. Availability of the above courses may alter over time as the majority of the courses listed are to commence in Spring 2013. The level of suggested background knowledge varies, and not all of the above courses will be appropriate for all employees. We have selected the courses based on a combination of our personal experience and our assessment of the range of data analysis needs that may be useful for a public sector employee, but we are unable to offer any comment or assurance as to the quality of the course content or delivery, or the continuing availability of any of these courses.

Course name	MOOC provider	Higher Education Institution	Estimated time commitment	Web link
Computational methods for data analysis	Coursera	University of Washington	10 weeks	www.coursera.org/#course/compmethods
Computing for data analysis	Coursera	John Hopkins University	3-5 hours per week (8 weeks)	www.coursera.org/#course/compdata

Note: no analysis has been done as to the quality of the above courses' content, which could vary significantly.

Traditional training will of course continue to play an important role, as well as interactive and workshop sessions, such as mash-up days, especially those involving external developers. These are useful for sharing knowledge and expertise and creating an environment which is conducive to experimentation and innovative thinking.

In addition, it may be that public sector organisations will wish to increase their recruitment of those with a scientific or numerate higher education background, so as to expand the pool of relevant talent upon which they are able to draw. This will vary according to the organisation in question and the appropriate skills mix for the performance of their public task; for this reason we do not make an across the board recommendation regarding recruitment. However, it is likely to be helpful for public sector organisations to consider the alignment of their current and future skills based with the increasing availability and use of information from across the public sector.

There will also be a role for Cabinet Office and other public sector organisations to build on notable achievements¹⁷⁶ in public sector information provision to foster a more conducive culture within the public sector to using public sector information innovatively. The benefits accruing from improved use of information within the public sector are likely to outweigh these. Adapting the McKinsey analysis used by Policy Exchange¹⁷⁷, improvements in efficiency of between 1 and 5 per cent can lead to annual savings of between £1 and £8 billion nationally and around £70 million in local government (on a smaller savings ratio).

Insights:

- While there may be gaps in the supply of 'data scientists', economic theory suggests that in the medium- to long-term, the number of data scientists will increase, filling the supply gap and reducing the current wage premium.
- However, in the short-term, this may mean public sector information is left under-exploited and value remains locked. The general scarcity and increasing competition for these skilled workers can make it harder to construct the infrastructure for world class public sector information and scale up efforts to exploit its value.
- There are some low-cost solutions that can be explored to quickly improve the skills base to be able to effectively manipulate and extract value from public sector information.

¹⁷⁶ These include (i) a process to drive the supply of open data from Whitehall, (ii) achieve the release of over 9,000 datasets on data.gov.uk, (iii) assist in the establishment of the Open Data Institute; and (iv) support the UK's efforts globally to improve open data and the wider transparency agenda of the Open Government Partnership and the G8.

¹⁷⁷ See www.policyexchange.org.uk/publications/category/item/the-big-data-opportunity-making-government-faster-smarter-and-more-personal

Format and reliability

The release of data in an unfinished form raises concerns for the public sector too because of concerns that the public may be provided with data that is inaccurate and potentially misleading, which can have negative consequences.

With respect to the format of public sector information, its importance may vary between customers. For casual consumers, it is clearly of the utmost importance, and they may most value format. In contrast, professional users may be more concerned with consistency of service, commitment to on-going supply of data and data and can accommodate changes to format. Equally, intermediaries may positively value low quality data as they can provide a service to improve the quality and format of public sector information for wider consumption.

Evidence as to how significant a barrier this currently forms to the use and re-use of public sector information is mixed. In general, there seems to be steady improvement in all the areas of format, although there remains work to be done – especially with regard to ensuring consistency in format and upgrading the star rating of datasets. Progress is also being made with data being updated more consistently. Examples of this include data released by the Met Office, and transport data released directly by Transport for London as well as through data.gov.uk.

Land Registry's release of PPI data is an example of good practice. The data was released in CSV (3 star format) then upgraded to linked data (4 star standard). The quality of the data was also added the organisation's list of Key Performance Indicators and anyone publishing the data is asked to include a link to a dedicated Land Registry team for reporting any inaccuracies in the data.

With respect to format, it is well established that data which is released should, wherever possible, be in machine-readable formats. Format is therefore linked to usability. For a commercial service to be based on an data source, there must be confidence that the data will be maintained and kept up to date. The ODI notes that "the sustainability of data publication is all important for those who build services on top of that data. So long as it is machine-readable (2-star) and openly licensed...the frequency, consistency and accuracy of the publication of a dataset is more important than the format in which it is published."¹⁷⁸ The reasons for this are easy to understand: if a business offers a service to customers, it cannot afford for that service to be unpredictably cut off or to offer out of date information. This is especially true of time-sensitive (volatile) information, such as travel updates and weather forecasts.

As highlighted in an earlier chapter, the working assumption of this report is that around 50 per cent of datasets are three stars or above (based on a review of data.gov.uk datasets and other sources including the data released by local government). The question is then whether this current level of usability acts as barrier to use and re-use of public sector information.

Conceptually, datasets that fail to meet the required standard may be constraining the ability to generate value from public sector information datasets in a number of ways:

- requiring users and re-users to have specialist proprietary software to view public sector information datasets. Examples include the early release of the COINS database, which due to its size and nature was difficult for users to manipulate;
- preventing users and re-users from exploiting the benefits of semantically linked datasets; and
- restricting access to certain public sector datasets and limiting their uses / re-uses.

Five star data offers significant advantages to users in terms of combining datasets together, which is one of the areas where this report has identified large potential benefits to be generated.

However, moving towards a five star standard could involve significant cost to PSIHs. Following discussions on the merits of having all data meeting this standard with members of the start-up

¹⁷⁸ See www.theodi.org/consultation-response/improving-local-government-transparency

community and other stakeholders, it appears that while it is considered desirable, the consensus seems to be that as long as data is machine-readable the lack of a five star rating is not currently a significant barrier for users and re-users. There is concern that pushing too hard for all data to be upgraded to five star status could dissuade information holders from releasing data at all, given the cost. In general, users have expressed a preference for data to be released earlier with a lower star rating, and then upgraded when possible.

In addition, consistency of format across comparable datasets is important, as this affects how easily they can be compared and combined. This is a particular issue where there are many organisations with their own outputs, such as local authorities and the NHS.

Of course, as noted earlier, a higher star rating cannot be viewed as a simple proxy for higher data quality. A dataset may be in a machine readable format and available as linked data, but still contain inaccurate or incomplete data. In general, stakeholders have reported that the quality (including reliability of delivery [see below] and accuracy) of the data is of a higher priority than upgrading it beyond two or three stars, since the quality affects the fundamental usefulness of the data.

The Open Standards Principles¹⁷⁹, the Standards Hub¹⁸⁰, the Open Standards Board and the public sector will be crowdsourcing, researching and implementing data standards for Government IT systems. The user challenges it will focus on are likely to cover standards relating to formats and meta-data that should help to provide data on reliability. Some of these open data standards may be made compulsory for central government use.

However, it should be noted that a minority of stakeholders have complained that datasets released by different PSIHs are not always easy to combine and work across, because of variations either in the content or the format. This is a particular issue with data produced by local authorities, with each local authority often adopting their own standards and procedures. This report recognises that there are on-going efforts, such as e-PIMS, to tackle this issue and secure a greater degree of standardisation across PSIHs.

Further, there appears to be scope for improvement is greater certainty and clarity over the publication schedule of public sector information datasets and what users can expect from PSIHs. In particular, certainty for businesses making investment decisions to use public sector (and other) information datasets could be improved through:

- having a clear articulation of each dataset's publication schedule;
- having a cover sheet setting out the limitations of the data, explaining outliers and providing links to previous analyses (greater use of **metadata**¹⁸¹);
- having a clear indication of when any given dataset may be discontinued; and
- the level of PSIH support that will be provided (perhaps at a charge).

The levels of support and explanations of the dataset could vary by dataset type.

Clearly there will be a cost involved in this for PSIHs (one-off and on-going), but the benefits to businesses from increased certainty could be significant.

Insights:

- **Improvements continue to be made on the quality, format and consistency of public sector information.**
- **There is scope for improvement, especially around the greater provision of metadata.**

¹⁷⁹ See www.gov.uk/government/news/government-bodies-must-comply-with-open-standards-principles

¹⁸⁰ See <http://standards.data.gov.uk/>

¹⁸¹ That is, broadly speaking, data about data.

Chapter summary

- This chapter has summarised the main barriers to the release, use and re-use of public sector information. These include
 - Legislative barriers: while current legislation, guidance and regulations are not hindering the public sector information market themselves, there is often a perception (wrongly or rightly) that some legislation, guidance and regulations prevent datasets from being released and shared thereby reducing the availability of datasets. The Open Government Licence is widely seen as an effective means of improving the availability of public sector information;
 - Economic barriers: there are significant data gaps around exactly what public sector information exists, making it difficult to reach decisions around its optimal provision. The issue of charging and funding models for public sector information is complex, with the lack of data making it hard, at present, to accurately conduct cost-benefit exercises on the extent that charged-for data might be delivered for free in future; and
 - Access barriers: improvements continue to be made to reduce fragmentation and improve the consistency of public sector information; however there remains scope for improvements in the accessibility of public sector information datasets to the general public. Equally, in some parts of the private and public sector, there is a skills shortage preventing the effective extraction of value from public sector information.

7. Conclusion

This report has considered, for the first time, in its entirety, the public sector information market in the UK. It has provided details on the supply and demand sides of the market, the legislative and regulatory framework, the current value of the market and identified a number of barriers that are preventing the UK from maximising the full value of public sector information. This report forms the evidence base for the Shakespeare Review, which is making recommendations to Government.

This concluding chapter sets out some closing thoughts on the subject of public sector information in the UK.

Closing thoughts

- Through a review of the literature and real-life examples of its use and re-use, it is clear that public sector information is generating significant value for the UK currently – in terms of financial, economic value and employment benefits, but also wider social benefits.
- The increasing availability and better quality of public sector information can lead to even more value being generated in the UK and also overseas through network effects.
- By being a global leader in the release of public sector information, the UK government is helping create an environment that is conducive for innovation and the development of new skills that could give UK businesses a competitive edge overseas¹⁸².
- This report has highlighted that while the availability and quality of public sector information has improved in recent years, which has helped the market evolve and grow, there remain certain barriers that may hinder or constrain the continued growth of the market. The policy insights raised suggest that, in some cases, there may be a role for Government to step in and address market failures.
- In addressing market failures, Government will need to balance the interests of different stakeholders, but equally it should not necessarily be constrained by the current fiscal environment in order to consider a much longer time-horizon of benefits and costs.
- However, it should not be forgotten that public sector information is only one component in the wider data landscape – the increased availability of more and better private sector data (open or otherwise) will also have significant benefits.
- Indeed, a key barrier identified has been the lack of information on how businesses use public sector information. Fostering a culture of greater openness and collaboration around data and its use will benefit all parties. For example, one way of collaborating and fostering more openness in future, providing it can be carried out in a not-too-onerous manner, might be providing data free to third parties on the pre-condition that the information flow becomes a two-way street for Government and policymakers, e.g. *“you can have our data, but we’d like to know how it is being used and re-used to give us insight, benefit us and in turn UK society”*.

¹⁸² Though this has not explored in detail as part of this report.

Appendix 1: Glossary

These glossary definitions have been taken from a variety of sources including the Open Data White Paper¹⁸³ (June 2012), the Advisory Panel on Public Sector Information (APPSI) Glossary website¹⁸⁴ and the OFT Commercial use of Public Information report¹⁸⁵ (the 'CUPI' report, December 2006). As noted in that document, given the relatively nascent nature of the market and rapidly changing landscape, these definitions will have an element of. At the time of writing (January 2013), Cabinet Office is consulting to gain a collective view on definitions to be used in forthcoming Transparency and Open Data publications. This is to be done via a 'wiki' site hosted by the APPSI¹⁸⁶.

For reference the source for each glossary term is listed. Where there is no term this refers to cases where the term has been defined specifically for this research.

For reference the source for each glossary term is listed. Where there is no term this refers to cases where the term has been defined specifically for this research.

Aggregated data	A form of anonymisation of unit records involving combinations such that individual records are not disclosed. <i>Source: APPSI Glossary</i>
Anonymised data	Data that has been adapted so that individual businesses, individuals and other organisations cannot be identified from it. <i>Source: adapted from APPSI Glossary</i>
Application programming interface	A specification intended to be used as an interface by software components to communicate with each other. An API may include specifications for routines, data structures, object classes and variables. <i>Source: adapted from APPSI Glossary</i>
Attribution licence	A licence that requires that the original source of the licensed material is cited. <i>Source: Open Data Handbook</i>
Asset list	A register of data and information items held by a PSIH which are of interest or value to the PSIH itself, and potentially to others. <i>Source: OFT CUPI</i>
Authoritative	Able to be trusted as saying something. Some information can be accurate but not authoritative, but it will be both if it comes from a source with authority to provide it. <i>Source: APPSI Glossary</i>
Big Data	Gartner ¹⁸⁷ describes Big Data as data defined as high volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.

¹⁸³ Available at: www.gov.uk/government/publications/open-data-white-paper-unleashing-the-potential

¹⁸⁴ See www.nationalarchives.gov.uk/appsi/appsi-glossary-a-z.htm

¹⁸⁵ Available at: www.of.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861

¹⁸⁶ See www.nationalarchives.gov.uk/appsi/open-data-psi-glossary-pilot.htm. At the time of writing the wiki had not yet been established.

¹⁸⁷ See <http://www.gartner.com/it-glossary/big-data/>

	<i>Source: Gartner</i>
By-product data	Data that is generated through the performance of regular or one-off activities, where the generation of the data was not the primary objective of the activity or part of its Public Task. Also referred to as 'exhaust data'.
Class licence	A licence that sets out standard terms and obligations, enabling the re-use of a particular class or category of material. <i>Source: OFT CUPI</i>
Click-use	The online licensing system for Crown and Parliamentary copyright information developed by the Office of Public Sector information in 2001. This has subsequently been superseded by the Open Government Licence and Open Parliament Licence, but remains historically significant. <i>Source: UK Government Licensing Framework</i>
Commercial use / re-use	Use or re-use that is intended for or directed towards commercial advantage or private monetary compensation. For clarity, commercial advantage applies either to reselling the data through products in any form or to internal use or re-use to improve business effectiveness. <i>Source: The National Archives Information Management Glossary with APPSI Glossary addition</i>
Consumer surplus	The value or benefit consumers of a product or service enjoy over and above the price they paid for it. It is the difference between the price consumers pay and the price they are willing to pay.
Copyright	Part of the family of intellectual property rights including trademarks, designs and patents. Copyright applies automatically when a work is created in a material form. Copyright applies to literary works, such as website articles/annual reports; artistic works maps, drawings, paintings and photographs; films; sound recordings; broadcasts; dramatic and musical works and typographical arrangements. The first owner of copyright will normally be the artist/author or organisation that created the work (except for Crown copyright). Copyright subsists in a work regardless of the level of artistic or literary merit. The standard term of copyright is the life of the author plus 70 years. <i>Source: The National Archives Information Management Glossary</i>
Core reference data	Authoritative or definitive data necessary to use other information produced by the public sector as a service in itself due to its high importance and value. <i>Source: Open Data White Paper</i>
Crown body	An organisation which acts on behalf of the Crown, meaning the sovereign acting in a public or official capacity. This includes most central government departments including government Trading Funds. In many cases the Crown status or otherwise is specified within the context of legislation. <i>Source: OFT CUPI</i>
Crown copyright	Crown copyright covers material created by civil servants, ministers and government departments and agencies. It is legally

defined under section 163 of the Copyright, Designs and Patents Act 1988 as works made by officers or servants of the Crown in the course of their duties. Copyright made by Her Majesty or by officers can also come into Crown ownership by means of an assignment or transfer of the copyright from the legal owner of the copyright to the Crown.

Source: APPSI Glossary

Crown copyright waiver

Categories of material on which the Crown asserts its copyright but waives it and which is not subject to formal licensing or payment.

Source: The Future Management of Crown Copyright White Paper

Customer insight Data

Data or information recording users' accounts of their experience, with an assessment of public service providers.

Source: Open Data White Paper

Data (singular or plural)

Qualitative or quantitative statements or numbers that are assumed to be factual and not the product of analysis or interpretation. Data can be structured or unstructured.

The term structured data refers to data that is identifiable because it is organised into a recognisable structure. The most common form of structured data (structured data records (SDR)) refers to a database where specific information is stored based on a methodology of columns and rows.

In contrast, unstructured data has no identifiable structure.

The terms data, information and knowledge are frequently used for overlapping concepts. The main difference is in the level of abstraction being considered. Data is a broad term, embracing others, but is often the lowest level of abstraction, information is the next level and, finally, knowledge is the highest level.

Source: adapted from APPSI Glossary and Open Data White Paper

Data controller

A person who (either alone or jointly with other persons) determines the purposes for which and the manner in which any personal data are, or are to be, processed.

Data discovery benefits

Benefits (economic or otherwise) that arise from the analysis of data and information itself.

Data enabler

An intermediary that facilitates public sector information or open data initiatives without actually publishing or consuming data themselves, usually through software, platform, data centre infrastructure etc. Service types include platform-as-a-service, infrastructure-as-a-service, and Software-as-a-service.

Data exploitation benefits

Benefits (economic or otherwise) that arise from using data and information as inputs into products and services.

Data holder

The public sector body holding public sector information. Also referred to as Public Sector Information Holders.

Data infomediary

An intermediary that aggregates, scrapes or collects publicly available data or data stores that enhance the raw data with visualization, completeness, accuracy, analysis, or accessibility.

Data mash-up	Combining different and distinct datasets to create new datasets and generate new insights.
Data owner	The data owner (organisation or individual) is responsible for understanding what information is brought into a system, assigning meanings to data collections and constructing and modifying data models.
Data processor	With respect to personal data, this refers to any person (other than an employee of the data controller) who processes the data on behalf of the data controller.
Data protection	<p>The Data Protection Act 1998 defines the ways in which data and information on living persons may be legally used and handled. The Act sets out the fundamental principles with which personal data must satisfy, including:</p> <ul style="list-style-type: none"> (a) be processed fairly and lawfully; (b) be obtained only for lawful purposes and not processed in any manner incompatible with these purposes; (c) be adequate, relevant and not excessive; (d) be accurate and current; (e) not be retained for longer than necessary; (f) be processed in accordance with the rights and freedoms of data subjects; (g) be protected against unauthorised or unlawful processing and accidental loss, destruction or damage; and (h) not be transferred to a country or territory outside the European Economic Area unless that country or territory protects the rights and freedoms of the data subjects. <p><i>Source: adapted from Data Protection Act</i></p>
Data scientist	An analytical role that is built on three core skills: data management, data analysis and business and policy insight. While the formal training of data scientists is typically in statistics, mathematics or computer science, they will also have strong communication skills and business/policy acumen. They have been described as “part analyst, part artist” by IBM.
Data sharing	<p>The transfer of data between different organisations to achieve an improvement in efficiency and effectiveness. This document assumes that data sharing will continue to operate in line with current domestic legislation and the UK’s international obligations.</p> <p><i>Source: Open Data White Paper</i></p>
Data subject	Under the Data Protection Act 1998, this refers to an individual who is the subject of personal data.
Dataset	<p>As defined in the Protection of Freedoms Act 2012: “ ‘dataset’ means information comprising a collection of information held in electronic form where all or most of the information in the collection:</p> <ul style="list-style-type: none"> (a) has been obtained or recorded for the purpose of

providing a public authority with information in connection with the provision of a service by the authority or the carrying out of any other function of the authority,

- (b) is factual information which —
 - (i) is not the product of analysis or interpretation other than calculation, and
 - (ii) is not an official statistic (within the meaning given by section 6(1) of the Statistics and Registration Service Act 2007), and
- (c) remains presented in a way that (except for the purpose of forming part of the collection) has not been organised, adapted or otherwise materially altered since it was obtained or recorded.”

It is important to note that when one talks about accessing public sector information or open data, a single page on a data portal such as data.gov.uk does not necessarily link to one item of data or dataset – there can be multiple datasets.

Source: adapted from Data Protection Act and APPSI Glossary

De-anonymisation

The process of determining the identity of an individual to whom a pseudonymised dataset relates.

Source: Open Data White Paper

Derived data

A data element adapted from other data elements using a mathematical, logical or other type of transformation.

Source: OECD Glossary of Statistical Terms

Disclosive

Data is potentially disclosive if, despite, the removal of obvious identifiers, characteristics of this dataset in isolation or in conjunction with other datasets in the public domain might lead to identification of the individual to whom a record belongs.

Source: Open Data White Paper

Executive agencies

A diverse group of organisations delivering a variety of services to internal and external customers. They are part of the Crown and do not usually have their own legal identity, but operate under powers that are delegated from Ministers and Departments.

Source: OFT CUPI

Exhaust data

Data that is generated as a ‘by product’ of an organisation’s activities, i.e. where data has been generated as a result of other activities but not as the primary purpose of these activities.

Full cost pricing

A pricing policy in which charges are set to recover the full resource costs of the activity.

Source: OFT CUPI

Free at point of use

Where there is no charge or fee to the end-user for the use or re-use of information.

Source: A Consultation on Data Policy for a Public Data Corporation – glossary

Freemium

A business model by which a product or service is provided free of charge, but a premium is charged for advanced features or functionality.

	<i>Source: APPSI Glossary</i>
Geospatial data	Data or information that identifies the geographic location of features and boundaries on Earth, such as natural or constructed features, oceans and more. <i>Source: adapted from APPSI Glossary</i>
Identifier	A particular element or reference in a dataset that allows individuals' or businesses' identity to be known.
Information	Output of such process that summarises, interprets or otherwise represents data to convey meaning. <i>Source: Open Data White Paper</i>
Information asset register	Registers specifically set up to capture and organise metadata about the vast quantities of information held by government departments and agencies. A comprehensive IAR includes databases, old sets of files, recent electronic files, collections of statistics, research and so forth. <i>Source: OFT CUPI Open Data Handbook and Open Data Manual</i>
Information fair trader scheme	A scheme to set and assess standards for public sector bodies in allowing the re-use of their information. Any public sector body may apply to become IFTS accredited. However, all Crown bodies that hold a delegation of authority from the Controller of HMSO must become IFTS accredited. IFTS measures members' performance against the six principles of maximisation, simplicity, transparency, fairness, challenge and innovation. It considers both the commercial re-use of public sector information and non-commercial citizen access to information. <i>Source: The National Archives Information Management Glossary</i>
Intellectual property	A set of property rights that grant the right to protect the materials created by them. Intellectual property comprises among other things copyright, designs, patents, database rights, certain confidential information and trademarks. <i>Source: Open Data White Paper and OFT CUPI</i>
Licence	Permission by the copyright holder to reproduce or re-use material protected by copyright. <i>Source: OFT CUPI</i>
Linked data	The term used to describe the recommended best practice for exposing, sharing and connecting items of data on the semantic web using unique resource identifiers (URIs) and resource description framework (RDF). <i>Source: APPSI Glossary</i>
Market failure	An instance where a market is not efficiently allocating goods and services. Market failures can include information asymmetries, non-competitive markets, principal-agent issues and externalities / public goods.
Marginal cost pricing	The cost of supplying another unit. Long run marginal cost is the full extra cost (both fixed and variable) of providing a further unit of output. Short run marginal cost measures how variable costs change when output alters. In practice, marginal costs are difficult to observe, and average variable costs are used as a substitute

for the concept of marginal costs.

Source: OFT CUPI

Metadata

Data that describes or defines other data. Anything that users need to know to make proper and correct use of the real data, in terms of reading, processing, interpreting, analysing and presenting the information. Thus metadata includes file descriptions, codebooks, processing details, sample designs, fieldwork reports, conceptual motivations, etc., in other words, anything that might influence the way in which the information is used.

Source: OCED Glossary of Statistical Terms and www.sasc.co.uk/Guides/metadata.htm

Mosaic effect

The process of combining anonymised data with auxiliary data in order to reconstruct identifiers linking data to the individual it relates to. Also referred to as the 'jigsaw effect' or 'cross-referencing'.

Source: adapted from Open Data White Paper

Non-commercial government licence

A legal solution to enable the provision and use of public sector information under a common set of terms and conditions at no charge for non-commercial use only. The main requirement for re-users is to attribute the information provider and source.

Source: UK Government Licensing Framework

Non-departmental public body

A body which has a role in the process of national government, but is not a government departments or part of one, and therefore operate to an extent at arm's length from Ministers.

Source: OFT CUPI

Open access

At its most narrow, this refers to the provision of free access to peer-reviewed academic publications and other information data to the general public.

Source: modified from Open Data White Paper

Open data

Data that meets the following criteria:

- (a) accessible (ideally via the internet) at no more than the cost of reproduction, without limitations based on user identity or intent;
- (b) in a digital, machine readable format for interoperation with other data; and
- (c) free of restriction on use or redistribution in its licensing conditions.

Open data can be provided by the public and private sector as well as individuals.

Source: adapted from Source: APPSI Glossary and Open Data White Paper

Open government data

Public sector information that has been made available to the public as Open Data.

Source: Open Data White Paper

Open Government Licence

The Open Government Licence (version 1.0), which forms part of the UK Government Licensing Framework, offers a legal solution to enable the provision and use of public sector information under

a common set of terms and conditions. It enables any public sector information holder to make their information available for use and re-use under its terms.

Source: modified from UK Government Licensing Framework

Personal data

As defined by the Data Protection Act 1998, personal data means “data which relate to a living individual who can be identified -

- (a) from those data; or
- (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,
- (c) and/or includes an expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

Source: adapted from Open Data White Paper

Price elasticity of demand

A measure of the responsiveness of the quantity demanded of a product or service following a change in its price.

Producer surplus

The value accruing to producers of goods and services when their output is purchased by consumers. In traditional supply and demand analyses, producer surplus is calculated as the difference between the lowest amount the producer would be willing to sell the good/service for and the price the producer actually sold it for. In many cases this is equal to profit.

Processing

Under the Data Protection Act, processing in relation to information or data, means obtaining, recording or holding the information or data or carrying out any operation or set of operations on the information or data, including –

- (a) organisation, adaptation or alteration of the information or data;
- (b) retrieval, consultation or use of the information or data;
- (c) disclosure of the information or data by transmission, dissemination or otherwise making available; or
- (d) alignment, combination, blocking, erasure or destruction of the information or data.

Pseudonymised data

Data relating to a specific individual where the identifiers have been replaced by artificial identifiers to prevent identification of the individual.

Source: Open Data White Paper

Public domain

Works that are publicly available and in which the intellectual property rights have expired or been waived.

Source: APPSI Glossary

Public good

A good or service provided by Government for public consumption to combat an actual or perceived market failure. Whilst these are generally free-at-the-point of use, for a good to be public it has to be non-excludable and non-rivalrous. The former meaning that no party can be excluded from the benefits conveyed, whilst the latter means that one person’s consumption does not prevent other persons from benefitting. PSI may be

	considered a public good in certain instances.
Public sector body	<p>The State, regional or local authorities, bodies governed by public law and associations formed by one or several such authorities or one of several such bodies governed by public law. (Directive 2003/98/EC on the reuse of public sector information, Art 2).</p> <p><i>Source: OFT CUIP</i></p>
Public sector information	<p>Public Sector Information covers the wide range of information that public sector bodies collect, produce, reproduce and disseminate in many areas of activity while accomplishing their public tasks.</p> <p><i>Source: adapted from BIS and APPSI Glossary</i></p>
Public sector information holder	<p>A public sector body that collects and/or holds information, data or content (as defined).</p> <p><i>Source: OFT CUIP</i></p>
Public task	<p>Public task information is that which a public sector body must produce, collect or provide to fulfil its core role and functions, whether these duties are statutory in nature or are established through custom and practice. The term 'public task' features in the Re-use of Public Sector Information Regulations 2005 (SI 2005 No. 1515) and the INSPIRE Regulations 2009 (SI 2009 No. 3157). The National Archives provides guidance that assists public sector bodies to define and publish a statement of their respective public tasks.</p> <p><i>Source: The National Archives</i></p>
Raw data	<p>In the context of public sector information, raw data is data collected which has not been subjected to processing or any other manipulation beyond that necessary for its first use. Raw data, i.e. unprocessed data, is a relative term; data processing commonly occurs by stages, and the 'processed data' from one stage may be considered the 'raw data' of the next.</p> <p><i>Source: A Consultation on Data Policy for a Public Data Corporation – glossary</i></p>
Refined data	<p>This is where unrefined information has been enhanced, manipulated and/or added to other inputs to create a retail product for businesses or consumers. The process of refining information can be undertaken by a PSIH, or viably in a commercial market by the private sector.</p>
Re-use	<p>Use of information other than for the purpose it was originally produced. This use could be for commercial or non-commercial purposes. (Re-use of Public Sector Information Regulations 2005 (SI 2005/1515), Reg.4(2)).</p> <p><i>Source: adapted from EU PSI Regulations</i></p>
Resource description framework	<p>A W3C standard that is the foundation of several technologies for modelling distributed knowledge and is meant to be used as the basis of the Semantic Web.</p> <p><i>Source: APPSI Glossary</i></p>
Semantic web	<p>A web of data that can be processed directly and indirectly by machines.</p>

	<i>Source: APPSI Glossary</i>
Sensitive personal data	<p>This refers to personal data consisting of information as to a person's:</p> <ul style="list-style-type: none"> (a) racial or ethnic origins; (b) political opinions; (c) religious beliefs or other beliefs of a similar nature; (d) whether they are a member of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992); (e) physical or mental health condition; (f) sexual life; (g) the commission or alleged commission by them or any offence; or (h) any proceedings for any offence committed or alleged to have been committed by them, the disposal of such proceeding or the sentence of any court in such proceedings.
Standard Industrial Classification	<p>First introduced in the UK in 1948, this is a framework for classifying business establishments and other statistical units by the type of economic activity in which they are engaged. There are a number of levels of the classification, with subsequent levels becoming more detailed.</p>
Standard Occupational Classification	<p>A common classification framework of occupational information for the UK on the basis of skill level and skill content.</p>
Star Rating	<p>In UK Linked Data, a system of ranking data sources that indicates ease of machine readability. APPSI subjective score for quality of a definition (qv).</p> <p><i>Source: APPSI Glossary</i></p>
Trading Fund	<p>A government department, executive agency, or part of department, established as a Trading Fund by a Trading Fund Order made under the Government Trading Funds Act 1973. A Trading Fund has authority to use its receipts to meet its outgoings.</p> <p><i>Source: adapted from HM Treasury Glossary (Managing Public Money)</i></p>
Unrefined data	<p>This is data which cannot be substituted directly from other sources. It relates to a PSIH's monopoly activities, where competition is very unlikely. Once a PSIH does something with the data which could be performed viably in a commercial market by the private sector it becomes refined information.</p>
Value-added data	<p>Raw data to which value has been added to enhance and facilitate its use and effectiveness for the user. Value can be added in a number of different ways including further manipulation, compilation and summarisation into a more convenient form for the end-user; editing and/or further analysis and interpretation; and commentary beyond that required for policy formulation by the relevant government department with policy responsibility.</p>

Source: adapted from A Consultation on Data Policy for a Public Data Corporation - glossary

Verbosity

The level of detail of a given dataset.

Velocity

How often data or information in a given dataset changes.

Appendix 2: Acronyms used in this report

API	Application Programming Interface
APPSI	The Advisory Panel on Public Sector Information
BIS	Department for Business, Innovation & Skills
CUPI	Commercial Use of Public Information
DfE	Department for Education
DfT	Department for Transport
DSB	Data Strategy Board
EC	European Commission
EU	European Union
HMG	Her Majesty's Government
HMT	Her Majesty's Treasury
IAR	Information Asset Register
IFTS	Information Fair Trader Scheme
MoJ	Ministry of Justice
NDPB	Non-Departmental Public Bodies
NHS	National Health Service
ODI	Open Data Institute
ODUG	Open Data User Group
OFT	Office of Fair Trading
OGL	Open Government Licence
OS	Ordnance Survey
PSI	Public Sector Information
PSIH	Public Sector Information Holder
SIC	Standard Industrial Classification

SOC	Standard Occupational Classification
SSNIP	Small but Significant and Non-transitory Increase in Price
TfL	Transport for London

Appendix 3: Literature Review

This Appendix surveys the literature that has analysed the role of public sector information in economic growth, both in the UK and internationally. Key themes are the direct economic value of public sector information, the less tangible impacts on economy and society, and debates over the most effective pricing models for public sector information.

Overview

“A new market for public service information will thrive if data is freely available in a standardised format for use and re-use...At present the market for information on public services is highly underdeveloped. Open Data across government and public services would allow a market in comparative analytics, information presentation and service improvement to flourish. This new market will attract talented entrepreneurs and skilled employees, creating high value-added services for citizens, communities, third sector organisations and public service providers, developing auxiliary jobs and driving demand for skills.”¹⁸⁸

In so saying, the Government’s *Making Open Data Real* consultation laid out the argument for the economic impact of increasingly the availability of public sector information for use and re-use. The argument is that for public and private sectors alike, more widespread availability of public sector information will drive efficiencies, boost innovation, lower barriers to entry and enable new insights into old problems. In so doing it holds the potential to reduce public expenditure, promote economic growth and make life more convenient for citizens. The objective of this Appendix is to review the arguments and evidence around this view.

There is a considerable body of literature supporting the view that opening public sector information to public use will boost innovation and economic growth. In this context, opening up public sector information is often synonymous with making it available free of charge. However, while the logic behind this argument is reasonably clear, the evidence to prove the point is often incomplete. As discussed in the main report, this paucity of evidence arises for two principal reasons.

The first is the complexity of measuring the impact of information, which permeates the economy in so many ways. The direct revenue generated through sale of public sector information may be measurable, where the data is available, but assessing the impact further downstream in the value chain is much more complex and inevitably rests on a range of more or less defensible assumptions. Where public sector information is released free of charge there is no direct revenue to measure. The question becomes still more complex when factoring in possible substitutes for public sector information.

The second reason is the difficulty of quantifying innovation. Part of the rationale for making public sector information publically available is that this will in effect ‘crowd source’ innovations, as thousands or even millions of users, it is hoped, experiment with new ways of using the data. This process is expected to generate innovative new products and services that will benefit individual users and society as a whole. While this has demonstrably taken place, and continues to occur, as ever greater volumes of public sector information are made available to users, the process of innovation is by its very nature unpredictable and, more often than not, there is a significant time lag between data release and crystallisation of benefits. It is therefore difficult to quantify the

¹⁸⁸ *Making open data real: a public consultation*, available at www.gov.uk/government/consultations/making-open-data-real.

benefit, in terms of enhanced innovation, that is likely to arise from making any unit of public sector information available to the public.

The purpose of this literature review is to provide an overview of the key literature on public sector information, including:

- attempts to size the economic contribution of public sector information;
- assessments of the less tangible social, environmental and political value of public sector information;
- studies on the optimal pricing of public sector information; and
- a series of case studies to illustrate these points.

The economic impact

That a link exists between public sector information and economic growth is widely accepted. A recent review of the literature on public sector information (**‘Review of recent studies on PSI re-use and related market developments’**) concluded that “knowledge is a source of competitive advantage in the “information economy”, and for this reason alone it is economically important that public information is widely diffused,” listing benefits from public sector information including:

- development of new products built directly on public sector information;
- development of complementary products such as new software and services;
- reduction of transaction costs in accessing and using information;
- efficiency gains in the public sector itself; and
- the crossing of different public and private information to provide new goods and services¹⁸⁹.

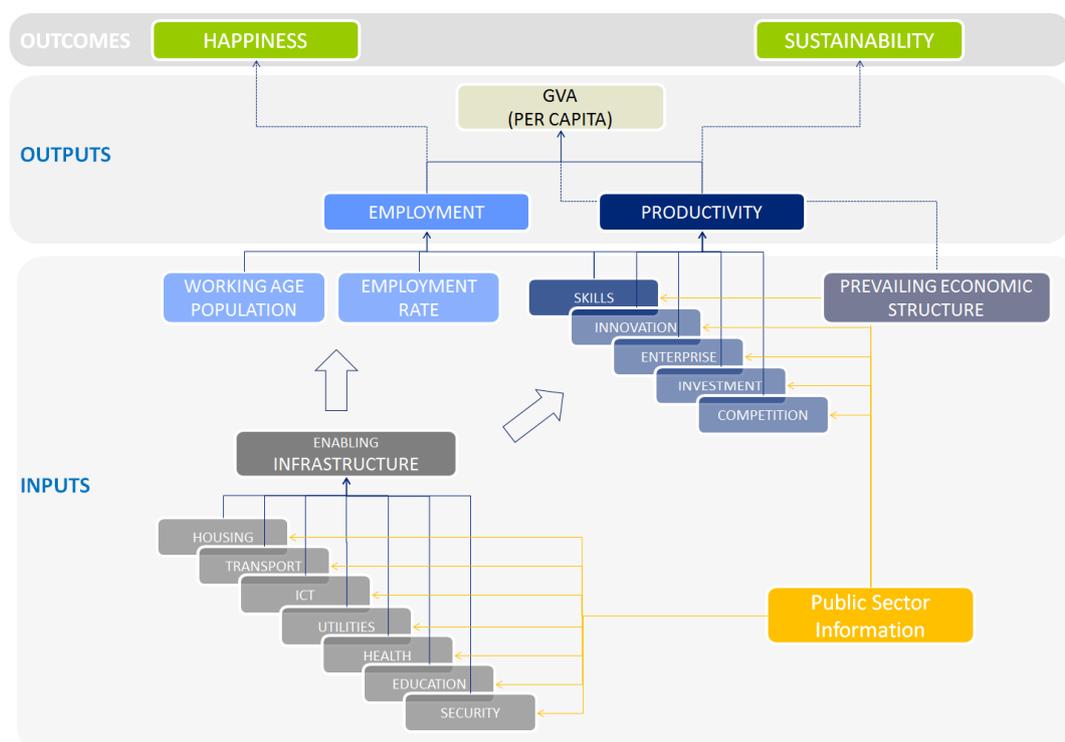
The economic importance of public sector information is seen to have increased radically with the spread of new communication technologies, most notably the internet, and the development of a ‘knowledge economy’ in which value is generated through innovation in information and services. These changes have had the effect of allowing the rapid diffusion of information to a large number of end users, who not only benefit (individually and as communities) from the educational, political and social advantages conferred by this information, but are also empowered to use it to create innovative value-added goods and services. Furthermore, barriers to entry are reduced for start-ups when they enjoy free or cheap access to a wealth of data as well as the tools to easily reach their target market. Although innovation cannot be manufactured by government, the conditions for it to emerge and flourish can be created: this means that “enlarging and systematically inviting serendipity can be argued to be an aim of government information policy making access to public sector information an important cornerstone in a comprehensive digitally driven innovation policy¹⁹⁰.”

Conceptually, one can see how public sector information can drive economic growth and wider prosperity. The simplified framework, shown below, illustrates the long-term drivers of economic growth. Measurable outputs (such as GVA, employment and productivity) as well as less easily measured outcomes (such as happiness and sustainability) are determined by a number of inputs to the economy. In the long-term and considering the supply side of the economy only, the economic output of the UK is a function of only two things: the number of people engaged in gainful employment and the amount each person in employment is capable of producing.

¹⁸⁹ Graham Vickery, “*Review of recent studies on PSI re-use and related market developments*” (2011)

¹⁹⁰ *Ibid.*

Figure A3.1: Long-term UK economic growth framework



Source: Deloitte analysis based on HMT and BIS analysis. See www.hm-treasury.gov.uk/d/ACF1FBD.pdf for more details.

Drawing on the HMT's research into the *Five Drivers of Productivity*¹⁹¹, the amount each worker produces is determined by skill levels; the extent of innovation in products and processes; the degree of investment in capital; entrepreneurial activity; and, lastly, levels of competition.

Underpinning employment and productivity are seven necessary enablers. These are related to the infrastructure required to facilitate long-term economic growth. A deficit in these enablers will equate to a supply-side constraint on economic growth in the long-run. This could be caused either by limiting the growth in working population (through insufficient housing capacity or supporting utilities) or by acting as a drag on productivity growth (through below-par ICT connectivity or a sclerotic transport system). Expanding and enhancing these enablers can therefore positively impact employment and productivity, which in turn generate economic growth and greater prosperity.

As Figure A3.1 shows, improvements in the availability and quality of public sector information can impact across all infrastructure enablers and productivity drivers – the following sections explain how.

The economic contribution of public sector information

Despite this consensus around the importance of public sector information in a modern 'knowledge economy', however, there is little agreement when it comes to quantifying the economic benefit contribution of public sector information. Attempts have varied widely both in the methodologies employed and in the conclusions reached, with the value added by public sector information assessed at figures ranging from millions to hundreds of billions of pounds. This is unsurprising given the difficulties mentioned in the introduction to this chapter: the ubiquitous presence of public

¹⁹¹ See www.bis.gov.uk/analysis/economics/productivity-and-competitiveness for further details.

sector information, making its impact difficult to disentangle from other factors; and the challenges in predicting and quantifying innovation.

Several recent studies have attempted to tackle this challenge. They have examined three levels of economic value generated by public sector information:

- **direct value:** i.e. revenue generated by government from selling access to public sector information;
- **commercial value:** i.e. the revenue generated by private companies through the use of public sector information; and
- **downstream value:** i.e. the value to users of products and the wider economic, social and environment benefits generated.

Assessing direct value is the most straightforward. A 2006 study by the Office of Fair Trading (OFT) (**'The Commercial Use of Public Information'**) estimated the direct revenues from around 400 UK PSI holders at £400m. The report extended this analysis by assessing the combined consumer surplus (the summed difference between the highest prices consumers would pay and the actual price) and the producer surplus (the difference between the price of the product and the cost of supplying it). This led to an assessment of economic value of £590m. The OFT estimated the total potential value of public sector information in the UK to be roughly double this figure, at £1.11 billion if three kinds of market distortions to be removed: unduly high pricing, distortion of downstream competition, and failure to exploit public sector information.

The OFT's approach avoids the danger, common with the top-down approach, of over-estimating value. However, it is likely that the results of its analysis significantly understate the value, both actual and potential, of public sector information. This is partially because some public sector information providers were outside the scope of the report, so that significant volumes of public sector information are excluded from the analysis. More fundamentally, the report did not consider the wider uses and re-uses of public sector information and the potential for enhancing the growth of existing companies and triggering new innovations and entrants into the market. By way of contrast, Vickery's 2011 literature review on reusing public sector information valued public sector information across the entire EU at EUR 140 billion.¹⁹² He derived this estimate by extrapolating figures for the total economic impact of geospatial information in Australia and New Zealand. Based on this calculation, the Government in its Autumn Statement 2011 estimated the current total economic value of public sector information in the UK at £16 billion.¹⁹³ The significant difference between Vickery's estimate and the OFT's figures demonstrates the impact of adopting a top-down approach to assessing value.

Other studies that have attempted to assess the commercial value of public sector information have on occasion reached even higher figures. The most relevant UK studies that have attempted to put a value on public sector information include:

- a 1999 Oxera study (**'The Economic Contribution of Ordnance Survey GB'**) which concluded that £79-£136 billion of Gross Value Added was dependent to some extent on products and services provided by Ordnance Survey. This study employed the value-added method, calculating the total value of all the products and services produced in the UK in which Ordnance Survey's products and services serve as an input.
- a 2007 study by PA Consulting (**'The Public Weather Service's Contribution to the UK Economy'**) which concluded that the public valued the services of the Met Office at £353.2 million, with a minimum additional contribution to the UK economy of £260.5 million based on three case studies.

¹⁹² Vickery (2011)

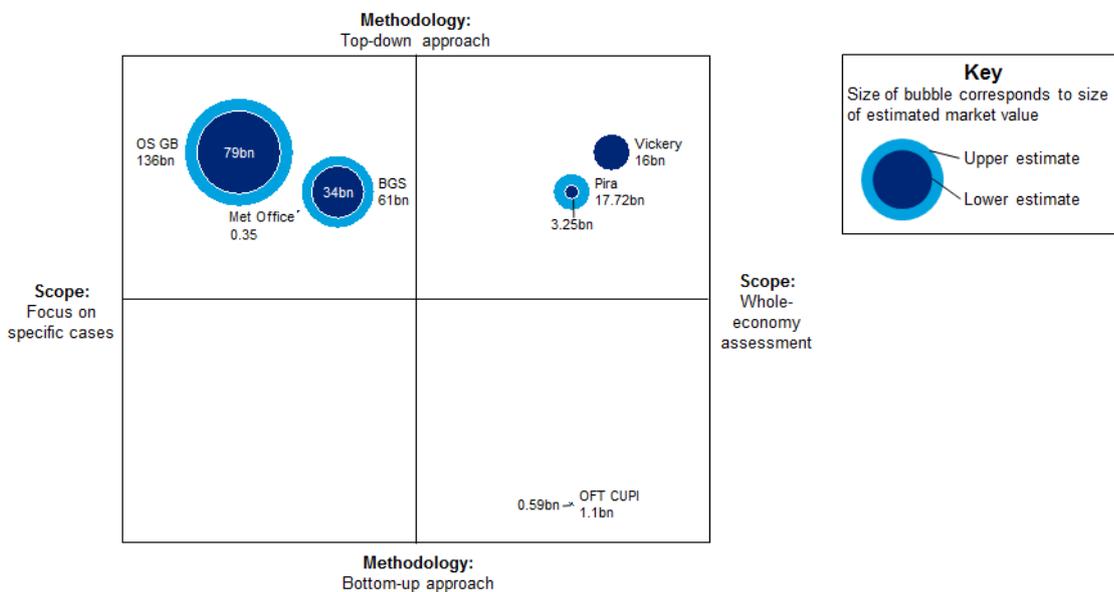
¹⁹³ NAO, *Implementing Transparency* (2011)

- a 2003 study on the British Geological Survey (**‘The Economic Benefits of the BGS’**) which estimated its contribution at between £34bn and £61bn, again using the value-added method.

An additional study of interest was carried out in 2010 by ConsultingWhere and ACIL Tasman for the Local Government Association (**‘The Value of Geospatial Information to Local Public Service Delivery in England and Wales’**). The authors calculated that the benefit to the local government sector from the use of geospatial data was £232 million over five years, with a GDP increase of £323 million for England and Wales (equivalent to 0.02 per cent of GDP for England and Wales) and increased taxation of £44 million. The report also calculated that geospatial data had led to an increase in labour productivity of 0.233 per cent among local public sector providers, equivalent to an additional 1,500 full time staff in England and Wales.¹⁹⁴ Since this relates only to one type of data in one sector, it is to be supposed that the overall effect of public sector information would be calculated to be much higher were the same methodology applied to all public sector information across the whole economy.

Figure A3.2 summarises the results, method and scope of these studies.

Figure A3.2: summary of public sector information market value in literature



Notes:

- This diagram is indicative rather than scientific – it is intended to comparatively illustrate the differing methodological approaches and conclusions of the reports discussed in this chapter
- In general the split is between reports which examine the impact of a specific provider of PSI, for example the Ordnance Survey, and reports which attempt to assess the impact of all PSI across the whole economy. In addition, there is a division between reports which adopt a top-down approach and those (notably the OFT report) which adopt a bottom-up approach
- The figures shown here should therefore not be treated as directly comparable – the OFT report was not attempting to measure the same things as, for example the Pira International report and this accounts for some of the difference in their conclusions

In 2000 a study was undertaken by Pira International (**‘Commercial Exploitation of Europe’s Public Sector Information’**) which estimated the national income attributable to economic activities based on the exploitation of public sector information at EUR 68 billion, or 1.4 per cent of EU GDP. Costs, or the government investment in collecting public sector information, were valued at EUR 9.5 billion, equating to a seven-fold return on investment. The UK’s share of this value was estimated to be EUR 11.2 billion (with an upper estimate of EUR 21.8 billion and a lower estimate

¹⁹⁴ ConsultingWhere and ACIL Tasman, *‘The Value of Geospatial Information to Local Public Service Delivery In England and Wales’* (July 2010)

of EUR 4 billion). These estimates were based not only on the direct revenues from the supply of raw public sector information, but also on value generated by products developed using public sector information as an input. While these Pira values are more moderate, and therefore perhaps more defensible than some of the figures reached by the studies mentioned above, the considerable range of the estimates produced by the report indicates that this is not a precise science.¹⁹⁵

The wide variation in estimated value results from differing methodologies, assumptions and economic models. The OFT report was critical of top-down assessments, arguing that this approach is prone to overstating the value of public sector information since it fails to account for potential substitutes. For example, it is clearly not true that without the information provided by Ordnance Survey there would be no mapping information available to users (a point Oxera acknowledge in their own report). Commercial customers have at their disposal a range of potential suppliers of this information – although it may be true that Ordnance Survey provides information that is more comprehensive and of a higher quality than would be possible for a private sector provider, given the investment required. It is therefore not reasonable to claim that all the value added to the economy by Ordnance Survey information would not be created were Ordnance Survey not to exist. It is also clearly problematic drawing parallels between countries, given variations in the supply and demand for public sector information as well as associated costs, licensing conditions, and other differences. Even drawing parallels between different types of data is fraught with difficulties, as it is clear that some types of data have generated more value than others. Nonetheless, as noted, the OFT's approach is also not without its drawbacks.

The costs of providing public sector information

As the NAO concludes in a recent report on the transparency agenda (**'Implementing Transparency'**), "when estimates of economic value vary this widely, it is difficult to assess the scale of effort or targeting needed to best build on that value."¹⁹⁶ This uncertainty matters because the cost of making public sector information available to the public must be offset against the forecast value created in order to create a robust business case. Unfortunately, just as there is very little certainty regarding the economic value of public sector information, there is also very little data available on the costs of public sector information provision, at least beyond the major trading funds and ONS. The NAO estimates additional staff costs of providing disclosure for pre-existing data range from £53,000 to £500,000 annually by department.¹⁹⁷ However, this does not take into account other costs, including those incurred by IT and other support functions, which are difficult to disaggregate.

Examples of costs associated with public sector information made publically available to date also vary. For example, the police crime map incurred set up costs of £300,000 and annual running costs of over £150,000, largely because the department repackaged the information to improve accessibility.¹⁹⁸ On the other hand, other cases such as the release of public weather service data have incurred much lower costs. This suggests that, as might be expected, improving the quality of data and providing an integrated platform for users is likely to require much greater investment of resources than simply releasing the data in its existing form. However, in value for money terms, this may under some circumstances prove more cost effective than releasing data which the majority of non-technical users are unable to exploit.

Providing value for money from public sector information

¹⁹⁵ It is also worth noting that the report concluded that commercial activities based on PSI have considerable scope for expansion in the EU. AS support for this view, it noted that value generated from the use of PSI in the United States is between two and five times greater than in the EU.

¹⁹⁶ NAO, *Implementing Transparency* (2011)

¹⁹⁷ NAO (2011)

¹⁹⁸ NAO (2011)

The investment into providing public sector information may be judged to provide value for money if improved quality means the data enjoys improved usage and therefore provides greater economic and social benefits. It is notable that releases of data have attracted varying levels of interest. For example, the police crime map website received around 47 million visits between February and December 2011. In contrast, the NAO reports that over 80 per cent of visitors to data.gov.uk leave without clicking on any links, and departments report limited interest in departmental spend data releases.¹⁹⁹ This disparity may be a consequence of the police crime map presenting data in a format that appears much more user-friendly than data.gov.uk for the average user, commensurate with the significant investment made. Similarly, government spending data has been released in a relatively unprocessed format and may therefore be challenging for the majority of users to exploit. In support of this view, the NAO records that “the Department for Education has reported an 84 per cent increase in the use of its comparative data on schools, compared with the same period last year, since it was consolidated in one location and data were made more accessible²⁰⁰.”

It therefore seems apparent that while making public sector information available is an important first step, it is unlikely to enjoy widespread use by the public unless it is presented in a user-friendly format, consolidated in one easy to find location: data availability is a necessary, but not always sufficient condition for use and value creation, It may be that the better prepared the data, the more likely it is to produce significant economic and social value as a wider range users are able to deploy it to improve accountability and develop innovative products. This requires greater investment of resources by public sector bodies responsible for publishing the data, but it is a reasonable hypothesis that, at least for some types of data, the investment will be rewarded by strong economic and social returns. As the NAO argues: “evidence on benefits should be considered alongside information on costs and risks to secure best value from the large stock of public data, match the range and presentation of data purposefully to fulfil specific objectives, ensure that risks are identified and mitigated and secure value for money²⁰¹.” However, in order to make the business case for investment, and ensure that this investment is appropriately targeted towards the most valuable PSI, it would be helpful to have a more comprehensive understanding of both the benefits and the costs.

Summary

There appears to be little dispute that public sector information generally delivers economic and social value in excess of the costs required to provide access to it. However, accurately assessing its economic contribution, both actual and potential, requires an improved understanding of the costs and benefits associated with public sector information, including:

- a better understanding of drivers of additional costs to release different types of public sector information;
- distinguishing between producer surplus and consumer surplus benefits from releasing more public sector information;
- a clearer means of determining demand, so as to prioritise the release of public sector information; and
- a robust method of evaluating the emerging effects of public sector information as it is released, so that efforts can potentially be focussed on high-value data.

It is important to note that all actors could deliver more in terms of transparency of information on the topic. Public sector bodies have, in some instances, not responded to survey questions posed

¹⁹⁹ It should be noted that data.gov.uk has been developed since 2011 and the situation may have developed since the NAO published their report

²⁰⁰ NAO (2011)

²⁰¹ NAO (2011)

as part of the study, and naturally, private sector organisations are unwilling to discuss and disclose sensitive financial information and the precise means by which PSI does or could improve performance.

Figure A3.3 below summarises the key studies that have attempted to value the contribution of public sector information. It should be noted that although summarised here for convenience, these reports were not all attempting to measure the same thing and so the figures are not directly comparable.

Figure A3.3: summary of UK public sector information studies

Study	Date	Author	PSI sources evaluated	Upper value estimated	Lower value estimated
OFT CUPI	2006	OFT	Various	£1.1bn (potential)	£590m
Vickery	2011	Graham Vickery	Various	c. £16bn (UK); EUR 140bn (EU)	-
MEPSIR	2006	HELM Group	Various	EUR 48bn (EU + Norway)	EUR 10bn (EU + Norway)
PIRA	2000	Pira International	Various	EUR 21.8bn	EUR 4bn
The Economic Contribution of Ordnance Survey GB	1999	Oxera	Ordnance Survey	£136bn*	£79bn*
The Public Weather Service's Contribution to the UK Economy	2007	PA Consulting	Met Office	£353m	-
The Economic Benefits of the BGS	2003	Roger Tym & Partners	British Geological Survey	£61bn*	£34bn*

Source: Deloitte analysis

* These figures are based on an analysis of the Gross Value Added dependent to some extent on PSI, and therefore differ from attempts to assess the value of PSI

The broader impact

Many of the studies discussed above attempt to quantify the actual or potential economic contribution of public sector information. In reality this is difficult to separate from other impacts, such as the social, environmental or political benefits conferred by access to public sector information. For example, if an individual's or a community's health outcomes are improved through use of public sector information, this will have a downstream economic benefit in terms of increased individual productive potential (through avoidance of loss of working hours from illness) and medical costs avoided. Due to the complexity of calculating these less tangible benefits, however, they are often omitted from quantitative studies of economic contribution.

However, it is also true that these benefits have an intrinsic value – good health is desirable irrespective of its positive economic impacts, and would remain so even if quantitative economic impacts could not be demonstrated. For this reason it is important to take such benefits into account when weighing up the costs and benefits of public sector information. Where it is not possible to calculate a financial value, case studies and other measures can be used.

One of the most important studies to address this area is the 2007 independent review by Ed Mayo and Tom Steinberg (**'The Power of Information'**), which cited a range of benefits that have resulted from sharing information²⁰²:

Figure A3.4: broad benefits of public sector information

Information use	Benefits
Online communities	In medical studies of breast cancer and HIV patients, participants in online communities understand their condition better and show a greater ability to cope. In the case of HIV, there are also lower treatment costs. This creates both a welfare benefit and a cost saving benefit.
'Wired' local communities	Studies of 'wired' local communities demonstrate that there are more neighbours who know the names of other people on their street. This is likely to create a safer, more cohesive and supportive community.
Restaurant food safety information	Sharing restaurants' food safety information in Los Angeles led to a drop in food-borne illness of 13.3 per cent, compared to a 3.2 per cent increase in the wider state in the same time frame. The proportion of restaurants receiving 'good' scores more than doubled, with sales rising by 5.7 per cent. This creates a welfare benefit, reduced medical costs and increases consumer spending.
Medical prescription information	By providing clear information when dispensing medication, pharmacists can improve patient adherence/persistence with medication advice by 16–33 per cent.

Source: Mayo and Steinberg, Deloitte analysis

Not all of these examples concern public sector information but they do indicate the range of benefits which can be delivered by widespread availability of information. They therefore hint at the potential benefits of making public sector information available to a wider range of users, given the volume and range of information held by public sector bodies. It is important to ensure that these less tangible and less easily quantifiable benefits are not excluded from any analysis of the costs and benefits of public sector information.

In a more recent study, Pollock (2011) estimated that the indirect benefits from greater availability of public sector information, such as reductions to transaction costs to users and re-users and other efficiency gains could imply gains of around £600m per year for the UK.

Maximising the impact of public sector information

Pricing models

An important aspect of the debate over the economic contribution and value of public sector information is the question of the correct level of pricing for data. Currently, while much public sector information is made available at no charge for use and re-use, some of the most valuable data – including, for example, some mapping data collected by Ordnance Survey, and data held by Companies House – is only available to paying users. There is some debate over whether these charging models are justified, given the costs involved in collecting the data, or whether this introduces counterproductive inefficiencies and market distortions.

²⁰² Mayo and Steinberg, *'The Power of Information'* (2007)

In one of the key papers on this topic (**‘Models of Public Sector Information Provision via Trading Funds’**), Newbery et al. identified four possible charging policies:

- Profit-maximization: setting a price that maximises profit given the demand for the data;
- Average cost (cost recovery): setting a price equal to average long-run costs (including fixed costs incurred by data production);
- Marginal cost: setting a price equal to the actual cost of the process of supplying data to a user; and
- Zero cost: releasing the data free of charge.²⁰³

The paper concluded that many Trading Funds’ products – primarily ‘refined’ data products – could not be analysed effectively due to data limitations, and that these should therefore by default be left with their pricing policies unchanged. However, Newbery et al argued that most ‘unrefined’ data should be made available at marginal cost, which owing to the low cost of providing digital data would effectively be zero. This would bring trading fund data into line with raw data charging policies in other areas of government. The basis for this argument is that the benefits to society of making this data freely available would outweigh the costs, although some trading funds, notably Ordnance Survey, would need additional taxpayer support.

The costs and benefits estimated by the paper are summarised below:

Figure A3.5: costs and benefits of Trading Funds and other PSIHs

Trading Fund	Gross benefit from releasing raw PSI	Cost to government	Net benefit
Companies House	£2.6m	£681k	£1.9m
The Met Office	£1.2m	£260k	£1.03m
Ordnance Survey	£168m	£12m	£156m
UK Hydrographic Office	£1.08m	£744k	£338k
The Land Registry	£2.3m	£1.1m	£1.2m
The DVLA	£4.3m	£582k	£3.7m

Source: Newbery et al, ‘Models of Public Sector Information Provision via Trading Funds’ (2008)

Clearly, this study raises important questions regarding the fairest way of financing trading fund data collection activities: should taxpayers in effect subsidise free access to data for users, or should users of data directly bear the costs associated with data collection and dissemination? Newbery et al’s figures indicate that free access to trading fund data, subsidised by the taxpayer, would in most cases lead to a net economic benefit. There is no guarantee, however, that this benefit will be equally spread across society. It is possible, for example, that most of the benefit might accrue to foreign-based companies which can use the data to boost revenues but not contribute to UK tax revenues. If it could be demonstrated that making the data freely available would be ‘revenue neutral’ – i.e. the associated tax income to government would at least equal the outlay in support of trading funds – there would be a much more economically and politically powerful case for taking this step.

In a 2009 paper (**‘Enhancing access to government information: economic theory as it applies to Statistics Canada’**) Kirsti Nilsen argued that in the case of public sector information the benefits are non-rivalrous and non-excludable, and so the principle that the beneficiary pays is a fallacy.

“The justification for cost recovery is often based on the so-called benefit principle: Those who benefit from a good should pay for it. However, it is very difficult to

²⁰³ Newbery et al, ‘Models of Public Sector Information Provision via Trading Funds’ (2008)

*determine the benefits of information. Information flows. It moves away from the initial buyer. So, what is the benefit? Who benefits? How do you apply the benefit principle? The assignment of benefit, like the assignment of costs, is an arbitrary exercise.*²⁰⁴

Nilsen argued that should Statistics Canada move from a cost recovery pricing model to a zero cost model, the following results could be expected:

- sales and licensing revenues for the public sector body would **decrease**;
- usage and reuse of the public sector information would **increase**;
- increased public sector information usage would provide **positive externalities**: information dissemination, more widespread usage of the data, and economic growth;
- tax revenues would therefore **increase**; and
- the public sector body's transaction and opportunity costs would **decrease**. This should offset the decline in sales and licensing revenues, in addition to the **increase** in tax revenues.

In considering these arguments, however, it is important to reflect on the diversity of the public sector information landscape and the dangers of applying a 'one size fits all' prescription. For example, the Trading Funds operate within a regulatory framework which limits their ability to cross-subsidise information provision from income generated in other areas of activity. From conversations with stakeholders it is clear that the picture is more complex than an initial examination might suggest, as are the implications of radical changes to the current model. Examples where this approach might be less appropriate are considered elsewhere in this report.

The impact of zero or marginal cost pricing on innovation

There is already a trend towards making at least some data freely available. This is clear across the full spectrum of PSIHs, from government departments through to trading funds such as Ordnance Survey, which now offers significant numbers of datasets through its Open Data initiative. A 2011 European Commission study (**'Pricing of Public Sector Information'**) found "a clear trend towards lowering charges and/or facilitating re-use"²⁰⁵ among public sector bodies. The study found that where PSBs moved to marginal or zero cost charging, the number of re-users increased by between 1,000 per cent and 10,000 per cent, indicating the potentially significant impact of removing cost barriers to re-use. Removing cost barriers was also found to attract new types of users, particularly SMEs, which may be a particularly innovative class of user and thus add additional value to the economy. The study further found that costs did not increase significantly, or even decreased, once prices were lowered. This is partly due to the fact that zero cost pricing greatly reduces transaction costs, as public sector information providers no longer need to devote resources to complex payment, licensing, supply and enforcement systems. In addition, enhanced volumes of users compensate for reduced prices.

The study also found that the availability of low cost data with clear re-use rules spurred the development of innovative public sector information-based apps, opening up a potentially significant new market. It cites examples such as MetroParis and London Tube apps which have jointly generated EUR 400,000 in revenues. However, when public sector information providers create value-added services in the form of their own apps this has been noted to have a detrimental effect on private sector app innovation, suggesting that they should arguably confine themselves to providing data for users and avoid competing directly through the development of value-added products. Some stakeholders, however, have made the point that private sector

²⁰⁴ Nilsen, *Enhancing access to government information: economic theory as it applies to Statistics Canada* (2009)

²⁰⁵ Pricing of Public Sector Information (2011)

providers did not develop apps and other value-added products taking advantage of the available data, prompting them to release their own to fill a gap in the market.

A significant body of work therefore supports the argument that the gross benefits from not charging, or charging at marginal cost, outweigh the added cost that may be borne by government. This has contributed to growing momentum towards reducing the cost of public sector information at the point of use. Despite the strength of Nilsen's arguments from the standpoint of economic theory, however, there may be political challenges in supporting data collection which is then freely provided to for-profit businesses, even if the wider benefits to society can be demonstrated to outweigh the direct cost to taxpayers. The challenge for advocates of open data, therefore, is to demonstrate that changes to pricing policy can be revenue neutral or lead to an increase in public sector revenues, through a growing tax base.

There is evidence indicating that changes to pricing policy can be revenue neutral. A 2011 Finnish study ('**Does marginal cost pricing of public sector information spur firm growth?**') surveyed the performance of 14,000 firms in the architecture and engineering sector (Standard Industrial Classification 7420) and pricing policies for public sector geographical information across 15 countries. It concluded that in those countries where geographical information was either free or priced at marginal costs, firms grew 15 per cent faster than in countries where information was priced at cost recovery. The study also found that an impact on company growth rates could be detected within a year of switching to a marginal cost pricing scheme, although the impact became more pronounced after two years.²⁰⁶

Importantly, the most significant impact on growth rates was experienced by SMEs rather than large firms. The author attributes this to high public sector information prices creating a barrier to entry for SMEs, which is removed by a switch to marginal cost pricing. He argues that for this reason, marginal cost pricing of public sector information is likely to create more competitive markets and thus lead to lower prices and a wider range of products, benefitting consumers. If this result could be demonstrated to be replicated across other types of public sector information, it would present a persuasive argument – from the point of view of encouraging dynamic markets and economic growth – for moving to a marginal cost pricing mechanism for public sector information. Nonetheless, such a change would need to be evaluated on a case by case basis, given the various types of information involved and the various markets in which public sector bodies which currently charge for information operate.

The table below summarises the key studies that have examined the pricing of public sector information.

Figure A3.6: pricing of public sector information studies summary

Study	Date	Author	Area examined	Conclusions
Models of Public Sector Information Provision via Trading Funds	2008	Newbery et al	UK Trading Funds	Significant net benefit from moving to marginal or zero cost pricing structure for raw data
Enhancing access to government information: economic theory as it applies to Statistics Canada	2009	Kirsti Nilsen	Statistics Canada	Zero cost pricing produces positive externalities across society while being revenue neutral due to increased usage and reduced transaction costs
Pricing of Public Sector Information	2011	Deloitte and others in association with	Charging models for PSI	Moving to marginal or zero cost pricing increases users by 1000-10,000 per cent, attracts

²⁰⁶ Heli Koski, *Does marginal cost pricing of public sector information spur firm growth?* (2011)

Study	Date	Author	Area examined	Conclusions
		the European Commission		new types of users including SMEs, and does not result in significantly higher costs
Does marginal cost pricing of public sector information spur firm growth?	2011	Heli Koski	Performance of 14,000 firms across 15 countries	Marginal cost pricing of geographical data leads to an average 15 per cent increase in the growth rate of SMEs

Case studies

The economic and social value potentially generated by public sector information is well illustrated by the case of **'wicked' problems**. These are challenges featuring complex interdependencies, which spill across jurisdictions and therefore have the potential to confound traditional, vertically organised structures such as government departments, whose expertise and mandate covers only one area of policy. Examples of wicked problems include climate change, youth unemployment and health issues such as obesity. By making public sector information available for re-use data from diverse sources can be linked and 'mashed', potentially producing new insights and generating unexpected solutions.

The value that can be generated by public sector information is also well illustrated through case studies on the impact on the public sector, business and wider society. These have been drawn from the UK and other countries that have been experimenting with different models for providing public sector information. While to some extent this is an anecdotal approach – it is not possible to forecast the number of new businesses or their value – it offers useful insights into the value that can be generated through public sector information.

Public sector efficiencies

Case study	Description
Spotlightspend <i>Helping public sector bodies publish data in a cost-effective manner</i>	<p>Spotlightspend is a managed service that is comprised of everything necessary to facilitate cost-effective publication of the spend and related information that is made available to the public.</p> <p>Spotlightspend is designed to:</p> <ul style="list-style-type: none"> • Cut costs by removing the need to add to the workload of current staff or increase headcount • Eliminate the complexity of becoming and staying compliant with policy • Reduce the risk of inadvertent breach of data protection legislation • Enhance the information published to improve its accessibility, relevance and value for the public
Barnet StreetPatrol <i>Delivering efficiency savings in local policing</i>	<p>Barnet, one of the largest boroughs in London, deployed a GPS system called StreetPatrol to help street wardens locate, identify and photograph issues including abandoned vehicles, graffiti, antisocial behaviour and fly-tipping. This enables them to send information immediately back to head office, ensuring a rapid and efficient response. Previously this process could take between three and four days.</p> <p>Using StreetPatrol, wardens are able to spend up to 70 per cent of their time on patrol, compared to 30 per cent for those without the system.</p> <p>Adoption of this system has delivered around £180,000 in efficiency savings. Improved speed of response also brings cost savings: an abandoned vehicle costs approximately £50 to recover; a burning vehicle nearly £4,000. By responding quickly wardens are able to tackle problems before they develop.</p>

Source: Deloitte: Open Data – driving growth, ingenuity and innovation; and Cabinet Office material

Business innovation

Case study	Description
<p>Red Spotted Hanky <i>Using rail industry data to offer customers low-price tickets</i></p>	<p>Launched in 2010, Red Spotted Hanky is an online ticket retailer which aims to offer customers an easier way to book without any administration or payment fees.</p> <p>Red Spotted Hanky relies on data from the rail industry to offer customers low-cost advance bookings. The business employs 13 people and is growing fast, with a loyalty scheme, a tie-up with Tesco and a 'price promise' for customers.</p>
<p>Duedil <i>Linking and aggregating information to improve business transparency</i></p>	<p>Duedil is a business information provider based in London's Soho. Duedil gives free access to governance and financial information for every company in the UK and Ireland, and combines this with data from online sources, Application Programming Interfaces, social networks and more. Duedil was launched in April 2011 by entrepreneur Damian Kimmelman. By late 2011 it was valued at £20 million and had attracted considerable interest from investors.</p> <p>Its aim is to make business more transparent, by opening up company information to make the due diligence and research process simple and intuitive. By aggregating and linking all the available information, users can gain a comprehensive understanding of businesses and the people who run them.</p>
<p>Parkopedia <i>Using local authority data to give drivers live parking updates</i></p>	<p>Parkopedia is an innovative open data company which fuses location and other data. A small UK-based business, it uses live data from local authorities to help drivers identify free car parking spaces. Parkopedia has grown to become the world's leading source of parking information covering more than 20 million spaces in 25 countries.</p> <p>Used by millions of drivers, Parkopedia's service includes a pre-booking tool which allows drivers to book parking online, and real-time parking space availability information. Parkopedia also works with other organisations to integrate its data into journey planner mobile applications and satnavs.</p>

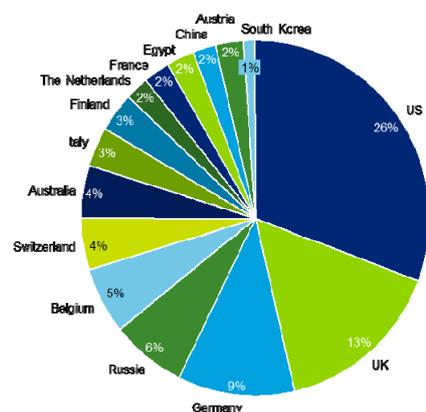
Source: Deloitte: *Open Data – driving growth, ingenuity and innovation*

Case study – the value of PSI in the mobile app market

One area where public sector information has generated significant innovation, some of which has been successfully monetised, is in the mobile app market. Deloitte's POPSIS study (2011) estimated the value of the mobile app market at around \$35 billion, of which 40% is contributed by apps that use public sector information. Based on a UK contribution of 13% of all apps, this suggests that the value of the PSI-based app market in the UK is some £1.13bn.

Furthermore, this is a rapidly growing market, with use of smartphones and tablets in the UK increasing rapidly year on year. The figure of £1.13bn is likely to understate the true value of public sector information-based apps to the UK economy, as it does not take account of the benefit to users (such as time savings and increased ability to access important information).

Country distribution of mobile apps in the sample



Public sector information-based app share

	Weather	Finance	Travel	All (average)
App #	130%	12%	29%	40%
Android	130%	10%	12%	
Or	130%	6%	2%	
Total (average)	100%	8%	12%	

Based on 'Apps' market snapshot of POPSIS study. The study considers a sample of ~500 apps from Android, Istore, and Ovi apps

Source: Deloitte, available http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/11_2012/apps_market.pdf

Country case studies

While international case studies are a helpful illustration of public sector information practices elsewhere, care must be taken when extrapolating the lessons. For example, the same proportion of benefits may not occur in the UK due to differences in economic structures, legislative frameworks and the culture and capabilities of consumers.

Denmark's Open Data Innovation Strategy²⁰⁷

In 2009 the Danish government launched the Open Data Innovation Strategy (ODIS) to provide easier access to PSI for businesses and other users. Although this is a recent development that is likely to have much as yet unrealised potential, efficiency gains in both the public and private sectors are already emerging. Certain sectors have been particularly quick to spot the opportunities presented by the availability of public sector information.

- **Financial services:** banks are working with the tax authorities to access payroll and pension data for clients of the banks. This has the estimated potential to save banks EUR 67 million per year in efficiency gains and reduced losses. Additional savings could be gained by accessing data on clients' employment conditions. The insurance industry is also interested in the potential for customer data to improve its risk assessments. In these cases, however, there remain issues to be resolved around customer consent for the sharing of these personal details.
- **The energy sector:** data on building specifications, as well as demographic data on residents, could be used to target energy-saving measures. Potential savings are estimated at EUR 0.54-2.7 billion.
- **The pharmaceutical and healthcare sector:** patient data could be used to select patients for clinical trials, improving the process of developing new drugs.

Denmark is in some ways an unusual case owing to its relatively small size and a legacy of robust public sector data collection and digitisation, which has left it well placed to improve economic and social outcomes through the release of public sector information. It is also clearly at an early stage in this process, with significant issues in data security and confidentiality yet to be resolved. Nonetheless it provides a striking example of the potential benefits to be realised by making public sector information available to business users, both in terms on financial savings and wider benefits to society.

Denmark: the value of address data²⁰⁸

In 2002, the official Danish address data was made available free of charge. This meant that any user could access the data without paying a fee to the Danish Enterprise and Construction Authority (DECA) charged with collecting and maintaining this data. Eight years later DECA commissioned a study to analyse the benefits of making the data free of charge.

The study concluded that between 2005 and 2009 the total direct financial benefits of the data were EUR 62 million, with costs of around EUR 2 million. The study also estimated that in 2010 total benefits would be EUR 14 million, with costs of EUR 0.2 million. The benefits were split at around 30 per cent in the public sector and 70 per cent in the private sector.

²⁰⁷ Vickery (2011)

²⁰⁸ 'The value of Danish address data' (2010)

As with many other studies, only direct financial benefits were measured, as downstream economic and wider social benefits were considered too difficult to quantify. However, the authors suggested that these benefits were likely to be of considerable value. They cite the example that 46 per cent of Danish families own a GPS navigation system, incorporating the address datasets, which has doubtless generated significant benefits – for example in travel time saved.

The study also highlighted barriers which have limited the benefits of the address data, caused by technical, traditional and legislative barriers which, for example, limit how quickly business addresses are registered and included in the database. This indicates the need for efficient data collection, processing and dissemination systems and the regulatory framework to support this.

In conversations with stakeholders, the caveat has been raised that some of the benefits identified in this study may not be transferable to the UK. In particular, the Public Sector Mapping Agreement means that sharing of geospatial data between public sector bodies in England and Wales²⁰⁹ is already relatively advanced compared to many countries. Utilities companies also have access to this information. Furthermore, the UK has an unusually dense post code network, which means that these are suitable for accurate navigation via GPS navigation systems. These caveats highlight the risks inherent in assuming that benefits experienced in one country will be directly transferable to another.

The study nonetheless highlights the important direct financial benefits that can result from making these key datasets freely available to users, and suggests that these direct financial benefits both in the public and private sector can considerably outweigh the cost to government.

Spain: The Aporta Project

In 2011 an analysis of the ‘infomediary’ business sector was carried out, defined as “the set of companies that create applications, products and/or added-value services for third parties, using public sector information.”²¹⁰ This was intended to be a comprehensive survey of all Spanish PSI activities. Headline findings were:

- business turnover directly associated with infomediary activities is EUR 550-650 million, 35-40 per cent of the total company activity of EUR 1.6 billion;
- activity by re-use field was as follows:
 - business/financial 37.6 per cent
 - geographic/cartographic 30.5 per cent
 - legal 17 per cent
 - transport 5.2 per cent
 - social data/statistics 1.9 per cent
 - meteorological 1.1 per cent
 - others 6.7per cent
- the re-used information comes mostly from national agencies, but half of the companies also re-use international information;
- re-use policies are valued, particularly to improve the quality and accuracy of information, improve understanding of the legal framework, and expand the amount and scope of information generated; and
- areas identified for improvement include standardization of formats, standardization and improvement in the regulation of licences for re-use, and pricing of information.

²⁰⁹ The One Scotland Mapping Agreement covers public sector bodies in Scotland.

²¹⁰ Vickery (2011)

This survey indicates the scale of the opportunity for business stemming from the re-use of public sector information, with the financial services and geospatial sectors seeing particularly significant opportunities. However, it also highlights the need to improve quality and accessibility, through standardised formats and improved licensing arrangements.

Australia: Office of Spatial Data Management and Geoscience Australia²¹¹

In 2001 the Australian Government launched the Commonwealth Policy on Spatial Data Access and Pricing, which provided free access to spatial data online, or charge no more than the cost of transfer for packaged data. This applied to all spatial data collected by government departments and agencies. Furthermore, in 2009 Geoscience Australia began licensing all data under the Creative Commons licence, allowing royalty-free use and re-use.

A 2010 report by PwC (**'Economic Assessment of Spatial Data Pricing and Access'**) valued the net welfare gain at \$4.7 million. Extrapolating this for all Australian Government spatial data, Houghton calculated net welfare benefits totally \$70 million.

Usage increased rapidly, with the total number of spatial datasets delivered by Australian government agencies increasing at 112 per cent per annum between 2001-02 and 2005-06, from around 75,000 to over 1.5 million. Even though the number of datasets rose, intensity of dataset use increased over the period, with the number of downloads per dataset rising 44 per cent per annum.

This case reinforces the view that making data available at zero or marginal cost generates welfare benefits in excess of the revenue lost, with Houghton calculating a benefit/cost ratio of 13.

New Zealand: spatial information in the economy – realising productivity gains²¹²

An extremely comprehensive 2009 study assessed the value of spatial data to the New Zealand economy. It concluded that the use and re-use of spatial information added an estimated \$1.2 billion to the New Zealand economy in 2008, equivalent to 0.6 per cent of GDP. In addition, it estimated the non-productivity benefits to be worth many times this figure, through, for example, planning 'smarter' cities and transport systems, aiding national security, or promoting social cohesion. However, the authors avoided calculating a financial assessment of these benefits because: there is too much uncertainty around likely events and probabilities and the underlying value approaches are controversial.

The figures presented by the report should therefore be treated as a 'lower estimate' which in reality could be substantially exceeded.

The report further estimated that barriers to the use and re-use of data cost \$481 million of productivity-related benefits, which would have generated at least \$100 million in government revenue. It recommended releasing the basic spatial information held by the government at marginal cost. A broader intervention to develop New Zealand's spatial data infrastructure would generate a predicted benefit-to-cost ratio of at least 5:1.

In conducting the productivity gains analysis, the report treated historical growth rates of the New Zealand economy as the base case. The authors then estimated the historical productivity of each sector without the productivity benefits of spatial information; and the potential productivity if adoption barriers for spatial information were removed (i.e. the potential unrealised benefit). This provides an interesting insight into the industry sectors most likely to benefit from the use and re-use of geospatial PSI.

²¹¹ John Houghton, *Costs and Benefits of Data Provision* (2011)

²¹² *Spatial Information in the New Zealand Economy: Realising Productivity Gains* (2009)

Figure A3.7: impact of adopting PSI

Sector	Quantifiable historical productivity without spatial information	Estimated productivity without adoption barriers
Crops	1.25	1.88
Bovine cattle, sheep and goats, horses	1.25	1.88
Other animals	1.25	1.88
Raw milk	1.25	1.88
Wool	1.25	1.88
Forestry	5.25	5.71
Fishing	3.44	3.44
Meat products	0.25	0.38
Dairy products	0.25	0.38
Other processed food	0.25	0.38
Coal	0	0
Oil	0	0
Gas	0.63	0.78
Electricity	0.63	0.78
Petroleum & coal products	0	0
Iron & steel	0	0
Other mining	0	0
Nonferrous metals	0	0
Non-metallic minerals	0	0
Chemicals, rubber, plastics	0	0
Wood and paper products; publishing and printing	0.25	0.38
Textiles and clothing	0.25	0.38
Other manufacturing	0.25	0.38
Water	0.63	0.78
Construction	0.75	1.13
Trade services	0.77	1.15
Transport	2.10	3.15
Communications services	0.82	0.82
Other business services	0.23	0.46
Recreational and other services	0.23	0.46
Government services	0.52	1.04
Dwellings	0	0

Source: ACIL Tasman calculations and estimates

Appendix 4: Further statistics

Companies House

Companies House is a statutory body largely funded by fees paid by customers to register a company. These fees provide the majority of revenue, with users of the data paying for dissemination costs only, which means the data is free of charge in some cases. Companies House serves as an ‘information exchange’ – unlike some of the trading funds, it does not provide value-add services to users, but rather exists to gather data in the form of company registrations and disseminate this to users.

Figure A4.1 below contains some download statistics provided by Companies House on usage statistics across its various distribution channels.

Figure A4.1: Companies House usage statistics

Companies House Direct

This is the Companies House online subscription service, for which users pay monthly. It offers free basic company information as well as company documents and reports for a charge. Usage statistics since 2010/11 are shown below. Companies House report that there were over 405 million free searches between January and December 2012, indicating high uptake of the free data available.

The number of subscribers to this service rose from 20,967 in 2010/11 to 21,653 in 2012/13 (YTD).

Webcheck

This is the Companies House ‘pay as you go’ service which allows users to check company name availability, search free basic company information and purchase company documents and reports.

The number of page hits grew from 380 million in 2010/11 to nearly 450 million in 2011/12, with strong growth posted again for the 2012/13 YTD.

Mobile App

Companies House have released a mobile app which gives free access to information including company address, status, company appointments and filing history. The number of downloads reported by Companies House (as of 10/12/12) is 12,362. 11,075 of these were for the iPhone version, and 1,287 for Android. The iPhone app was released in September 2012 while the Android version was released in November.

URI company searches

Companies House runs a Uniform Resource Identifier service (URI) for all companies on the register. This allows free searches which will return basic details for the company. Companies House report that since the introduction of this service in October 2011 it has seen over 111 million data requests.

DVD ROM

Companies House offer a product stored on a DVD ROM, for a fee, containing PDF images of company documents filed since 31 December 2002. There has been a decline in both one-off requests for the DVDs since 2010/11 reflecting the increased availability of information via the website. One-off requests fell from 931 to 898 between 2010/11 and 2011/12 and subscriptions fell from 664 to 637 over the same period.

Source: Companies House

Land Registry

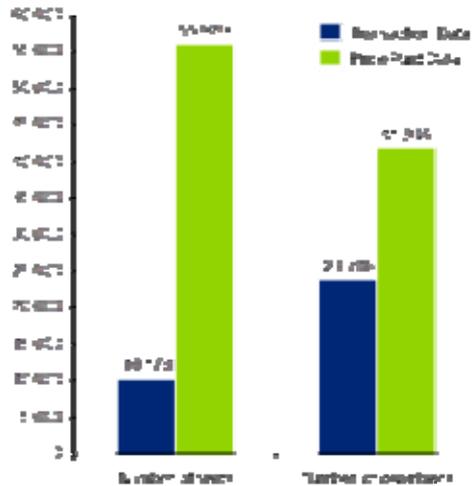
Land Registry’s main purpose is to register ownership of land in England and Wales and to record dealings with land once it is registered. The relevant powers and responsibilities are set out in the Land Registration Act 2002. There are also statutory responsibilities under the Agricultural Credits Act 1928 and the Land Charges Act 1972.

Under the Land Registration Act 2002, Land Registry has the power to provide ‘consultancy and advisory services’. The income from these services is used to invest in Land Registry and helps to ensure that fees for normal registration services are kept to a minimum.

Figure A4.2 below shows usage statistics for the data held by Land Registry.

Figure A4.2: Land Registry usage statistics

Land Registry released their Transactional Data in January 2012 and their Price Paid information in March 2012 in open data format. The chart below shows the number of visits and downloads received between their release and the end of October 2012.



A Land Registry survey of Price Paid Data users between March and November 2012 indicated that the data was being used to:

- Help businesses, for example to provide housing market updates
- Assist research, for example into local housing needs and affordable housing policy
- Help people buy and sell houses
- Help people monitor properties
- Present the data in new ways and carry out new analysis

Source: Land Registry

The Met Office

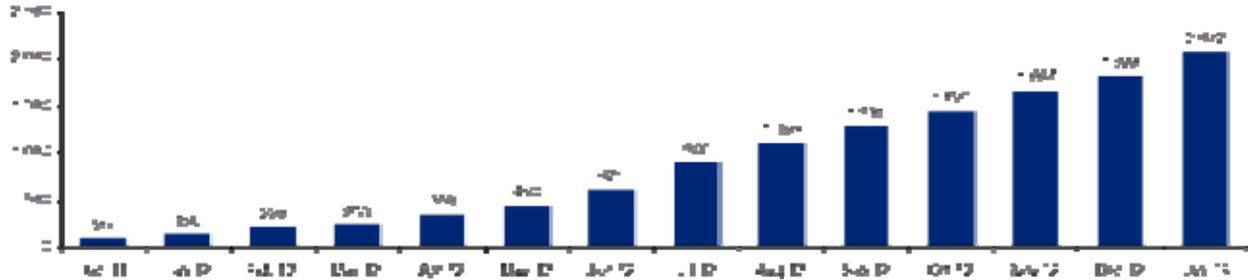
The Met Office uses its data to produce a range of value-added services, some of which are non-competed services to governments, and others which it sells into competed markets, either to government or the private sector. Examples of the former include the services provided to defence and its climate modelling work. Examples of the latter include services provided to aviation clients and insurance companies. Most paid-for services are therefore value-added, as even wholesale data requires a high level of processing to enable re-use.

Figure A4.3 contains statistics received from the Met Office on the number of requests and downloads received by its Data Point service.

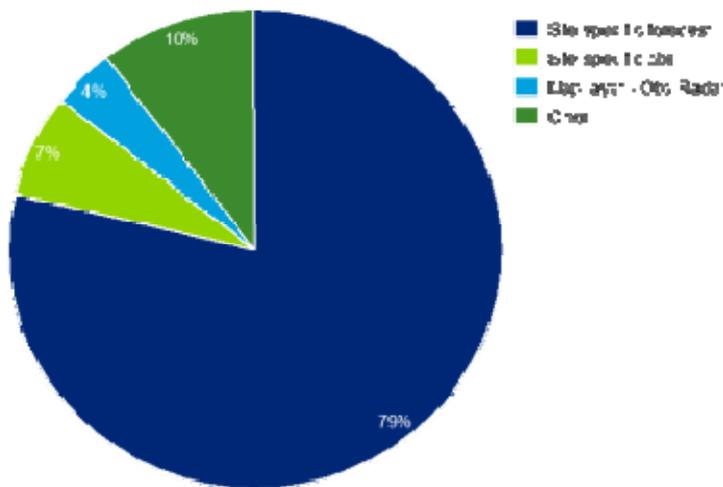
Data Point

Portal for accessing Met Office reusable data for Web and App developers.

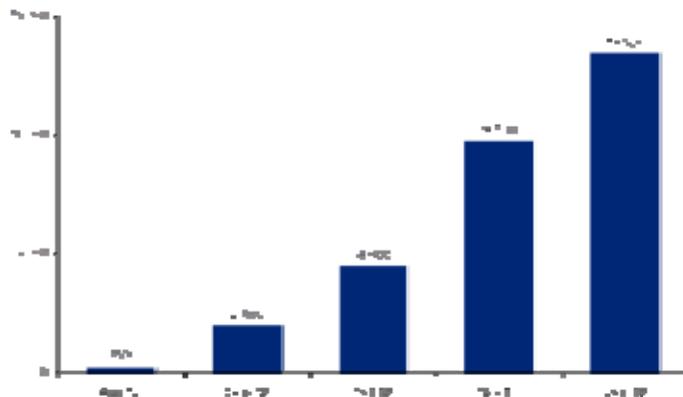
The number of registered users has also increased significantly between December 2011 and January 2013, from around 100 to over 2000



The Met Office has supplied Deloitte with the number of requests received by data feed in December 2012, which may be taken as a recent indication of the level of demand – around 2.4 million. The proportions of each request by data type are shown in the chart below.



As is clear from these figures, the Met Office experiences very high and continual demand for across its data feeds, with site specific forecasts accounting for over three quarters of the demand, at over 1.9 million requests. The average daily download of data from the Met Office site, in terms of Mb downloaded, has increased rapidly over the last six months, as the chart below shows.



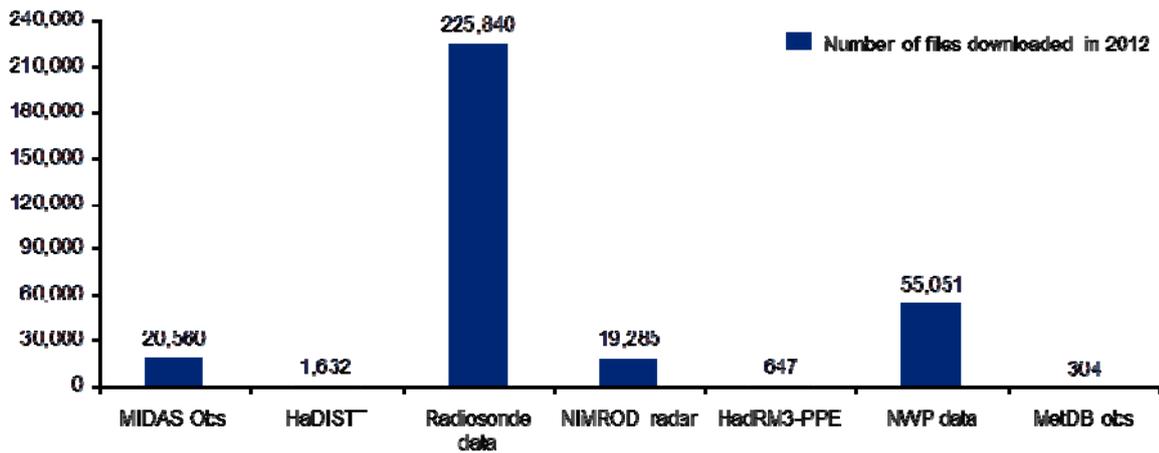
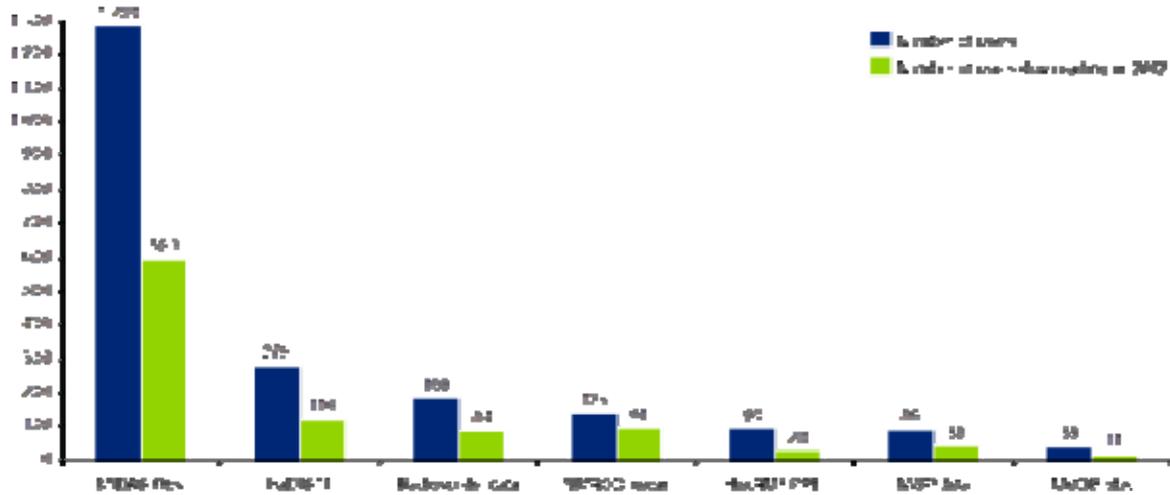
The British Atmospheric Data Centre

The British Atmospheric Data Centre (BADC) is NERC's dedicated data centre for the atmospheric sciences. As such it has a crucial role in providing long-term curation and access to datasets that support scientific research. Whilst its main role is to archive the outputs of NERC's own research activities it also accesses datasets from many other sources in response to the needs of the research community.

- MOHC climate simulations contributing to the CMIP5 archive: 58Tb, multiple models and experiments

- MIDAS Observations: 258Gb, all years, all observations

The number of users and the number of files downloaded in 2012 is shown in the charts below.

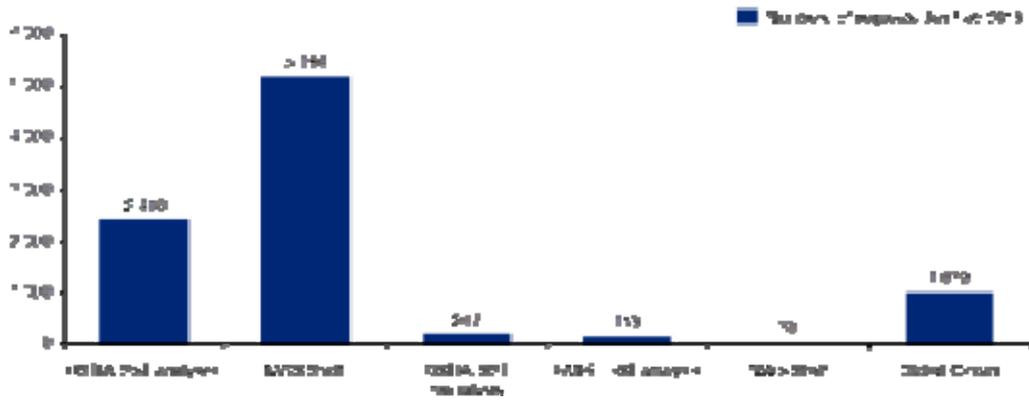
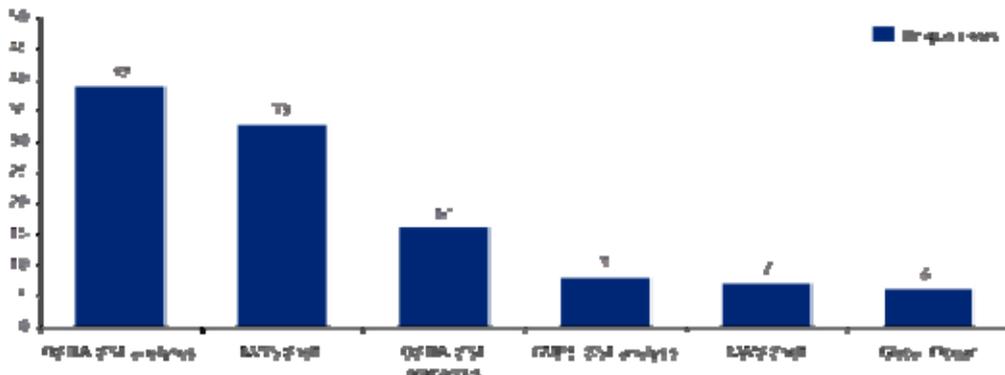


MyOcean

The main objective of MyOcean is to deliver and operate a single European Ocean Monitoring and Forecasting system of the GMES Marine Service (OMF/GMS) to users for all marine applications.

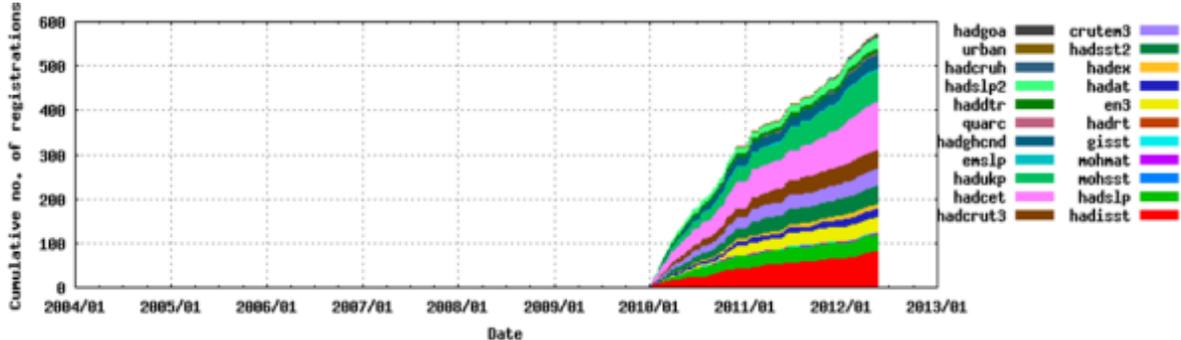
OSTIA SST is also made available through the GHRSSST portal and has many more users than through my oceans as registration is not required.

The charts below show the number of unique users and the number of requests in Jan/Feb 2013.



Hadley Centre Climate Observation Datasets (HadObs)

Researchers at the Met Office Hadley Centre produce and maintain a range of gridded datasets of meteorological variables for use in climate monitoring and climate modelling. This site provides access to these datasets for scientific research and personal usage only. Registered users are a mix of national and international. It should be noted that one person may be registered for more than one dataset.



Other sources of data include:

- Wholesale (ECOMET) Catalogue - The primary objectives of ECOMET are to preserve the free and unrestricted exchange of meteorological information between the NMS's for their operational functions within the framework of WMO regulations and to ensure the widest availability of basic meteorological data and products for commercial applications. There are currently 35 Customers purchasing our Wholesale data, the majority of whom are resident outside the UK.
- WMO Resolution 40 - The meteorological and related data and products described as the minimum set of data and products which are "essential" to support WMO Programmes and which Members shall exchange without charge and with no conditions on use.
- On-line from the Met Office website – various downloadable data related to content available to view. Includes historical station data, regional climate data and marine observations.
- Weather Observations Website (WOW) – A portal allowing users to upload weather observations, this also includes the Met Office official monitoring sites. Data can be downloaded based on the content viewed.

- Library requests – About 300 requests a year for historical observation datasets of various sizes provided under personal or research use only terms of use.

It is important to note that the outputs of the Met Office reach many millions of people through channels that are not represented in these figures. For example, the forecasts of the Public Weather Service broadcast by the BBC on TV and radio are estimated to reach an audience of between 10 and 20 million per week. The audience through other channels are estimated at around four million. An additional 5-10 million are estimated to access forecasts through digital channels such as the web and mobile apps. These, it should be noted, are figures for 'value-added' services rather than raw data. The Met Office also provides a large range of other services, many on a commercial basis, meaning that it is complex to estimate the number of users across the full range of data and value-added services.

Source: Met Office

Ordnance Survey

In the case of Ordnance Survey, the collection of data is the organisation's *raison d'être* and this public sector information is therefore not a by-product of other activities. Most revenue, therefore, is generated from charges levied on users which cover not only dissemination costs but also the cost of collection. Where value-added services are provided, these are priced to cover the cost of development. Ordnance Survey products and services are priced at market rates to avoid undercutting competitors. Access is provided on the same terms for all parties.

The following data was received from Ordnance Survey for this report:

The approximate split between open O/S data and O/S data available under licence and the number of downloads of O/S data by different channels

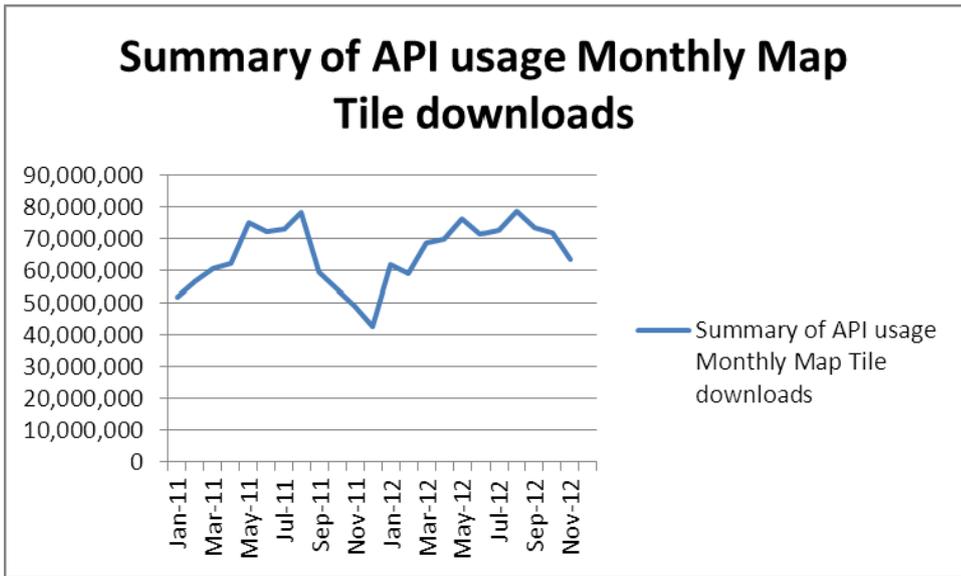
- For OS OpenData, in 2012 we processed 34,500 orders. Nearly 95 per cent of orders were for download, the exceptions tending to be for national datasets of OS VectorMap District and OS Streetview where the volume of data is very high. The top three products by items ordered were OS VectorMap District, OS Streetview and Code-Point Open. The very different nature of the products means that a comparison by data volume and number of downloads does not provide clarity. For example OS VectorMap District may be ordered as a national set (one order) or as individual 100km x 100km tiles, whereas Code-Point Open is available only as a national set.

Number of O/S license holders for last three years

- We do not keep a periodic record, but at present we have approximately 250 licensed partners who take the data and on-sell it in value-add products or services. The trend has been for the number of licensed partners to grow and for the number of direct customers to shrink, as Ordnance Survey focuses on the key users in specific areas such as Energy and Infrastructure, while encouraging further development of value-added products by a growing range of partners

API usage statistics

- Below is a graph showing the volume of data on a monthly basis utilised through OS OpenSpace since January 2011.



Source: Ordnance Survey

Appendix 5: Empirical methodology

This appendix provides further details of the quantification methodology used for the current value of public sector information in the UK. The appendix covers how the report has defined the term ‘value’ and the different modelling assumptions used.

Final estimates and ranges can be found in the main report.

General approach and limitations

As has been discussed in the main report, attempts to accurately quantify the value of current or future public sector information in the UK are restricted by the lack of data. To recap, all things being equal, a standard valuation exercise for public sector information would draw upon the following pieces of data to reach estimates:

- an estimate of the total number of public sector information datasets available and not available;
- the relative value of different datasets to one another;
- the price paid for different datasets;
- how different datasets are used;
- the different costs incurred in supplying different datasets;
- the number of jobs in the UK attributable to public sector information or involved in its supply or use/re-use;
- the willingness to pay by consumers for different datasets;
- the financial value of the different types of benefits (economic and societal) enjoyed by consumers using and re-using public sector information; and
- price elasticities of demand for different datasets and awareness of the shape of consumers’ demand curves.

Further, in order to reach a robust estimate of public sector information value, data or knowledge of linkages would be required to establish the direction of causality between use and re-use of public sector information and their impacts in order to properly account for additionality effects.

While elements of the above data requirements are available to differing degrees, the majority of the above is not available for 2011-12 (the financial year this report focusses on), or for that matter for earlier years.

In the time available for this research project this report has sought to acquire some of this data from government officials, PSIHS, consumers and other stakeholders. However, in the absence of any extensive primary research exercise (including, but not limited to, consumer surveys, data audits and experiments) there remain significant data gaps. Accordingly, the modelling exercise is based on a pragmatic approach, making conservative assumptions where necessary and relying on previously collected data, e.g. on price elasticities. Similarly, the methodology builds on previous methodologies rather than creating a new one from first principles. The authors of the

report have discussed, and received feedback on, the methodology with the DSB and a number of other government officials²¹³.

The approach has sought to give an indication of the depth and breadth of the value of public sector information in the UK rather than a definitive statement of its value today or tomorrow. Given the public sector information is rapidly evolving, we believe it is better to give a broad indication of the sources and quantum of value (complete with the appropriate caveats) rather than a figure that has spurious accuracy.

It is acknowledged this approach is not ideal and could be improved with the addition of new primary research to update assumptions and fill data gaps.

Definition of value

It is important to clearly define what is understood by the term value. From the literature, it is possible to disaggregate value into four distinct components:

- **the direct value of public sector information to producers and suppliers (the PSIHs):** these are the benefits accruing to producers and suppliers of public sector information through the sale of public sector information or related value-added services;
- **the indirect value of public sector information arising from its production and supply:** the benefits accruing up the supply chain to those organisations interacting with and supplying PSIHs (but not directly using or re-using public sector information), and the benefits accruing to those organisations where employees of PSIHs and supply chain organisations spend their wages;
- **the direct use value of public sector information to consumers of public sector information:** the benefits accruing to businesses, civil society, individuals and the public sector from directly using and re-using public sector information for a variety of purposes; and
- **the wider societal value arising from the use and re-use of public sector information:** the benefits to society of public sector information being exploited, which are not readily captured elsewhere.

The first three types of value can be grouped as **economic value** or narrow economic value and can be measured using standard economic methodologies to derive a monetary estimate for value and the associated employment figure. The final type of value is harder to measure as it captures **wider benefits to society** arising from the use of public sector information – these are typically not measured in monetary terms. The literature discusses a number of ways in which public sector information can have broader impacts:

- **increasing democratic participation:** giving citizens and businesses access to public sector information allows them to perform their own analyses of salient issues, make more informed choices about public service providers and interact with policymakers to challenge their assumptions and improve the policymaking process;
- **promoting greater accountability:** for example through the scrutiny of costs of public service provision and benchmarking comparable services;
- **greater social cohesion:** for example, by providing more information on the provision and distribution of services, public sector information can be used to dispel myths on who receives certain public services;
- **generating environmental benefits:** such as reducing congestion and pollution through the release of better traffic and transport data which helps drivers to better plan journeys; and

²¹³ Although they have not endorsed the methodology used.

- **identifying previously unknown links between different policy areas:** through data-mash ups it may be possible to develop system-wide solutions that holistically seek to address the root of policy challenges.

Thus, for the purposes of this evidence base, the value of public sector information is defined as the benefits accruing to the suppliers, users and re-users of the information and data in terms of profits generated, jobs created and supported and the wider societal benefits.

It is important to clarify that value defined in this way is not the same as the **market value** of public sector information. Market value typically refers to the volume of sales multiplied by the price of the product. In this way it equates to revenues or thus includes labour costs, capital costs, profits and the intermediate costs of production which are **not** included in our definition (see below for more details). Conversely, market value will not capture indirect and induced effects and wider societal benefits.

Methodologies considered

As highlighted in the literature review appendix, there is no consistent methodology used in previous research to quantify the value of public sector information which has led to a large variance between estimates (ranging from £590 million to £16 billion - a factor of almost 30). The underlying reasons for these different figures appear to be:

- whether the quantification approach has been ‘top down’ or ‘bottom-up’;
- how the term ‘value’ has been defined;
- the underlying data used; and
- the category of public sector information being valued.

Below are summarised some of the different methodologies used in the literature.

Figure A5.1: Value of public sector information methodologies

Author	Approach summary	Data requirements	Assumptions used
OFT / DotEcon (2006)	Bottom-up or welfare approach. Value = net surplus = consumer surplus + producer surplus	<ul style="list-style-type: none"> • Revenue from PSIHs • Number of OPSI licences 	<ul style="list-style-type: none"> • Linear demand curves • PSIHs split into three categories and then subdivided in public sector information type • Calibrated elasticity estimates • Value-added public sector information is priced competitively • For 'free' public sector information, value is estimated relative to usage of value added public sector information
Houghton (2011)	Welfare approach and returns to expenditure approach	<ul style="list-style-type: none"> • Website usage and number of downloads 	<ul style="list-style-type: none"> • Assume a lower bound of 20% return on expenditure • Useful life of public sector information knowledge is 5 years • Elasticity varies over time
Koski (2011)	Econometric approach examining impact of marginal cost pricing on real sales growth	<ul style="list-style-type: none"> • Information on marginal cost pricing • Various control variables for firms 	<ul style="list-style-type: none"> • Limited to firms within SIC 7420
MEPSIR (2006)	Survey of PSIHs asking size of PSI market and sum of turnover of individual re-users minus cost of acquisition	<ul style="list-style-type: none"> • Public sector information turnover data • Number of users and re-users 	<ul style="list-style-type: none"> • Average turnover per user calculated from survey • Ratio of re-users per sub-domain was 9.5 (mean) and 8.5 (median)
Pollock et al (various)	Gains from moving from marginal cost pricing is $\frac{2}{5}(F\lambda\epsilon)$	<ul style="list-style-type: none"> • F = total revenue from sales of public sector information 	<ul style="list-style-type: none"> • Multiplier varies between 5 and 8 and elasticity 2 to 3.5
ACIL Tasman (2009)	Productivity based approach for use and non-use values – Computable General Equilibrium (CGE) model	<ul style="list-style-type: none"> • Extensive macro- and microeconomic data required for CGE modelling process 	<ul style="list-style-type: none"> • Various assumptions underpin CGE model

Author	Approach summary	Data requirements	Assumptions used
Cebr (2012)	Future efficiency-gain approach – combination of a literature review tracing impact of R&D and impact on profits of start-ups due to reduced entry barriers	<ul style="list-style-type: none"> ONS economy data by sector 	<ul style="list-style-type: none"> Assumptions on adoption rates from literature Impacts via business efficiency, innovation and creation
Cowi A/S (2010)	Gains from making address data free of charge quantified comparing use of data before and after policy change multiplied by price	<ul style="list-style-type: none"> Volumes of data usage 	<ul style="list-style-type: none"> Counterfactual of what would have happened anyway
Deloitte Access (2011)	CGE modelling to analyse the productivity improving from addressing the 'information glut'	<ul style="list-style-type: none"> Extensive macro- and microeconomic data required for CGE modelling process 	<ul style="list-style-type: none"> Various assumptions underpin CGE model
McKinsey (2011)	Combination of an index on value potential and index on the ease of capture; scaling up drivers in case studies; and identifying the addressable share of market for efficiency gains	<ul style="list-style-type: none"> Spending and taxation data Graduate and employee data Capital stock data Bespoke case study data 	<ul style="list-style-type: none"> Various underpinning each calculation
PIRA (2000)	Investment cost estimates and demand side estimates built from cost of time, access price, price paid for public sector information etc.	<ul style="list-style-type: none"> Annual reports Statistical agencies data 	<ul style="list-style-type: none"> Two categories of public sector information: public sector information for final users and public sector information for intermediate users Costs allocated accordingly
PwC (2010)	Welfare approach	<ul style="list-style-type: none"> Proportion of fixed to total costs Administrative costs Proportion of purchases by firms 	<ul style="list-style-type: none"> Assumptions on multipliers and elasticity
Roger Tym & Partners (2003)	Surveys on willingness to pay and case studies	<ul style="list-style-type: none"> Survey of customers 	
Oxera (1999)	Value-added approach that charts the different ways public sector information affects the UK economy	<ul style="list-style-type: none"> GVA and revenue contribution 	<ul style="list-style-type: none"> Various

Following a review of the different methodologies used, two broad ways of quantifying the current²¹⁴ value of public sector information were considered: a bottom-up approach along the lines of the OFT/DotEcon methodology and a top-down approach using a SIC-SOC matrix.

The top-down approach began by identifying the number of jobs directly involved in the supply, use and re-use of public sector information. This was on the basis of Standard Occupational Code (2010) (SOC) categories of jobs and the numbers employed in each. The categories identified as being involved in public sector information were:

- science, engineering and technology associate professionals;
- information technology and telecommunications professionals;
- IT business analysts, architects and system designers;
- IT and telecommunications professionals not elsewhere classified;
- management consultants and business analysts;
- actuaries, economists and statisticians; and
- business and related research professionals.

While this might not capture all jobs involved in public sector information, it is a useful first approximation. The next stage was to allocate these jobs across different sectors of the economy using a bespoke SIC²¹⁵-SOC matrix which shows how SOC jobs are distributed across different sectors of the economy. Having an estimate of the number of public sector information-related jobs in each sector of the economy allowed the estimation of the level of turnover attributable to public sector information (using worker productivity ratios) in each sector and then adjust to focus only on value.

However, as has been noted elsewhere²¹⁶, such a top-down approach, while straightforward, has a number of drawbacks which reduce its viability. Namely:

- it does not properly account for the counterfactual or the additional impact of public sector information; and
- accordingly there is a risk of over-stating the value of public sector information.

While it may be the case that public sector information may underpin a number of products and services in a given sector, not all the value in that sector can be traced back to public sector information. Further, such analysis does not explicitly consider what value the sector could generate using substitutes to public sector information if it were no longer available.

For these reasons, this report uses a bottom-up approach modified from the OFT/DotEcon.

The quantification approach for the current value of public sector information

The approach has two main stages:

- **Stage 1:** estimates of the value of public sector information to PSIHs and value to direct users of public sector information using a consumer and producer surplus approach; and

²¹⁴ The focus was on deriving a methodology for the current value of public sector information rather than the potential future value in the first instance. As is discussed later, the methodology for quantifying the potential value of public sector information was more ad hoc given uncertainties over how the market will develop.

²¹⁵ The Standard Industrial Classification splits the UK economy into over 600 different sectors.

²¹⁶ For example by the OFT (2006).

- **Stage 2:** estimates of the value of associated indirect and induced impacts to PSIHS using Input-Output model multipliers (this stage was not carried out by DotEcon).

This report has not sought to quantify the value of the wider societal impacts of public sector information using a systematic methodology. There is a lack of reliable evidence on the linkages between the use and re-use of public sector information and the impact on environmental, political, social and other indices – though, where possible, the report has made ad hoc estimates of the value generated from particular datasets.

Stage 1: consumer and producer surplus approach – the welfare approach

The welfare approach can be estimated from summing the current net consumer surplus²¹⁷ with the total producer surplus²¹⁸ from the supply of public sector information. The approach differs according to whether the public sector information in question is paid-for or free. By focusing on consumer surplus, which takes into account the presence of substitutes, the report is able to accommodate the presence of substitute public sector information datasets and in doing so captures the additionality impact.

Paid-for public sector information

One of the key drivers for the size of value arising from paid-for public sector information is how demand varies with price changes (its price elasticity of demand) – this will influence the size of consumer surplus. When demand is more elastic, the economic value is likely to be lower as customers will switch to substitute products in response to price rises. When price elasticity of demand is low or inelastic, economic value will be higher as customers will be less inclined to switch to substitutes – perhaps because there are few true substitutes.

Clearly, a critical assumption will therefore be the price elasticity of demand for different categories of public sector information and PSIHS. The value of the price elasticity of demand will depend on the benefit users and re-users currently enjoy from public sector information in 2011-12 and will differ by dataset and it unlikely to be stable as new benefits and uses emerge. Additionally, price elasticity will be influenced by the availability of substitutes and context (e.g. some data is only valuable at a particular point in time).

In the time and resource available it has not been possible to survey customers to generate new price elasticities of demand for different public sector information datasets. On this basis, the report has relied on the long-run elasticities presented in the DotEcon report for the OFT. It is acknowledged that the elasticity values will have changed since 2006 especially given the greater availability of open data from PSIHS. However, the other key driver of elasticity – the availability of substitutes – has not changed significantly since the report as there has been no emergence of new competitors in the provision of public sector information in most cases. Clearly, this is an area where new primary research is required, but for the purposes of this evidence base this is considered to be a sufficient starting point for indicating the value of public sector information.

Following the approach by DotEcon, the report uses three broad classes of elasticity for different types of PSIHS and public sector information datasets:

- **Low elasticity:** -0.3
- **Medium elasticity:** -0.8
- **High elasticity:** -1.5

²¹⁷ That is the amount consumers of public sector information are prepared to pay over and above the price they current pay for access.

²¹⁸ That is the extent to which revenues exceed the costs of supply – profit.

These elasticity figures are treated as the baseline. However, as is good practice, the report considers adjusted alternative values for each category following research by Davies, Slivinski, Bedrijvenplatform, Lazo and others²¹⁹.

Figure A5.2: preliminary elasticities assigned to PSIHs supplying pay-for public sector information

Case	Baseline	Alternative 1	Alternative 2
Low	-0.3	-0.5	-0.2
Medium	-0.8	-1.2	-0.6
High	-1.5	-2.0	-1.2

Source: Deloitte analysis drawing on OFT/DotEcon research

These elasticity figures are then matched to a number of key PSIH suppliers in the dataset.

Figure A5.3: preliminary elasticities assigned to PSIHs supplying pay-for public sector information

PSIH	Elasticity
HM Land Registry	Low
Registers of Scotland	Low
Companies House	Low
Ordnance Survey	Medium
UK Hydrographic Office	Low
Environment Agency	Medium
Met Office	Medium
DVLA	High
Office of National Statistics	Low

Source: Deloitte analysis

It is acknowledged that these are not the only suppliers of pay-for public sector information in the UK. However, as noted by Newbery et al (2008), it is estimated that the Trading Funds listed above comprise around 70 per cent of the estimated total income from UK PSIHs, and this captures a sufficiently large dataset population (and the subsequent calculations 'gross-up' the figure to capture the full value).

The choice of elasticity assumption assigned to each PSIH is based on a combination of the understanding of the public sector information held and disseminated by each PSIH, its use and re-use, the level of substitution, discussions with stakeholders and the nature of customers – they have been adjusted from the 2006 mapping. There is, by necessity, an element of abstraction due to the diversity of datasets supplied for a fee by each of the above PSIHs.

To accurately calculate total consumer surplus for the pay-for public sector information market, it is necessary to understand the shape of the demand curves for each dataset. Attempting to derive individual demand curves for 4,000 plus datasets is clearly not feasible and the report has instead posited aggregated demand curves for each individual PSIH. For simplicity, and following previous quantification approaches, the report assumes linear demand curves for pay-for public sector information. This appears justifiable in this instance along similar lines to those set out by DotEcon:

²¹⁹ Annexe G *Economic value and detriment analysis*, December 2006, p.32

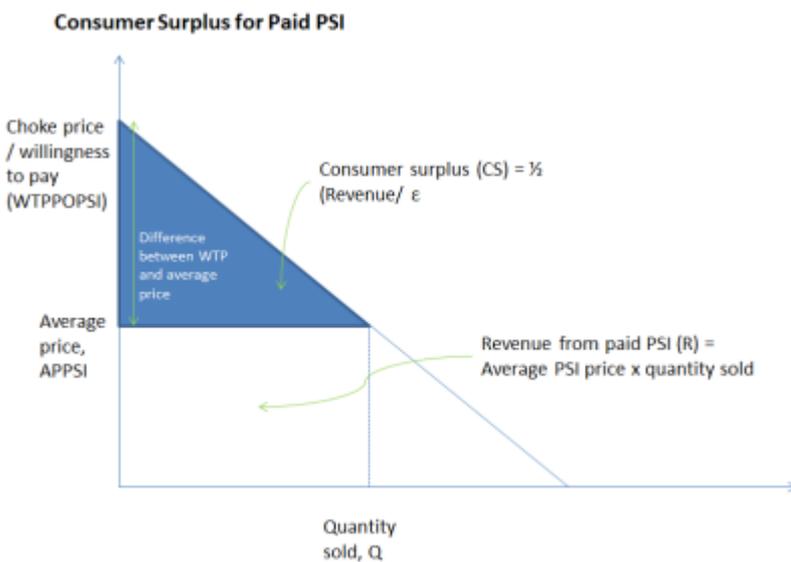
“While in practice demand curves may not be linear, assuming a linear demand can reasonably be argued to give a lower bound on consumer surplus, as real-world demand curves are often convex (as shown in the dotted line) and thus result in a greater difference between willingness to pay and costs. With more complex specifications of demand, this formula is more complex, but there is still a one-to-one relationship between revenue and consumer surplus.”²²⁰

There is perhaps more scope for debate as to the shape of the demand curves when it comes to free public sector information – this is discussed in more detail below.

Thus, using the assumption of linear demand curves, it is possible to recover the consumer surplus (i.e. value to the direct users of public sector information) calculated using the standard formula:

Formula (1): Consumer surplus = $\frac{1}{2}$ (Revenue / Price Elasticity)

Figure A5.4: consumer surplus for paid-for public sector information



Source: Deloitte analysis

Formula 1 is used for each PSIH in turn. Revenue data is sourced from annual reports from PSIHs and adjusted following discussions with stakeholders. In particular, it should be noted that a proportion of Trading Funds’ revenue is not related to the provision of public sector information and should accordingly not be included.

Figure A5.5: PSIH revenue and revenue attributable to public sector information

PSIH	Total revenue, 2011-12 (£)	Of which, is public sector information related (£)
HM Land Registry	360,000,000	72,000,000
Registers of Scotland	58,000,000	12,000,000
Companies House	66,000,000	15,00,000
Ordnance Survey	142,000,000	142,000,000
UK Hydrographic Office	19,000,000	13,000,000

²²⁰ Ibid, p.35

PSIH	Total revenue, 2011-12 (£)	Of which, is public sector information related (£)
Environment Agency	417,000,000	834,000
Met Office	196,000,000	121,000,000
DVLA	459,000,000	1,376,000
Office of National Statistics	29,000,000	29,000,000

Source: Deloitte analysis based on Public Data Group publication *Approach to Charging* and making adjustments for revenue that is not public sector information related, December 2012. Numbers rounded.

Note the above figures have not been ‘grossed up’ for the missing Trading Funds.

The assessment of producer surplus from paid-for public sector information also follows the DotEcon approach. As they note:

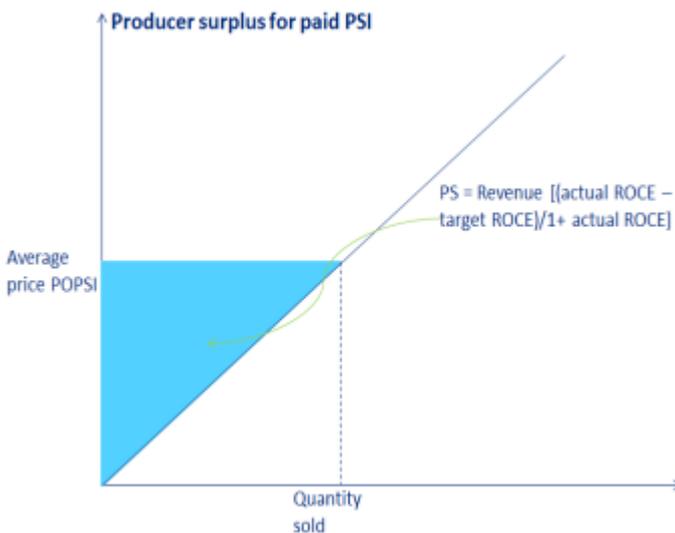
“In order to estimate the producer surplus or loss, we would require estimates on the costs of collecting and supplying the data and then compare this to revenue. However, estimates on the costs of supplying [public sector information] data for each group are not readily available.

In addition, many of the costs of data collection and processing may not be related to the supply of [public sector information] to the private sector, but might result from other public policy objectives that already take such costs into account.”²²¹

Accordingly, when PSIHs do not report a target rate of return on capital employed (ROCE), losses and profits are assumed to be accounted for in the setting of public policy objectives are left out of the assessment (i.e. the National Office of Statistics). In contrast, where a ROCE is stated, this reflects the cost of capital is taken into account when setting public policy objectives and producer surplus can be estimated as the difference between actual and target ROCE²²².

Formula (2): Profit = Revenue ([actual ROCE – target ROCE] / [1 + actual ROCE])

Figure A5.6: producer surplus for paid-for public sector information



Source: Deloitte analysis

Again, the report has used financial data from annual reports and stakeholder discussions to calculate producer surplus (profit) for each PSIH where relevant. For most Trading Funds, the

²²¹ Ibid. p.43.

²²² Following the DotEcon methodology it is assumed all costs are a function of the capital employed and target ROCE. For further details of the assumptions underlying the producer surplus methodology see Annex G of the OFT report.

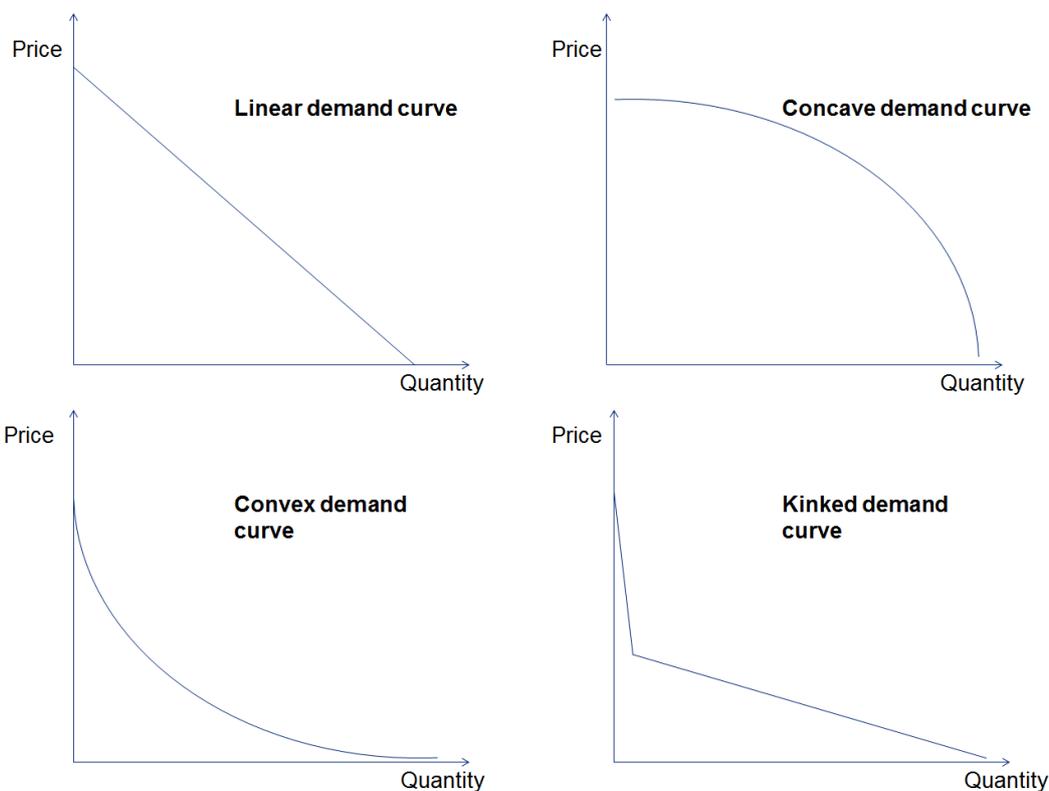
ROCE target is the standard 3.5 per cent figure – actual ROCE data for each Trading Fund can be found in their annual reports. Given the Office of National Statistics is structured differently and does not have to make a return it is not included in producer surplus figures.

Free public sector information

Clearly where there is no price for public sector information the above methodology becomes more challenging – using revenue data to estimate consumer surplus is not feasible as there is no PSIH revenue attributable to free public sector information. Further there is no producer surplus to capture as the information is made available for free by PSIHs.

There are also valid questions as to whether the demand curve for free public sector information datasets can still be assumed to be linear – they may be convex or concave. The different potential shapes for the free public sector information demand curves are shown below.

Figure A5.7: potential demand curves for free public sector information



Source: Deloitte analysis

In all cases, given the 'price' for free public sector information is zero, consumer surplus is the entire area under the demand curve and can be found taking the integral of the curve.

However, as is known from behavioural economics literature, consumer behaviour can move in unexpected directions as price moves from zero to a nominal amount and vice versa²²³. In the left-hand corner, assuming a linear demand curve suggests that as price moves away from zero, consumers react by reducing the quantity demanded proportionally. This may not be the case if consumers react to the price becoming non-zero by not reducing their consumption proportionately

²²³ Indeed, one might argue that at zero price the quantity of public sector information demanded should be infinite. This may not be the case if there are transactional costs involved in the acquisition of the data. For simplicity (and due to the lack of data) transactional costs are not included in the analysis.

– this may suggest a convex, concave or a kinked demand curve. The shape of the demand curve will therefore have a significant bearing on the size of consumer surplus²²⁴.

In the available time and resource, it has not been possible to collect primary data on consumer behaviour to changes in price from zero. In the absence of such data, any specification for the demand curve will be assumption driven: in the case of linear demand curves that consumers proportionately adjust quantity demanded to changes in price, in the case of convex demand curves that the integral sufficiently approximates a discrete summation and is not biased by the units of scale. Following convention, and as a simplifying assumption (and for reasons of consistency with pay-for public sector information), the report continues to use a linear demand curve formulation. However, for completeness an alternative demand curve formulation is considered.

Beginning with the linear demand curve formulation, in the absence of detailed willingness to pay data on different types of free public sector information, it is necessary to make a series of assumptions to recover a provisional willingness to pay for free public sector information.

The DotEcon methodology suggests one approach for this: assuming the choke price (or maximum willingness to pay) for raw data (taken as a proxy for public sector information) is equal to the choke price of value-added (pay-for) public sector information minus the current price of value-added public sector information²²⁵. Using linear demand curves for modelling simplicity, this methodology assumes that value-added public sector information is competitively priced in the main and that the willingness to pay for this would not exceed the willingness to pay for raw data (a proxy here for free public sector information) plus the cost of value-added features – unless the consumer would be better off obtaining the raw data and performing the value-add services themselves.

Using geometry it is possible to recover the choke price for pay-for public sector information. However, in order to do this it is necessary to use average prices for pay-for public sector information. This is not straightforward as prices vary by dataset, volume, usage and consumer type and we have been informed there is no such thing as an ‘average price’ per dataset. The report has sought to recover a proxy average price using annual report data of Trading Funds and then used geometry to calculate willingness to pay estimates for different types of free public sector information. The average value for free public sector information willingness to pay is estimated to be around £1,300 with the figure varying between £500 and over £2,000. While this figure may appear high, it should be noted it encompasses the willingness to pay across the economy and all types of users and re-users, ranging from individuals looking up civil servant pay levels to SMEs developing apps that harness bus timetables to large multi-nationals downloading free economic statistics. This appears to be a reasonable first approximation for these indicative estimates. It would be advisable to conduct more research to ascertain a more accurate, disaggregated figure.

Having recovered willingness to pay for free public sector information, quantity or usage of this information is still required. As discussed in an earlier chapter, data on the number of downloads and page views from a number of major data portals such as data.gov.uk and ons.gov.uk has been gathered – approximately 2.7 million downloads over a twelve month period. Based on the assumption that of these downloads around two thirds of the datasets are actually used or re-used (approximately 1.8 million uses and re-uses). Having quantity used / re-used and the willingness to pay allows us to generate revenue and using the previous linear demand elasticity assumptions it is possible to recover consumer surplus for free public sector information.

²²⁴ In particular, compare the formula for consumer surplus under a linear demand curve which can be rewritten as $CS^L = 1/4p^cD$ with the formula for consumer surplus using a Cobb-Douglas convex demand curve: $CS^D = p^c \ln(D)$ – the value of consumer surplus in the convex case will be much lower.

²²⁵ A fuller description can be found in Annex G of the OFT report.

The alternative calculation of surplus assumes a Cobb Douglas demand curve with customers maximising their utility of consumption from public sector information and all other goods subject to a budget constraint. Through a series of algebraic manipulations it is possible to recover the formula for consumer surplus as:

Formula (3): Consumer Surplus = price*ln(demand)

The estimated quantity (demand) and price (willingness to pay) figures can then be used to calculate consumer surplus in this instance.

The final estimates have taken an average of the linear and Cobb-Douglas estimates.

Stage 2: indirect and induced value of public sector information

Estimates of the value impacts accruing through the business-to-business supply-chain, and employees spending associated wages, are based upon the UK Domestic Use Matrix for 2005 (latest available tables accounting for import leakage) from ONS and the direct use estimate derived above.

The direct use estimate of producer surplus (analogous to operating profit) is converted into expected gross output (GO) for each relevant industry on the basis of information contained in the UK DUM. This process uses the inverse of the industry average ratio of operating profit to GO.

GO TYPE I and TYPE II multipliers are then used in an Input-Output setting to consider the upstream business-to-business purchasing effects (indirect) and consumer spending effects (induced). The average GO multiplier used in this process was estimated to be in the order of 3.0.

To allow a comparable estimate of value to the original producer surplus (and thus allow aggregation with other quantified elements), the results are converted back from GO into GVA and then, in the same manner, operating profit, by definition, providing a comparable surplus estimate.

Again this is achieved by considering the ratios of operating profit to GVA and GO, relative to the incremental GO for business-to-business spending and consumer spending.

The 'surplus multiplier' – the ratio of indirect and induced surplus to the direct use surplus – was estimated to be 3.4 (1 + 2.4). For each £1 generated as producer surplus in direct use, a further £2.40 is generated via indirect and induced effects.

Per worker productivity estimates sourced from ONS through the Annual Business Survey are then used in conjunction with estimates of GVA to provide an indicative level of employment supported in organisations supplying inputs to the PSI supply chain and supply goods and services to consumers.

Appendix 6: Transport sector case study

Over the last decade the release of ever greater volumes of detailed and real-time information has had a considerable impact on the transport sector. This appendix focuses on quantifying the impact of releasing transport data in London, where particularly large volumes of information have been made available. It also discusses impacts further afield and the future opportunities for increased use of information. It is therefore an extended version of the transport case studies presented in Chapter 5.

Overview

Over the past decade the UK transport sector has seen the release of ever greater volumes of detailed and real-time information. This has allowed travellers to plan and adapt their journeys in response to real-time updates on conditions across the transport network, as well as allowing improved public scrutiny of the performance of the transport system.

This section highlights some of the ways in which information has impacted the transport sector. There is a particular focus on London, which – partly owing to the extent of its transport network – has seen particularly high volumes of information released, and a correspondingly large number of products and services developed to allow travellers to access this information and use it to enhance their journeys. As part of this case study there is an attempt to quantify part of the impact of releasing this data, by assessing the value of time saved through better access to up to date travel information.

This case study also assesses the impact of information on the transport sector outside London: in other cities and on inter-city transport by road and rail. Where possible it offers some estimate of the value generated by information in these areas, although any such figures should be treated as high level assessments based on reasonable assumptions, in the absence of the data that would be needed to conduct more detailed calculations.

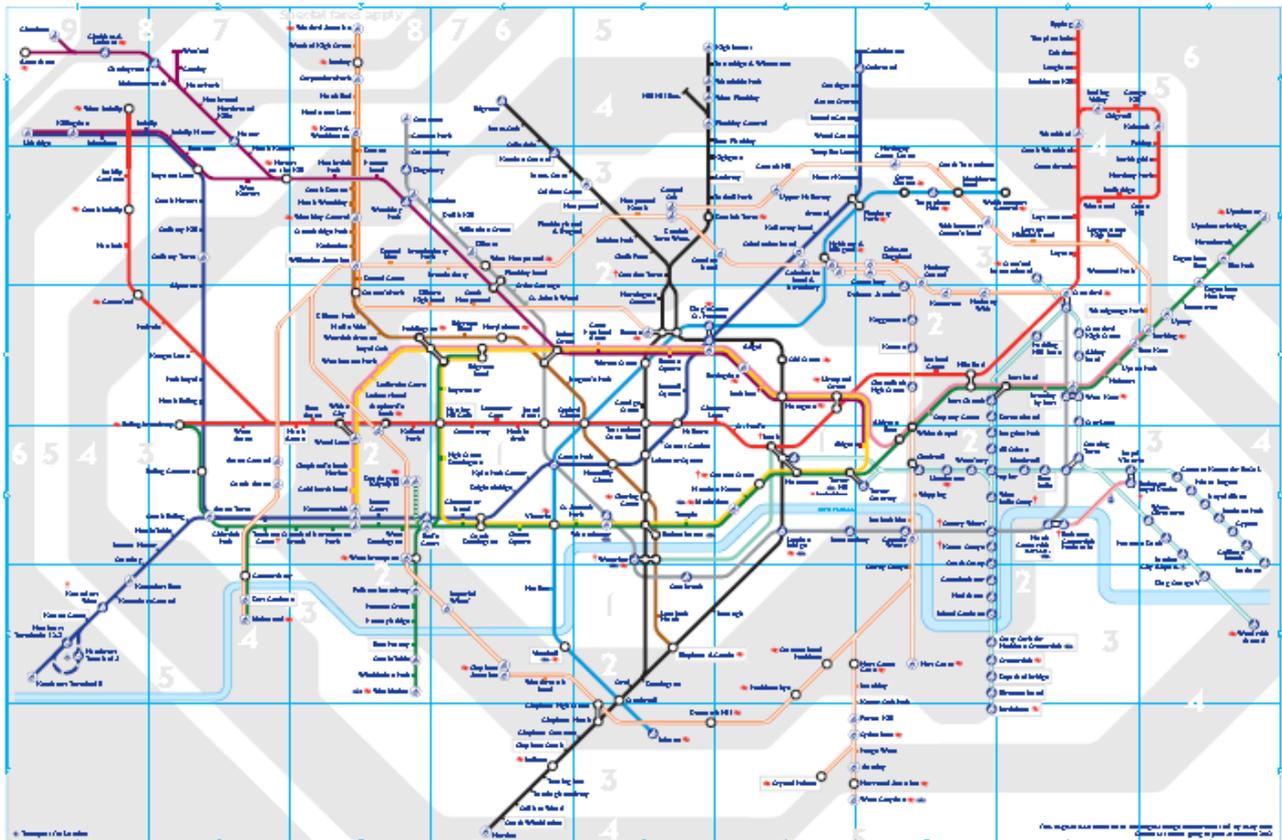
Information released by Transport for London

London is unique among the UK's cities in the extent and complexity of its transport network. Its underground network alone comprises over 400km of track, with over 500 trains operating across the network at peak times. Each train travels over 184,000km a year, carrying 1.171 billion passengers in 2011/12. This is in addition to approximately 19,500 bus stops served by around 7,500 buses²²⁶; an extensive overground rail network; light rail networks including the DLR and London Tramlink; and a riverboat service operating from Greenwich to Millbank. In addition, London has an extensive road network, with private motor vehicles accounting for more journeys than the underground and buses combined.²²⁷

²²⁶ Source: www.tfl.gov.uk/corporate/modesoftransport/1548.aspx

²²⁷ Source: www.tfl.gov.uk/corporate/modesoftransport/2794.aspx

Figure A6.1: TfL rail network map



Source: www.tfl.gov.uk

This complexity means that detailed and current information is vital to enable travellers to efficiently navigate the network. This includes maps, timetables, and up to date information on closures, disruptions and delays. Transport for London (TfL) has signed up to the transparency agenda and provides a wide range of information to users and re-users - principally through its website but also through other channels. In so doing it builds on the 2010 Mayor’s Transport Strategy, which included among its commitments “improving the provision of real time and other journey planning information, including upgrading the TfL web-based journey planner, allowing further improvements to its real time performance, accuracy and personalisation.” TfL writes that through transparency it hopes to:

- “Enable our stakeholders to hold TfL to account;
- Deliver better value for money; and
- Enable businesses and non-profit organisations to develop innovative applications using our data.”²²⁸

Accessing information

The information collected and released by TfL is largely made available to developers and the general public through their website. For developers, at the time of writing, there are 29 data feeds

²²⁸ See www.tfl.gov.uk/transparency/

available in TfL's Developer Area.²²⁹ For information, the complete list of data feeds available is included at the end of this appendix.

This data is not published under the OGL, and the licensing terms for developers set by TfL include some restrictions, including branding conditions. The data is, however, available free of charge, and TfL encourage its re-use in innovative ways, subject to licensing conditions. The TfL website contains an extensive guidance system providing contextual information and assistance for each feed, as well as suggestions for its use by developers.²³⁰

For users other than developers, TfL provides information directly through its website under the 'getting around' and 'live travel news' sections. The information available includes:

- a range of maps;
- station locations;
- accessibility information;
- live service updates;
- live departure information; and
- information on planned works and weekend closures.

Other ways of accessing the information include a free mobile service alerts service for the tube and DLR. In addition, there are several free tools available for non-developer users, including a journey planner.

Usage levels

In the absence of usage figures, the analysis here presented is based on the apps available in the UK through the Apple App Store and Google Play (for Android apps) which use TfL data. This analysis suggests that in the UK across both platforms there have been around 500,000 downloads of apps relating to bus services, and around 2.6 million downloads of apps related to tube services. In addition there have been nearly 900,000 downloads of apps covering both the bus and tube networks. This means that in total there have been nearly four million downloads of apps using TfL data.²³¹

It should be emphasised that these figures give only a very rough estimate of the number of users of TfL data. Many users will access the information directly through the TfL website, through another website, or through other services. There is also no guarantee that an app download translates into regular use, as users may download an app but then decide either that they do not like it (and download an alternative), or fail to integrate it into their daily routine. Nonetheless, these figures are indicative of the high level of uptake of TfL data.

It is also important to recognise that much of this information has long been available through other forms. For example, even before the invention of the internet it was possible to access information about disruptions to travel through the media, such as the television and radio news. Drivers have long listened to radio traffic updates to help them avoid congestion. However, this information is now significantly easier and more convenient to access at a time that suits the user rather than the distributor of the information.

The use of smartphone apps allows instant access to detailed up-to-date information on the status of all aspects of the transport system, as well as enabling users to plan an alternative route if necessary. There is likely to be value in this added convenience, both in terms of a greater number of users of the data, and also due to improved access to the latest information when on the move.

²²⁹ See www.tfl.gov.uk/businessandpartners/syndication/16492.aspx

²³⁰ See www.tfl.gov.uk/businessandpartners/syndication/16493.aspx

²³¹ See www.xyo.net; based on a search for 'London' within the transit category in the UK site (November 2012)

Benefits

Conceptually, it is easy to see how the information provided by TfL has generated a range of benefits. The most notable benefits are likely to be derived from:

- **Live arrival and departure times for trains, buses and the underground**

Developers are now able to offer smartphone apps which allow travellers to see how long they need to wait for the next bus, or when their next train home will be departing. This means that they are able to make more efficient use of their time, with less time wasted waiting, and are able to make informed decisions about the most efficient way of completing their journey. For example, if there is a 12 minute wait for the next bus, it may be more time efficient to walk five minutes to the tube station instead. With this information, travellers are empowered to make better decisions

- **Traffic cameras and congestion data for the road network**

Access to live traffic cameras for drivers and other road users means that it is now possible to plan a route in advance avoiding areas with heavy traffic and incidents such as road accidents. This information is available both via the web, as with the BBC Travel News website shown below, and on the move via smartphone apps. It is also possible to access information on congestion and average traffic speeds along a route.

Figure A6.2: the BBC Travel News page



Source: www.bbc.co.uk/travelnews/london/trafficcams

- **Oyster Card data allowing travellers to track their usage**

By registering an Oyster Card online, users are now able to view a history of their journeys on TfL services, including how much each journey has cost them and the start and end times. This allows users insight into their usage patterns, including the time and money they spend on transport, and might potentially encourage behaviour change, for example changing their travel patterns in order to travel outside peak time and take advantage of lower fares, where possible.

- **Oyster Card data giving insights into the flows of travellers through the transport network**

A by-product of the Oyster Card system is an extensive database containing details of every journey taken on the London public transport network, allowing TfL to monitor journey times and volumes of passengers across the network. This should enable the transport network to be run in a more resilient way, revealing common bottlenecks and allowing the impact of delays and closures to be modelled across the network, improving contingency planning.

That the changes described above will benefit both users and operators of the transport network seems clear. In many cases it is difficult to identify and isolate the precise impact of each benefit, beyond incremental improvements to efficiency and user convenience. However, there follows details of a quantification of some aspects of these benefits, notably time savings to travellers.

Quantification of benefits

TfL 'lost customer hours' data

TfL track the performance and reliability of all their transport services. However, each mode of transport reports its performance in different ways. London Underground publishes performance reports covering each 28 day period throughout the year. To understand the overall level of delay caused by a disruption on the network, London Underground uses a measure known as Lost Customer Hours. This is generated by multiplying the duration of a disruption by the number of people estimated to be affected, based on the severity of the incident and expected customer demand at different times and places. Lost Customer Hours therefore accounts both for those who faced delayed journeys, and for those who were unable to travel at all. Over the year to October 2012 the underground network averaged just under two million lost customer hours per month.

London buses do not provide information on lost customer hours in their performance reports. Annual lost customer hours have been calculated by taking the figure for the average 'excess wait' (beyond that timetabled) and multiplying it by the number of annual passenger trips made by bus.

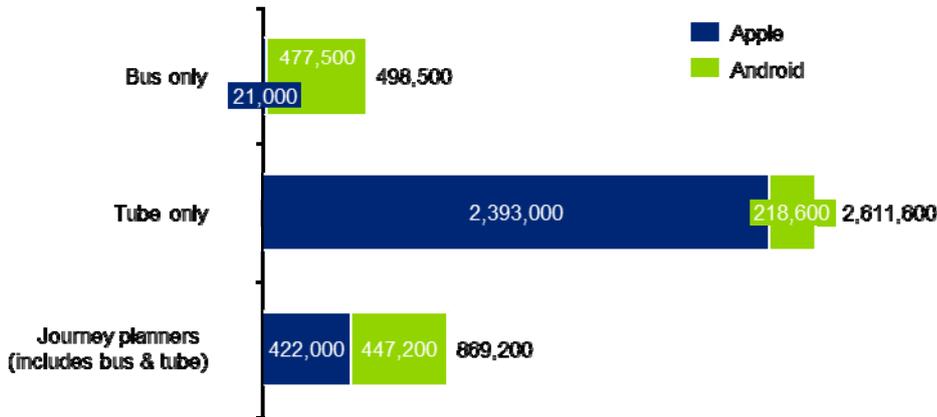
For private road users, London Streets benchmarks performance by measuring 'journey time reliability' (JTR): the proportion of journeys completed within an allowable excess of five minutes for a standard 30 minute journey during the morning peak. The analysis is based on taking the JTR figures and multiplied them by the annual number of driver and passenger journeys, and separately the number of taxi journeys, to work out the number of journeys experiencing delays in excess of five minutes. An assumption regarding the average time lost per delay is multiplied by the number of delayed journeys to arrive at a figure for lost customer hours for private road users.

The approach to rail travel (London Overground) is similar: the reliability figure of 96.6 per cent for 2011-12 from the Travel in London Report and is used to work out the number of delayed journeys. An assumption regarding the average time lost per delayed journey is then used to calculate a figure for lost customer hours for London rail travel. Note that this refers only to London Overground, the TfL-operated rail network. It does not include rail travel to and from locations outside of London; nor does it include non-TfL services run by national rail operators in London.

Apps using TfL data

An online resource (www.xyo.net) has been used to estimate the number of downloads of apps based on TfL data, as described above. Although these figures should not be assumed to be fully accurate, they give a sense of the level of demand for apps using TfL data.

Figure A6.3: download figures for apps based on TfL information

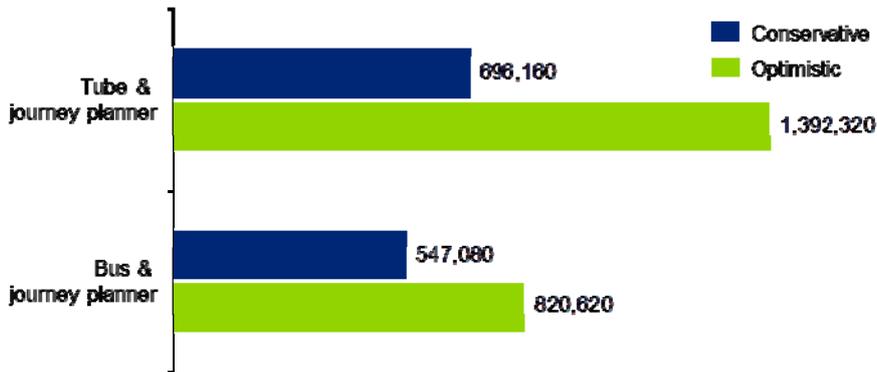


Source: Deloitte analysis of data from www.xyo.net

It is unlikely that all app downloads lead to regular use of the app. It may be that a user downloads an app to try it out, but decides they do not like it and deletes it in favour of an alternative. Equally, it is likely that some people download an app but then fail to develop the habit of consulting it regularly. This is accounted for in the model.

The model includes both a conservative scenario, in which 20 per cent of app downloads lead to regular use, and an optimistic scenario in which 40 per cent of app downloads lead to regular use. These percentages are then applied to the total app download figures to generate conservative and optimistic figures for regular app use. Note that the multi-function journey planner apps are included in the totals for both tube and bus, as they are equally applicable to both. These figures apply only to bus and tube passengers.

Figure A6.4: scenarios for numbers of users of app based on TfL information



Source: Deloitte analysis of data from www.xyo.net

It should be noted that in addition to the benefit to users of apps (who are second order users of the data), which is the focus of this analysis, there will be first order benefits from the data through contributing to an app economy which provides jobs, tax revenue, and other benefits. Even where apps are not charged for they may create jobs, as there are other mechanisms for revenue generation (for example through advertising). This ‘infomediary economy’ is an additional benefit from the release of TfL data. This aspect is not quantified in this analysis, but it should be borne in mind as another wider economic impact of releasing the data.

The model

By making some assumptions about the number of passenger hours saved through better access to information, and the value of an hour, it is possible to estimate the time potentially saved, and the value of that time, owing to the information released by TfL.

The model uses Department for Transport figures for the value of time, published in October 2012. They make a distinction between working time, which is a business cost, and non-working time (including commuting) which is time to which each individual attaches a value. The value of working time in the DfT analysis varies by mode of transport. The figures provided by the DfT are for 2010; in this analysis they have been inflated at the prevailing rate of CPI to estimated 2012 values. The full DfT cost of time figures are provided for reference at the end of this section.

To work out the number of passenger hours saved, the model is based on TfL figures for customer hours lost, or calculations of these where the data is unavailable, as described above. The proportion of passengers likely to be using an app based on TfL data to access travel information is then calculated, based on the download figures.

In addition to the two scenarios regarding the level of regular app use, it is important to recognise that not all travellers would be able to adapt their route even if they knew that they faced disruption on their planned route. For example, someone needing to travel from High Barnet to Waterloo would be likely to have no viable alternative to the Northern Line, even if this meant enduring significant delays. The model therefore factors in two additional scenarios: a conservative assumption in which ten per cent of users are able to alter their route in response to information on delays, and an optimistic scenario in which 25 per cent of users are able to alter their route.

Although DfT provide a range of values of time, this model uses the lowest value – for leisure/commuting – rather than the higher working time values. Adopting this approach generates conservative estimates, in the absence of more detailed information on the types of travellers affected by delays. It is also reasonable to assume that the majority of journeys do not occur with business as the primary journey purpose. However, it should be noted that introducing the value of working time into the model would mean that the estimated value of time saved would become significantly higher due to the large differential between leisure/commuting and working time in the DfT values of time.

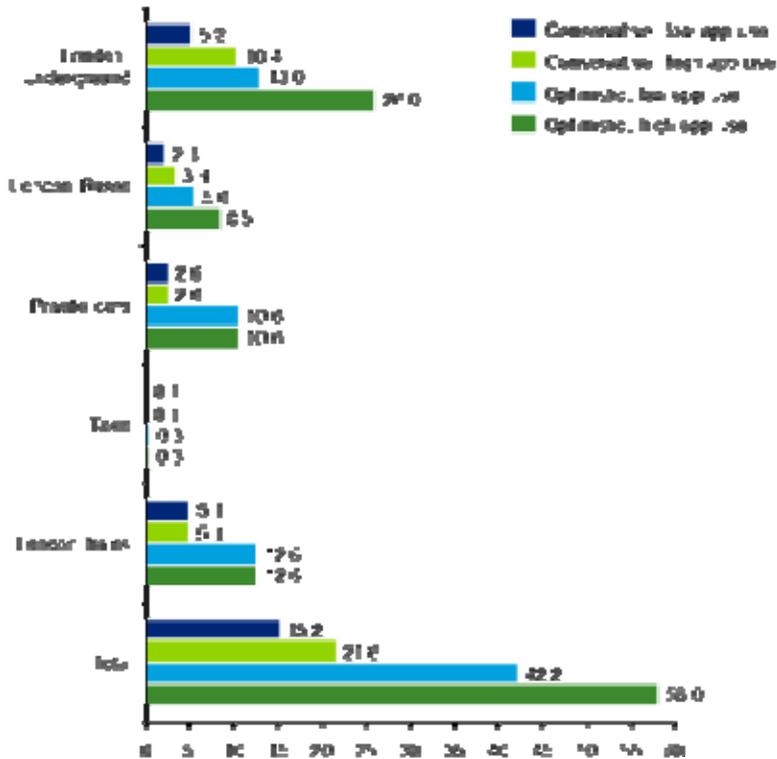
Findings

The estimate of the value of time saved through use of TfL data generated by the model ranges from around £15 million to nearly £58 million per annum (based on 2012), depending on the scenario for level of app use and ability of travellers to change their journey routes.

Of this value, around a third is accounted for by savings for underground passengers, and another third by savings for passengers of London trains (over-ground). The remainder is split roughly equally between bus passengers and private vehicles. Taxis account for a relatively small level of savings, due to the low number of passengers relative to other modes of transport. It should be noted, however, that if the values of working time are used the savings for taxi users become relatively more significant, as the DfT accords taxi users the highest value of time of any mode of transport.

These figures are shown in the chart below.

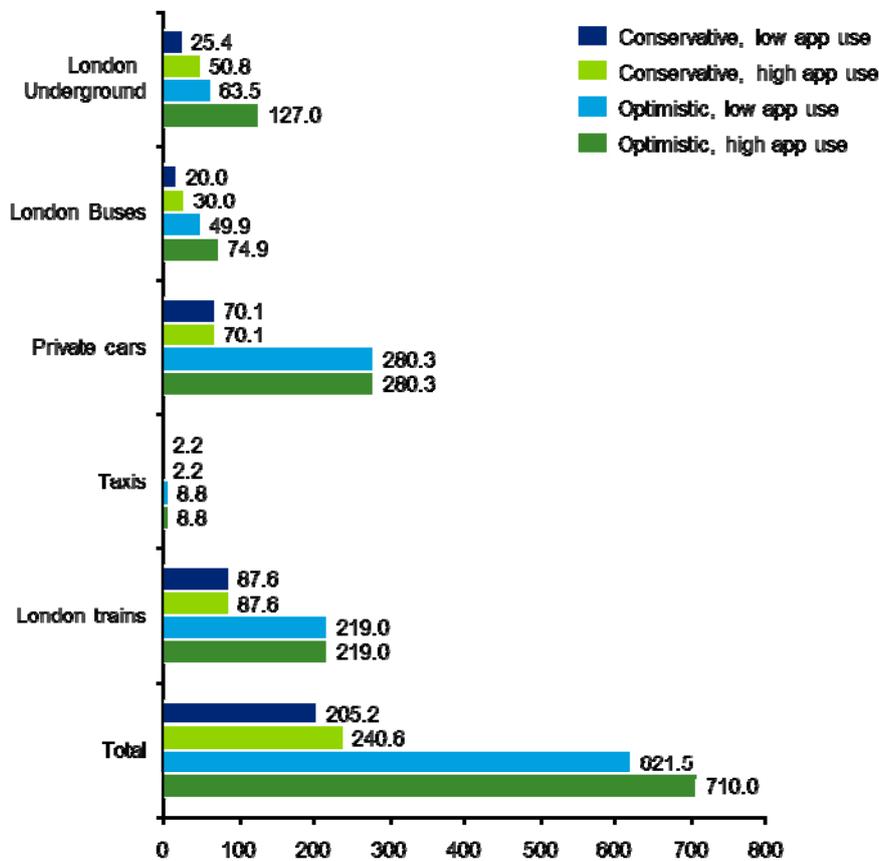
Figure A6.5: the value of time saved through use of TfL information, 2011/12 (£ million)



Source: Deloitte analysis

The number of beneficiaries, i.e. passenger journeys which save time as a result of travellers using TfL information, is shown in the graph below (millions of journeys). Under the most conservative scenario around 200 million passenger journeys per year are estimated to save time due to TfL information, while under the most optimistic scenario the figure is over 700 million passenger journeys. The highest number of beneficiaries are private car users, indicating that under this modelling approach there are a large number of car users experiencing savings to their journey time, but that the time saving to each journey is relatively small.

Figure A6.6: the number of passenger journeys experiencing time savings through use of TfL information, 2011/12 (millions of journeys)



Source: Deloitte analysis

Summary

It is important to note the limitations of these figures. They attempt to model only time saved due to travel disruption avoided – they omit the time potentially saved in everyday travel, for example by allowing commuters to time their exit from the office so as to catch the next bus.

They are furthermore based on a range of assumptions regarding traveller behaviour, specifically assumptions regarding app use and ability to alter their route of travel, as well as estimations of the value of time that may not be accurate for all travellers.

For these reasons the approach here adopted is the most conservative that seems reasonable. The figures represent the lower bound impact of TfL data – in other words, the true impact is likely to be significantly higher. By way of comparison, the much fuller HS2 benefit assessment estimated annual time saving benefits at a net present value to 2043 of £7.3 billion.²³² On a per annum basis, over the 17 years between 2026 and 2043 (before which the line is not scheduled to be operational) this equates to £417 million in annual journey time savings, or £440 million in 2012 prices.²³³

²³² See <http://highspeedrail.dft.gov.uk/sites/highspeedrail.dft.gov.uk/files/hs2-economic-case.pdf>

²³³ There are a number of shortcomings to this estimate, notably that it calculates the benefits on a straight line basis. This report has not conducted an in depth analysis of the benefits of HS2: the figure is included in order to illustrate a point, which is that even making conservative modelling assumptions, the annual time saving benefits delivered by

This is based on many fewer passengers but a much greater journey time saving per journey. As such, the provision of exhaust PSI in London and its environs, at what is a relatively low marginal cost to Government, may create as much as 13 per cent of the annual monetised time savings of a major infrastructure investment such as HS2. In short, information is time, and time is money.

Finally, these figures do not take account of the broader welfare benefits – for example social and environmental – which is likely to results from reducing friction and improving efficiency for travellers around London. These benefits, although beyond the scope of this study and difficult quantify, are likely to be significant.

TfL statement

In the course of the work on this case study, discussions were held with TfL. They have provided the following statement for inclusion in the report.

“Every day, millions of Londoners and visitors rely on the information provided by Transport for London (TfL). Our online journey planning and service update tools are essential sources of travel advice used by 8 million customers. And through traditional media such as radio, television and newspapers, our travel information reaches millions more.

But in our aim to deliver world leading customer service, we are now realising the enormous potential of opening up our data sources to the wider community. Our digital strategy, formulated in line with the UK Open Government agenda and the Mayor of London’s open data policies, sets out our commitment to free data, updated in real time where possible, to encourage web and app developers to create the tools and services our customers want. Over 5,000 developers have already signed up to our data feeds, supporting hundreds of travel apps and helping millions of end users – all achieved at much less cost than were the same services created using TfL’s own resources, and while supporting the digital economy.

The wider economic benefits of releasing data are clear. If transport is disrupted for some reason, real-time customer information can alleviate the impact by helping people choose alternative routes. The case study presented here estimates that this benefit alone may be worth tens of millions of pounds per year. Yet even when transport is running smoothly, TfL’s journey planning information helps people select the quickest or least congested route, while recent innovations, such as our release of real time bus arrival data, enable customers to adopt a ‘just-in-time’ approach to their travel, reducing waiting time and freeing up time for more productive activities.

TfL has also exploited the opportunities of customer information, both open source and through our own channels, to tackle some of the biggest challenges a transport operator can face – from alerting people to travel hotspots during the London 2012 Games, when 60,000 people followed dedicated Games travel advice on Twitter, on top of TfL’s usual 400,000 followers, to supplying real time service updates to keep London moving during critical transport infrastructure upgrades. Moreover, it’s not just TfL’s customers that benefit. Providing high quality real time information also helps us, as a transport operator, to provide a better service through people avoiding already crowded stations or roads, and to recover quickly from disruption by giving our customers choices.

While providing substantial benefits, the continued release of data does incur some up front and ongoing management costs, so we would encourage developers using our feeds to share analytics with us on how customers use their products, to improve our ability to release the most useful data and drive the greatest value for money.

In summary, we recognise the substantial advantages resulting from open transport data, for customers, for operators, and for the wider economy. We welcome the findings of this case study into the impact for disruption mitigation, and will continue to seek ways of making our data even more accessible and useful.”

releasing transport data for London are a significant proportion of the estimates for journey time savings delivered by a major national infrastructure project such as HS2, for much lower investment of resources.

The estimated savings for HS2 include values of working time. If only the value of leisure/commuting time is included (as is the case for the calculations of the value of savings in London), the value of time saved by HS2 falls to £105 million per year. This figure is not used as it is likely that HS2 will be used more heavily for business travel, whereas London transport is more heavily used for commuting and leisure purposes. However, this illustrates that the relative significance of the savings due to release of information are likely to be greater than the 13 per cent of HS2 savings suggested above.

Information relating to transport beyond London

This case study focusses on London and specifically the information released by TfL. London provides a particularly compelling example of the benefits that can be generated by releasing information. This is partly due to the large volumes and variety of information released by TfL, and the user-friendly formats in which it is provided. It is also because London is unique in the UK both in its size as a conurbation and the scale and complexity of its transport network.

It is reasonable, however, to anticipate that benefits from the release of information in other cities in the UK also exist, either actually or potentially; as well as benefits to inter-city rail and road transport. Anecdotally, the transport systems of other major UK cities such as Manchester and Leeds are not well served by tools such as smartphone apps, either because the information has not been made available or is not in an adequate format, or because there are not a sufficiently high number of users to justify creation of an app. The latter seems unlikely, given the number of apps available and the very limited use base to which some of them cater: if nothing else, one would expect an app to be created as a service to the community.

An example of an initiative in this area is the Future City Demonstrator, for which Glasgow won £24 million of government funding early in 2013. In its press release the ODI noted that “open data will be at the heart of the programme. People will be able to monitor traffic levels on the road, before beginning their journeys. They will also be able to check whether bus and train services are running to time.”²³⁴ In many ways this is similar to the services already available in London, indicating that other cities are interesting in developing a similar ‘open data infrastructure’ to support and improve the operations of their physical transport infrastructure.

Due to smaller transport networks and fewer passenger journeys, the scale of the benefits realised are unlikely to match that experienced in London. For example, the Glasgow subway contains 15 stations and a route length of ten kilometres, just 2.5 per cent of the length of the London Underground. Nonetheless, all things being equal it is to be expected that there will be benefits in line with the scale of the transport systems involved. There will also be a broader social and environmental value to the more efficient running of these transport systems.

In addition to benefits in other cities, there are potential benefits to be realised in inter-city transport. With nearly 490 billion vehicle kilometres recorded on major roads in England and Wales in 2011, and just 81.9 per cent of journeys completed on time in the year to October 2012, there appears to be significant scope for improvements.²³⁵ The case study below highlights one example of how the release of information is capable of improving inter-city transport.

Traffic England

Traffic data is available through Traffic England, a service provided by the Highways Agency’s National Traffic Information Service. The information covers most of the motorways and major A-roads in England. Data includes real time information on traffic speed for each section or motorway or road, details of disruptions and closures, and other useful information such as weather conditions. Users are able to employ the information provided through this service to:

- monitor regular commutes;
- avoid unnecessary queues and delays;
- see how busy the roads are by viewing live traffic cameras;
- identify roadworks and whether or not they are causing delays; and

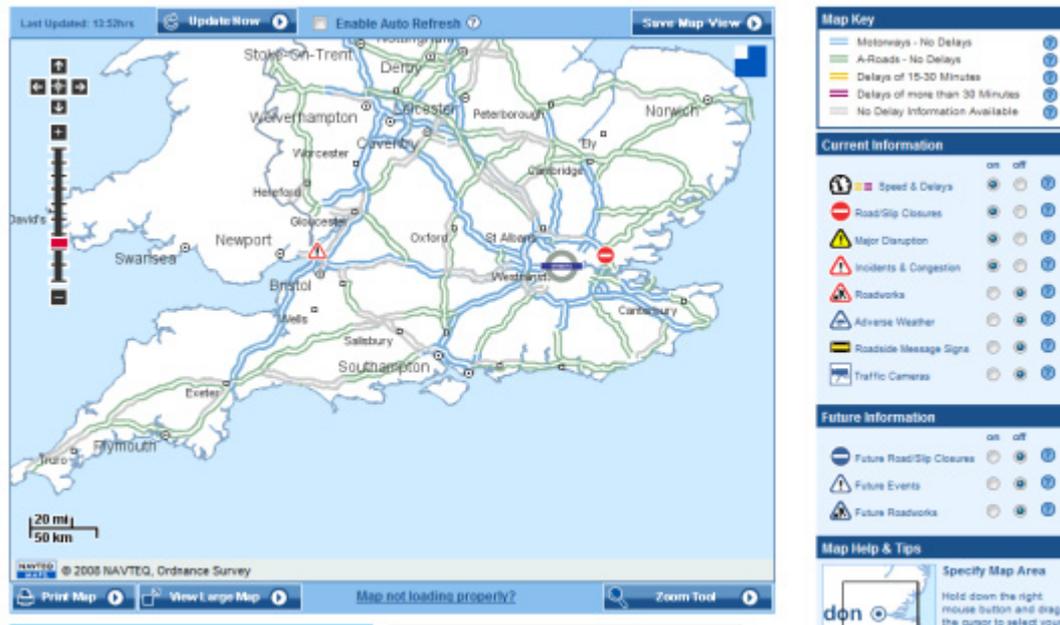
²³⁴ See www.theodi.org/news/glasgow-wins-%C2%A324m-boost-after-recognising-%E2%80%9Crevolutionary-impact%E2%80%9D-open-data

²³⁵ DfT road congestion statistics

Traffic England

- plan business trips, deliveries and holidays by checking future roadworks, events and forecast traffic conditions that are expected to cause delays.

Traffic England - Motorway Traffic Flow



Source: www.trafficengland.com/motorwayflow.aspx?ct=true#mf

Traffic England notes that the website receives around 960,000 visitors per month.²³⁶ As an indicative estimate, if 50 per cent of these monthly users were to save ten minutes of journey time, using a low value of time of £6.80 per hour (based on DfT figures), this would generate time savings worth over £544,000 per month, or over £6.5 million per year. These are very high level estimates based on broad assumptions about the level of time savings achieved by users, but provide an indication of the scale of benefits, in terms of time saved, that could be expected to be currently generated by this service.

It seems safe to assume that not all road users who could benefit from this information are currently aware of or using it. It is therefore likely that the potential or 'latent' benefit from increased exploitation of the information is even greater than the current benefit in journey time savings. In 2011 there were nearly 490 billion vehicle km travelled on major roads.²³⁷ If even one per cent of these could made more efficient or timely through the use of better information, this would improve journeys totally nearly five billion km. Assuming an average speed of 35 km/h, this would equate to 140 million hours of journey time affected. Note that this is not time saved, but serves to illustrate the potential scale of the impact of improved access to information to road users if even very conservative assumptions are applied.

In addition to the financial value of time saved, there will be other financial savings, including fuel savings, and avoidance of cost to delivery services and the businesses they serve owing to delays. There will also be wider non-financial benefits, including environmental benefits from reduced transport emissions, and benefits arising from reduced road congestion.

Further to the 'latent' benefits that could be gained from increased use of the available information under current conditions, there is likely to be an additional benefit from future use as congestion becomes more prevalent across the UK road network. The DfT estimates that in 2010 eight per cent of traffic travelled in 'very congested conditions'. It forecast that by 2035 this figure would be 17 per cent, rising to 42 per cent for London (if no further investments in road infrastructure are made beyond those set out in the 2010 Spending review). This equates to 32 lost seconds a mile for all traffic by 2035, and 140 lost seconds a mile in London.²³⁸ This increased risk of congestion and delays seems likely to increase the need of drivers for detailed and up to date

²³⁶ See www.trafficengland.com/faq.aspx#a28

²³⁷ See www.gov.uk/government/publications/road-traffic-estimates-quarter-2-2012

²³⁸ See <http://assets.dft.gov.uk/publications/road-transport-forecasts-2011/road-transport-forecasts-2011-results.pdf>

Traffic England

traffic flow information, and increase the amount of travel time that could potentially be saved through use of information.

In addition to increased exploitation of the information, it is likely that its value to road users will increase as more information is made available, and the tools for accessing and manipulating the information become more powerful. Traffic England is currently running a Traffic Map Beta which offers improved and more resilient information on speeds by drawing data from a wider range of sources, as well as extended coverage. As the information improves it is likely to generate higher levels of value for users.

One of the challenges for the release and use of information relating to the national road network is the lack of cohesion amongst the responsible bodies. The House of Commons Select Committee on Transport has summarised the organisational structure:

- the Secretary of State has responsibility for overall Government policy on roads, puts the relevant legislation in place, sets the strategic framework for new developments in traffic management, and establishes financial parameters;
- the Highways Agency is an executive agency of the Department for Transport (DfT) and, on behalf of the Secretary of State, operates, maintains and improves the strategic road network - most motorways and all-purpose trunk roads - in England;
- local highway and traffic authorities - County Councils, Metropolitan Borough Councils, Unitary Authorities, London Boroughs and Transport for London - are responsible for all other public roads (including non-trunk 'A' roads, 'B' and 'C' roads) and a small number of short, motorway standard 'A' roads in major urban areas; and
- Integrated Transport Authorities (ITAs) (which replaced the six English Passenger Transport Authorities in 2009) have full responsibility for local transport plans in their cities and can modify governance arrangements within their areas.²³⁹

Given this complexity, information is often split across different jurisdictions and may be unavailable in a consolidated form or a single location – unhelpful for road users whose journeys may well take them across several administrative boundaries. The case study below examines the role of roadworks.org in improving access to data by combining data held by multiple authorities.

Improving access to fragmented information – roadworks.org

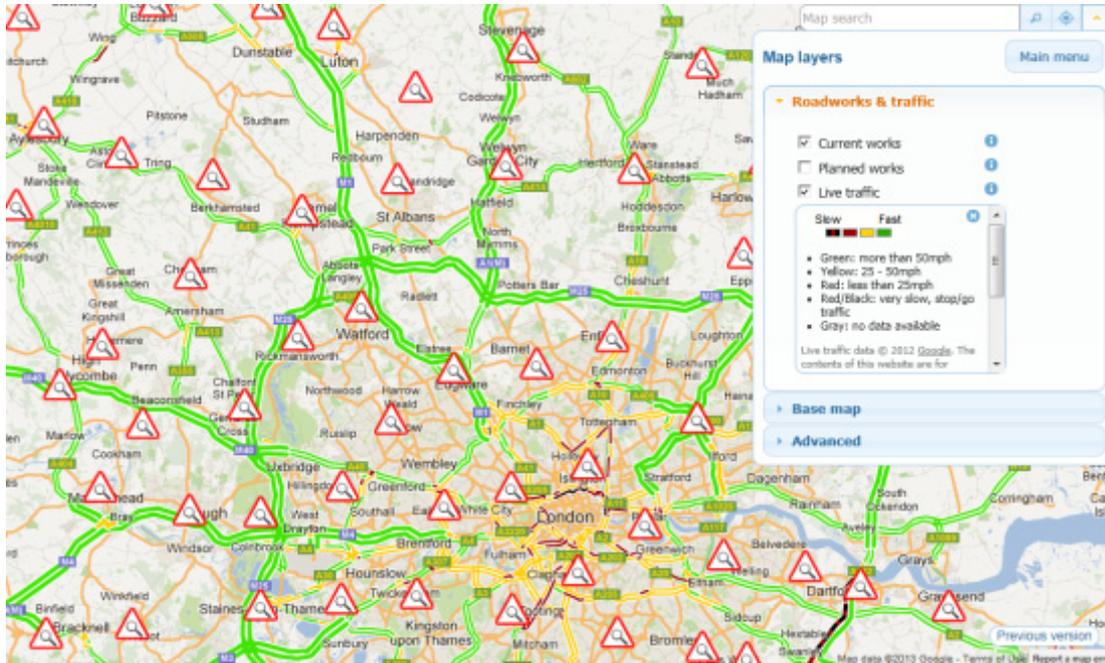
Information on when and where roadworks will take place has the potential to allow road users to plan their journeys better, in order to avoid congestion and the inconvenience and economic cost this incurs. However, in England and Wales roadworks information is split across 175 local Highways Authorities. Where it is published by local authorities this is done using a variety of bespoke platforms, which are costly to administer and cover only the area within the administrative boundaries of that local authority. The diverse formats of this disaggregated information make it difficult to incorporate into real-time systems with national coverage. This situation therefore incurs higher than necessary costs for local authorities while failing to serve the needs of road users.

Roadworks.org is an attempt to provide a solution to this problem. It currently publishes details of over two million roadworks annually, from over 140 Highway Authorities in England and Wales.

²³⁹ See www.publications.parliament.uk/pa/cm201012/cmselect/cmtran/872/87204.htm#a1

Improving access to fragmented information – roadworks.org

Roadworks.org



Source: www.roadworks.org

This aggregated national database has demonstrated significant benefits including:

- Many utility companies and their contractors, including BT Openreach, have embedded roadworks.org within their works management systems, enabling pre-coordination of planned works with those local authorities that have implemented roadworks.org. This reduces clashes and the disruption and congestion caused by works.
- Roadworks.org provides a data feed API to over 60 companies, enabling them to use up to date and broad coverage roadworks data to develop innovative services.

A recent report by ELGIN estimated the total benefits at £25 million per annum. This includes tangible benefits of £6.3 million arising from costs savings to each local authority owing to the greater efficiency of the service as compared to individual bespoke systems.

Tangible savings	Savings (£ per LA)
Better coordination	15k
Communication with stakeholders	20k
Public enquiries	7.5k
Duplication of systems	25k
Fixed Penalty Notice revenue	5k

The remaining £19 million savings are calculated as 'intangible savings' from the benefits of reduced congestion owing to better coordination of works; and greater operational efficiency for utilities companies and local authority Highway Maintenance departments. These estimates are likely to be less accurate but give an indication of the scale of the benefits achievable.

There are costs associated with this system. The paper calculated the total costs of operating the roadworks.org system to be £700,000 per year (based on 2011/12 costs), implying a subscription fee of £4,000 per local authority if all 175 local authorities were to subscribe. This is set against estimated tangible savings of £72,500 per local authority, in addition to the intangible benefits, meaning a net tangible saving of £68,500 per local Highway Authority (or £12 million p.a.).

Improving access to fragmented information – roadworks.org

This case demonstrates an important point regarding information: there is generally a cost to making it available in the most useful form for users and through an effective distribution channel. However, where the cost benefit analysis is favourable, this is an investment worth making – even before the additional downstream benefits are estimated. This case also demonstrates the role that the private sector can play in helping public sector organisations make the best possible use of their data, in order to derive the maximum benefit for the lowest cost. It is also a unique example of governance to protect the a critical national dataset which is Open but crated by private risk capital

Source: Elgin, 'A new public-private model for creating a national database of local roadworks' (March 2013)

There are also benefits to be gained from opening up data in the rail industry. There is considerable effort currently being expended to increase the availability of data in this space, as the case study below demonstrates.

Disruptive innovation in rail travel – Placr and other providers

Several companies are focussing on using information to disrupt the rail industry through innovative products and services. The objective is a liberalised data market which will allow a range of value-added services to arise, with the potential to considerably enhance the customer experience and disrupt the industry status quo – either through providing more accurate information on the nature of the service and any delays, cancellations or route changes, or by allowing customers to find more competitively priced ticket options.

It should be noted that much of the data discussed in this section may fall outside the strict definition of public sector information. The train operating companies are private, while Network Rail, as a statutory corporation, operates the physical rail infrastructure. This means that the status of data relating to privately operated trains running on a quasi-public network data is unclear. Given, however, that the rail industry receives billions of pounds of government subsidy annually, it could be argued that there is a public interest in all data relating to the rail industry being treated as public sector information. In theory, therefore, the government could make provision of information by the train operating companies as if it were public sector information a condition of operating a franchise on the network. For the purposes of this analysis we treat all information discussed as public sector information. We consider that to omit this information would be to ignore a vital part of the picture.

One company operating in this space is Placr, a start-up focussed on location and transport data, currently being incubated by the Open Data Institute. Its website states: "Our chief objective since foundation has been the creation of a single UK source of transport information by unification of timetable, live departure and disruption information for bus, rail, metro and ferry services."²⁴⁰

Barriers

Placr has identified a range of both opportunities and barriers. Industry data streams that they would like access to include:

- short term cancellations;
- rolling stock formation, i.e. how many carriages does a train contain; and
- cycle policies.

In addition, data that is made available by the Association of Trade Operating Companies (ATOC) under the Rail Settlement Plan incurs a charge for use, including a license fee of £5,005 annually for data supplied daily or weekly, and an annual quoted charge of up to £27,430 for daily fares, timetable and routeing guide data.²⁴¹ This is sufficiently high to act as a potential disincentive for a start-up that has yet to generate a reliable revenue stream. ATOC does provide trial data for proof of concept and test purposes, but this is not necessarily up to date and therefore may be inappropriate to be used for commercial purposes.

Licensing conditions for rail data are also said by some stakeholders to be confusing for developers and have been used to shut down services. For example, in 2009, in following a dispute between Kizoom and ATOC, the latter withdrew Kizoom's licence to use to use train department information for its free MyRailLite app²⁴²

²⁴⁰ See <http://placr.co.uk/history.php>

²⁴¹ See www.atoc.org/about-atoc/rail-settlement-plan/data-feeds

²⁴² See <https://mocko.org.uk/b/2010/10/29/national-rail-have-killed-my-train-times-app/>

Disruptive innovation in rail travel – Placr and other providers

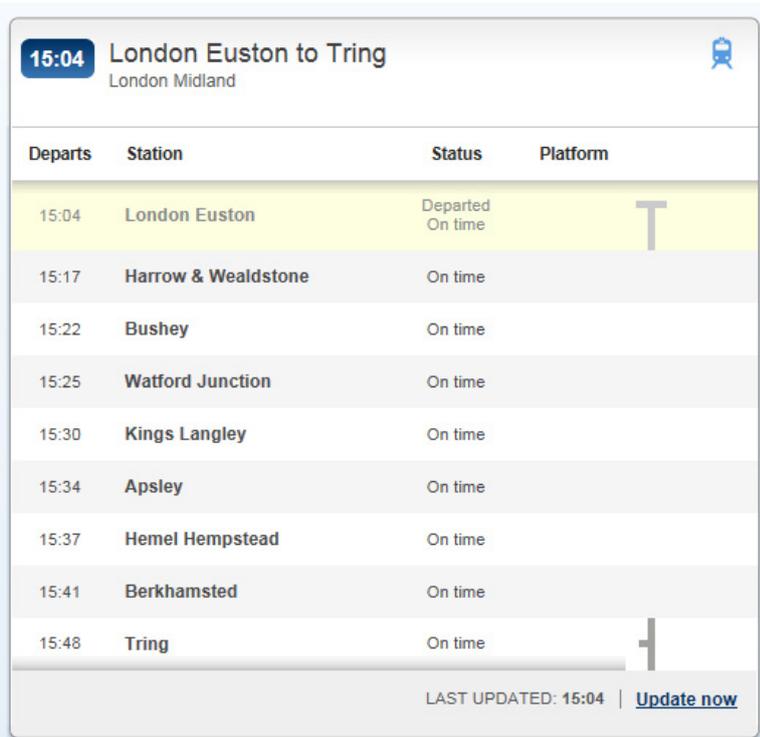
Developers may therefore face some challenges in accessing data in the rail industry.

In terms of the opportunity, Placr identify three areas in which they believe increased availability of data can disrupt the rail industry and improve the experience of passengers.

Providing live service updates

Increased availability of data has the potential to allow passengers current information on the status of rail services. Some of this information is already available through the National Rail website, TOC websites and mobile apps, which provides live departure boards and allows users to track a train's progress throughout its journey.

National Rail Enquiries



Departs	Station	Status	Platform
15:04	London Euston	Departed On time	T
15:17	Harrow & Wealdstone	On time	
15:22	Bushey	On time	
15:25	Watford Junction	On time	
15:30	Kings Langley	On time	
15:34	Apsley	On time	
15:37	Hemel Hempstead	On time	
15:41	Berkhamsted	On time	
15:48	Tring	On time	T

LAST UPDATED: 15:04 | [Update now](#)

Source: <http://nationalrail.co.uk/times/fares/ldb/>

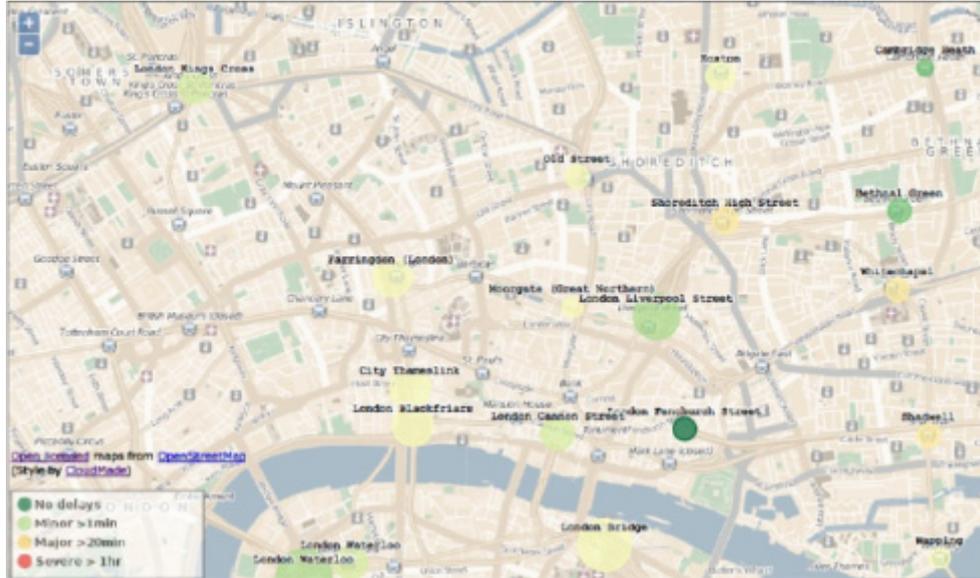
There are other, potentially more interesting ways of tracking service quality. For example, the train lateness map available at transportapi.com represents 'lateness' as coloured bubbles on a map, allowing users to intuitively see where the delays are and plan their journey using the best-performing route.

²⁴³ See www.whatdotheyknow.com/request/atoc_written_advice_to_dft_on_im

²⁴⁴ For example, see <http://www.dailymail.co.uk/news/article-2271415/Earl-Attlee-defends-train-fares-admitting-expensive-forced-cancel-day-motor-show.html>

Disruptive innovation in rail travel – Placr and other providers

Train lateness map



Source: www.scribd.com/doc/123365071/Friday-Lunchtime-Lectures-at-the-ODI-How-can-Open-Data-Revolutionise-your-Rail-Travel

By presenting the data in innovative formats, service users are able to interact with it in new ways and gain additional insight into the performance of travel services and the best way to complete their journey. Travel becomes more intuitive and users gain enhanced understanding of the travel options at their disposal, which should translate into faster and more efficient journeys.

Generating financial savings for passengers

It is hoped that fare data will soon be made open. This data, and the innovative services that arise from it, is likely to be of benefit to passengers. Part of the potential value of this data to passengers has been demonstrated by the example of TfL, which now allows holders of a registered Oyster Card to view their journeys online, and also offers to send monthly statements detailing all journeys and credit top-ups during the period. This gives users the ability to track their expenditure patterns on public transport, and also allows them to identify any journeys where they believe they have been wrongly charged and apply for a refund. The data is therefore of clear benefit to users.

New and potentially disruptive services may also be built on this data. An example of using data to generate savings for travellers is Tickety Split, a service run by Money Saving Expert. This service is based on the insight that buying separate tickets for constituent parts of a journey is often cheaper than buying one ticket. Although the journey may take longer and require changing trains, it may be that the traveller values the financial saving more highly than the opportunity cost of additional time spent on the journey. Train operating companies do not generally offer their customers this choice, and searching manually for the best split ticket combinations can be cumbersome. Tickety Split searches available fare data to calculate the cheapest combination. In the example below, a ticket from London to Leeds costs £97, but this can be reduced by £13.30 to £83.70 if intermediate tickets are bought from London to Grantham and from Grantham to Leeds.

Disruptive innovation in rail travel – Placr and other providers

Tickety Split

The screenshot shows the Tickety Split interface. At the top, it displays the standard rate of £97.00 for a journey from London Kings Cross to Leeds. Below this, it states that the current offer is only for walk-on fares and that booking in advance is usually cheaper. The main offer is to split the journey into two parts: London Kings Cross to Grantham for £41.90 and Grantham to Leeds for £41.80, totaling £83.70, which saves £13.30 compared to the standard rate. The ticket types are identified as OFF-PEAK DAY 5 and SUPER OFFPEAK 5.

Split 1	
From	Start
London Kings Cross	13:35
To	Price
Grantham	£41.90
Ticket type: OFF-PEAK DAY 5	

Split 2	
From	Start
Grantham	14:40
To	Price
Leeds	£41.80
Ticket type: SUPER OFFPEAK 5	

Source: <http://splitticket.moneysavingexpert.com/>

Information on the impact of fare-splitting on revenues is not publically available.²⁴³ It is, however, possible to roughly estimate the scale of the savings that might be achievable through this technique, based on the following:

- franchised passenger revenue in 2011/12 was £7.229 billion;
- around 75 per cent of this revenue was generated by ordinary fares (assuming that season ticket holders are unable to secure savings through ticket splitting); and
- assuming an average saving through ticket splitting of 13 per cent.

Based on these assumptions, passengers could save an estimated £705 million per year through ticket splitting. The full extent of these estimated potential savings are unlikely to be realised, due to varying levels of awareness and adoption among passengers. Even if only 25 per cent of passengers were to take advantage of split ticketing, this would still equate to **an estimated £176 million saving per year**. This indicates that increasing accessibility of fare data holds large potential benefits both for individual passengers and for the travelling public overall. In addition to the financial gains to passengers, this example also demonstrates the ways on which data can be used to empower consumers. In so doing it is likely to induce train operating companies to run more efficient, consumer-focussed operations and to offer more competitive fare pricing.

Future releases of data could offer the potential for additional innovative services which would increase potential savings to passengers. For example, the data from a smart ticketing system such as Oyster could be linked to real time data on train movements to provide a service whereby automatic refund claims are filed on behalf of passengers when their trains are delayed. This could have a genuinely disruptive effect on the industry, forcing train operating companies to focus on delivering a reliable service, as a poor level of service would have a direct impact on revenues. Were such a service to be developed, it would depend on access to data that can be used to monitor reliability.

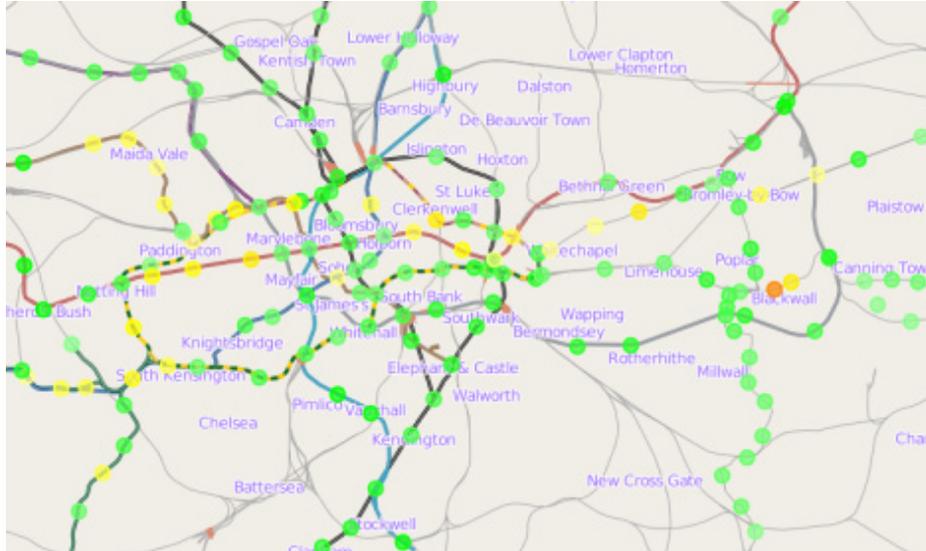
Creating independent streams of data about rail network performance

Many of the benefits to users discussed above will depend on data that can provide an independent view of the performance of the railways. For example, Tube Radar (based on Open Street Maps) provides a view of performance (time intervals between trains) across the network, as compared to normal performance. Placr also cite the example of being asked by the press to offer a view on levels of service during a recent strike, in

Disruptive innovation in rail travel – Placr and other providers

the face of conflicting reports from TfL and the unions.

Tube Radar



Source: <http://tube-radar.com/>

One source of this data is smartphones carried by passengers, which can be used to monitor the progress, frequency, acceleration/deceleration and other details regarding train operations. By crowd sourcing this data, an additional window on the performance of the rail network is opened up, increasing pressure on train operating companies to improve standards and provide a high quality service to their customers.

Conclusion

Placr and other start-ups operating in this space offer numerous ideas for products and services that could collectively transform the experience of rail travel in the UK, by empowering passengers and enabling them to hold rail operating companies to account. As average rail fares rise year on year, with many users perceiving rail travel as offering poor value for money,²⁴⁴ data offers a mechanism ensuring that service users and taxpayers receive value for money from train operators.

Some of the data required to drive this transformation remains either unavailable or subject to restrictive pricing and licensing conditions, although there are signs that this situation may gradually be changing. There is demonstrable value to be generated from releasing this data, both in terms of financial savings to users, convenience in journey planning, and in driving accountability (and from thence efficiency) in the rail industry. While not strictly public sector information, given the level of taxpayer contribution to the rail industry and its status as infrastructure of national importance, there is a clear responsibility for government to ensure that the industry operates in an open and accountable manner. This seems likely to be an area in which increasing the openness of data will generate rapid and measurable returns to rail service users, taxpayers and the wider economy.

Source: much of the content of this section is derived from a session at the ODI, "How can open data revolutionise your rail travel?" (01/02/2013). Slides available at <http://www.scribd.com/doc/123365071/Friday-Lunchtime-Lectures-at-the-ODI-How-can-Open-Data-Revolutionise-your-Rail-Travel>

Figure A6.7: additional data 1

Syndicated feeds available to developers through the TfL website

Syndicated feeds available to developers through the TfL website

The full list of syndicated feeds offered by TfL is included below. TfL state “before we give permission to use any feeds, we need to know how they will be used, where they will be used and how many people are likely to view them.” Gaining access to these feeds requires:

- Providing personal and/or company contact details
- Providing information on intended use, target audience and estimated audience numbers
- Agreeing to the terms and conditions

The full list of feeds available is provided below (correct as of January 2013).

- Live traffic camera images v2
- Live bus arrivals API (instant)
- Source London Charge Point data dictionary
- Journey Planner API Beta
- Tube station accessibility data
- Rolling origin and destination survey
- River services timetable
- Tube departure boards, line and station status
- Journey Planner Timetables
- Coach Parking sites/timetables
- Licensed private hire operators – Find-a-ride
- Pier locations
- Bus routes
- Tube – this weekend
- Station facilities
- Live bus arrivals API (stream)
- Live Roadside Message Signs v2
- Source London Charge Point Location Data
- Barclays Cycle Hire availability
- London Underground passenger counts
- Public transport accessibility levels
- Barclays Cycle Hire statistics
- Oyster card journey information
- Live Traffic Disruptions
- Dial a Ride statistics
- Station locations
- Oyster Ticket Stop locations
- Bus stop locations
- Tube – this weekend v2

Source: www.tfl.gov.uk/businessandpartners/syndication/16492.aspx

Figure 6.8: additional data 2

DfT value of working time per person (£ per hour)		
Vehicle occupant	Market price (2010 prices and values)	Market price (2012 prices and values – Deloitte analysis)
Car driver	33.74	35.4
Car passenger	24.17	25.4
LGV (driver or passenger)	13.00	13.7
OGV (driver or passenger)	13.00	13.7
PSV driver	13.00	13.7
PSV passenger	25.81	27.1
Taxi driver	12.47	13.1
Taxi/minicab passenger	57.06	59.9
Rail passenger	47.18	49.6
Underground passenger	45.90	48.2
Walker	37.83	39.7
Cyclist	21.70	22.8
Motorcyclist	30.53	32.1
Average of all working persons	34.12	35.8

Figure A6.9: additional data 3

Appendix 3 – DfT value of non-working time per person (£ per hour)		
Purpose	Market price (2010 prices and values)	Market price (2012 prices and values – Deloitte analysis)
Commuting	6.46	6.80
Other	5.71	6.00

Source: Department for Transport, 'TAG Unit 3.5.6: Values of Time and Vehicle Operating Costs', October 2012

Appendix 7: Further case studies

This appendix contains further details on the healthcare and public sector case studies outlined in Chapter 5.

The healthcare and life science sector

This section considers examples of how the release of public sector information datasets relating to the healthcare sector can have a beneficial impact. This may be in terms of identifying new efficiencies in the NHS, or it may be in improving patient outcomes, which may itself result in cost savings either directly to the health service or to the wider economy.

Mastodon C – identifying NHS prescription savings from big data

Case study summary

By using data on prescribing practice across England, variations in spending on different classes of drugs can be identified. It is then possible to calculate the potential savings to be achieved by moving from prescribing branded to generic drugs.

Size of the prize

- For statins alone, the NHS could save around **£200 million per year** by reducing prescriptions of branded in favour of generic versions.
- When extended to all classes of drugs, the total potential savings could amount to **£1.4 billion per year**.

NHS prescribing data is released at GP level through the Health and Social Care Information Centre. These are large datasets, with around 10 million lines of data released every month – an example of ‘big data.’ Mastodon C, a start-up company currently being incubated by the Open Data Institute and describing itself as an ‘agile big data specialist’, saw an opportunity in this data to identify potential efficiency savings for the NHS.²⁴⁵

Prior to this analysis there was already an awareness in the sector of the possibility of achieving savings through changes to prescribing practices. The British Medical Journal has published research indicating that the potential savings to the NHS of switching from branded to cheaper (but in many cases equally effective) generic drugs could total £1.4 billion²⁴⁶. There have been previous and ongoing attempts to achieve savings through increasing prescription of generic drugs, notably through the work of prescribing advisors. However, these attempts have hitherto met with limited success. GPs may habitually prescribe a branded drug without considering the cost implications; equally, patients may be accustomed to a branded version and feel that a generic is an inferior substitute. For these and other reasons, it has proved difficult to bring about significant change in behaviour.

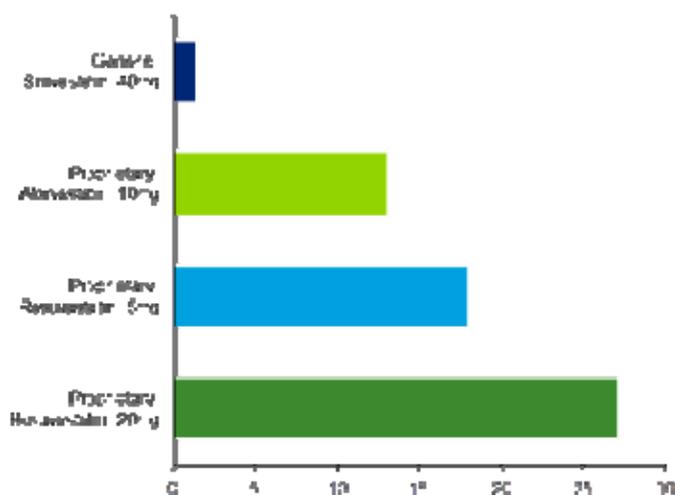
Mastodon C, working with Open Healthcare UK, wanted to try a new approach to raising awareness. They used big data to demonstrate regional differences in the cost of prescriptions, hoping thereby to drive change by allowing GPs and PCTs to compare their performance to those of GP practices and PCTs across England. By highlighting seemingly unwarranted variation, GPs and PCTs could identify where they were underperforming compared to their peers in terms of keeping prescribing costs as low as possible.

²⁴⁵ Source: Mastodon C presentation at the ODI, 18th January 2013. Slides available at www.scribd.com/doc/122981341/Friday-Lunchtime-Lectures-at-The-ODI-Big-Data-Comes-to-the-NHS. See also <http://blog.mastodonc.com/> and <http://prescribinganalytics.com/>

²⁴⁶ See www.bmj.com/content/341/bmj.c6449

The team chose statins as the focus of their project, because there is widespread agreement that the generic version is in most instances as effective as the branded versions (based on the guidelines provided by NICE). The price difference is also significant: £1.30 for a generic version as opposed to £20 or more for many branded statins, as shown in the Figure A7.1.

Figure A7.1: prices of generic and proprietary statins, £ (prior to June 2012)²⁴⁷



Source: www.scribd.com/doc/122981341/Friday-Lunchtime-Lectures-at-The-ODI-Big-Data-Comes-to-the-NHS

The team's decision highlights an important point: users of big data and public sector information need to approach the analysis responsibly and in an informed way. The Mastodon C team maintains that data itself is not a silver bullet to understanding a problem. The team needed to fully understand the clinical guidance attached to each drug, so as to avoid, for example, recommending savings where there were strong clinical reasons for choosing a branded drug rather than the generic version. In their own words, "good domain knowledge usually beats super-smart algorithms." In this case, while the data itself was crucial, the insights became possible only through working closely with GPs and other healthcare professionals.

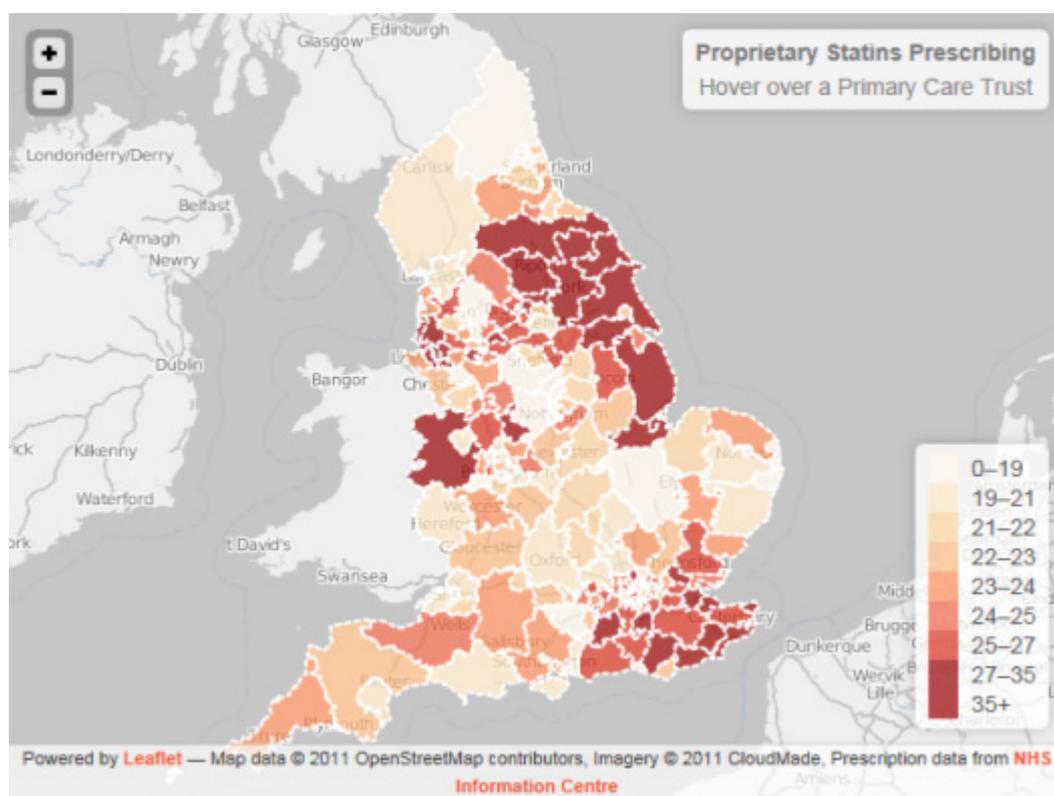
This issue of responsible use of the data also applied to the level of granularity of the analysis. Although the data would allow comparisons to be made at a GP practice level, the Mastodon C team opted to map their findings at a PCT/CCG²⁴⁸ level. This decision arose out of concerns that presenting the results at a GP level could lead to distortions in the data involving very small practices, as these might handle only a few relevant cases. This could mean that their prescribing costs cost appear exceptionally high. The team were concerned that their work could lead to a GP practice being unjustly labelled 'the worst in Britain' in the media. This might then create a backlash and build resistance to future analyses and releases of data among the healthcare profession.

By comparing prescribing practice across England, the Mastodon C team identified around £200 million worth of potential savings. These could be achieved if the behaviour of the GPs with the highest level of branded prescriptions was brought into line with the behaviour of the GPs prescribing the highest level of generic versions. The mapping of results by PCT revealed notable differences between PCTs across England.

²⁴⁷ Note: these prices have subsequently changed as Atorvastatin came off patent in June 2012 and the price has now dropped to £1.25 for 10mg: www.ppa.org.uk/ppa/edt_intro.htm. Prescribing Analytics reports that this was accounted for in the analysis – see <http://prescribinganalytics.com/analysis>

²⁴⁸ Primary Care Trusts (PCTs) commission primary, community and secondary care from providers. Clinical Commissioning Groups (CCGs) are the bodies that are to assume most of the commissioning responsibilities of PCTs under the Health and Social Care Act 2012.

Figure A7.2: percentage of proprietary statin prescribing by CCG Sept 2011 – May 2012



Source: <http://prescribinganalytics.com/>

The team also identified some potentially effective levers of influence over prescribing practice. Cambridge PCT had very low prescribing costs for statins: the team found that this was because the PCT had taken a very strong line on prescribing the generic version, and this had changed behaviour at the GP practice level. The achievement of considerable cost savings in Cambridgeshire demonstrates that action at the PCT level may be the most effective approach to achieving savings through changes in GP prescribing behaviour.

It should be noted that work in this field is already undertaken by the government. For example, the NHS Prescription Services, part of the NHS Business Services Authority, provides the NHS with a range of drug, financial and prescribing information. The IT tools allow NHS organisations to look at prescribing patterns for a range of medicines right down to individual GP level, so they can target work with practices to help them improve. At a local level, NHS organisations have developed systems to incentivise prescribers to prescribe more cost effectively. Many local NHS organisations have also invested in software which prompts prescribers on the most cost effective prescribing choice.²⁴⁹ These tools have helped the NHS to achieve one of the highest rates of generic prescribing in Europe, with the overall prescribed generic rate calculated at 83 per cent in 2011.²⁵⁰

This highlights that in some instances where data is released, there may have been previous and ongoing efforts by public sector organisations to use this data to develop insights. The work of private sector organisations may be able to build on this work and enhance the value derived from

²⁴⁹ Based on Department of Health comments on an earlier draft

²⁵⁰ NHS Information Centre, *Prescriptions Dispensed in the Community: England, Statistics for 2001 to 2011* (July 2012)

the data, but this is likely to be most effective, and avoid duplication, where communications between public and private sector organisations are strong.

In summary, the work of Mastodon C in this area demonstrates a number of key insights into the value that may be derived from making more and better use of public sector information in the health sector. These insights include:

- it is not generally enough to have access to data – also required is the contextual and sector knowledge to use it intelligently and responsibly;
- large efficiency savings may be hidden in areas of regular practice. There are likely to be other areas both in the NHS and the wider public sector where changes in behaviour could generate significant financial savings without compromising outcomes;
- the insights from data can help drive change in areas that previously resisted policy solutions. For example, in this case a key insight was that change in GP prescribing behaviour might be best driven from the PCT level;
- there is a need for communication between public and private sector organisations to ensure that where data is released, the analysis undertaken is complementary and builds on previous and ongoing efforts to use the data to gain insights.

A patient database for the NHS – the challenges to extracting value from large datasets

Case study summary:

A central NHS patient database could offer significant savings to the NHS as well as improving standards of care and the patient experience. However, there are significant technical, privacy and cultural hurdles to overcome if this is to be made a reality. This case illustrates both how attractive the prize of harnessing the power of large public sector information datasets can be, but also the difficulties these can present.

Size of the prize:

Difficult to estimate, but if the system is delivered as planned the savings are likely to be billions of pounds. An initial illustrative estimate identified **£4.4 billion of potential savings**, although these are not all directly related to the patient database.

The idea of a central NHS patient database has existed for some time, but previous attempts have encountered a variety of issues owing to the challenges of building a system with the requisite capabilities and scale. However, early in 2013 the Health Secretary, Jeremy Hunt, launched a plan to store patient records in a cloud-based system by 2018, with the ultimate goal of transitioning towards a 'paperless NHS'.²⁵¹

While the details have not been fully established, the broad outline of the scheme is that each patient would have an individual electronic health record which would be accessible from any point within the NHS – whether by the patient's GP during a consultation, by a hospital consultant preparing to carry out an operation, or an ambulance crew responding to an emergency call-out.

Outwardly at least these proposals have much to recommend them, and appear to be a good example of an opportunity to achieve increased efficiency and improved outcomes through increased exploitation of public sector information. Patient records offer a rich source of data with the potential to both improve the standard of care received by the patient, and to enhance the efficiency and effectiveness of the NHS. However, despite this promise, some have viewed the proposals through the lens of past failures of NHS database projects, as well as privacy and confidentiality concerns.

²⁵¹ Jeremy Hunt's announcement (16th January 2013) is available to view at: <http://mediacentre.dh.gov.uk/2013/01/16/16-january-2013-jeremy-hunt-policy-exchange-from-notepad-to-ipad-technology-and-the-nhs/>

Commentators have also raised concerns that this database may threaten patient confidentiality. In part this is simply a technical issue: a cloud-based, highly networked system will need to be rendered secure against unwanted incursions by non-authorised agents, given that much of the information held will be of a highly personal and sensitive nature.

If the information is released, either to approved users as is currently the case with the National Pupil Database, there will be additional concerns regarding confidentiality²⁵². As discussed below, it may be possible to anonymise datasets to an extent that the level of risk of an individual record becoming known is acceptable. However, healthcare information may pose particular challenges. In the case of very rare diseases where, for example, there are only one or two cases per GP practice area - it may be impossible to effectively anonymise the data. This indicates that the approach to protecting patient privacy will need to be rigorous and adapted to the constraints of the data.

In his announcement the Health Secretary acknowledged these concerns, but maintained that the Government had learnt from past failures and would not be adopting a top-down approach in an attempt to construct a monolithic central IT infrastructure. Instead of a centralised approach akin to constructing an aircraft carrier, he argued that “most systems won’t necessarily need to be replaced, just updated or adapted so they can talk with each other. A thousand different local solutions linking together using common standards.”²⁵³

The announcement argued that this would deliver considerable financial benefits, based on a report which identified potential savings amounting to £4.4 billion²⁵⁴. It should be noted that these benefits were calculated from a wide range of changes which were estimated to deliver a range of incremental benefits.

In addition, the Health Secretary’s announcement identified broader welfare benefits that include improved outcomes, ultimately meaning lives saved, as well as an improved user experience of engaging with the healthcare system at all levels. For example, if staff in Accident & Emergency have details of a patient’s health history and current medication, they should be able to respond more appropriately to the emergency, with potentially life-saving consequences. Equally, if clinicians have access to up to date health records for all patients this means that patients would not have to verbally repeat their medical history at each stage of the process. This is likely to save time and ensure that clinicians have access to accurate records.

There is some international precedent for a centralised patient database. The most notable example is Denmark, which currently makes hospital records available to patients online, and which is in the process of making GP records available. However, Denmark has a population numbering just ten per cent of the population of England alone and a population that is, *prima facie*, more homogenous, meaning that a patient database in the UK will need to operate on a much larger scale. This makes it a more challenging technical and logistical proposition.

This case throws into sharp relief the tensions around using large datasets concerning individuals. On the one hand the potential for a range of benefits, from efficiency savings to enhanced user convenience and new insights, is extremely tempting. On the other hand there are valid concerns around privacy and confidentiality and the potential misuse of sensitive information. Effectively harnessing the potential of data to transform the NHS, as well as other areas of the public sector, will require these tensions to be reconciled in a manner which satisfies security and privacy concerns while nonetheless permitting effective use of the data.

²⁵² Needless to say, where release of personal information outside the NHS is intended, full consent of the patient(s) concerned will be needed. Without such safeguards, it will be difficult to secure clinician and patient support for such a database. This is particularly the case given that the information could affect the ability of patients to secure health insurance, employment, and subject them to marketing attention, if released.

²⁵³ <http://mediacentre.dh.gov.uk/2013/01/16/16-january-2013-jeremy-hunt-policy-exchange-from-notepad-to-ipad-technology-and-the-nhs/>

²⁵⁴ This can be accessed at www.wp.dh.gov.uk/publications/files/2013/01/Review-of-use-of-Information-and-Technology.pdf

Publication of mortality rates following cardiac surgery

Case study summary:

Publishing data on mortality rates following adult cardiac surgery appears to be associated with a decline in mortality. There are various theories as to why this is the case, including competitiveness among surgeons which leads to a rise in performance, increased awareness among healthcare professionals, and public pressure for higher standards. However, there are also concerns that the apparent decline in mortality may reflect 'gaming' of the mortality data.

Size of the prize:

A decline in mortality rates is itself desirable. The economic value depends on the value attributed to a statistical life. Taking a median value from a range of recent studies, the value of lives saved among those undergoing adult coronary artery surgery in NHS centres in north-west England in 2005 was around **£55 million**. This suggests that the total value of lives saved per annum could **exceed £400 million** for England and Wales, if similar benefits were observed in all regions.

Since 2005, data on mortality rates following adult cardiac surgery has been made publically available in the UK. Subsequently, studies have attempted to gauge the impact that the release of this data has had, both on outcomes of surgery and the willingness of surgeons to accept high-risk cases.

There is a reasonable volume of evidence suggesting that publication of this data is associated with a decline in mortality. A 2007 research paper published in the journal *Heart* found that over an eight year period (from 1997 to 2005) observed mortality decreased from 2.4 per cent to 1.8 per cent. The authors argue that while data was not made publically available over the majority of the period covered by the study, over this period it became clear to surgeons that public release of the data was a matter of time, following the Bristol Public Enquiry in 2001.²⁵⁵ The findings of the paper, moreover, appear to be corroborated by other studies, including a study which found a 41 per cent reduction in risk-adjusted mortality rates during the first four years following publication of outcomes data.²⁵⁶

An improvement in clinical outcomes that leads to an increase in lives saved is of itself desirable from a wider social welfare perspective. The economic value attributed to this decline in mortality from surgery depends on the value of a statistical life, which varies widely. To take an example, drawing on a survey of recent studies²⁵⁷ produces a median value for the total value of a statistical lifetime of around £3 million. Using this value means that the decline in mortality can be valued at around £55 million, just for north-west England in 2005. Assuming similar benefits were experienced across England and Wales, the total value of lives saved in 2005 alone is estimated to have exceeded £400 million, as compared with 1997 mortality rates. If similar benefits can be realised in other areas of clinical practice, the annual benefits in terms of lives saved are likely to be valued in terms of billions of pounds.

There have, however, been fears that public disclosure of outcomes could trigger risk-averse behaviour among surgeons, meaning that they would not accept cases where there was an increased risk of mortality.²⁵⁸ There is little evidence that this has in fact occurred, with the *Heart* study finding that the number of patients classified as high-risk actually increased over the period. However, a response to the article argued that surgeons have a powerful incentive to 'game' the system by over-assessing the risk profile of patients, thereby making their own performance

²⁵⁵ See www.ncbi.nlm.nih.gov/pmc/articles/PMC1955202/#ref16

²⁵⁶ See Hannan E L, Kilburn H, Jr, Racz M. *et al Improving the outcomes of coronary artery bypass surgery in New York State*. JAMA 1994

²⁵⁷ See The Government Office for Science, 'Reducing Risks of Future Disasters: Priorities for Decision Makers' (2012)

²⁵⁸ www.ncbi.nlm.nih.gov/pmc/articles/PMC1955009/

appear stronger.²⁵⁹ There is some evidence that similar effects may have occurred elsewhere, with one survey cited in the *Heart* study finding that 79 per cent of New York cardiologists reported that publication of mortality statistics had influenced their decision about whether to perform angioplasty on individual patients.²⁶⁰

These concerns notwithstanding, the evidence seems to indicate that transparency can be a powerful driver of accountability and improved standards in the healthcare sector. As the authors of the *Heart* paper suggest, “if public disclosure can drive data collection and analysis, but does not

create significant risk-averse behaviour, its introduction may be beneficial in other areas of

medicine.”²⁶¹ Releasing data on standards, where appropriate and subject to the appropriate monitoring mechanisms, therefore appears to be an ‘easy win’ for the use of public sector information to drive tangible economic and welfare benefits.

The public sector

One of the most interesting areas for the improved exploitation of public sector information is greater sharing of and access to information within the public sector itself. By removing barriers to the flow of information between public sector bodies – either because potential users are unaware of what information is available, or because there are physical or legal constraints on the sharing of this information – policy formation could be based on much richer and more complete information. All things being equal, this should lead to better informed and therefore more effective policies.

A report by the Administrative Data Taskforce published in December 2012 highlighted a number of areas of research and policy where sharing of data could generate new insights and lead to improved policy outcomes. These are summarised below.

Area of research and policy	Description
Social mobility	Linking data on education, training, employment, unemployment, income and benefits
Causal pathways over the life course	Linking data on education, health, employment, income and wealth
Support for the elderly	Comparative analysis of access to and provision of social care support for the elderly
Poverty	Linking data on housing conditions, health, incomes and benefits
Social care for children	Linking indicators of parental employment, social background and childcare

²⁵⁹ Ibid.

²⁶⁰ See Hannan E L, Siu A L, Kumar D. *et al Assessment of coronary artery bypass graft surgery performance in New York. Is there a bias against taking high-risk patients?* Med Care 1997

²⁶¹ See www.ncbi.nlm.nih.gov/pmc/articles/PMC1955202/#ref16

Area of research and policy	Description
Offence and re-offence	Linking data on offending and re-offending behaviour, income, benefits, health and mental health

Source: *The UK Administrative Data Research Network: Improving Access for Research and Policy* (December 2012)

This list serves as an indication of the many policy areas which could benefit from increased sharing of information between government departments. Given the level of government spending, and the wider economic significance of these policy areas, if sharing leads to enhancement of policy in any one of these areas the economic and broader welfare impacts should be large. Below is an example of one area in which progress is already being made in the sharing of information.

Announced in May 2012, the Social Mobility Transparency Board is tasked with pursuing “smarter use of data between the Department for Education, the Department for Business, Innovation and Skills (BIS) and HM Revenue and Customs.”²⁶² The role of the Board is to improve sharing between key information holders and external researchers, building connections and establishing procedures to ease access to data for research.

The value of public sector information to social mobility researchers

From conversations with key stakeholders it is clear that public sector information is crucial for researchers focussing on social mobility. Hitherto, most work in this field has been built upon birth cohort studies which follow a group of subjects and update key indicators every ten years.

Data held by public sector bodies holds the promise of far richer and more robust insights into key indicators of social mobility, potentially extending over the entire population. For researchers to use this data effectively, however, it needs to be shared and linked across the various information holders.

The organisations so far identified as holding the information most useful to researchers in this field are:

- the Department for Education;
- the Higher Education Statistics Agency;
- the Department of Work and Pensions; and
- Her Majesty’s Revenue and Customs.

This being a relatively recent initiative, tangible benefits are still emerging. However, stakeholders point to a number of incremental benefits. There is steady improvement in the level of understanding of the drivers of social mobility, especially how higher education works as a driver of social mobility and understanding of the factors driving participation in higher education. Insights into the role of socio-economic group, school type, family income and other factors in influencing participation in higher education continue to deepen as a result of increasing access to data.

Opportunities in the local public sector

The local public sector spends around £70 billion per year and employs approximately two million people, providing many of the essential services across the UK. It therefore seems reasonable to expect that there will be many opportunities to realise efficiencies and improved outcomes through the more effective exploitation and sharing of information.

The local public sector also poses unique challenges owing to its great diversity. It includes county and district councils (in two-tier ‘shires’), unitary authorities, London and metropolitan boroughs, and *sui generis* authorities such as the City of London and Isles of Scilly. These administer a

²⁶² See *Open Data White Paper*, June 2012

diverse range of services under a variety of delivery models. Any changes in the use of public sector information at a local level is therefore likely to be incremental and involve, at least initially, either individual local authorities or small groups working together.

This diversity, however, is also a significant strength, since it makes local authorities laboratories for the use and re-use of information. In the course of our research Deloitte has encountered many examples of local authorities exploiting information in ways that are often highly innovative, with lessons not only for other local authorities but also for central government departments and other public sector bodies. The individual efficiency gains and improvements in outcomes may in many cases be small, but collectively they represent a quiet revolution in the delivery of services at a local level. Should these experiments grow in number and the best examples become widely adopted, the national impact could be significant.

There are also numerous other examples of local authorities using information in innovative ways, including:

- Lambeth Borough Council have set up a site called “Lambeth in Numbers” to help inform their Food Strategy work. It brings together data from various sources including central and local government on a map.
- Bristol City Council Air Quality data
- Trafford Council – Breakthrough Fund on stimulating data sets
- Hampshire County Council work on land supply

The example below illustrates an example of improvements in policy formation and delivery through information sharing between local authorities and other public bodies.

An example of data sharing at a local level – families with complex needs

Examples of data sharing between local councils and other public bodies are proliferating across the UK. These are often responses to the twin pressures of deep funding cuts and intractable problems involving multiple agencies.

An example of this sort of problem is the case of families with complex needs, often referred to in the press as ‘troubled families’. These families may combine issues such as mental health problems, children out of school, and long term worklessness and benefit dependency, meaning that they fall within the remit of multiple public sector bodies including social services, the Police, and the local and national welfare services. These services are estimated to cost around £75,000 per family per year.²⁶³

Increased coordination between local councils and other local partners may prove both more cost efficient and more effective in resolving the problems faced by such families. The councils of Greater Manchester, Leicestershire and Bradford are working together to improve information sharing and management in this area.

The project aims to develop a single toolkit for information sharing, combining existing guidance and approaches. This is intended to be applicable to both the councils and the agencies working with families with complex needs.

In addition the project is intended to lay the foundations for a culture more conducive to information sharing, addressing issues such as different professional cultures of sharing, lack of training and expertise, and differing interpretations of legislation.

In practice, the steps needed to achieve this can appear prosaic but are nonetheless potentially powerful enablers of an environment in which information can more easily be shared within and between organisations.

As the project is still ongoing it is too early to judge its effects, whether in terms of reduced costs or

²⁶³ Association of Greater Manchester Authorities, ‘Improving Information Sharing and Management: A National Exemplar Project’

An example of data sharing at a local level – families with complex needs

improved outcomes. Nonetheless, this appears to be a positive example of the potential of greater sharing and exploitation of data between councils and other public sector bodies in response to a policy problem which has proved unresponsive to a siloed approach.

The Local Government Transparency Survey

The Local Government Association (LGA) conducted a survey in September 2012²⁶⁴ covering the publication of open data, the impact of open data on councils and how data is used locally. 37 per cent of local authorities responded to the survey. In their responses they identified the ways in which they are using open data, which give an indication of the impact it is having. A selection of these responses is included below.

- engaging with community groups to create API makers and ways to make data useful;
- using their own open data and that of other councils to gain insight into the characteristics of people in the area and, accordingly, their needs in relation to the services the council currently and intends to provide or commission;
- planning to utilise the Police, NHS, public health and other public sector partners' open data to produce a single view of the borough;
- seeing a reduction in FOI requests;
- using data to help customers 'self-serve' online in their reporting e.g. fly tipping, tree issues, damage to street furniture; and
- undertaking service reviews and improvement – benchmarking, performance, spend, organisational structures and pay scales.

These responses indicate the diversity of ways in which local authorities are making use of public sector information from a wide range of sources, in order to achieve efficiency savings and improve service delivery and the ways in which citizens interact with local government.

That said, the survey also identified barriers to local authority use of data. The barriers most commonly cited by respondents were:

- lack of resources to prepare and publish data (69 per cent);
- issues around data protection and the release of personal information (32 per cent);
- organisational and cultural barriers (30 per cent);
- technical barriers (30 per cent); and
- lack of skills to prepare and publish data (27 per cent).

Respondents also cited a range of other barriers, including:

- a lack of a definitive list to help them know what information should be published;
- a lack of clear guidance on data standards; and
- a lack of mature offerings from suppliers to provide open data as part of standard operations.

Given how diverse the landscape is, in terms of both the extent to which local authorities are using and releasing public sector information, the ways in which they are using it, and the barriers they

²⁶⁴ See Local Government Association, 'Local Government Transparency Survey 2012'

are facing, it makes sense to build on current efforts to share best practice across the local public sector. This would provide a way to capitalise on local authorities as ‘laboratories’ for the use of information. Where one local authority achieves efficiency savings or improves services, this may be applicable to many other local authorities and in this way the benefits can be scaled up. It would seem sensible to have a discussion about how and through what channels this can be achieved.

The LGA has already begun some work in this area, along with DCLG, through initiatives such as the Local E-Government Standards Board (LEGSB) and the Local Public Data Panel. There appears to be support for such initiatives from local authorities themselves: in the survey, around two thirds were in favour of the LGA providing a framework for publishing data through by way of a transparency strategy, and indicated that they would like to see case studies and guidance on how to publish open data.

Summary of opportunities in the local public sector

The examples given above indicate some of the ways in which the local public sector is beginning to exploit information to deliver both efficiencies and improved outcomes for policies and services, leading to welfare gains for citizens. These examples are likely to be a leading edge for significantly expanded future benefits, if the current trend of increasing sharing and exploitation of data continues.

The report by ConsultingWhere and ACIL Tasman on the use of geospatial information in the local public sector identified the following benefits for 2009:

- GDP was £323m higher than it would otherwise have been (an increase of 0.02 per cent);
- government revenue from taxation was £44m higher than it would otherwise have been;
- the delivery of goods and services by local public service providers was £232m higher than it would otherwise have been; and
- an increase in labour productivity equivalent to 1,500 full time staff across England and Wales, owing to the effects of improved citizen and business contact with local service providers.²⁶⁵

These benefits have arisen from the use of just one type of information (geospatial) by some, but not all local authorities. Given that the local public sector collectively accounts for expenditure of some £70 billion per year²⁶⁶, and is often the primary point of contact for citizens and businesses dealing with issues as diverse as education, social services and planning permission, the potential for further gains – both to local public sector efficiency and the wider economy – is likely to be significant.

As this section has demonstrated, the strength of the local public sector is its diversity, making it a laboratory for the use of information. In order to facilitate the potential of information release and sharing, it is important to ensure that the necessary frameworks (legal, regulatory and in terms of culture) are in place, and that local authorities have the freedom to experiment while also receiving suitable guidance. Under these conditions the potential for public sector information to transform local public service delivery and efficiency is likely to be significant.

²⁶⁵ ACIL Tasman and ConsultingWhere, ‘*The Value of Geospatial Information to Local Public Service Delivery in England and Wales*’ (July 2010)

²⁶⁶ Ibid.

Appendix 8: Government Officials informal consultation

Consultation questions

Background

Deloitte has been engaged by the Data Strategy Board to conduct a market assessment of Public Sector Information (PSI). The results of study will feed into the independent Shakespeare Review which is examining ways to widen access to Public Sector Information (PSI) and consider new and innovative opportunities for it.

Survey

1. Please could you provide your name; organisation; job title; role in the collection/dissemination of PSI in your organisation; and your contact details (telephone and email address)?
2. Approximately what proportion of the data that your organisation collects is made available to the general public (either freely or for a fee)? *Choose one option only.*
 - a. None
 - b. Between 0 and 25%
 - c. Between 26 and 50%
 - d. Between 51 and 75%
 - e. More than 75%
 - f. All data is made available
 - g. Don't know
3. Approximately what proportion of the data made available to the general public by your organisation is at no cost? *Choose one option only.*
 - a. 100% is free
 - b. Between 75 and 99% is free
 - c. Between 50 and 74% is free
 - d. Between 25 and 49% is free
 - e. Less than 25% is free
 - f. None is free
 - g. Don't know
4. How does your organisation make data available to the general public? *Indicate all that apply.*
 - a. Own website
 - b. On data.gov.uk
 - c. On another data portal (please specify)
 - d. Upon receipt of special requests
 - e. Other (please specify)
5. Approximately what proportion of your organisation's staff are directly involved in the collection, processing and dissemination of data? *Choose one option only.*

- a. Less than 5%
 - b. Between 5 and 10%
 - c. Between 11 and 25%
 - d. More than 25%
 - e. Don't know
6. Approximately what proportion of your organisation's budget is spent on the collection, processing and dissemination of data? *Choose one option only.*
- a. Less than 5%
 - b. Between 5 and 10%
 - c. Between 11 and 25%
 - d. More than 25%
 - e. Don't know
7. Would you say that the data collected by the department is primarily used by: *Choose one option only.*
- a. Your own organisation
 - b. Other Government organisations and agencies
 - c. Third parties – please specify
 - d. Don't know who the main users / re-users of data are
8. What sort of requests do you receive from the general public regarding open data? *Indicate all that apply.*
- a. Requests for other data that your organisation collects to be made public
 - b. Requests for your organisation to provide new data that is currently not collected
 - c. Requests for clarification of various issues around datasets or to improve the quality of datasets
 - d. Requests to correct different aspects of datasets
 - e. Requests for the data currently available publicly to be made available in different formats
 - f. Other – please specify
9. Are you able to give an indication of the types of users downloading your organisation's datasets? *Indicate all that apply and provide an approximation of the proportion of users they represent.*
- a. Other Government organisations and agencies
 - b. Other public service providers
 - c. Not-for profit organisations including researchers
 - d. For-profit private sector organisations
 - e. Individuals
 - f. Other – please specify
10. Can you give an indication of how the data is used and re-used?
11. Does your organisation plan to make available more data to the general public in the near future (within twelve months)?

- a. Yes
- b. No
- c. Don't know
12. If the above answer is yes, how much more data does your organisation plan to make available to the general public?
- a. Less than 25%
- b. Between 26 and 50%
- c. Between 51 and 75%
- d. More than 75%
- e. Don't know
13. What are the key challenges you see in making more data available to the general public?

Consultation responses

The following organisations responded to the survey:

- HMRC
- Home Office
- Defra
- DH
- DH Statistics Function
- DH MHRA
- FCO
- DCLG
- MoD
- MoJ
- CO
- BIS

The following table shows the responses to each questions, by percentage of respondents selecting each option. Note that in some cases it was possible to select more than one response, meaning that the sum of responses for these questions exceeds the total number of respondents.

Question	a	b	c	d	e	f	g
1	n/a						
2	8%	25%	0%	0%	17%	8%	33%
3	42%	42%	0%	0%	0%	0%	8%
4	92%	83%	67%	75%	33%	0%	0%
5	33%	0%	0%	8%	50%	0%	0%
6	42%	0%	0%	0%	50%	0%	0%
7	50%	8%	17%	33%	0%	0%	0%
8	67%	50%	75%	33%	58%	8%	0%
9	50%	42%	50%	42%	42%	33%	0%
10	n/a						
11	83%	8%	0%	0%	0%	0%	0%
12	42%	8%	0%	8%	25%	0%	0%

Question	a	b	c	d	e	f	g
13	n/a						

Appendix 9: Bibliography

Academic Papers & Reports

ACIL Tasman, 2009. *Spatial Information in the New Zealand Economy*. Wellington: Land Information New Zealand.

Administrative Data Taskforce, 2012. *The UK Administrative Data Network: Improving Access for Research and Policy*. London: Economic and Social Research Council.

Association of Greater Manchester Authorities. *Helping to improve information sharing and management: a national exemplar project*.

Capgemini Consulting, 2013. *The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data*. London: Capgemini Consulting.

Cebr, 2012. *Data equity: unlocking the value of big data*. London: SAS.

Cebr, 2012. *The Economic Costs of Gridlock*. London: Cebr.

Centre for Technology Policy Research, 2010. *Open Government: some next steps for the UK*.

ConsultingWhere and ACIL Tasman, 2010. *The Value of Geospatial Information to Local Public Sector Service Delivery in England and Wales*. London: Local Government Association.

Danish Enterprise and Construction Authority, 2010. *The Value of Danish Address Data*. Copenhagen.

Dekkers, M., Polman, F., te Velde, R. and de Vries, M., 2006. *Measuring European Public Sector Information Resources*. Helm Group of Companies.

Deloitte Access Economics, 2012. *The Economic Impact of the "Information Glut"*. London.

Deloitte Analytics, 2012. *Open data: driving growth, ingenuity and innovation*. London.

Deloitte Consulting and Tech4i2, 2011. *Pricing of Public Sector Information Study*.

Department for Transport, 2011. *The Economic Case for HS2: The Y Network and London –West Midlands*. London: Department for Transport.

Department for Transport, 2012. *Road Transport Forecasts 2011*. London: Department for Transport.

Department for Transport, 2012. *Values of Time and Vehicle Operating Costs*. London: Department for Transport.

Earnst & Young, 2011. *Making Energy Efficiency Your Business: understanding the potential of the non-domestic Green Deal*. London: Earnst & Young.

Elgin, 2013. *A new public-private model for creating a national database of local roadworks*

- e-skills UK, 2013. *Big Data Analytics: an assessment of demand for labour and skills, 2012-2017*. London: SAS UK.
- Genovese, E., Roche, S., Caron, C. and Feick, R., 2010. *The EcoGeo Cookbook for the assessment of Geographic Information Value*. International Journal of Spatial data Infrastructure Research.
- Goodwin, P., 2004. *The Economic Costs of Road Traffic Congestion*. London: University College London.
- Graves, A. *The Price of Everything but the Value of Nothing*. Office of Fair Trading presentation.
- Heseltine, *The Rt Hon the Lord of Thenford CT, 2012. No Stone Unturned in Pursuit of Growth*.
- HM Government, 2012. *Open Data White Paper: Unleashing the Potential*. London.
- Houghton, J., 2011. *Costs and Benefits of Data Provision*. Victoria: Centre for Strategic Economic Studies, Victoria University.
- Huijboom, N. and van der Broek, T., 2011. *Open data: an international comparison of strategies*. European Journal of ePractice.
- Kalapesi, C., Willersdorf, S. and Zwillenberg, P., 2012. *The Connected Kingdom: how the internet is transforming the UK economy*. London: Boston Consulting Group, Inc.
- Kundra, V., 2012. *Digital Fuel of the 21st Century: Innovation Through Open Data and the Network Effect*. Joan Shorenstein Center on the Press, Politics and Public Society.
- Lippert, C., 2010. *Public Sector Information Reuse in Denmark*. European Public Sector Information Platform Report No. 20.
- Local Government Association, 2012. *Local Government Transparency Survey 2012*. London: Local Government Association.
- Mandela, A., 2009. *PSI re-use: UK perspectives*. Locus Association.
- Mayo, E. and Steinberg, T., 2007. *The Power of Information: An Independent Review*.
- McKinsey Global Institute, 2011. *Big data: the next frontier for innovation, competition, and productivity*. McKinsey & Company.
- National Audit Office, 2012. *Implementing transparency: cross-government review*. London: National Audit Office.
- National Fraud Authority, 2012. *Annual Fraud Indicator*. London: National Fraud Authority.
- O'Hara, K., Whitley, E. and Whittal, P., 2011. *Avoiding the jigsaw effect: experiences with Ministry of Justice reoffending data*. University of Southampton, London School of Economics, and Detica.
- Office of Fair Trading, 2006. *The Commercial Use of Public Information*. London: Office of Fair Trading.
- Open Data User Group, 2013. *Further Benefits of an Open National Address Set*. London: Open Data User Group.
- Oxera, 1999. *The Economic Contribution of Ordnance Survey GB*. Oxford: Oxera.

- PA Consulting Group, 2007. *Met Office: The Public Weather Service's Contribution to the UK Economy*. London: PA Consulting Group.
- Pira International, 2000. *Commercial Exploitation of Europe's Public Sector Information*
- Pollock, R., 2011. *Welfare Gains from Opening up Public Sector Information in the UK*. Cambridge: University of Cambridge
- Pollock, R., Newbury, D. and Bently, L., 2008. *Models of Public Sector Information Provision via Trading Funds*. Cambridge: University of Cambridge.
- PriceWaterhouseCoopers LLP, 2010. *Economic Assessment of Spatial Data Pricing and Access*. Canberra: ANZLIC – The Spatial Information Council.
- PriceWaterhouseCoopers LLP, 2013. *A review of the potential benefits from the better use of information and technology in Health and Social Care*. London: PriceWaterhouseCoopers LLP.
- Public Data Group. *Business Case*.
- Roger Tym & Partners, 2003. *The Economic Benefit of the British Geological Survey*. London: Roger Tym & Partners.
- Rothenberg, J., 2012. *Towards a better supply and distribution process for open data: international benchmark*. Logica Business Consulting.
- Stauffacher, D., Hattotuwa, S. and Weekes, B., 2012. *The potential and challenges of open data for crisis information management and aid efficiency: a preliminary assessment*. ICT4Peace Foundation.
- The National Archives, 2009. *How we deal with exceptions to marginal cost pricing*. London: The National Archives.
- The National Archives, 2010. *The United Kingdom Report on the Re-use of Public Sector Information 2010: unlocking PSI potential*. London: The National Archives.
- The Office of Public Sector Information, 2011. *Information Fair Trader Scheme Report: Ordnance Survey*. London: The National Archives.
- The Office of Public Sector Information, 2012. *Information Fair Trader Scheme Report: Land Registry*. London: The National Archives.
- Transport for London, 2011. *Travel in London, Supplementary Report: London Travel Demand Survey*. London: Transport for London.
- Transport for London, 2011. *Travel in London: Report 5*. London: Transport for London.
- Transport for London, 2012. *London Buses Performance: Financial Year 2011/12*. London: Transport for London.
- Transport for London, 2012. *London Underground Performance Report: Period 4 2012/13*. London: Transport for London.
- Transport for London, 2012. *Syndication Developer Guidelines: Transport for London Data Service Version 2.0*. London: Transport for London.

Transport for London: London Streets, 2012. *Performance Report: Quarter 1 2012/13*. London: Transport for London.

Uhlir, P. F., 2009. *The Socioeconomic effects of public sector information on digital networks: towards a better understanding of different access and reuse policies (workshop summary)*. US National Committee for CODATA, and Board on Research Data and Information.

Vickery, G., 2011. *Review of recent studies on PSI re-use and related market developments*. Information Economics.

Yiu, C., 2012a. *A Right to Data: Fulfilling the Promise of Open Public Data in the UK*. London: Policy Exchange.

Yiu, C., 2012b. *The Big Data Opportunity: Making government fast, smarter and more personal*. London: Policy Exchange.

Departmental Open Data Strategies

Cabinet Office, 2012. *Open Data Strategy*. London: Cabinet Office.

Defra, 2012. *Open Data Strategy June 2012 – March 2014*. London: Defra.

Department for Business, Innovation & Skills, 2012. *Open Data Strategy 2012-14*. London: Department for Business, Innovation & Skills.

Department for Communities & Local Government, 2012. *Open Data Strategy April 2012-April 2014*. London: Department for Communities & Local Government.

Department for Culture, Media and Sport, 2012. *Open Data Strategy 2012-15*. London: Department for Culture, Media and Sport.

Department for Education, 2012. *Open Data Strategy*. London: Department for Education.

Department for International Development, 2012. *Open Data Strategy April 2012 – March 2014*. London: Department for International Development.

Department for Transport, 2012. *Open Data Strategy*. London: Department for Transport.

Department for Works and Pensions, 2012. *Open Data Strategy*. London: Department for Works and Pensions.

Department of Energy and Climate Change, 2012. *DECC's Open Data Strategy*. London: Department of Energy and Climate Change.

Department of Health, 2012. *The Power of Information: putting all of us in control of the health and care information we need*. London: Department of Health.

Foreign and Commonwealth Office, 2012. *The FCO's Open Data Strategy*. London: Foreign and Commonwealth Office.

HM Revenue & Customs, 2012. *Open Data Strategy*. London: HM Revenue & Customs.

HM Treasury, 2012. *Open Data Strategy 2012-2015*. London: HM Treasury.

Home Office, 2012. *Open Data Strategy for the Home Office 2012-13*. London: Home Office.

Ministry of Defence, 2012. *Open Data Strategy 2012-2014*. London: Ministry of Defence.

Ministry of Justice, 2012. *Open Data Strategy 2012-15*. London: Ministry of Justice.

Other Documents

Association of Train Operating Companies, 2012. *RSP licence fee and datafeeds charge*.

Cabinet Office, 2012. *Making open data real: a government summary of responses*. London: Cabinet Office.

Department of Business Innovation and Skills, 2012. *Annual Report and Accounts 2011-2012*. London: Department of Business Innovation and Skills.

HM Government, 2011. *Making open data real: a public consultation*. London: HM Government.

Met Office, 2012. *Annual Report and Accounts 2011/12*. Exeter: Met Office.

Open Data User Group, 2012. *Open Data User Group response to the consultation on the code of practice for datasets and beta charged licence*. London: Open Data User Group.

Crown copyright 2013

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. Visit www.nationalarchives.gov.uk/doc/open-government-licence, write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

This publication is available from www.gov.uk/bis

Any enquiries regarding this publication should be sent to:

Department for Business, Innovation and Skills
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000

If you require this publication in an alternative format, email enquiries@bis.gsi.gov.uk, or call 020 7215 5000.

BIS/13/743