Deloitte.

# Measuring the Value of Statistics
# A Report to ONS

16 September, 2016

# Contents

# Key Findings

**PLEASE NOTE**: elements of this paper were put together with external users of ONS data. Their comments and all information they provided were collected on the basis of non-disclosure. As such, we have redacted those elements of this paper to publish.

### Introduction and context

- The way statistics are being generated, collected and used is rapidly changing. Providers and collectors of statistics, such as the Office of National Statistics (ONS) are under increasing pressure to respond to changing user requirements and demonstrate they offer value for money and contribute to improvements in socio-economic outcomes.
- The 2016 Bean Review of economic statistics and the Royal Statistical Society have separately highlighted a number of areas for improvement for the ONS. Both make recommendations that are likely to require changes to ONS's working practices and require funding for improvements in services and skills. In essence the aim is to provide better statistics in a more efficient manner.
- This report explores the ways in which ONS can measure the value or benefit of ONS economic statistics to users and society. This will assist ONS in making the case for additional resource allocation to these, and other, statistics, or prioritising spend, by showing how changes to quality, coverage, frequency and other statistical dimensions benefit the wider fabric of the UK.
- The three economic statistical domains considered here are GDP, labour market and price inflation statistics. Using a hybrid method made up of market and non-market approaches ONS economic statistics across the three domains are valued between £130 and £155 million per annum.

### Evidence and analysis

**Economic statistics are a critical input to policymaking and organisational decision-making. They underpin a variety of products and services and can promote transparency and accountability…**

- Economic statistics contribute to an evidence base upon which policy and business decisions can be made more effectively. In turn this leads to better outcomes for society. For example, labour market statistics can inform policy interventions in the job market by identifying where interventions are required

and can be used in evaluations to see which interventions were successful in the past.

- Economic statistics can be combined with other data and statistics to form the basis of new products and services, or supplement existing ones. For example, price inflation data could be used in pricing models for retailers or for determining real changes in financial assets; GDP data can be used in economic forecasts sold to customers across the economy.
- Economic statistics, typically combined with other statistics, can also be used to hold decision makers to account. They can be used, through the media, to challenge the government of the day over their record in generating jobs, keeping price inflation on target and creating economic growth.

**…which in turn can contribute to raising productivity and wealth creation…**

- ONS activities (inputs) create the three statistical domains (outputs) which are used by to affect business and policy (outcomes) and in turn generate value for the UK (impacts), such as:
  - Business revenues
  - Costs saved
  - Output and Productivity
  - Job creation and new business creation
  - Tax receipts
  - Welfare gains; and
  - Option value from more choice

**…and this is well recognised by stakeholders across the public sector.**

- Stakeholders from a number of public sector organisations including the Department for Business, Innovation and Skills (BIS), the Office for Budget Responsibility (OBR), the Bank of England, the ONS itself and academia also recognise these benefits, noting:
  - There are a number of accepted arguments for treating economic (and other) statistics as public goods.
  - When considering individual domains in isolation rather than all statistics together, there is a risk that the confidence and utility UK statistics generate when bundled together is overlooked and impacts are understated;
  - The value of statistics come not just from the numbers themselves and usage but also the supporting collateral (metadata, method and customer service) provided to accompany the statistics. This improves user and re-user value by contributing to better understanding, data analysis and decisions;

- Simply providing more numbers without context or supporting materials risks destroying value through mis-use rather than creating it – here statistics provide value by being regarded as a universal view of the truth, if, and only if, they are understood and well-interpreted;
- The extent of the user and re-user base itself is a significant demonstration of statistical value in its own right, even if not a scientific measure of aggregate value;
- In particular, if organisations are taking the time to model and forecast with ONS data, then by implication the statistics are of value to them directly, as well as indirect beneficiaries who are either willing to pay for insight or benefit from better policy decisions;
- The importance of ONS statistics to stakeholders should not be understated and the costs of provision in a world without ONS would be shifted to users, without the guarantee of commensurate levels of quality; and
- The implications/costs of statistical errors is dependent on the statistic in question and how it is used.

**There is a wealth of research exploring the value data creates more generally (as opposed to statistics specifically – which by definition are produced to national standards and are badged with a quality measure), using a variety of valuation methodologies each with their own advantages and disadvantages…**

- Reviewing UK and international literature reveals a number of different economic, econometric, statistical and other methodologies used to estimate the monetary value or benefit of data and statistics. These methodologies include:
  - Perception of value estimates using stated or revealed preference surveys to estimate customers' willingness to pay for data provision or to avoid loss;
  - Return on investment approaches that calculate the cost involved in generating the statistics and then compare this to the expenditure made by direct customers (users, re-users and redistributors);
  - Market-based approaches which seek to put a market value on the benefits of the use, re-use and redistribution of data using available price data;
  - Non-market-based approaches which consider changes in specific externalities and then quantify them using relevant techniques;
  - Avoided cost approaches that seek to estimate the costs saved by users and re-users in using the data rather than the benefits created by it per se;
  - Computable General Equilibrium and Input-Output models that build representations of the economy and through a series of assumptions model

the ‘shocks’ caused by data across the macro-economy, in income shifts and productivity changes;

- Dynamic Welfare Approaches that build on willingness to pay frameworks, but look to further model efficiency aspects of data in competition and innovation over time and often use scenarios to capture the range of expected benefits; and
- Case studies that use concrete examples to highlight the mechanism through which impacts can be realised, and which are then sometimes scaled up for aggregate impacts.

**… with the most valid approaches for the three statistical domains depending upon on a series of criteria pertinent to ONS’ use of any valuation.**

- The criteria applied to ensure that the method chosen is fit-for-purpose for end-use by ONS are:
  - Simplicity: the method should be simple to use, articulate and understand. Any ‘black box’ solution which churns out values that are not understood by users or readers runs the risk of misinterpretation;
  - Transparency: any assumptions required of a method should explicit and transparent. Moreover, the more assumptions required (in place of data) for a method, the greater scope for error in outputs;
  - Flexibility: the method should be sufficiently flexible to allow for sensitivity testing and future refinement. HMT Green Book business cases require sensitivity testing as part of the analysis of risks;
  - Replicability: the method should be replicable across a range of statistical types;
  - Specificity: the method should be capable of distinguishing between benefit streams for the same statistic delivered in different ways – e.g. the quality of ONS outputs;
  - Data-availability: the method chosen must have a minimum level of information available for it to be of actionable, and of use. We have interrogated the data provided to date (summarised in chapter 4), but we would like to source more data if possible;
  - Robustness: Related to all other factors, robustness is concerned with the actual and perceived standard of outputs produced by the chosen method. It is critical that the method should stand up to scrutiny from internal and external stakeholders, such as HMT; and

  - Cost effectiveness (financial & other): the method should not create high marginal costs for ONS each time it is used, nor require a major initial investment to set up.
- As ever in options analysis given the trade-offs inherent, no single option is likely to score highest on each measure, and ONS will need to decide which options present the best solution for them.

**A market/non-market-based approach was agreed and the chosen method uses a range of information from ONS and third-parties in conjunction with shadow prices to value statistics on the basis of usage by a range of organisations.**

- The method was discussed in detail at a workshop in early May 2016 attended by ONS and Deloitte.
- The market/non-market based approach builds on available public domain and ONS data along with relevant shadow prices from market-based provision to estimate the value of ONS statistics, as well as considering externalities. Shadow prices can come from private companies supplying data, or ONS data that is paid for, or has been paid for in the past, with adjustments to reflect the fact that it is generally provided at cost.
- As an example, the prices paid for CIPS PMI data, a leading indicator for GDP, can be used as a proxy for the value of GDP statistics, because organisations are willing to pay for a timelier indicator of economic output/growth.
- In isolation market and non-market based approaches offer a viable, low-cost option for the assessment of aggregate benefits by statistic or domain, but are not especially suited to use in a business case, unless relevant adjustments can be made for quality factors.
- This approach was ranked as second in our options analysis. However, given constraints on ONS (notably in terms of the level of stakeholder consultation already in train, and cost considerations), it was decided that such an approach is the best one to follow in this study.
- The method subsequently employed, therefore, uses estimates of usage in conjunction with shadow prices and willingness-to-pay adjustments to yield estimates of value across the three domains for ONS Statistics.

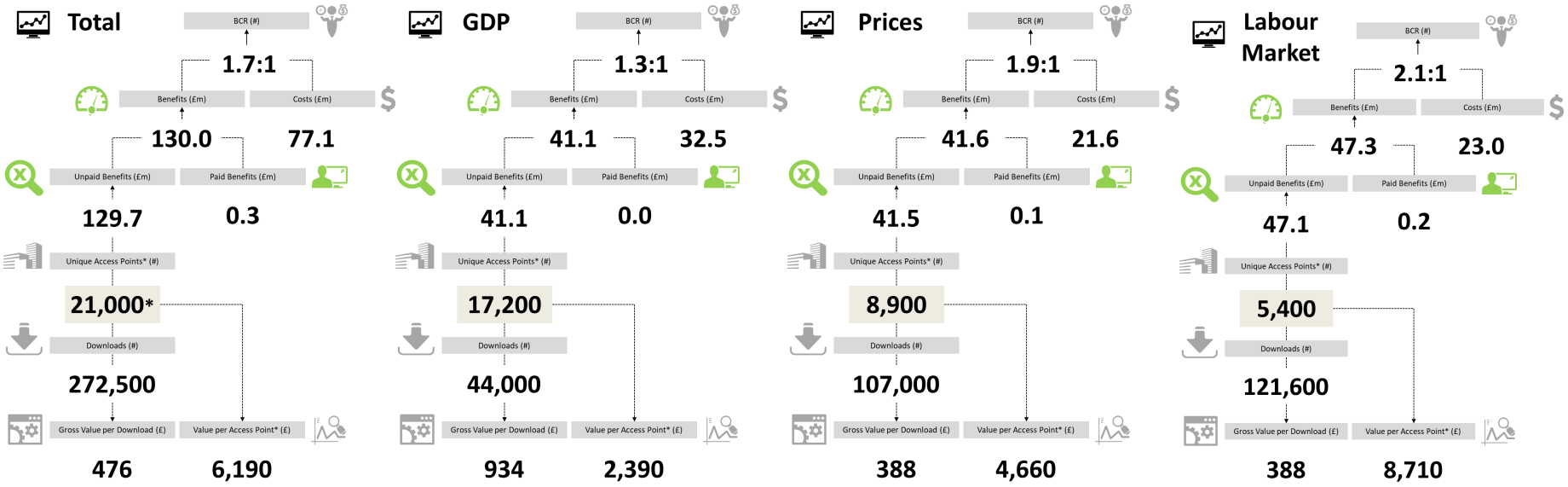
## Results and conclusions

- Using a shadow price methodology, we estimate that the combined value of the three statistical domains (GDP, inflation and labour market), lies **in the range £130 million to £155 million for 2014**, and yields a BCR (benefit cost ratio) of c. 1.7:1 based on costs of £77.1 million in 2014.
- These estimates are based upon the use and re-use of statistics in organisations known to obtain and use ONS statistics, either from ONS directly; from other Government websites; or from third parties. They do not include so-called 'wider benefits'. Neither do they include the 'greater-than-the-sum-of-all-parts' benefits that stem from the holistic provision of all national statistics.
- Management information detailing ONS website use and downloads by 'unique access points' – broadly corresponding to organisations – is used together with information on usage from data.gov.uk; the costs of statistical production; and shadow prices from third party providers of proxy data to estimate a value for statistics in each domain, as well as in aggregate.
- The estimates of value by domain (as defined in chapter 8) are detailed overleaf, but in summary:
    - **GDP** statistics generate a net value of **£41 million** and a BCR of 1.3:1;
    - **Inflation** statistics generate a net value of **£42 million** and a BCR of 1.9:1; and
    - **Labour market** statistics generate a net value of **£47 million** and a BCR of 2.1:1.
- These results do not factor in measures of quality or include all future potential uses and re-uses. Willingness to pay based on shadow prices captures the costs users would pay, but this does not capture the quality differential that many users state as a differentiator. As a consequence, they are likely to understate benefits and could be improved in a number of ways.
- Applying ready reckoners to account for wider benefits implies that the total benefits from these three domains might be in the order of £0.5 billion in total – with an associated BCR of 6.7 – though we urge interpretation of this figure with caution.
- Chief amongst any improvements would be conducting a survey to understand user activities, products and preferences. Feeding results from a large-scale 'soft-factor'-style exercise would ensure future decisions are based on quality factors as well as usage and cost only.
- Other improvements could include making more of the management information from ONS new website for 2015 and 2016 when a full year's data

is available, and conducting an econometric exercise to consider aggregate statistical contribution to UK economic output in the same vein as Haskel et al on Big Data, referenced in section 2.3 of this report.

- In the interim, there are a number of ways this analysis can be used in business case work, although this will require some level of assumption. Two examples include:
  - BCRs might be used as base-case ready reckoners for new ways of statistical provision. As an example, benefits might calculated on the basis of existing costs, and then held equal as costs are reduced through novel ways of provision. This will imply a higher BCR from efficiencies for any given statistical output; or
  - For a given BCR, ONS could make an assumption that an increase in quality might create an increase in the benefit that is at least equivalent to the increase in cost. This means that where a new method or statistic is associated with greater cost, the greater costs are likely to provide equivalent, if not more than proportionate levels of benefit.
- As a final point of note, consultees suggest that the extent of usage in itself demonstrates the significant value these and other ONS statistics generate. In summary, in 2014:
  - Over **21,000 unique access points** were used to obtain ONS data across the three domains. This understates total unique users as it includes internet service providers with multiple users.
  - The best indication of scale is downloads, and there were over **270,000 downloads** of free-at-the-point of use statistics and publications across the three domains.

Figure 1.1a: Aggregate and domain-specific statistical value, 2014




Source: Deloitte Analysis, excludes wider impacts. Totals may not sum exactly due to rounding (nearest 100 or 10 depending on scale to prevent 'spurious accuracy'). Results explained in detail in Section 8.

* denotes that the sum of unique access points in each domain cannot be aggregated to give the total across domains, because many users are common to two or more domains.

# 1 Introduction

**PLEASE NOTE**: elements of this paper we put together with external users of our data. Their comments and all information they provided were collected on the basis of non-disclosure. As such, we have redacted those elements of this paper.

## 1.1 Study scope and objectives

Deloitte has been commissioned by the Office of National Statistics (ONS) in March 2016 to explore the ways in which the monetary value of official government statistics can be measured. Specifically, the research examines and determines an appropriate methodology or methodologies to estimate the value of the contribution that three domains of ONS statistics make to the UK economy and wider society: GDP statistics, labour market statistics and price inflation statistics.

The study objectives include:

- Analyses of the benefit and value inherent in ONS statistics across three domains: GDP, price inflation and labour markets;
- An understanding of how to increase the governance, quality and delivery of UK government statistics and the additional value that will be generated through improved statistics; and
- The provision and articulation of a replicable methodology that may be applied to other types of ONS statistics, where appropriate data allows.

The approach agreed for this research includes a literature review, an examination of the available data from ONS and public sources, stakeholder discussions and a preliminary modelling exercise across the three statistical domains.

This document forms the third project deliverable, and final report, at the end of three months' work setting out a detailed literature review, chosen method and a preliminary modelling exercise.

The agreed study scope includes:

- Consultation with internal and external stakeholders offering insight on the three specific domains;
- Analysis of quality and timeliness of data, and the way in which value might erode in time – an important consideration for business cases;

- Consideration of the counterfactual of substitutes for ONS data on GDP, labour market and inflation;
- Production of a replicable methodology that can be used to support business cases for collection, production and dissemination of statistics; and
- A final deliverable allowing ONS to easily estimate the likely value of economic statistics to inform the benefit line in NPV calculations.

The study scope expressly excludes:

- Provision of a specific 'model' for use by ONS; and
- Primary survey work.

## 1.2 Statistical domains in scope

The scope of this study involved the exploration of three domains of statistics, namely, GDP, Price Inflation and labour market. To clarify scope these are defined in greater detail below.

Gross Domestic Product (GDP) statistics included preliminary, second and final estimates of GDP released over a quarter as more data becomes available. All constituent parts of GDP are included and are in scope for this analysis. The final estimate is published in the Quarterly National Accounts. GDP is the main measure of UK economic growth based on the value of goods and services produced during a given period. Here we are considering the aggregate GDP series as a whole.

Inflation is defined as the rate of increase in prices for goods and services. Measures of inflation and prices includes consumer price inflation, producer price inflation, house price inflation and industry inflation. Here we are considering the full range of inflation statistics defined as primary (rather than constituent) series.

The full ONS range of labour statistics is in scope for this study. This involves a broad range of statistics that include estimates of employment, unemployment and economic inactivity. This is the most diverse of the domains considered.

In addition to these statistical domains at national level, regional dimensions were also considered. The current government's devolution agenda creates a focus on regional statistics. Regional and local level labour market statistics are available through the ONS website and third-party provision such as data.gov.uk and NOMIS. Some regional GVA/GDP and Inflation statistics are also published by ONS, for example, the regional housing index statistics, but some are experimental and these do not cover the same dimension and granularity as labour market statistics at local level. More widely, there are other domains within the ONS that are not included in our analysis here. Many of these have no links to the domains assessed here. Some, such as the Census Programme do not have

specific links to inflation or GDP but are used for labour market statistics, and some, such as Wellbeing statistics are used as an alternative measure to GDP.

### 1.3 Study method

The research has focused on the theoretical underpinnings of statistics valuation and the formulation of fit-for-purpose options by triangulating available data, expert opinion and available methodologies.

The scope, objectives and aims of the project were assessed and refined in collaboration with ONS to ensure that the approach yields a methodology that fulfils the scope of the study. The methodology to date has included:

1. **Literature Review**
   An end-to-end review of major studies was considered in order to contextualise the landscape of the literature and various methodological approaches. After narrowing focus, subject to client requirements, the review isolated key opinions in the literature that attempted to quantify the value of data or statistics in the UK. Finally, thought pieces that analysed the value of statistics or data for government bodies such as ONS were assessed. This can be found in chapter 1 with a summary of methodologies in chapter 2.
2. **Data collection and analysis**
   Data collected from ONS included web metrics and KPIs (such as visits, downloads and unique IPs across a range of data, Media metrics, Media performance, Media KPI's and information on paid data, information requests and NOMIS Licensing). The study also considered change in the fields of data collection and analytics and how future trends might impact this area. This was to fulfil the study objective that the model could be used to support subsequent assessments on the value of statistics for ONS. The analysis is split between chapters 4 and 8, with much analysis not recorded in this report. A large scale exercise was conducted in SQL to refine and reorganise ONS management information.
3. **Options analysis**
   Methodologies were assessed on the basis of the literature review, available data and stakeholder requirements. Sift criteria were determined from the evidence and then applied to the methodologies to short-list likely candidates. In conjunction with ONS, and on the basis of preferences, costs and timescales the most appropriate methodology was selected. The process and outcomes of this exercise can been seen in chapters 5 to 7.
4. **Quantification of the value of statistical domains**
   Estimates of statistical value for ONS domains, are based upon the use and re-use of statistics in organisations known to obtain ONS statistics,

either from ONS directly, from other Government websites, or from third parties. These are presented in chapter 8.

5. **Conclusions**

   The report concludes in chapter 9 by identifying potential next steps.

## 1.4 Study context

This project has been undertaken against the broader background of Sir Charles Bean's Independent Review of UK Economic Statistics[1], suggestions for the improvement of ONS by the Royal Statistical Society[2], exponential growth in technology, increasing pressure on public bodies to demonstrate that they offer value for money and the wider government productivity and growth agenda.

In July 2015, Sir Charles was tasked with leading an independent review to assess the UK's future economic statistics needs, the efficacy of the ONS in delivering these statistics and what governance framework might be most appropriate to support the production and delivery of world-class economic statistics. The Review reiterated the importance of high quality and reliable economic statistics as key inputs underpinning business decision-making and policymaking.

Sir Charles reported back with his final report in March 2016. Broadly speaking, the Review highlighted two key areas of concern. First, that measuring the economy has become even more challenging in recent times, in part as a consequence of the digital revolution. As well as technological advancement challenging existing data categories (e.g. SIC codes). This raises questions around how ONS can take advantage of the full range of new data sources becoming available. Secondly, users expressed a belief that ONS's performance has deteriorated in recent years. Some of this criticism stemmed from the size and frequency of revisions, though depending on how such revisions are measured, at least some of this criticism is shown to be unfounded.

The Royal Statistical Society (RSS), the professional body for statisticians and data analysts also recommended how ONS can evolve to meet changing requirements. Its principal recommendations are the need for ONS to acquire greater expertise in national accounting, increased resources on economic statistics and improved access to administrative data from both the private and public sectors. It further emphasised the efficiency gains that could be realised from a closer working relationship between ONS, Bank of England and HM Treasury.

More broadly, there is a view that GDP is used in a way that it was not designed for. Whilst this study very briefly touches on alternatives to GDP that provide a better measure of societal progress as raised in discussions, it is beyond scope to consider changes in statistical provision in this specific way.

[1] The final report can be downloaded at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf
[2] The final report can be downloaded at:
http://www.rss.org.uk/Images/PDF/influencing-change/2015/RSS-written%20evidence-to-Bean-Review-Sept-2015.pdf

# 2 Literature review

A review of literature has been undertaken to assess how to best quantify the aforementioned change in quality that adapting to recommendations and technology could bring. Analysis of papers gathered illustrates the limitations of theories and methodologies and helps to formulate areas for further research. This literature review includes an appraisal of the nature of statistics, the benefits and value of statistics and previous studies estimating the value of data.

## 2.1 The nature of statistics

*Statistical categories*

The literature categorises different data into different types, of which statistics are a quantitative subset. These categories are summarised below – the categories are agnostic to the statistics' content (e.g. they could cover economic, demographic, geo-spatial and all other subject areas).

Figure 2.1.a: Statistical categories

| **Access route** | **Collection method and supplier** | | | | | | |
|---|---|---|---|---|---|---|---|
| ***Open*** Free (or nominal cost) and unrestricted use | ***Data supplier*** | | | | | | |
| | ***Government data*** | | ***Research or science data*** | | ***Private sector data*** | | |
| | *Survey based* | *Administrative* | *Survey based* | *Research outputs* | *Survey based* | *Administrative* | *Research outputs* |
| ***Mixed / shared*** Limited public access or with conditions on use, may be restricted to particular groups | ***Data supplier*** | | | | | | |
| | ***Government data*** | | ***Research or science data*** | | ***Private sector data*** | | |
| | *Survey based* | *Administrative* | *Survey based* | *Research outputs* | *Survey based* | *Administrative* | *Research outputs* |

| ***Closed*** Paid for data or data that is for internal use only | ***Data supplier*** | | | | | | |
|---|---|---|---|---|---|---|---|
| | ***Government data*** | | ***Research or science data*** | | ***Private sector data*** | | |
| | *Survey based* | *Administrative* | *Survey based* | *Research outputs* | *Survey based* | *Administrative* | *Research outputs* |

Source: Deloitte, based on literature reviewed

Statistics that are open, are open in terms of their accessibility and their use conditions. The Open Knowledge[3] network defines open data (or statistics) as when:

- the statistics are available as a whole and at no more than a reasonable reproduction cost[4] in a convenient and modifiable form;
- the statistics are provided under terms that permit re-use and redistribution, and the intermixing with other datasets; and
- everyone is able to use, re-use and redistribute the statistics, i.e. there are no 'non-commercial' restrictions that would prevent 'commercial' uses.

Much of the data under consideration in this study is open per this definition. Some data accessed at a given level of detail requires a bespoke table to be produced by ONS. This is typically provided at cost and is therefore open. Although it is not provided free-at-the-point of use, the cost to the user represents a reasonable reproduction cost.

Mixed or shared statistics share many of the characteristics of open statistics, but may have some restrictions on re-use or redistribution, or only a partial element is available at no or minimal cost. In contrast, closed data is typically paid-for data, or only available to internal stakeholders or users.

A limited amount of data under consideration here may be classified as mixed/shared, an example being where users have to 'sign' a notice for local level statistics to prevent disclosure when using the statistics. No data under consideration for this study is closed as national statistics are, by definition, public goods.

The next set of dimensions of statistics refer to how they are collected. As the Bean Review highlights, traditionally many ONS statistics have come from surveys of businesses and individuals. Increasingly, more administrative data is becoming available – defined as data collected by public and private bodies in the course of their day-to-day actions. Research or science output statistics are data funded or collected from publicly or privately funded research. As suggested in both the Bean Review and by the Royal Statistical Society, ONS need better access to both

[3] Defined at: http://opendatahandbook.org/guide/en/what-is-open-data/
[4] There remains debate whether open data can only truly be open if it is free.

public and private data to provide a better and wider range of statistics more cost-efficiently. New legislation should be put into place to give ONS the right of access to confidential personal and business data and a closer relationship between ONS, Bank of England and Treasury should be encouraged.

In terms of the three economic domains under consideration in this research, our understanding is that they all fall under the open access category, although there are paid-for elements that are available to the public that would fall under the open category – as ONS charges are at cost fees. The data is predominately currently collected from surveys rather than administrative sources.

*Statistical characteristics*

The previous sub-section categorised the ways statistics are made available and are collected. The benefit or value of these statistics will, to some extent, be influenced by the access route and cost, but the key determinants of value, according to the literature, will be the characteristics of the statistics themselves. This section examines in more detail the nature of this value. The figure below sets out the different ways statistics can be characterised or assessed.

Figure 2.1.b: Statistical Characteristics

| Characteristics | Description |
|---|---|
| **Verbosity or complexity** | The level of detail the statistics go into. For example, simpler statistics may only carry observations at an aggregated level, whereas other statistics (which might be classed as big data) could have a high level of disaggregation across multiple dimensions. |
| **Velocity and timeliness** | How often the statistics are refreshed/updated, and whether they are released in a timely and regular fashion relative other available statistics. |
| **Coverage (geography and timespan)** | The area (regional and local) and period the statistics cover, and in what level of detail reflective to other available statistics. |
| **Robustness and quality** | The level of accuracy of the statistics, can be proxied by the frequency and extent to which the statistics are revised. This can also include the extent to which the statistics have missing observations or redactions and suffer from sample size issues. Quality in the aggregate, especially perceived quality, is more subjective than these more specific characteristics. |
| **Level of metadata and** | The amount of additional information available to explain the statistics and support available to users reflective to other statistics. |

| | |
|---|---|
| **support available.** | |
| **Compliance with five star rating (the format of the statistics)** | How well the data meets open data and usability standards relative to other data and statistics[5]. |
| **Nature of the statistics** | Which thematic areas the statistics cover: these can include economics, demographics, geo-spatial, meteorological, personal finance, transport, health and social care, education, crime and justice, environment, energy, housing and agriculture, among others. |
| **Distribution channels** | The different ways users can access the statistics ranging from internet downloads, controlled access to data terminals, data sent out on CDs or in hard copy. |

Source: Deloitte, based on literature reviewed

Many of these themes were explored in the Bean Review, especially robustness and quality, and recommendations were made on these themes.

In terms of the three statistical domains under consideration in this research, they can be characterised as follows.

Figure 2.1.c: Characteristics of selected economic statistics

| Characteristics | Labour market | Price inflation | GDP |
|---|---|---|---|
| **Verbosity or complexity** | People in work, out of work, hours worked, earnings, employment and employee types, workplace pensions, redundancies, economic inactivity, out of work benefits , labour market flows | Various indicators including consumer (CPI, CPIH, RPI, RPIJ etc.), producer and service prices, PPIs, services indices, construction prices, house prices and more detailed component | GDP values and growth rates |

[5] See http://5stardata.info/en/ for more details.

| | | | |
|---|---|---|---|
| | | indices at sectoral level (e.g. aerospace or electronics) | |
| **Velocity and timeliness** | Monthly, quarterly and annually at scheduled times | Monthly, quarterly, at scheduled times | Quarterly at scheduled times. With two preliminary estimates each quarter and various releases of component measures released at different intervals |
| **Coverage (geography timespan and industry level figures)** | UK-wide and regional/local sectoral, occupation and other breakdowns | UK-wide | UK-wide (with experimental local GVA statistics) and industry level figures |
| **Robustness and quality** | Is revised | PPIs are occasionally revised back to the start of the year but generally prices are not revised | Is revised |
| **Level of metadata and support available.** | Available | Available | Available |
| **Compliance with five star rating (the format of the statistics)** | Excel files, CSV files, Image files, PDF files | Excel files, CSV files, PDF files | Excel files, CSV files, structured text |

| | | | |
|---|---|---|---|
| **Nature of the statistics** | Economic | Economic | Economic |
| **Main Distribution channels** | ONS website, bespoke commissions | ONS website | ONS website |

Source: Deloitte, ONS website

While the literature identifies these characteristics of statistics, it is much harder to objectively measure many of these and then link these to overall value generated. In some cases, actual measurement data or proxies do exist, e.g. the five star rating or velocity, but for others, such as robustness and quality, measurement is much more subjective. As a result, many studies estimating the monetary benefits of statistics tend to normalise all characteristics of statistics to one, i.e. they do not model differences in statistical dataset's characteristics beyond their thematic content and usage.

However, as data improves, it may be possible to compare changes in characteristics within the *same* statistical dataset over time and their impact on value/benefits. For example, if price inflation statistics increased in regularity or expanded their input data to include more timely commercial data, one might be able to do a 'before and after' analysis on usage and satisfaction to show the expected change in benefits to the economy and society.

*Customers of economic statistics*

The literature identifies a range of 'customers' of data (generally – as opposed to statistics). These include businesses who use public and private data to develop and refine products and services, often as aggregators or developers, and individuals who use the data to hold organisations to account or to make more informed decisions.

For example, a Financial Services Industry analyst could use inflation statistics to inform a business decision directly or they could package up inflation statistics with other statistics and redistribute it, often with paid-for value-added services.

There is, however, a distinction as to how a 'customer' uses data and their purpose for accessing this data. Customer personas have been identified by ONS and offer insight into the different aims of customers. For example, an intellectually curious citizen looking at GDP figures to observe the growth of the UK economy has very different needs to a public body who are assessing detailed labour statistics to develop government policy.

To take account of customer aims and usage, this study views 'customers' through the vehicle of the 3:3 matrix below.

Figure 2.1.d: ONS customers of economic statistics

| Type of ONS customer | | Direct user | Re-user | Redistributor |
|---|---|---|---|---|
| | | Covers customers who use the economic statistics produced by ONS directly to inform decision-making and policy. The statistics are largely used directly without any modification or transformation. | Covers customers that do not directly use economic statistics for pricing, labour market and GDP analyses, but re-use it in other ways, potentially with other data sources and to augment/refine products and services. | Covers customers that package up the economic statistics with other statistics and redistribute it, often with paid-for value-added services. |
| **The Information Forager** | Covers customers who use the economic statistics produced by statistical agencies inform internal decision-making and policy formulation.<br>This customer wants to look for data that can be used to make practical, strategic decisions for his/her business. The information forager wants to see high level summaries, narratives | These customers are likely to include:<br>• Central government departments<br>• Devolved, local and other arms of government<br>• Central Bank, regulators and other arms-length public sector bodies | These **customers** are likely to include those relying on policy analysis and outputs from analysis of the data:<br>• Ministers<br>• NDPBs<br>• Media organisations<br>• Citizens<br>• E.g. what BIS said yesterday | N/a |

| Type of ONS customer | | Direct user | Re-user | Redistributor |
|---|---|---|---|---|
| | and key charts that provide context for deeper understanding. He/she may occasionally download datasets for simple analysis to support arguments in funding applications and strategy reports and may look for time series and comparison data in order to be able to predict future opportunities. | • Universities, think tanks and other research agencies<br>• The financial services sector<br>• The business and professional services sector | | |
| **The Expert Analyst** | Covers customers that do not directly use economic statistics for pricing, labour market and GDP analyses internally, but re-use it in other ways, potentially together with other data sources and to augment/refine the products and services they offer These users typically do not redistribute the data itself, rather the insight from it. | These customers are likely to include:<br>• Central government departments<br>• Devolved, local and other arms of government<br>• Central Bank, regulators and other arms-length public sector bodies | These **customers** could include:<br>• Those buying research, statistics or forecasts from businesses to use in their decision-making | These **customers/providers** could include:<br>• Organisations providing open data based on statistics<br>• Media organisations<br>• Citizens with a political interest<br>• Citizens with a personal financial interest in economic implications (e.g. for |

| Type of ONS customer | | Direct user | Re-user | Redistributor |
|---|---|---|---|---|
| | Their key goals tend to be to find a particular excel spreadsheet to download without being distracted by similar sounding information.<br>The expert analyst tends to find exactly what she/he wants, but can be frustrated by not being able to find it quickly on the ONS website. The user will phone the ONS for help in finding specific data or querying methodology and tends to access ONS website from desktop PC in office they may be critical about mistakes and shortcomings in the provision of statistics. | • Universities, think tanks and other research agencies<br>• The financial services sector<br>• The business and professional services sector<br>• Data service providers and data analytics firms<br>• Software providers | | pensions and investments) |
| **The Inquiring Citizen** | Covers customers whose key goals are to find out the unbiased 'truth' about economic indicators in order to be able to make informed | These customers could include:<br>• Citizens with a political interest | N/a | N/a |

| Type of ONS customer | | Direct user | Re-user | Redistributor |
|---|---|---|---|---|
| | decisions about pensions and investments or to find about newsworthy topics such as immigration, house prices, inflation and economic growth. | • Citizens with a personal financial interest in economic implications (e.g. for pensions and investments) | | |

Source: ONS based on personas developed for ONS website and Deloitte analysis

It is important to note that a single individual may shift between persona types. Academics might be expert analysts for their research, but may forage for information on a number of issues outside of their expertise to support their teaching.

It is further work noting that the personas of 'Expert Analyst', 'Information Forager' and 'Enquiring Citizen' are based on research for ONS website in 2013 so are not specific to this study. These personas are employed to ensure the output of this study is consistent with ONS methodologies.

The literature notes, and Deloitte's experience also suggests, that whilst it is possible to identify categories of customers, it is much harder to understand how customers use the data and the extent to which the data is instrumental or incidental in generating value/benefits.

The next section outlines what these benefits might look like, with a subsequent section considering the different approaches seen in the literature to measuring the monetary value of these benefits.

## 2.2 The benefits and value of statistics

There is a consensus amongst economists that statistics used under the right circumstances can enhance economic growth through improved decision-making and policy-making. In practice, the relationship is complex and hard to disentangle.

This link is based on economic assumptions and evidence. Information is shown to have public good characteristics. In general, the private sector will tend to under-produce such goods as it is difficult to realise their full value. This justifies the public sector supply of information and in particular statistics[6].

The literature sets out a number of stylised benefits of data generally, from which the benefits of economic statistics can be inferred. These are summarised below:

**Better decision making and outcomes**. Economic statistics contribute to an evidence base upon which policy and business decisions can be made more effectively. In turn this leads to better outcomes for society. For example, the Bank of England has a specific inflation target designed for macroeconomic stability that relies on statistics. Equally, labour market statistics can inform policy interventions in the job market by identifying where interventions are required and can be used in analyses determining which interventions were successful in the past and are likely to be successful in future. Moreover, economic statistics influence societal outcomes by changing consumer and business expectations and behaviour, in turn impacting real variables such as consumption and investment, and thus GDP.

[6] Available at http://www.chass.utoronto.ca/datalib/misc/Nilsen%20Economics%20Paper%202007%20final%20version.pdf

**Creating new products and services and refining/sustaining existing ones.** Economic statistics can be combined with other data and statistics to form the basis of new products and services, or supplement existing ones. For example, price inflation data could be used in pricing models for retailers or for determining real changes in financial assets; GDP data can be used in economic forecasts sold to customers across the economy; and labour market data is used to feed into, for example, multi-modal transport models considering the efficient movement of people at very local levels.

**Enhancing transparency and accountability.** Economic statistics, typically combined with other statistics, can be used to hold decision makers to account. Most obviously, they can be used to challenge the government of the day over their record in generating jobs, keeping price inflation on target and creating economic growth.

Other benefits typically identified from open data such as lower transaction costs, new platforms and improved service delivery are not found to be especially relevant for the economic statistics under consideration here.

Clearly these identified benefits are not static and should be expected to change according to the characteristics of the economic statistics in question and also the customer using the statistics (i.e. not all customers will receive the same benefits). It is the characteristics of the statistics that are likely to be the drivers of benefits.

One may hypothesise that as, say, robustness and quality improve, so will the quantum of benefits, all things remaining equal. However, the increase in benefit may not be proportional or linear.

### Applying a theory of change in assigning monetary values to the benefits of economic statistics

As we move toward ascribing a monetary value to the benefit of economic statistics, it is useful to posit a theory of change, which charts how these statistics contribute to outcomes and then impacts. An initial theory of change is set out below.

Figure 2.2.a: Theory of change for economic statistics

| Inputs | Description | Outputs | Description | Outcomes (benefits) | Description | Impacts (measurable) | Description |
|---|---|---|---|---|---|---|---|
| **Data collection via surveys** | ONS business as usual data collection on prices, national accounts and labour market | **GDP statistics** | Downloadable datasets from ONS website, plus also paid-for data and associated meta-data | **Better decision making** | Benefits as described above accruing across direct customers | **Business revenues** | Additional revenue generated from products and services that use economic statistics (Gross Output) |
| | | **Price inflation statistics** | | **New products and services** | | **Costs saved** | The value of costs saved through the use of economic statistics |
| | | | | | | **Output and Productivity** | Changes in output (Value Added) and productivity attributable to |

| Inputs | Description | Outputs | Description | Outcomes (benefits) | Description | Impacts (measurable) | Description |
|---|---|---|---|---|---|---|---|
| | | | | | | | economic statistics |
| | | | | | | **Job creation and new business creation** | New jobs and businesses attributable to the use of economic statistics |
| | | **Labour market statistics** | | **Refinement of existing products and services** | | **Tax receipts** | Direct and indirect tax receipts to HMRC due to the application of economic statistics |
| | | | | | | **Welfare gain** | The value of change in wellbeing through use of economic statistics, e.g. lower prices |

| Inputs | Description | Outputs | Description | Outcomes (benefits) | Description | Impacts (measurable) | Description |
|---|---|---|---|---|---|---|---|
| | | | | | | | transferring surplus to consumers |
| | | | | **Enhancing transparency and accountability** | | **Option value from more choice** | The value derived from customers of additional choice in economic statistics |
| | | | | | | **Other specific benefits** | Benefits that might only arise in very specific circumstances with specific statistics, that are not accounted for elsewhere |

Source: Deloitte, based on literature reviewed

The benefits of economic statistics can be traced in the above, preliminary, theory of change. It is important to note that while the outcomes affect users (i.e. those actually using/re-using/redistributing the statistics), the impacts are society- and economy-wide, i.e. the downstream impacts benefit the public at large as well as users.

The above table should not be taken to mean the relationship between inputs, outputs, outcomes and impacts is linear – there will be feedback loops and discontinuities depending on the statistical dataset and user. Subsequent analyses must also take into account the counterfactual, i.e. in absolute terms what would have happened anyway in the absence of the ONS producing these economic statistics. Or, in relative terms, where proposed changes are tested, what the do-nothing/do-minimum outcome would be.

Each of the metrics lend themselves to monetisation to measure the size of the impact. Prospective metrics include, inter alia:

- Revenue, £
- Costs avoided, £
- Change in producer surplus (profit), £
- Delta in productivity growth, %
- Delta in GDP growth, %
- Return on Investment/NPV, £
- Benefit to Cost Ratio (BCR), ratio
- Tax receipts, £
- Change in consumer surplus, £
- Option value, £

Whilst the majority of these might be termed 'market benefits' – in that they can be quantified as a result of market activity (through the observation of quantity and prices where they exists), consumer surplus and option valuation are 'non-market benefits' because they are externalities not embedded in market valuation/transactions.

This is of relevance to the study in a number of ways, and most pertinently because the statistics ONS produced are mostly free-at-the-point-of-use, meaning the initial value of the statistics have to be determined as a 'non-market' benefit. The next section explores the different ways in which the quantum of the impacts can be measured through the literature review.

## 2.3 Previous studies estimating the value of data

We have conducted a high-level literature review on different methodologies to assess the value of data. These methodologies could subsequently be combined to produce a hybrid method, with complimentary use of qualitative and quantitative approaches highlighting the various dimensions of value and quality.

We consider perception of value estimates, return on investment approaches, market-based benefit approaches, non-market-based approaches, avoided cost approaches, computable general equilibrium models, dynamic welfare approaches and case study approaches. A summary of the methods follows at the end of the study review.

Below we provide brief outlines of the main studies identified;

1. **Bank of England (2006) Cost-benefit analysis of monetary and financial statistics, UK**

The Bank of England used a cost-benefit analysis framework to discuss a monetary estimate of the benefit of financial statistics.

The project took into account the wide variety of users of economic statistics. Given the inherent difficulties in putting a monetary value to the benefits of statistics, attention in the project focused on assessing relative benefits.

As a first step, a survey of users from different parts of the Bank of England was undertaken in which views were sought on the relative importance of various uses of the Banks monetary and financial data. The survey asked users about the importance of a number of different activities and about the contribution to those made by monetary and financial data.

The form took account of the following dimensions:

- Policy use: (percentage weight up to 25) the highest marks are given to data that contributes to the assessment and maintenance of monetary and financial stability, or that are used directly in the National Accounts, in line with views from the internal survey;
- Policy relevance: (percentage weight up to 25) this captures the importance of data to the principal policy use(s) and decisions identified under the previous criterion. This is a subjective judgement that will vary dependant on the precise policy use and the information concerned;
- Meeting international standards and additional uses: (percentage weight up to 25) these are given as additional marks to capture the incremental benefit where data is required by law;
- Value added: (percentage weight up to 15) this section captures the gain from these data over and above what is available elsewhere; and

- Quality: (percentage weight up to 10) this section looks at the underlying statistical quality of the data – how good is the data are there frequent revisions, do they correspond well with other data series?

This approach avoids the need for monetary valuation by focusing on whether the benefit from a particular collection is high or low. But it does not avoid the need for a subjective judgement on the relative importance of different users and on the contribution of monetary and financial data to those uses. The information from the survey also only gives a partial and indicative picture of the benefits from these statistics.

**Insight**

This study shows the **benefit of surveys** to quantify value but could benefit from utilising **additional methodologies to reduce subjectivity**.

2. **Danish Enterprise and Construction Authority (2010) The Value of Danish address data: Social benefits from the 2002 agreement of procuring address data and other data free of charge, Denmark**

The Danish Enterprise and Construction Authority used an avoided cost approach to value Danish address data.

The study is based on usage information from 22 data distributors which disseminate free of charge the address data via the Public Data Server (the PDS). The authors only included the direct financial benefits for the more than 1,200 parties receiving address data from a PDS distributor. The assessment included the savings made because enterprises and municipalities no longer have to use resources to procure data.

The value calculations in the assessment are based on an assumption that the economic value of the free-of-charge addresses in the individual IT solution corresponds to the price users actually paid for municipalities address data before the free-of-charge agreement. The study further clarified these figures by benchmarking against the current price of similar data in the EU.

The supplementary financial benefits that arise in later parts of the distribution chain have not been included in the analysis. For example, this may be applicable if the party receiving data from a distributor is a supplier of data for GPS systems. The financial benefits linked to reuse of free address data in the third, fourth, fifth and subsequent link as described above, have not been included in the value assessment, however they are likewise expected to be of a considerable size.

3. **Pollock, Stephan and Valimaki (2010) The Value of the EU Public Domain, EU**

Pollock adopted a bottom-up approach to estimating the economic value of copyright and public domain material in the EU and saw the net economic value of material as the willingness to pay for material minus the cost of supplying it.

Data included in this study included royalty data, music sales data, music recordings data, and music distribution data.

To estimate the value of this data, the study looked at three distinct but related areas. First, the usage of public domain material, that is how much public domain material is bought, broadcast, downloaded etc. Second, the price differences between copyright and public domain material. Third, the differences in usage corresponding to those differences in price.

The authors approach was to first estimate the deadweight loss of copyright (conversely the value of the public domain) generally. That is, Pollock estimated this as a function of a few key parameters defining the demand curve. When data was insufficient to estimate the full demand system Pollock employed a reduced form approach in which he sought direct estimates for key variables (price changes, demand elasticities etc.) and then combined these with a particular parameterisation for the demand curve. Combining these averages allowed estimates for impact on the distribution of sales across works and time. Finally, putting this together with figures for the size of the public domain allowed Pollock to obtain an overall estimate for the total value.

The primary assumption in this study is that the direct estimates for key variables are valid.

**Insight**

This methodology shows how **willingness to pay can be estimated** through key variables such as price changes and demand elasticities.

4. **Gueber and Trip (2011) Economic Impact of the Human Genome Project, US**

For this Battelle Memorial Institute funded research, Gueber and Trip used an Input-Output approach to explore the economic impact of the Human Genome Project on the US economy. The study demonstrated that data generated real value, even if potentially overstated. To evaluate genomics-enabled industry impacts in the U.S., Gueber and Trip constructed a “from the ground-up” database

of individual companies engaged within the sector. The employment of this industry base was used as the foundation for an input/output analysis to quantify the total impacts of these firms (in terms of direct and indirect output, employment and their multiplier effect).

Data involved in this study includes historical R&D data, historical employment data using the National Establishment Time-Series (NETS) database, personal income, economic output state and local tax revenue and federal tax revenue.

In terms of assumptions, It is worth noting that the study suggested a very high multiplier of 141 (i.e., benefits of $141 for every $1 of US Federal Government funding).

**Insight**

This approach illustrates how **employment numbers** of the relevant industry base can be used to **demonstrate value**.

5. **Beagrie and University of Victoria (2012) Economic Impact Evaluation of the Economic and Social Data Services (ESDS), UK**

The conceptual framework and starting points for the wider assessment of benefits in this study was the Keeping Research Data Safe (KRDS) Benefits framework. The KRDS Framework is a tool for identifying, assessing and communicating the benefits from investing resources from research data. It breaks benefits into direct and indirect outcomes and internal and external stakeholder benefits.

The economic analysis used a range of approaches starting with the most immediate and direct measures of values and moving outwards to estimates of the wider economic benefits. These included investment and use value, contingent value (the amount users would be willing to pay to access ESDS data and services), consumer surplus, net economic value and efficiency gains.

The report made use of various assumptions. For example, activity times have been converted to costs using the annual average salaries for academic staff and graduates reported in the most recent Times Higher Education Salary Surveys. Willingness to pay of users is collected through a survey and in terms of an annual fee and pay-per-access basis and it is assumed that these figures are accurate.

**Insight**

This approach **shows the benefits of a hybrid approach** in terms of assessing the wide impacts of value.

6. **Deloitte (2013), Market Assessment of Public Sector Information for BIS on behalf of the Shakespeare Review into Public Sector Information, UK**

Deloitte used a willingness to pay methodology alongside sensitivity analysis to value the contribution of public sector information to the UK economy. The report went on to use case studies to quantify downstream impacts beyond the direct customer.

The report adopted a three-stage approach to valuation of public sector information. Stage one estimated the value of public sector information to Public Sector Information Holders (PSIHs) and the value to direct consumers (users and re-users) of public sector information using a bottom up approach that quantified consumer and producer surplus. Stage two estimated the value of associated indirect and induced impacts to PSIHs using Input-Output analysis and stage three estimated a ready-reckoner value of wider value based on other available research.

Quantity or usage of information was based on the number of downloads and page views collected. Estimates of the value accruing through the business-to-business supply-chain and employees spending associated wages are based upon the UK Domestic Use Matrix (DUM). Estimates of producer surplus were converted into expected gross output (GO) for each relevant industry on the basis of information contained in the UK DUM. Per-worker productivity estimates were sourced from ONS through a combination of national accounts data, the Business Register and Employment Survey and the Annual Business Survey.

A primary assumption for this study was that for free-at-the-point-of-use public sector information, the choke price (the lowest price at which quantity demand for an item is equal to zero) for raw data was assumed to be equal to the choke price of paid-for public sector information minus the current price of data.

**Insight**

This study demonstrates how the **price of paid-for-public sector information can be used to estimate** the value of open data and statistics.

7. **ConsultingWhere and ACIL Tasman (2013) OS OpenData Economic Value Study, UK**

This study used a Computable General Equilibrium (CGE) model to evaluate the value of OS OpenData to the UK.

The primary input into the study was quantitative data from interviews with 20 firms. Further Inputs into the study included download records whereby information supplied covers, date of download, breakdown by sector (according to OS categorisation), single use figures of product and routine use figures or the number of instances where repeated downloads of the same product have been made and time spent evaluating the data downloaded.

A "bottom-up" technique was used for the inputs to the study based upon case studies and market intelligence which were cross checked with information provided by OS and publically available market statistics and data. Given that the experience of previous studies indicated the greater difficulty in identifying benefits than costs, the study adopted a considered but cautious approach and chose not to include wider social welfare benefits within the CGE model but to present these qualitatively.

In the feasibility review of this report it was identified that the assumptions used in "grossing up" the sub-sector impacts, based on the case studies were critical to establishing the credibility of the study.

The primary assumption that this report makes is that the 20 firms interviewed are representative of those who are downloading data from OS. However, in practice, as they are those who can articulate the benefits they may be generating greater benefit than the average user.

For each case study, the paper assumed that a high download count of the major businesses (by turnover) particularly where the business has made repeated accesses shows that "grossing up" is likely to be valid and that the case study is not unique. However there is limited evidence underpinning the choice of sensitivities applied to this analysis.

**Insight**

CGE's models rely on **multiple assumptions** and these leave them **susceptible to criticism** in business cases of the nature required by ONS.

### 8. McKinsey Global Institute (2013) Open data: Unlocking innovation and performance with liquid information, US

McKinsey utilised perception of value estimates to estimate the impact of open data in the US. Their approach focused on examining microeconomic industry trends to better understand the broad macroeconomic forces affecting business strategy and public policy.

McKinsey utilised a variety of inputs including annual spending of sectors impacted by open data, visits to open data sources by sector and organisation, paid for open data requests by sector and organisation, a range of open data sources by sector and investment and projected investment into open data sources by sector.

To quantify the impact of Open Data. McKinsey focused its efforts on seven sectors, namely, education, transportation, consumer products, electricity, oil and gas, health care and consumer finance. For each of these sectors, McKinsey identified ways that open data might create economic value, explored potential barriers to adoption and considered which actions would be required for capturing value with open data. McKinsey estimated how much annual value to the economy each sector might help enable through case studies and interviews.

McKinsey assumes that perception of value can be subjectively and accurately quantified by the user. This approach does not account for wider societal benefits.

**Insight**

McKinsey demonstrate how **case studies and interviews can be used to demonstrate value** although this may not be fit for purpose in terms of developing a mechanism for valuation that can be easily re-used.

### 9. PwC (2015) Open Data Challenge Series Final Report, UK

PwC used a Return on Investment (RoI) methodology to evaluate the value of an Open Data Challenge Series programme to the UK economy.

The approach involved an assessment of benefits and costs through a review of applicants, participants and secondary data sources such as research papers and the valuation metrics in the Cabinet Offices' Unit Cost Database.

Inputs into the methodology included application data, SIC classification of applicants, cost of resources allocated to the programme, staff resources, funding given to participants and event space hire.

PwC further built on its existing methodology by including a Dynamic Welfare methodology. Inputs into this area included the three year sales projections and employment plan of product users, the time period to finalising the business case, the expenditure profile of product users, other funding secured by product users and the customer base and profiles of product users. In order to provide an alternative set of outcomes (given uncertainty regarding the finalists impact projections) two potential scenarios were considered. These were in the worst

case that all of the projects fails and second that the failure rate of projects is around 40%.

There are various limitations in this approach and these include the assumption that all finalists will continue operations regardless of whether they received Open Data Challenge Series Funding. In addition, the project applied impact measurement estimates of only 5% of the beneficiaries identified. Due to the aims of this study in creating a model that can be used in HM Treasury Businesses Cases a Green Book approach is preferred.

**Insight**

PwC's methodology may be challenging to apply for ONS as primary users are government bodies and larger businesses who **may not be willing or able to share confidential information.**

### 10. London Economics (2015) Met Office, General Review, UK

London Economics utilised a market-based benefit approach and avoided cost approach to produce an aggregate estimate of the Met Office's potential impact on the UK economy over the next ten years.

Inputs into this methodology included planned life expectancy of Met Office assets and planned investment into Met Office assets, Met Office spend on observation data, staff and communication of forecasts and services and Met Office changes in observation, reach and scientist, modeller and forecaster inputs.

Three main areas of sensitivity analysis were identified for the investigation. These are how estimates of net economic value react to variation in the frequency of high impact weather events, how the key inputs, outputs, and outcomes, in relation to weather analysis impact economic benefit and sensitivity analysis around levels of investment. Two primary counterfactual scenarios to the base case were analysed, to identify the marginal benefits and cost savings of a standard versus the current work-leading weather and climate service and a service which separates weather and climate reporting.

Perception of value estimations are calculated through surveys that capture public perceptions of value. Where this is not possible the analysis used avoided cost approaches. The analysis created a base case which compares to a 'do-nothing' counterfactual where there is no Met Office. In some cases there are significant benefits which the study has been unable to quantify. In which case these benefits are identified and qualitative evidence for their potential relative scale is given.

Whilst the non-existence of the Met Office is a strong assumption, it is consistent with previous studies and allows the full impact of the Met Office to be estimated.

**Insight**

An avoided cost approach may be difficult to use for ONS as it will be **challenging to quantify a base case in which ONS does not exist** or statistics are so bad as to cause real issues for customers. Perception of value estimates are preferred although these may be **less timely to execute** as they rely on surveys.

### 11. Beagrie (2016) The Value and Impact of the European Bioinformatics Institute, UK

Beagrie uses a stated preference methodology to estimate the value and economic impact of the European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI) which manages public life science data making its resources freely available to the global life science community.

The quantitative economic approaches used included: estimates of access time and use value, contingent valuations using stated preference techniques, an activity-costing approach to estimate the efficiency impacts of EMBL-EBI data and services, and a macro-economic approach that seeks to explore the impacts of EMBL-EBI use on returns to investment in research. Survey responses to questions about time to access and obtain the last data used were collected as was willingness to pay information. User registration was estimated by combining information from user survey's and external studies. To this end the Beagrie study used a hybrid approach, which saw some aggregation but some presentation of results from competing methods.

**Insight**

This study shows the **benefits of hybrid approaches** in that they present results from competing methods and **through ranges allowing for greater accuracy** in quantifying value.

### 12. Goodridge and Haskel (2015) How does big data affect GDP? Theory and evidence for the UK

An assumption of the study is that activity times can be converted to costs by assigning each respondent to the Beagrie survey to a salary group. This is primary based on the UK Times Higher Education Salary Survey and information from the UK Department of Education for 2014-2015, then scaling to include non-wage labour costs using a 30 per cent uplift based on the HM Treasury

Green Book method. For students, school leavers and graduates, average salaries reported in the UK Complete University Guide were used, to reflect the opportunity cost of earnings forgone. Non-academic respondents were allocated to a comparable academic staff levels and salaries. Goodridge and Haskel present a conceptual approach to measuring the impact of Big Data on GDP and GDP growth. To do so they employ an upstream-downstream framework/production function that considers two upstream sectors, these sectors are the data building sector and the knowledge creation sector, as well as users.

Data used in the study includes data from ONS: market sector value-added, and sector data from nine broad industries (this excludes the public sector, sectors that are involved in private delivery and the real estate sector). For labour composition and hours worked ONS Quality-adjusted labour input (QALI) data is used and tangible capital labour variables are based on estimates from Oulton and Wallis (2014). The National account measure, publicly available social media data and industry expenditure data on R&D from the Business Enterprise R&D survey are also all used.

The approach taken by Goodridge and Haskel is to assume that it is not Big Data per se that affects output but the knowledge gleaned from Big Data. Big Data is therefore treated as an intangible asset that contributes to output and spending on creation and knowledge generation as investments into that intangible asset. In other words, data is a derived demand – derived from the desire for knowledge and insight. Goodridge and Haskel start by arguing that investment in Big Data can be thought of as having two stages (a) data-building and (b) knowledge creation. In the first stage raw records are transformed into "information" that is data in a usable format. The second analysis of such data produces "knowledge" that is useful insights from the information. The knowledge asset is then used as an input in final production of goods and services.

Goodridge and Haskel also use new estimates of employment and investment in Big Data as set out in Chebli, Goodridge et al (2015) and Goodridge and Haskel (2015a). This is not an exact science as investment data is generated via spending on workers who are producing knowledge assets based on Big Data. Due to the wide-spread usage of Big Data, exact figures of spending on workers and exact calculation of those employed in the Big Data sector is open to discussion.

Goodridge and Haskel place of value of c. £1.6 billion on the value of big data use in GDP terms (technically above and beyond that which is already captured in GDP), and suggest that big data has added £152 million to UK GDP growth over the period 2005-12. They expect these contributions to increase over time as data becomes even more pervasive.

This study shows how an academic approach can be taken to quantify the contribution of data/statistics through the specification and econometric analysis of a **production function**.

Insight

### 13. UK Data Service (2016), Communicating for Impact 2012-2017, UK

The UK Data Service is a comprehensive resource funded by the Economic and Social Research Council (ERSC) to support researches, teachers and policymakers who depend on high-quality social and economic data. UK Data Service is noted as following best practise in quantification of value (by Eurostat), which is why it is included in this review in spite of not being completed at the tie of publication. They use case studies and interaction to quantify their impact and are planning on using the ESDS economic impact assessment, assessed earlier on in the literature review, in 2017 to quantify their value.

To quantify impact UK Data Service have a programme called Discover that details more than 13,000 research outputs based on ESRC research project data. Each Data Collection record in Discover also lists selected primary and secondary use publications arising from that particular collection. UK Data service further uses a unique Digital Object Identifier to search for all publications citing international data to quantify impact. Finally, UK Data Service have collected over 100 case studies that demonstrate how data has been used in research studies and teaching materials. Impact of the service for is further measured through number of unique visitors, number of page views, and number of registered users.

The UK Data Service will conduct an end-of contract economic impact assessment in 2017. This is expected to draw of the principles scope and design of the 2011 ESDS economic impact assessment which is covered earlier on in the literature review. The ESDS economic impact assessment broke economic benefits into direct and indirect outcomes and internal and external stakeholder benefits. The economic analysis used a hybrid approach which included investment and use value, contingent value (willingness to pay), consumer surplus, net economic value and efficiency gains. The report made use of various assumptions including converting activity time of users to cost using the average annual salaries for academic staff and graduates.

Insight

The UK Data Service has been mentioned by Eurostat as following best practise in its efforts to quantify the benefit of statistics. This suggests that ESDS economic impact assessments, case studies and hybrid approaches could be well supported.

## 2.4 A case study – borrowing from transport

**Borrowing from transport appraisal: a valuation study with Soft Factors and WTP.**

It is common practise for business cases in transport and particularly those considering public transport, to consider 'soft factors' as monetised benefits arising from transport usage.

As an example, soft factors include vehicle quality, driver quality and ease of use or real-time information provision. Typically stated preference techniques are applied through a range of user surveys which yield monetised values, or willingness-to-pay (WTP) for these factors such that trade-offs in these factors can be understood and fed into business case analysis. Typically these studies only consider users, rather than non-users who might switch to public transport.

Recent work in the field has been conducted by specialist transport consultancies such as AECOM and Steer Davis Gleave (SDG) to update soft factor values. Academic and Government studies from the 1990's/2000' such as the former DETR's paper are a useful starting point to understand the methods and limitations involved, and DfT's WebTAG (web-based transport analysis and guidance) has some useful guidance on use in business cases. The methods are not without criticism, but they continue to be used and we understand a refresh of soft factors for business usage is underway and due late 2016.

Typically a study will revolve around a stated preference survey of users split into two components. This includes questions around trade-offs between soft factors on public transport as well as trade-offs between competing modes of transport. This can be complemented by revealed preference analysis as a separate way of valuation or as a sense check.

Results are typically analysed in a statistical setting (typically a multinomial logic? model or similar) to provide monetised benefits of 'quality attributes' and or equivalent 'journey time savings' for use in transport models that use soft factors.) Valuing statistics in this way is in some ways easier and in some ways harder than the transport example. The user/non-user/mode-switch distinction is not important with statistics (given a much bigger restriction on substitutes) and the attendant complexities are therefore not relevant. However, public transport is largely paid for at the point-of-use, which gives market values against which to consider the soft factors. Statistics are largely free at the point of use and are thus a special case, which means that novel techniques would still be required to yield monetised benefits.

Nonetheless, stated and revealed preference techniques yielding values of quality in statistics may be a useful method for ONS to employ.

## 2.5 Summary of Literature Review

As attempts to quantify the benefits of data and statistics have matured there has been an increase in the use of hybrid models to assess this value of statistics. This is an approach that ONS could take as it limits reliance of assumptions inherent in some methodologies. Due to the large scale of ONS some of the approaches identified are more suited for statistics that have smaller impact and therefore it is likely that a bespoke solution will need to be identified and implemented for ONS. The studies are summarised overleaf.

Figure 2.5.a: Summary of studies

| Study | Methodology | Data | Assumptions | Criticisms |
|---|---|---|---|---|
| 1.Cost-benefit analysis of monetary and financial statistics | Cost-benefit analysis | Survey data the contribution that statistics make in terms of policy use, policy relevance, international standards, value added and quality | That users of statistics make can objectively assess their contribution | The information from the survey only gives a partial and indicative picture of benefits from statistics |
| 2.The Value of Danish address data | Avoided-cost approach | Usage information on free-of-charge Danish address data from 22 data distributors | That the economic value of free-of-charge address data corresponds to its price before the free-of-charge agreement | Supplementary financial benefits that arise in later parts of the distribution chain have not been included in this analysis |
| 3.The Value of the EU Public Domain | Willingness to pay minus the cost of supplying data | Royalty, sales, recordings and distribution data | Direct estimates for key variables are valid | That direct estimates for key variables (price changes, demand elasticities etc.) do not sufficiently accurately predict a consumers demand curve |
| 4.Economic Impact of the Human Genome Project | Input-Output approach | Historical R&D data, historical employment data using the National Establishment Time- | A multiplier of 141 (i.e., benefits of $141 for every $1 of US Federal | That the multiplier assumed was too high and not sufficiently justified |

| | | | | |
|---|---|---|---|---|
| | | Series (NETS) database, personal income, economic output state and local tax revenue and federal tax revenue. | Government funding) was assumed | |
| 5.Economic Impact Evaluation of the Economic and Social Data Services | A hybrid approach involving immediate and direct measures, investment and use value, contingent value, consumer surplus, net economic value and efficiency gains | ESDS database, download data, ESDS Annual Reports, case studies, user experience data and numerous other data elements collected through surveys | A primary assumption is that activity time can be converted to costs using the annual average salaries for academic staff and graduates reported in the most recent Times Higher Education Salary Surveys | Willingness to pay of users is collected through survey information and these figures may not be accurate |
| 6.Market Assessment of Public Sector Information | A hybrid approach of willingness to pay, sensitivity analysis and case studies | Download and page view data, UK Domestic Use Matrix, national accounts data, the Business register and employment survey and the annual business survey | For free public sector information, the choke price for raw data was assumed to be equal to the choke price of paid-for public sector information minus the current price of data | Criticisms focused around the simplification of the wider impacts that public sector information may have |
| 7.OS Open Data Economic Value Study | Computable General Equilibrium (CGE) model | Quantitative data from interviews with 20 firms and download records | The primary assumption that this report makes is that the 20 firms interviewed are | There is limited evidence underpinning the choice of sensitivities applied to the analysis |

| | | | | |
|---|---|---|---|---|
| | | | representative of those who are downloading data from OS | |
| 8.Open Data: Unlocking innovation and performance with liquid information | Perception of value estimates | Annual spending of sectors impacted by open data, visits to open data sources by sector and organisation, paid for open data requests by sector and organisation, a range of open data sources by sector and investment and projected investment into open data sources by sector | Perception of value is sufficiently quantified through theoretical considerations underpinned by user's perception of the value that open data generates | McKinsey's focus on seven sectors impacted by open data is too narrow |
| 9.Open Data Challenge Series | Return on Investment and Dynamic Welfare methodology | Participants and secondary data sources such as research papers and the valuation metrics in the Cabinet Offices' Unit Cost Database | The assumption is made that all finalists will continue operations regardless of whether they received Open Data Challenge Series Funding | The project did not undertake an HM Treasury Green Book compliant 'additionality' analysis of the projected outputs and outcomes |
| 10.Met Office, General Review | Market-based benefit approach and avoided cost approach | Life expectancy of Met Office assets and planned investment into | Whilst the non-existence of the Met Office is a strong assumption it is | In some cases there are significant benefits that |

| | | | | |
|---|---|---|---|---|
| | | Met Office assets, Met Office spend on observation data, staff and communication of forecasts and services and Met Office changes in observation, reach and scientist, modeller and forecaster inputs. | consistent with previous studies and allows the full impact of the Met Office to be estimated | this study has been unable to quantify |
| 11.The Value and Impact of the European Bioinformatics Institute | Hybrid approach including a stated preference methodology | Estimates of access time and use value, efficiency impacts, user registration | An assumption of the study is that activity times can be converted to costs by assigning each respondent to the survey to a salary group | Willingness to pay figures and user data based on surveys may not be accurate |
| 12.How does big data affect GDP? Theory and evidence for the UK | Upstream-downstream approach using a production function and to estimate GDP/welfare benefits | Selection of official statistics on output and employment by sector married with other sources (including previous academic research and databases) | The authors assume that it is not Big Data per se that affects output but the knowledge gleaned from Big Data. Usage determines value. | Openly recognises some assumptions are required to arrive at a set of estimates. |
| 13.UK Open Data Communicating for Impact 2012-2017 | Case studies, interaction and ESDS economic impact assessment | more than 13,000 research outputs based on ESRC research project data, 100 case studies , user experience | Various assumptions including converting activity time of users to cost using the average annual salaries for | By only making use of case studies and interaction to data a value of the data has not |

| | | | | |
|---|---|---|---|---|
| | | data and numerous other data elements collected through surveys | academic staff and graduates | yet been conclusively estimated |

Source: Literature Studies and Deloitte Analysis

# 3 Available valuation methodologies

To allow the selection of a valuation methodology for ONS it is necessary to triangulate the findings of the literature review and stakeholder opinion on the relative efficacy of the various methodologies, with the aims and objective of ONS.

Before we do this in chapters 5-6, this chapter summaries the findings of the literature review with respect to the task at hand.

The list below begins to draw out the methodology benefits/burdens of potential approaches as identified in the literature.

1. **Perception of value estimates:**
   - Using stated preference or revealed preference surveys to estimate consumers' willingness to pay/accept for data provision/loss – it focuses on the impact component of the theory of change;
   - These techniques are particularly useful where the data is provided free of charge;
   - A comparison between willingness to pay and what was actually paid (even if zero) can be made to derive consumer surplus, a measure of welfare;
   - However, perception of value estimates will typically miss or under-estimate downstream benefits of data unless explicitly asked about. Such approaches are also static and average, rather than dynamic/marginal; and
   - Most applications of stated preference are intended to identify estimates of relative utility weights rather than absolute values and under these conditions worries about potential over or indeed under-statement are related to any nominal values to which the weights are applied.

**Insight**

This approach estimates users' willingness to pay or avoid costs and is especially **useful for free at the point of use statistics** such as GDP and inflation.

2. **Return on Investment approaches:**
   - These approaches calculate the cost (capital and operating) involved in generating the statistics (the input component of the theory of change) and then compare this to the expenditure made by direct customers (users, re-users and redistributors) – thus it requires data on the price paid and other expenditure made by users of statistics ;
   - To capture indirect or downstream customers, assumptions are made to estimate the expenditure that would have taken place;
   - This top down approach then calculates the return on investment as the ratio between total expenditure on data and the cost of its collection/generation; and
   - Another criticism levelled against this approach is that it can overstate the value of data as it does not explicitly account for the counterfactual unless treated as a separate option/scenario.

**Insight**

Return on Investment approaches can be useful in a hybrid approach but **rely on multiple assumptions** and require detailed information on user costs

3. **Market-based benefit approaches:**
   - These approaches also focus on impacts in the theory of change, seeking to put a market value on the benefits upstream and downstream of the use, re-use and redistribution of data;
   - This approach uses price data (where available) to capture impacts beyond the price charged (or not charged) for the data;
   - For example, this may include applying a multiplier to account for supply-chain impacts or data re-use;
   - In the case of unpaid data, the approach relies on valid proxies being available to give an indication of value for substitute products and services; and
   - One criticism of such approaches is that they are susceptible to double-counting if a number of different market-based approaches are used to evaluate different benefits associated with the same dataset.

**Insight**

This approach can focus on the **wider societal aspects** of value but is reliant on a number of assumptions around paid-for data and statistics being related closely to the value of unpaid statistics.

4. **Non market-based approaches:**
   - In the same way prices can be used as a means of imputing benefits; changes in non-price specific outputs can also be quantified using relevant tools and techniques;
   - These 'non-market' benefits are externalities benefiting/harming individuals outside any market transaction. For example, in the case of meteorological data, if weather data is harnessed to improve travel flows, benefits can be estimated in the form of time saved from shorter journey times across an aggregation of travellers not paying any price for the improvement in outcomes; and
   - This approach explicitly considers the counterfactual by estimating the change in outcomes and can be added to market-based estimates, but this approach usually relies on some form of guidance or assumption for monetisation in the absence of market prices.

**Insight**

This method is particularly useful for quantifying **wider societal impacts** and can be added to market impacts to generate an estimate of welfare including market transactions and externalities

5. **Avoided costs approaches:**
   - In some ways, this is the flipside of the market based approach, in that it seeks to estimate the costs saved by users and re-users in using the data rather than the benefits created by it per se;
   - For example, again using meteorological data examples, data can be used to more accurately predict weather events and patterns, which in turn can allow for effective mitigations to be put in place quicker or have more accurate insurance models, which in turn save costs if risk is appropriately mitigated against;
   - As highlighted before, this is one interpretation of the way ONS has conducted business case analysis in the past – minimising costs rather than considering benefits. The main difference being a full avoided cost approach includes the costs to third parties and not just ONS; and
   - By inverting a negative avoided cost it may be treated as a positive benefit stream. By definition this approach also has to include the counterfactual. In the case of statistics, the counterfactual can be as simple as the (increased) cost to users of having to create/source equivalent information for use instead of the statistics (which may be an inferior product).

**Insight**

This approach may be difficult to apply in practice as it is likely to be **difficult to understand a counterfactual situation** without ONS, without extensive consultation. Consultation may yield better outcomes as part of a stated preference type approach – e.g. willingness to avoid

6. **Computable General Equilibrium (and Input-Output) models:**
   - The previous 5 methodologies are largely static and do not easily pick up impacts such as increased productivity in a dynamic setting. In contrast, CGE models can characterise the use/re-use/redistribution of data as 'shocks' to the macro-economy, the income shifts and productivity changes of which are then modelled;
   - These models are typically calibrated using a large set of assumptions on how data is used and who uses it, typically based on a sample of case studies and interviews. While these approaches can be comprehensive, they are subject to a number of criticisms around the choice of assumptions, their nature and sample size;
   - Such an approach would likely mean recalibration of inputs based on each specific case, which may be an issue for ONS in this setting;
   - A further issue of relevance is the relative scale of impact and suitability of a large macro model to tease out impacts; and
   - A fully specified CGE model of the UK will work in trillions or billions for any given year, when in reality the impact of changes in statistical provision are probably of a low-millions magnitude.

**Insight**

The CGE methodology relies on multiple assumptions and therefore might result in **a lack of buy-in from stakeholders**. Whilst the modelling system is robust, the extent of assumptions required in a complex system might not be

7. **Dynamic welfare approaches:**
   - These can build on willingness to pay frameworks, but look to further model efficiency aspects of data in competition and innovation over time and often use scenarios to capture the range of expected benefits;
   - Modern welfare economics adds to the dynamic welfare approaches by putting incentive constraints at centre stage. This analysis dispenses with the assumption that lump-sum transfers are feasible because of the incentive problems they create;
   - The dynamic welfare approach is not universally accepted, because different choices of the way to aggregate individual utility functions may yield different equilibria. The parameters of optimization in a dynamic welfare approach may be arbitrarily chosen or restricted by the rules of the game (e.g., by law); and
   - Welfare economic approaches to the policy process have been criticised by those operating in the public choice tradition for failing to consider how actual policy choices are made. Thus, even if optimal policies are identified, there is no guarantee that the decisions making institutions that we observe in reality will bring them about. The public choice critique of welfare economics says that, by failing to model government, dynamic welfare approaches provide a misleading view.

**Insight**

The dynamic welfare approach is controversial due to the **planners control over parameters** and has been criticised for **failing to model government**.

8. **Case studies**
   - Case studies provide concrete examples and often highlight the mechanisms through which impacts can be realised;
   - They are widely used in the evaluation of research facilities and activities and can focus on the scientific, economic and/or wider social impacts; and
   - They are, however, limited because it is not always possible or robust to scale up a case study to estimate overall impacts. Consequently, case studies are often combined with broader economic estimates and/or formal frameworks for analysis to add colour to quantitative analysis rather than replace it.

**Insight**

Case studies can provide useful examples and are most often **combined with other methodologies** in order to quantify value.

Assessing the value and impacts of statistics is a relatively new consideration and no single approach dominates or stands out. Therefore, the methodology chosen is the one most suited to the needs and requirements of ONS and their users.

In order to select the optimal methodology the following sift criteria are considered in chapter 5; simplicity, transparency, flexibility, replicability, specificity, data availability, robustness (buy-in) and cost effectiveness (financial & other).

# 4 User and usage data

## 4.1 Purpose

To fully appraise the long-list of methodologies available to ONS, it is necessary to develop an understanding of the data available from ONS management information that pertains to statistics and their use.

This is subsequently used as one of eight sift criteria to analyse the suitability of the eight broad methods covered through the literature review conducted in chapter 2.

## 4.2 ONS data availability

The management information made available by ONS provided us with data bracketed into the following categories:

- General statistical usage indicators including:
  - statistical type/domain/webpage;
  - website visits by web area;
  - downloads;
  - unique IP addresses;
  - download formats;
- All users across sectors and organisation (but not re-users), including:
- Paid-for data requests and licensing data (where relevant), including, inter alia:
  - the revenue from data provision;
  - ONS output code;
  - the statistical directorate/division which provided the data
  - the cost centre/area which provided the data;
  - the thematic area the statistics pertain to;
  - the customers paying for the data;
  - specific data pertaining to NOMIS labour market statistics
- Media analysis and KPIs (as a simple proxy for quality), which include:
  - article coverage/mentions by publication
  - negative/positive articles
  - other, more qualitative information.

## 4.3 Users of statistics

The user base of GDP, inflation and labour market statistics is wide and crosses central and local government, banking, professional services and a number of other major sectors, as well as the media and general public.

A single source of users is not available without significant analysis but ONS have confirmed to Deloitte their main users across the relevant domains, based upon user group and contact information. These organisations are not exhaustive across or within all sectors, but they do provide useful insight.

In our analysis, we have taken the major users identified by ONS, and added to the list with organisational turnover/budget, employment and SIC code. We have also added in other organisations where we can justify their inclusion on the basis of usage. These are predominantly organisations providing economic forecasts. We have chosen to add in those providing forecasts to HMT's monthly consensus report as they very obviously use all three statistical domains of GDP, inflation and labour market.

A picture starts to emerge of key users across the three domains in Government departments, other public bodies, banking and finance firms, and general and specialist consultancies. Together these have a combined (identified) turnover of £150 billion and are thus significant in scale to the UK economy.

Another useful way to consider the scale of type organisations using labour market statistics is to infer that those willing to pay for bespoke labour market commissions, are also users of non-paid data. There were 70 plus organisations who paid for labour market statistics in 2014, with a significant proportion of these organisations management consultancies or niche economic consultancies, who will be reliant to different extents on data and ONS statistics.

### 4.4 Unpaid data usage

In 2014 the alpha website disseminated almost 4 million statistical downloads to users **across all ONS statistics** rather than just the three domains of interest. Of these 2.3 million were machine readable data files such as .xls or .csv, with a further 1.6 million downloads of .pdf files that will have included statistical releases and reports in non-machine readable format and/or associated metadata/methodologies. In comparison, January 2016 alone saw 330,000 downloads on the alpha website, with 220,000 coming in data form rather than .pdf form.

It is clear from the alpha website usage data that not all visits culminate in a download, and of the total number of visitors, a significant proportion are return users. Over 3 in 10 visits seem to culminate in a download. Some users will return later to obtain statistics, some users will not have found what they need. Of course, even after download and examination by users there is no guarantee that the statistics downloaded will be used and/or create value. Nonetheless, downloads are likely to be the best proxy for end usage in any valuation methodology carried forward.

Nearly 90% of all website visits across the ONS suite of websites are on the alpha ons.gov.uk site, and over 75% of downloads also come from the alpha site. Downloads on the visual and open geography platforms amount to around 10,000 per month on this evidence.

This means that 80,000 downloads are made each month on the neighbourhood statistics and NOMIS platforms. Both platforms disseminate labour market data to varying degrees, with NOMIS designed for labour market statistics but also disseminating census statistics, and neighbourhood statistics giving some, if not perhaps as many statistical options from labour market sources.

We have additionally analysed 2014 ONS website data for the in scope statistical elements on a detailed basis as part of the quantification methodology. This involved significant amounts of data covering every ONS website and unique access point for each website in the three domains. More detail is given in chapter 8, but a unique access point is taken from an IP address and represents either an organisation, or an internet service provider where individuals are accessing ONS data. To this end, a unique access point does not equal a unique user.

We were able to identify many additional organisations that have downloaded data from ONS or viewed ONS webpages than those listed above. Further details of this and the major users by domain from their management information systems have been provided to ONS separately.

Chapter 8 also summarises relevant KPIs used in the quantification methodology for each domain and in aggregate across the three domains.

## 4.5 Paid data usage

Paid data analysis gives some indication of the willingness to pay for different types of data, and ONS collects information on revenues from clients across all data types in a systematic way due to the need to invoice and collect payments. It is thus a richer source of usage by user than unpaid data.

### Paid data: aggregate summary

At least 548 unique identifiable customers paid in some form for statistics produced and packaged by ONS in FY 14-15.

Net income to ONS for these paid statistics totalled £27.3 million over this period (which includes some adjustments in the form of creditors). £4.9 million in revenues is for a client or collection of clients who are not disclosed. Therefore the mean payment by each identifiable organisation was c. £41,000 – a significant amount.

The top 20 paying clients by value (including the null returns as a single entity), accounted for 88% of all paid-data revenue for ONS in FY 14-15. The top 50 paying clients accounted for 98.3% of revenues, and the top 100 paying clients accounted for 99.6% of revenues. There is therefore a long-tail of organisations willing to pay for small-scale bespoke data.

Turning attention to the top 100 clients by revenue, leads to some interesting findings. 73 of the top 100 organisations by revenue are defined as public sector (either departments, non-departmental public bodies or other government bodies) and these account for c. £21 million in revenue. 23 of the top 100 by revenue are private or third sector organisations with the majority of these 23 organisations being universities, accounting for £6m in revenue.

DWP alone accounted for nearly £6 million in bespoke data commissions, with BIS leading 5 other Government organisations all spending more than £1 million each of bespoke data. The private organisation spending the most in this space was ETC Venues, who spent in excess of £0.5m on bespoke commissions. Southampton University were the highest spending third sector/charity organisation – perhaps because they are a leading player in big data science.

This aggregate analysis tells us that organisations are willing to pay to obtain statistics relevant to their organisation or business needs. However, the payment could be as a result of two quite different reasons. Firstly the data might not be available at all or in the detail required if it is not paid for, or secondly the cost of paying ONS to extract the existing data may be lower than the cost to the organisation of having an employee extract and tabulate the data.

In essence the data tells us that clients are willing to pay for data for either reason; that they value the data enough to do so, but not how much they are willing-to-pay for the data above 'at cost' pricing, or the value it adds.

**Paid data: three domains**

In terms of the three statistical domains of interest to this study, there is little or no payment made to obtain GDP statistics given the availability of the data. Labour market and price inflation statistics are often paid for in bespoke commissions.

Bespoke data commissions for both labour market and inflation statistics both exceeded £100,000 in value for ONS in FY 14-15. IS and HMRC account for a significant amount of the payments made for labour market statistics, c. £100,000 from £113,000 in FY 14-15. Professional services firms also feature strongly along with universities.

For inflation data the main user(s) is not disclosed, but HM Inspectorate of Constabulary, the Bank of England and HMRC are the other main users along with the Royal Institute of Chartered Surveyors.

**A final point on value**

Of relevance later in the study when we consider willingness to pay, the average payment made by each identifiable organisation across all statistical types in FY14-15 was £41,000.

For price inflation statistics the equivalent estimate (adjusted for rebates) was c. £5,000, and for labour market commissions the equivalent estimate was c. £500. The labour market estimate contains Chancellor's Notices for FY14-15 at £60 each, which works to reduce the paid average because 106 of the 127 identifiable organisations paid some multiple of £60 for access to the data.

Whilst not suitable for use as a proxy for price, these do give some rough-order-of-magnitude parameters as to what organisations might be willing to pay for access to statistics more broadly, and at the very least that these organisations value the use of statistics to such an extent that they are willing to pay for them at all.

# 5 Sift criteria

This short chapter outlines best practice in options analysis, a discussion as to how best to account for quality in any quantification method, and the sift criteria used to determine which options are worth further consideration.

## 5.1 Best practice principles for options analysis

Government guidance on options formulation methodologies as well as the wider requirements of the appraisal process can be seen in "The Green Book: Appraisal and Evaluation in Central Government."[7]

Though not a Green Book exercise to choose the methodology for benefit quantification, it is worth using in considering the options available to ONS.

In particular, the Green Book recommends the following actions:

- *Research existing reports and consult widely with practitioners and experts to gather the set of data and information relevant to the objectives and scope of the problem.*
- *Analyse the data to understand significant dependencies, priorities, incentives and other drivers.*
- *From the research, identify best-practice solutions, including international examples if appropriate.*
- *Consider the full range of issues likely to affect the objective.*
- *Identify the full range of policy instruments or projects that may be used to meet the objectives.*
- *Develop and consider radical solutions. These options may not become part of the formal appraisal but can be helpful to test the parameters of feasible solutions. Well-run brainstorming sessions can help to generate such a range of ideas.*
- *A shortlist of options may be created, partly to keep the appraisal process manageable, usually at the preliminary stages of a policy appraisal. There is a risk that the process of short-listing will eliminate the optimal solution before it is given full consideration. Therefore, shortlists should still try to cover a wide range of potential action.*
- *The shortlist should always include the 'do minimum' or 'do nothing' options. Reasons behind the rejection of each excluded option should be recorded.*

Relevant evidence has been considered as part of the research documented in chapters 1 to 4. In order to move from a long-list to a short-list of options, we have identified a set of pre-determined sift criteria, against which each option can be evaluated.

The sift criteria reflect the above principles, in light of the specific requirements of ONS on this project. The criteria are applied to each option to allow a ranking of options in the long-

[7] See: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf

list. Potential sift criteria are discussed in this chapter, before relevant criteria are applied to the long-list of options in chapter 6.

## 5.2 A note on accounting for quality

As well as considering mechanisms through which statistics generate market and non-market value, and methods available to us for quantification, our analysis must be cognisant of the end-purpose of the research.

ONS will use this research and associated methodologies to value statistics as part of business cases for the provision of statistics. To this end, any valuation of statistics needs to be relative rather than solely absolute in nature. By way of an example, ONS may decide to change the way a particular statistic is generated, disseminated, or assured, but the statistic itself may be identical in name and format for users. In this case, a static analysis of usage which does not account for the changes made would yield the same benefit stream for that statistic irrespective of those changes.

In HMT Green Book parlance, this means the benefits in the numerator of the Benefit-to-Cost ratio (BCR) are constant and the BCR increases if the same statistic can be delivered at lower cost. As part of any future business case work a series of options for statistical provision need to be referenced against a 'do-nothing' option where current benefits have to be quantified. The benefit stream for each option then needs to account for changes driven by the factors outlined in the framework.

In essence, this is the challenge laid down to ONS by the Bean Review. Make statistics 'better' as well as cheaper to produce using new and novel sources and that confer efficiency – or, in other words, provide greater benefit at lower cost.

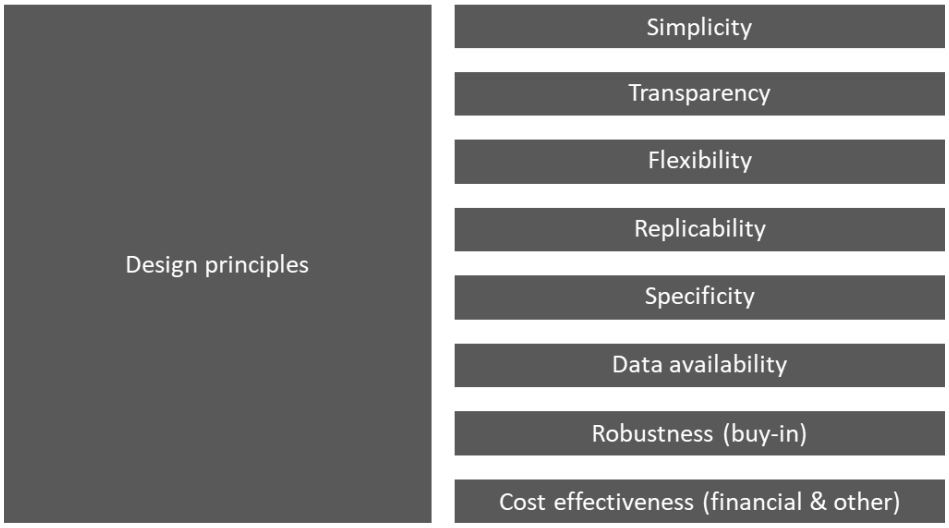
As a corollary of this requirement, some of the methods that are available may not be fit-for-purpose, if they are not capable of distinguishing between user and re-user benefits and quality in different options.

## 5.3 Sift criteria summary

The sift criteria below were developed through desk-based research and limited consultations with ONS and external stakeholders.

The sift criteria when considering design principles are:

Figure 5.3.a: Sift criteria




Source: Deloitte Analysis

## 5.4 Sift criteria rationale

This section introduces the reasoning behind the prospective sift criteria for quantification solutions capable of appraising statistical value. Naturally, there are significant trade-offs between criteria, such that shortlisted options will be the ones that provide the best fit with ONS' requirements.

The criteria to ensure that the method chosen is fit-for-purpose for end-use by ONS are:

**Simplicity**

The method should be simple to use, articulate and understand. Any 'black box' solution which churns out values that are not understood by users or readers runs the risk of misinterpretation. There is a trade-off between simplicity and robustness, with some of the more simple methods available unlikely to be sufficiently robust for HMT Green Book compliance.

**Transparency**

Any assumptions required of a method should be explicit and transparent. Moreover, the more assumptions required (in place of data) for a method, the greater scope for error in outputs.

**Flexibility**

The method should be sufficiently flexible to allow for sensitivity testing and future refinement. HMT Green Book business cases require sensitivity testing as part of the analysis of risks. Future refinement could lead to improved estimates, for example as better/more data becomes available to ONS to use in benefit quantification.

**Replicability**

The method should be replicable across a range of statistics. This includes the three domains used in this initial study – GDP, Labour Market and Inflation. It should also be capable of being applied for individual series and aggregations of statistics as required – providing the relevant input data is available, and special circumstances (such as the data being explicitly paid for) do not preclude the analysis.

**Specificity**

The method should be capable of distinguishing between benefit streams for the same statistic delivered in different ways – e.g. the quality of ONS outputs. See section 6.2 for a more detailed discussion of quality in the quantification process. If the chosen method is only able to distinguish between quality drivers on the basis of assumptions, the method is not likely to stand up to scrutiny.

**Data availability**

Data availability is a pre-requisite. The method chosen must have a minimum level of information available for it to be of use.

**Robustness**

Related to all other factors, robustness is concerned with the actual and perceived standard of outputs produced by the chosen method. It is critical that the method should stand up to scrutiny from internal and external stakeholders, such as HMT.

**Cost effectiveness (financial & other)**

The method should not create high marginal costs for ONS each time it is used, nor require a major initial investment to set up.

As ever in options analysis given the trade-offs inherent, no single option is likely to score highest on each measure. Chapter 6 ranks each long-list option against these criteria to allow ONS to trade off the options in deciding a preferred approach for quantification.

# 6 Options analysis

## 6.1 Introduction

This chapter presents the results of the options analysis.

We assess each identified option from chapter 3 against the sift criteria in chapter 5. A discussion of the results and how they help to determine the short-list of options for further discussion/analysis is presented in 6.5.

Hybrid options are also considered here in section 6.3. Chapter 7 subsequently details our take on the shortlisted options.

## 6.2 Options scoring method

Given the need to quantify outcomes in a common currency, the 'scoring' mechanism is based upon a 5 to 1 best-to-worst scale and is also presented using a high/medium/low (HML) framework. Naturally the approach we have taken is subjective, but is firmly based upon the work conducted as part of this study and our understanding of ONS requirements.

The approach uses 'Best/Worst' as the appropriate framework because some criteria would otherwise work against each other if least/most were used instead. As an example, a highest cost approach would be scored as a 5, and a highest robustness approach would also be scored in the same way, leading to non-consistent outputs and spurious results.

The matrix uses colours for the H/M/L scoring. This allows readers to take in the information in the table and compare and contrast the options more easily.

This second visualisation has the added benefit of moving away from potential debate on precise, and arbitrary, numbers for each option, though we do provide both outputs for consideration at this stage for transparency purposes and to facilitate further discussions on the option to carry forward.

## 6.3 Hybrid options

As chapter 4 showed, a number of studies have chosen to combine elements and value data using hybrid methodologies, either additively or to provide alternative estimates. To that end, whilst we have appraised each methodological approach separately as a distinct option for simplicity, in practice they may be combined.

Hybrid options for ONS include, but are not limited to:

- **Combinations of stated and revealed preference techniques**: Asking users to explicitly value data and using conjoint analysis (or similar) to force users to reveal preferences to value data that way;
- **Combinations of market and non-market valuation techniques**: Using shadow pricing to approximate the value of non-paid for data, and augmenting this (where appropriate) with non-market methods, such as value of time;

- **Combinations of case studies and other techniques**: Using case studies to explain or quantify specific instances of value generation to add 'meat to the bones' of other techniques;

Chapter 7 considers shortlisted options in terms of individual and hybrid options.

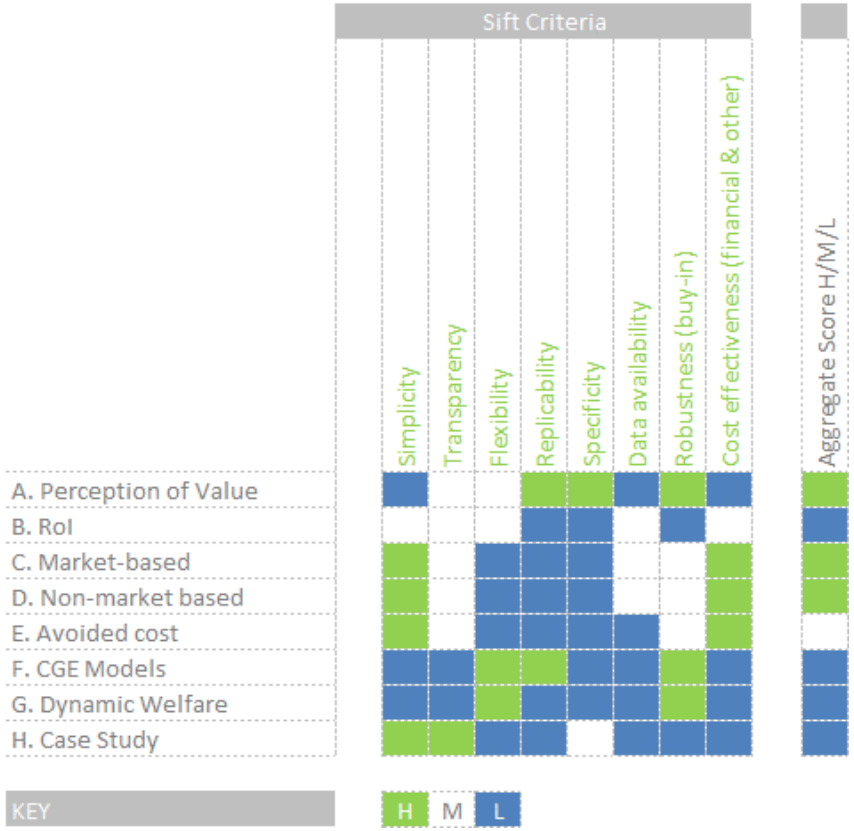
## 6.4 Options scoring

Figure 6.4.a (overleaf) shows the scoring from the analysis.

As the key in figure 6.4.a shows, a high score is best and is denoted by a green element; a medium score is average and denoted by a white element; and a blue element shows a low score, and the worst outcome.

Figure 6.4.a: Options analysis (note 2 types of scoring)

| | Sift Criteria | | | | | | | | Scoring (1 = worst; 5 = best) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Simplicity | Transparency | Flexibility | Replicability | Specificity | Data availability | Robustness (buy-in) | Cost effectiveness (financial & other) | Total Score / 40 (Unweighted) | Average Score / 5 (Weighted) | Rank |
| Weights | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | | |
| A. Perception of Value | 2 | 3 | 3 | 4 | 4 | 2 | 4 | 2 | 24 | 3.0 | 1 |
| B. RoI | 3 | 3 | 3 | 2 | 1 | 3 | 1 | 3 | 19 | 2.4 | 5 |
| C. Market-based | 4 | 3 | 2 | 2 | 1 | 3 | 3 | 4 | 22 | 2.8 | 2 |
| D. Non-market based | 4 | 3 | 2 | 2 | 1 | 3 | 3 | 4 | 22 | 2.8 | 2 |
| E. Avoided cost | 4 | 3 | 2 | 2 | 1 | 2 | 3 | 4 | 21 | 2.6 | 4 |
| F. CGE Models | 1 | 2 | 4 | 4 | 1 | 2 | 4 | 1 | 19 | 2.4 | 5 |
| G. Dynamic Welfare | 2 | 2 | 4 | 2 | 1 | 2 | 4 | 2 | 19 | 2.4 | 5 |
| H. Case Study | 5 | 5 | 1 | 1 | 3 | 1 | 1 | 2 | 19 | 2.4 | 5 |
| Average Score | | | | | | | | | 21 | 2.6 | |




Source: Deloitte Analysis

## 6.5 Non-shortlisted options

The options that the scoring mechanism suggests **should not** be considered for Phase 2 of the analysis are shown below:

- **RoI approach**: This option yields average scores on simplicity, transparency, flexibility, data availability and cost effectiveness but falls down on specificity to the task at hand and robustness/suitability for the task. The main issue with a return-on-investment approach for ONS is that the estimated benefits will be used in a formal cost-benefit analysis setting. Using a 'cost-plus' approach to obtain benefits means at least some circularity in the estimates. In this setting at least, this method would not deliver robust results.
- **Dynamic welfare approach and CGE**: These options are taken together as they yield reasonably similar outcomes and the same aggregate score. The methods are robust and flexible after the point at which a model is built. However, they are inherently complex and costly given the need for significant preparatory work in each case. On top of initial cost, much work is needed to calibrate models and then apply assumptions to them. We contend that whilst such models are usually suited to macroeconomic analysis and, for instance, assessing the impact of changes in tax rates, they are not best suited to analysing the impact of specific data/statistics at the microeconomic level.

  **Important note:** We found it difficult to categorise the work conducted by Haskel on Big Data into our 8 options. It probably fits here, and whilst it would be our chosen method for arriving at a single estimate of statistical contribution/value, it would not provide the type of outputs required to feed into a business case, so we discount it along with other similar econometric methods.
- **Case study approach (in isolation)**: Whilst simple and transparent, a case study or series of case studies, whether qualitative or quantitative are not sufficient to fulfil requirements. Grossing up a small sample of case studies, or applying findings from an unrelated case study to another theme are not likely to be robust enough to fulfil HMT Green Book appraisal requirements.

These options are all 'scored' as low, being below average in both scoring and rank.

Chapter 8 details the methods carried forward for consideration, and provides more detail as to how they may be implemented.

# 7 The short-list

## 7.1 Shortlisted options

In no particular order, the options that the appraisal mechanism infers should receive further attention from ONS for use in Phase 2 of the analysis are:

- Perception of value approaches;
- Market and non-market value approaches; and
- Avoided cost approaches.

The pros and cons of each shortlisted approach, and greater detail on how each approach might be used to quantify the value of each statistical domain appear below. In practice, there may be some instance of borrowing estimates from other methodological options as required to ensure the analysis makes best use of all available data and does not exclude permissible benefits.

**Perception of Value**

The appraisal, based on our understanding of ONS requirements, suggests that asking users about their perceptions of value, either directly, or indirectly, is the best method to use. This is especially true for robustness and specificity criteria but such an approach does come with attendant costs.

The major benefit of this approach is that it can be tailored to provide either an explicit valuation or a series of 'soft factor' values to apply to given situations to yield composite benefits of statistical provision. This could take account of generalised trade-offs in user requirements, such that business cases reflect the value arising from user needs.

Any willingness to pay (WTP) survey would yield the surplus generated by statistics above the cost of provision. The survey would also be used to tease out the 'counterfactual' to ensure economic benefits are 'additional' net benefits, as required by HMT Green Book guidance.

Within a broad perception of value/WTP approach there are two approaches available (as well as some combination of the two methods as appropriate). These are:

- A simple stated preference survey: asking users explicitly about the monetary values they put on the statistical domains as a whole, and the factors that determine quality to yield estimates that can be used in further analysis; and
- A more complex revealed preference survey: asking users a series of questions, using techniques like conjoint analysis, that work to reveal their preferences without them having to make a valuation.

There are distinct costs and benefits to each approach. A simple stated preference survey may be sub-optimal given the difficult and abstract concept of valuing something with no price. This may affect the quality of stated preference responses/values and lead to very different assessments of value from users.

A fuller and more complex revealed preference survey would be costly in terms of financial cost, consultant cost and the increased burden of time on participant organisations. However, this could tie in well with wider moves by ONS to better understand its clients and provide better solutions to them, whilst simultaneously demonstrating value to HMT.

Even with this extent of primary data collection, the perception of value methods would still require secondary source information. This would include data on usage and some price 'anchor point' for unpaid data if preferences were not monetised directly in any survey. This would be required to extrapolate total benefits where values are not provided explicitly. A complex conjoint analysis would also have higher 'back-office' costs in terms of using the results of the survey and analysing them to obtain meaningful valuations for further use.

Both methods require a careful consideration of the sample for any survey. In some instances, willingness to pay studies consider non-users, though in this instance it is not clear that this would be either useful or cost effective. It may also be useful to consider a limited number of re-users and redistributors to ensure that different perspectives are taken for different points in the value chain.

**Ranked 1st**

**Conclusion**: this is the only real option for assessment of marginal benefits of statistics in a business case, but highest cost of the shortlisted options. Subsequent decisions on type/depth of analysis still need to be made. It would likely require the use of a survey, this is beyond the project scope and there are not the available resources to action this.

### Market and non-market value

This approach would build on available public domain and ONS data along with relevant shadow prices from market-based provision to estimate the value of ONS statistics. Shadow prices can come from private companies supplying data, or ONS data that is paid for, or has been paid for in the past, with adjustments to reflect the fact that it is generally provide at cost.

As an example, the prices paid for CIPS PMI data, a leading indicator for GDP, can be used as a proxy for the value of GDP statistics, because organisations are willing to pay for a timelier indicator of economic output/growth. An adjustment could be made to reflect relative revisions between the two series as some form of 'quality multiplier'

For this approach data availability is the best of all options, and without the need for extensive consultation and primary research this represents a very cost-effective solution. The option is also simple to understand and reasonably transparent. However, this method suffers in terms of a relative lack of specificity, robustness and flexibility.

Given the lack of information capable of feeding into the method to account for changes in quality, this method is perhaps better suited for arriving at an aggregate estimate of the benefit of a large group of statistics or data, rather than differentiating between different means of providing similar statistical outputs. In other words, this method is unlikely to adequately account for changes in quality.

A lack of bespoke inputs from primary research means that robustness is only average relative to other options – this might affect the gross benefits estimated as well as any

'additionality' assumptions required to form the counterfactual or 'do nothing' option in any business case setting.

**Ranked 2nd**

**Conclusion**: a viable, low-cost option for the assessment of aggregate benefits by statistic or domain, but not especially suited to use in a business case, unless relevant adjustments can be made.

**Avoided cost**

Rather than considering benefits per se, this method would concentrate on estimating the costs associated with poor or no provision of statistics and thus already includes the counterfactual. To that end, if used in a business case setting, this approach would not require a 'do nothing' option.

As our studies have shown, in some instances previous research could be used to imply the value of avoided costs, e.g. with respect to MPC decisions on interest rates, changes in provision of inflation statistics, or inaction on the financial crisis.

This method is bound by the (generally inconsistent) information available, unless, for example, HMT would be able to run a series of hypothetical counterfactual scenarios based through their macro model on misinformed decisions and somehow link these back to statistical provision.

However, it is worth noting that in a limited number of cases (outwith the domains in the scope of this study) e.g. the flow of funds business case – an avoided cost, or augmented avoided cost approach is entirely suitable and appropriate.

In the flow of funds case, the analysis considered the significant cost to the economy of the financial crisis in 2008-09 and estimated the change in the probability of the crisis **not happening** that would be required to offset the costs of the flow of funds project. As the change in probability was infinitesimally small, the case for investment in flow of funds was proven using an avoided cost approach.

**Ranked 3rd**

**Conclusion**: as with market/non-market based methods, this approach is probably not best suited to application is a business case setting when appraising a range of options, unless there is a clear case to do so. Exceptions are likely to occur in the case where a probability of averting a doomsday scenario would only have to change marginally in an expected value setting, to make an investment in statistics worthwhile.

## 7.2 The chosen option

A workshop was held on May 4th 2016, at ONS in Newport, to discuss the work to date, decide upon a preferred option and agree the subsequent work programme to fulfil project requirements.

The three options detailed at 7.1 were discussed at length and it was decided to seek further data on usage from ONS' management information systems to allow a 'market and non-market value' approach to be used. It was also agreed that no further consultations would

take place at this juncture to avoid ‘over-consultation’ with users in the wake of recent reviews.

In practice this means that a hybrid methodology is taken forward, that is based exclusively on data that is available from ONS and in the public domain.

The approach chosen for this report uses both market and non-market valuation techniques to consider in this benefit quantification of GDP, price and labour market statistics as domains and in aggregate. Whilst an activity costing approach was considered (seen as best practice by Eurostat and academics), this method was not pursued due to a lack of available survey information and the above constraints.

Chapter 8 details the method used and the results of this analysis.

# 8 Benefits quantification

## 8.1 Method

The modelling adopts a three-stage approach to valuation of ONS economic statistics. All analysis pertains to the year 2014 and is based on ONS management information, usage data from data.gov.uk, and price data from third-party commercial organisations.

The three stages are:

- Stage 1: estimating the value of economic statistics through the value to direct consumers (users and re-users). This uses a bottom up approach based on usage metrics for ONS and data.gov.uk to quantify consumer surplus above the zero price level for unpaid data using shadow prices for paid data;
- Stage 2: adding in known paid for data produced by ONS; and
- Stage 3: subtracting the cost it takes to produce the statistics from this figure to generate as estimate of net benefits.

The report does not seek to systematically quantify the value of the broader social impacts of economic statistics. There is a lack of reliable evidence on the causal linkages between the consumption of economic statistics and economic, democratic, social, environmental outcomes. Previous studies (including our own) have provided ready-reckoners for the value of this important outcome set, but for use in a business case setting such estimates are unlikely to be seen as valid core metrics by HMT and other stakeholders. For this reason, we present estimates of wider benefit as an additional analysis rather than as core to our estimates. Equally the analysis cannot include the 'greater-than-the-sum-of-all-parts' benefits that stem from the holistic provision of all national statistics, and these are left out completely.

It should be stressed that the lack of reliable and systematic data on the usage of ONS economic statistics make the quantification of value difficult and reliant on assumptions. To reflect this inherent uncertainty the estimates contain upper and lower bounds.

Further research (and especially primary data collection) is required to develop more comprehensive estimates that include provision for statistical quality.

**Stage 1: Contingent value – the welfare approach using shadow prices**

Management information data from ONS was delivered to Deloitte as statistical elements, each from a common domain, and with a sub-set of web addresses to show the type of product offered to users. Each web address by user was aggregated to the statistical element, and the statistical element was then aggregated up to the GDP, Price and Labour Market domains in scope for this study.

The statistical elements considered in the study can be seen below by type. These represent the majority of statistics in the three domains but we have not checked that these are collectively exhaustive:

| | |
|---|---|
| GDP | Balance of Payments |
| GDP | Business Investment |
| | Output of Construction |
| GDP | Industry |
| GDP | E-Commerce |
| GDP | Economic Review |
| GDP | Family Spend |
| GDP | FDI |
| GDP | GDP First Estimate |
| GDP | GDP Second Estimate |
| | GDP and the Labour |
| GDP | Market |
| GDP | IOP |
| GDP | IOS |
| GDP | IT IS |
| GDP | MQ5 |
| GDP | National Balance |
| GDP | QNA |
| GDP | Retail Sales |
| GDP | TOPSI |
| GDP | Trade in Goods |
| Labour | BRES |
| Labour | IOLC |
| Labour | Labour Market |
| Labour | Labour Productivity |
| Labour | Low Pay |
| Labour | NEET |
| Labour | Public Sector Finances |
| Labour | RLM |
| Labour | AWE |
| Labour | Tax and Benefits |
| Price | CPI |
| Price | Aerospace |
| Price | HPI |
| Price | PPI |
| Price | SPPI |

Some important assumptions were made regarding the data provided by ONS:

- For some of statistical elements there were gaps in the data and in this case the data was aggregated up to a complete year by extrapolating numbers based on monthly averages;

- A further assumption involved including all webpages where the title had the word download, xls, csv, pdf or xlsx in the total download count for each statistical element; and
- Statistics are also commissioned by ONS customers, and in this case are paid for. The cost of these statistics to customers are included in the overall value, as they implicitly show a minimum value of these statistics to users, but the aggregate costs of providing all statistics (paid and unpaid) in each domain are removed to avoid any double counting.

The model uses input data on downloads attributable to economic statistics domains by specific webpage. Assumptions are made relating to:

- price elasticity of demand for different dataset categories;
- the shadow price of ONS economic statistics;
- the proportion of users going on to actively use and re-use downloaded datasets; and
- the shape of the demand curve for organisations of different types and sizes.

The 'welfare' approach seeks to quantify the consumer surplus derived from the use and re-use of economic statistics and the producer surplus to ONS. The welfare approach can be estimated by summing the consumer surplus with the producer surplus. In this instance, because ONS is not a profit making organisation, the producer surplus is approximately equal to zero (operating surpluses can be reinvested in future periods). Technically, this means that the total surplus can be approximated by taking the total imputed value of the statistics and netting off the costs of producing the statistics.

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 economic statistics when they carry a charge, the demand curve may exhibit non-linearity's and discontinuities when the economic statistic is available free of charge – and the consumer surplus may be much larger or smaller when the demand curve is not linear.

Shadow prices as to the value of ONS economic statistics were estimated from available information, some of which is commercially sensitive and cannot be shared externally as it shows pricing policies for commercial organisations (in this case Deloitte). The shadow price information was collected by considering Deloitte's own subscriptions as well as requesting quotes for data services we do not subscribe to.

**Price and labour market domain shadow price**

For the price and labour domains the shadow price was estimated from the amount organisations pay to access Bloomberg and Factset statistics of the same type. According to both public and private organisations who utilise ONS economic statistics this is the nearest paid for comparator to ONS price and labour statistics.

This throws up an interesting conundrum however, because these are ONS statistics within the domains that are essentially paid for through a platform. Technically it should be possible to speak with these data aggregators and distributors to understand usage through the platforms, before adding this usage to paid and unpaid ONS data provision. In practice, this would lead to some element of double counting because these organisations already source the data from ONS, and they might well be unwilling to share data on users to allow a full analysis. To this end, we do not add in usage from third-party commercial sources.

### GDP domain shadow price

For GDP, CIPS PMI data price was considered as a result of organisations using this data to predict ONS GDP statistics. Using CIPS PMI does not take into account the value of the multiple statistical elements in GDP and thus delivers a lower shadow price for GDP than might be expected. This was noted by stakeholders who indicated that using CIPS prices for GDP without any adjustment, would likely lead to an underestimate of the value of GDP statistics.

We initially considered an arbitrary adjustment to the shadow price used, to reflect this thinking. However, on reflection, there is also a timeliness premium at play. CIPS data is not as detailed as GDP and would not be fit-for-purpose in the same way, but one of the reasons users are willing to pay for the data is that it precedes official release. To include a premium would ignore this. We err on the side of caution and use CIPS pricing in its purest form.

### Adjusting for willingness to pay

Different organisations have a different willingness-to-pay for economic statistics dependant on preferences and available resources, and it is these differing factors that determine the shape and slope of the demand curve. Without related information, and given that the data is predominantly free-at-the-point-of-use, it is necessary to assume a demand curve, or at least how organisations differ in their willingness to pay.

This study assumes that larger organisations with an annual turnover of greater than £10 million are willing to pay 100% of the shadow price used. Whereas medium organisations with an annual turnover of less than £10 million but greater than £1 million are willing to pay a quarter of the shadow price. Finally, smaller organisations with an annual turnover of less than £1 million are willing to pay only 1 per cent of the shadow price. This effectively assumes that the 'long-tail' of users outside large organisations value the data less than large organisations, and would simply not use the data or equivalent if it were charged for. These adjustments effectively assume organisational size determines willingness-to-pay and give a stepped (and non-continuous) demand curve.

Using the web-page specific data we have estimated the number of organisations that are using ONS data. This number of Organisations differs from the number of "Unique Access Points" (UAPs) in the data received which would include Internet Service Providers (ISPs). ISPs can represent a number of individual users and smaller organisations. In total we found c. 21,000 Unique Access Points in the data across the three domains (adjusted for those using more than one domain) and we estimate this to be representative of c. 76,000 organisations using ONS information. We found our 21,000 UAPs by aggregating across GDP, Price, and Labour. Once we had a list of each page accessed across all domains, we observed c. 21,000 Unique Access Points. In some cases UAPs would represent users accessing information across all 3 domains, whereas in others they may only access one of the three domains. In either case the UAP has only been counted once.

We reached an estimate of 76,000 organisations by analysing the activity of named organisations and apportioning the activity captured from Internet Service Providers to match the activity of the organisations we were able to identify from their access point, this gave us an estimate of usage per ISP and unknown access point and a total of c. 76,000 organisations using ONS information. In cases where there is less activity than the average

usage among organisations in an unknown Unique Access Point, it is assumed this is one organisation.

Finally, based on Deloitte analysis of organisations downloading in scope statistics on ONS websites in 2014 an overall shadow price was calculated for GDP, Price and Labour on the basis that 80% of organisations downloading economic statistics were large organisations, 15% were medium organisations and 5% were small organisations.

A range was estimated by also calculating modelled revenue if all organisations downloading data were willing to pay 100% of the shadow price.

The estimates generated in stage one are gross estimates of value, in that they are not adjusted for the cost of producing the statistics, which needs to be factored in to account for the opportunity cost of doing so.

**Stage 2: Paid data**

The estimates of paid data by domain are added to the gross estimates of value from stage one. These are simply the values for each domain as provided by ONS.

**Stage 3: Costs**

The starting point for the cost side of the model was the investment in producing ONS economic statistics, calculated through survey costs and direct and indirect staff costs funded by Grant Aided Expenditure (GAE), as supplied by ONS.

## 8.2 Overall results

Figure 8.2.a summarises the results of the modelling exercise for both upper and lower bound estimates (which vary on the basis of willingness-to-pay assumptions).

This gives an effective value **range across the three domains of c. £130 million to £155 million**. As a reminder, these estimates include the use and re-use of statistics, but do not consider broader socio-economic-environmental outcomes stemming from the use of statistics. They should thus be viewed as a comparatively narrow measure of value. The lower bound estimates are our suggested core estimates, because they make a rudimentary adjustment for willingness to pay on the basis of organisational size.

Figure 8.2.a: Results Summary

| | Upper Bound Value (£m) 100% WtP | Lower Bound Value (£m) Adjusted WtP | *of which* | Gross Value Unpaid Data (£m) | *plus* | Paid data (£m) | Costs (£m) |
|---|---|---|---|---|---|---|---|
| GDP Domain | 50.0 | 41.1 | | 41.1 | | 0.0 | 32.5 |
| Price Domain | 49.5 | 41.6 | | 41.5 | | 0.1 | 21.6 |
| Labour Market Domain | 56.3 | 47.3 | | 47.2 | | 0.2 | 23.0 |

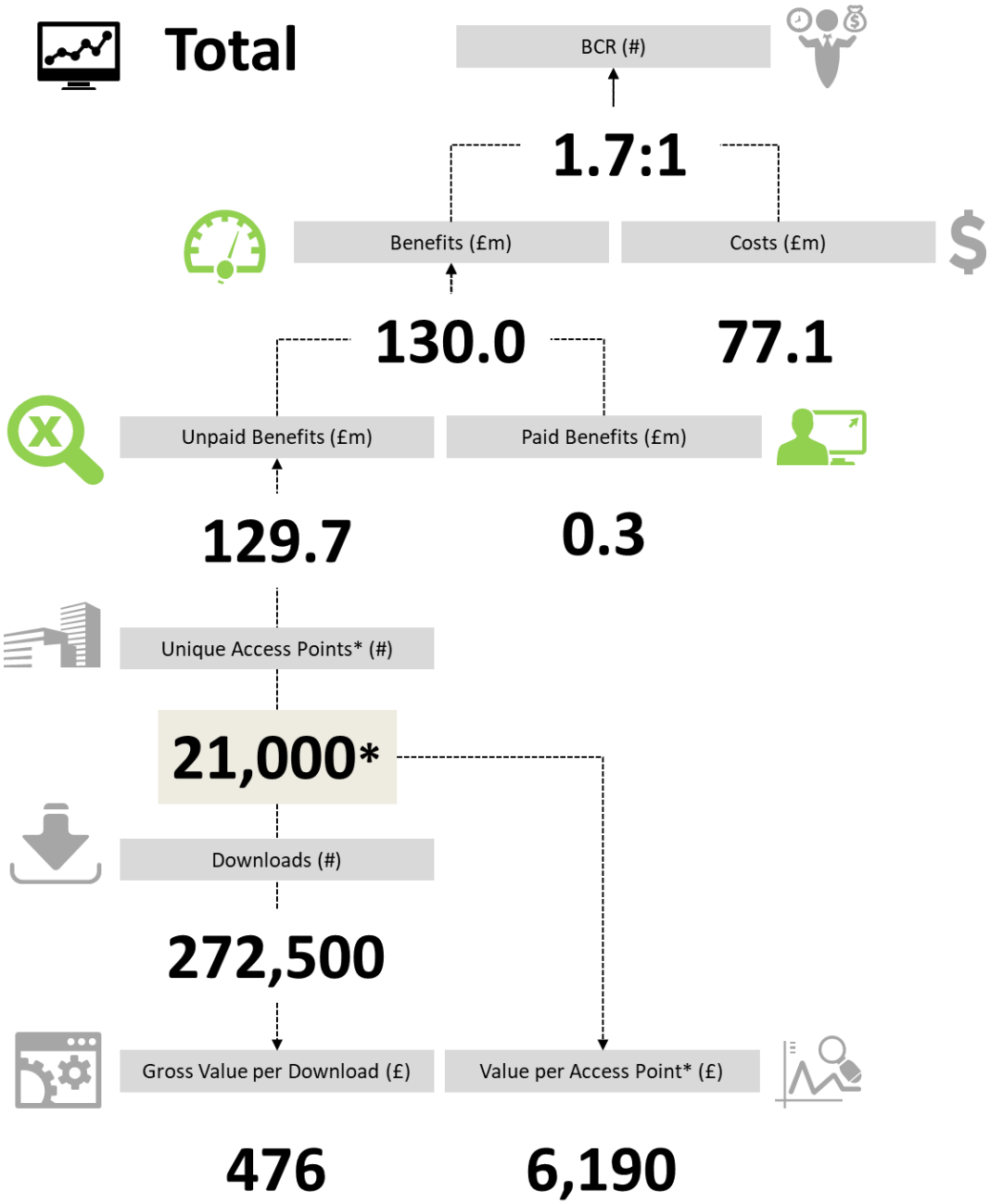
| | | | | | |
|---|---|---|---|---|---|
| Aggregate | **155.8** | **130.0** | **129.7** | **0.3** | **77.1** |

Source: Deloitte Analysis

We estimate that GDP statistics contribute c. £41 million (32%) to the aggregate value of statistics, with price statistics contributing £42 million (32%) and labour market statistics contributing £47 million (36%). These outcomes are driven by the parameters in the model, usage estimates and shadow prices, all of which are detailed for each domain and in aggregate over the following pages.

The costs of production are significantly higher for GDP statistics – 42% of total costs across the three domains – with 28% of costs in the price domain and 30% in the labour market domain.

Figure 8.2.b: Aggregate Value of Statistics (3 domains), 2014




Source: Deloitte Analysis

Our interrogation of ONS management information suggested that there were around 21,000 unique access points across the three statistical domains. These include the registered organisations as well as individuals and organisations counted as one ISP. Some access points accessed only one domain, some two domains and some all three domains. Further, we estimate that there were at least 72,000 unique users of statistics across these access points, but it is impossible from the data available to estimate a precise amount because it is not feasible to understand how many users go through ISPs (and how many are also going through company/organisational routes). This is the main reason why download is the only permissible means of generating an estimate of value at present.

We estimate that these 21,000 access points were used to download statistics on 272,500 occasions in 2014. The shadow prices used in the analysis gave an average gross value per download of £476, which yields a net value of £130 million across the three domains once paid data are factored in. This equates to £6,190 in value for each access point across the three domains.

When the costs of provision (£77.1 million in 2014) are used to benchmark the benefits, this leads to a benefit to cost ratio (BCR) of 1.7:1. In other words for each £1 spent in production/dissemination by ONS, the statistics give a value of £1.70 to users.

As a sense check, our lower bound estimate of £130 million in value represents 8 per cent of the equivalent £1.7 billion estimate of value for all public sector information (PSI) in the UK (at today's prices) we estimated as part of the Shakespeare Review of UK PSI.

This study made use of a similar method and included the significant revenues accruing to the trading funds, all other ONS statistics and a wide range of departmental and local government data outputs. Using the recent Census Transformation Programme OBC, a simple annual average of 10-year benefits suggests that a 2014 equivalent value for the Census would be in the order of £300 million – meaning that the three domains together account for over 40% of like-for-like benefits of Census statistics.

### Wider benefits

As we have demonstrated in this report, there is much literature pertaining to the wider benefits associated with data, beyond the use and re-use covered in our core estimates. These effects are based on the notion that data confers non-market benefits (externalities) to third-parties not privy to the use and re-use of that data. As such it is understood that the total social value of public sector information is likely to be significantly greater than the narrow economic value implied here.

Here, our core estimates only include the value to those using and re-using data through direct association with ONS and statistics, and this often excludes the end beneficiary – such as the member of the public benefiting from better allocation of public monies.

If we include a wider benefits 'ready reckoner', as we did in the 2013 study for BIS on the value of public sector information, the wider benefits associated with statistics might be expected to be four times higher than the core estimates. The figure below shows the effects of including 'wider benefits' in the analysis.[8]

Figure 8.2.c: Results Summary accounting for wider benefits of use in total benefits, 2014

| | Core Benefits Value (£m) Adjusted WtP | Core BCR | Total Benefits Value (£m) (inc. WBs) | Total BCR |
|---|---|---|---|---|
| Aggregate | 130.0 | 1.7 | 520.1 | 6.7 |

Source: Deloitte Analysis

In aggregate (the multiplier used does not vary by domain), total benefits across the three domains might be in the order of £0.5 billion, which is equivalent to a BCR of 6.7 – up from 1.7.

The ready reckoner, is more art than science, and our conclusions in chapter 9 suggest ways in which the overall value of statistics might be improved.

[8] Our work for BIS was based upon insight from a range of case studies across diverse areas of the economy, such as transport, real estate, health, meteorological, and more.

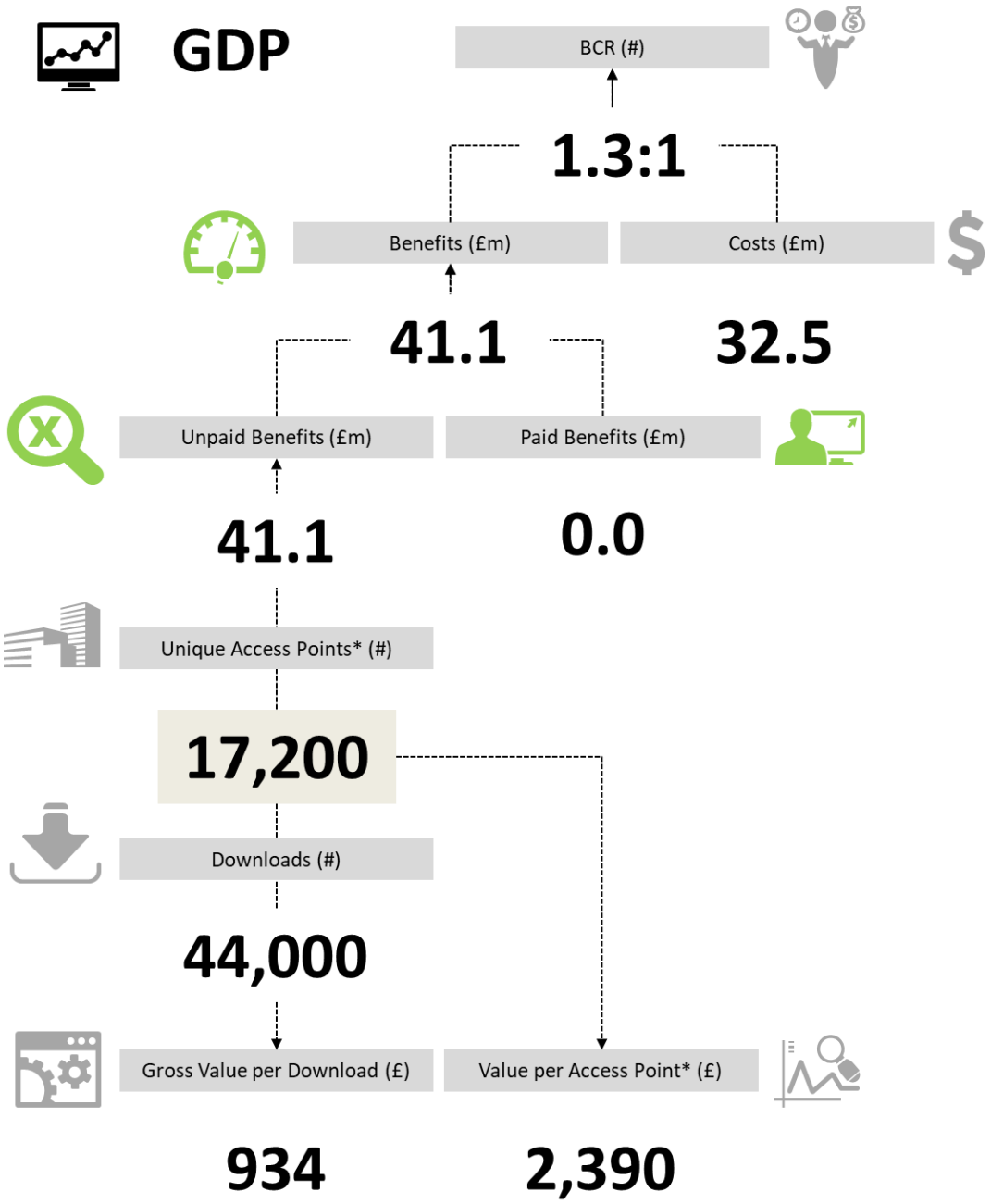
## 8.3 GDP results

For the GDP domain, our interrogation of ONS management information indicates that 17,200 UAPs made 44,000 downloads related to GDP statistics in 2014. This was the lowest number of downloads by domain, and the highest number of access points. This implies that each user makes a small number of downloads (relative to other domains) but that there are a high number of users (as proxied by access points). Whilst users may be only assessing headline figures for GDP, this would not necessarily undervalue this statistic as it is still being used to inform the user or to develop further analysis using the headline data.

We estimate that each access point realised £2,390 in value for each of these interactions, with a value of £934 per download – the highest of each of the domains. GDP can therefore be viewed as high value on a 'per transaction' basis, but with low usage on a large user base.

This implies a benefit of unpaid data of c. £41.1 million (and the same total for all data as GDP data was not paid for under any circumstances in 2014).

Figure 8.3.a: Value of Statistics, GDP Domain, 2014




Source: Deloitte Analysis

In relative terms, the BCR associated with GDP statistics is 1.3:1, the lowest of each of the domains. This is reflective both of the relatively lower number of downloads, and the significantly higher costs of production for GDP. Nonetheless, this means that for each £1 spent on producing GDP estimates, users and re-users realise £1.30 of benefits from the statistics in the domain.

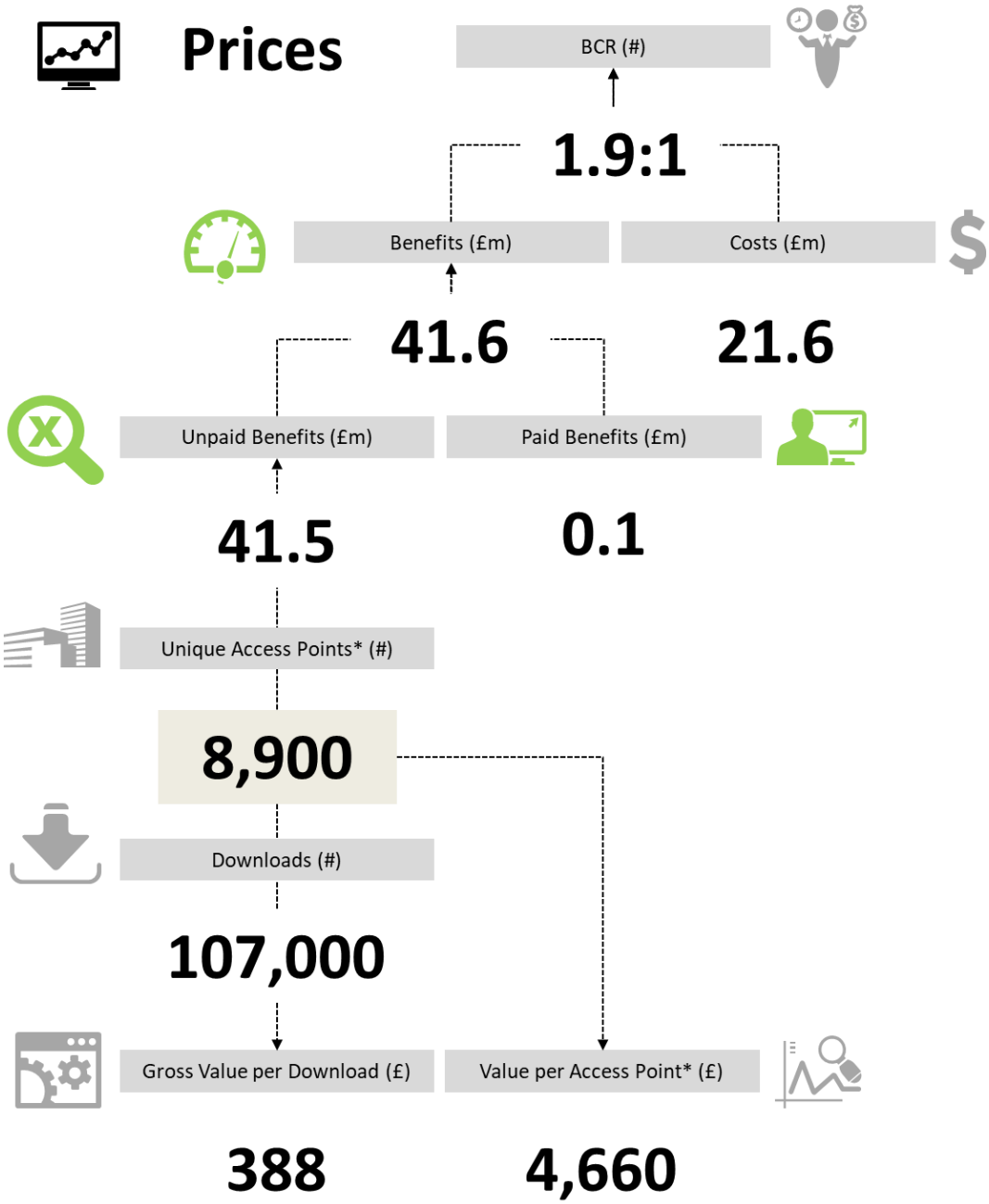
## 8.4 Inflation results

For price inflation statistics there were 8,900 UAPs downloading a combined 107,000 times in 2014. This equates to 12 downloads per UAP, which is much greater than the equivalent figure for GDP statistics of three downloads per UAP. This divergence could be driven by the frequency of the statistics or a range of other factors and we have not investigated this further.

With a gross value per download of £388 (lower than GDP), and a value per access point of £4,700, this yields net benefits of £41.6 million for users of inflation statistics (also including a small amount of paid data - £0.1 million).

With costs of production of £21.6 million, the BCR is of the order of 1.9:1 for inflation statistics – generating £1.90 in benefits for each £1 expended.

Figure 8.4.a: Value of Statistics, Inflation Domain, 2014



Source: Deloitte Analysis

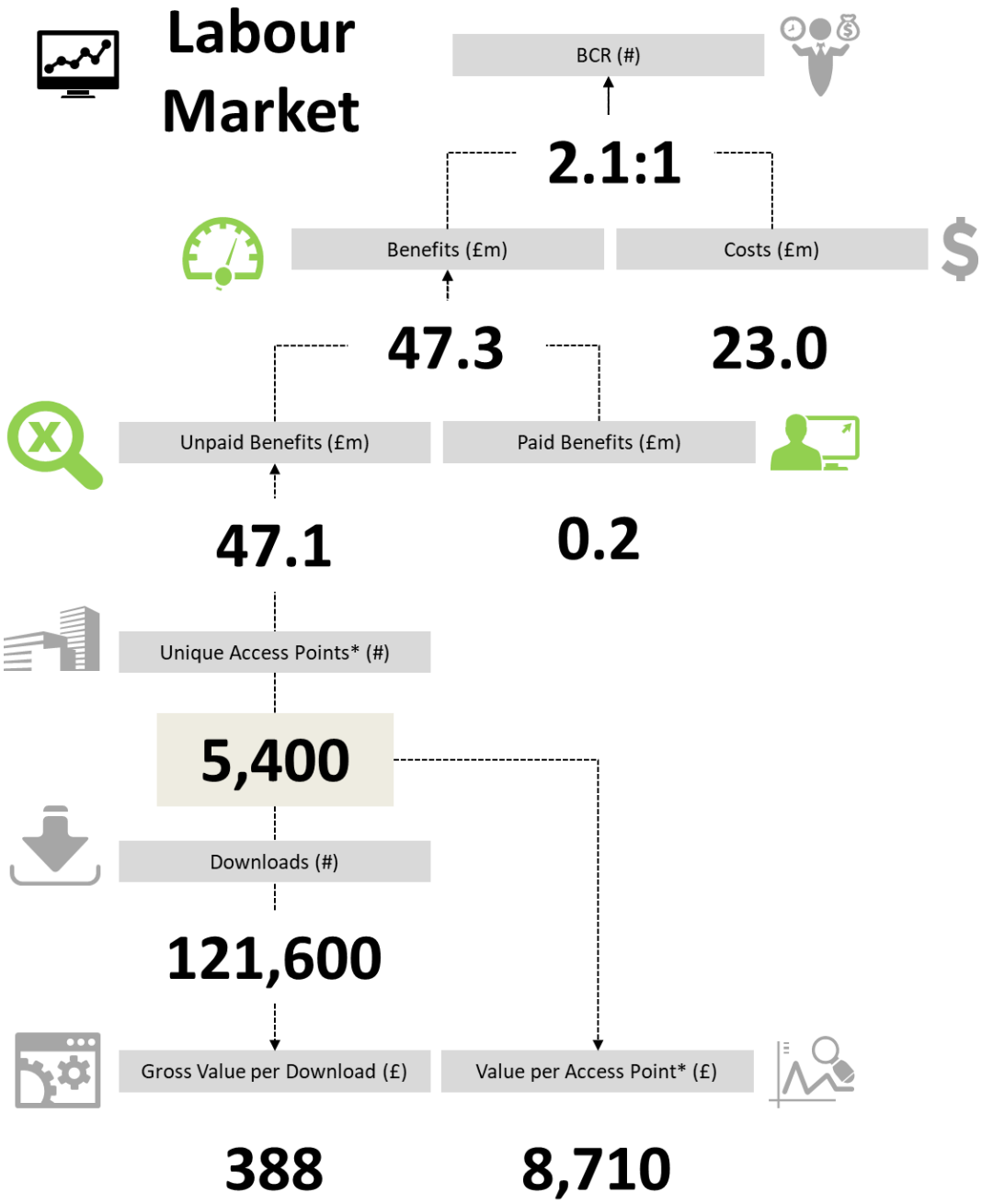
## 8.5 Labour market results

Our estimates suggest that labour market statistics generate the highest absolute benefits of any of the three domains - £47.3 million in 2014. Of the £47.3 million in benefits, £0.2 million are for paid data (including chancellors' notices on the Business Register and Employment Survey).

As well as having the greatest absolute benefits, the labour market domain also has the second-lowest cost base, which means that the BCR associated with it is 2:1:1, higher than the GDP and Inflation domains on this measure.

The value side of this equation is driven by 121,600 downloads across 5,400 UAPs. This means each UAP made 23 downloads in 2014, almost double that of inflation statistics and eight times that of GDP statistics. This yields the highest value per access point of each of the three domains £8,700. Even though the gross value per download is less than GDP and equal to that of inflation statistics, the usage levels drive higher value for labour market statistics.

Figure 8.5.a: Value of Statistics, Labour Market Domain, 2014




Source: Deloitte Analysis

Much of this value premium is likely due to the breadth and depth of coverage for labour market statistics. Whereas GDP and Inflation statistics are aggregated to the national level, labour market statistics can be sourced at varying levels of geographic detail and these statistics form the mainstay of regional and local policy analysis.

# 9 Conclusions

## 9.1 Replicability and use

In our view the estimates of statistical value provided in chapter 8, are a reasonable attempt to quantify value on the basis of the information and resources available.

We are likely to underestimate some aspects of statistical value by omitting certain types of impact, and overstate others by assuming that usage and observable prices are relevant proxies for value (when some users may not attain the same utility from statistics as others). We have made a number of comparisons to equivalent estimates in chapter 8 and the results are defensible in this context.

In keeping with previous analyses and our expectations, our work shows that statistics – as a public good provided largely free-at-the-point-of-use – confer values in excess of the costs of provision. The analysis also shows that the ratio of benefits to costs (so-called BCRs) are likely to differ markedly by statistical domain and statistic, based on a number of factors including usage and the cost of production.

The method can be replicated and used for further analysis either to calculate benefits for a particular statistic or to use BCRs on existing costs to get a ready-reckoner of benefits. In the interim, there are a number of ways this analysis can be used in business case work, although this will require some level of assumption. Two examples include:

- BCRs might be used as base-case ready reckoners for new ways of statistical provision. As an example, benefits might calculated on the basis of existing costs, and then held equal as costs are reduced through novel ways of provision. This will imply a higher BCR from efficiencies for any given statistical output; or
- For a given BCR, ONS could make an assumption that an increase in quality might create an increase in the benefit that is at least equivalent to the increase in cost. This means that where a new method or statistic is associated with greater cost, the greater costs are likely to provide equivalent, if not more than proportionate levels of benefit.

In doing so, those using the method should be aware of the following limitations:

- There are a number of **simplifying assumptions** used to turn ONS and third-party data into valuations. These included assuming that each download triggers a shadow payment and by definition the data is used; and that most organisations are willing-to-pay a price equivalent to that of similar data from commercial providers.
- The **quality of statistics** (which can be defined in a number of ways) is not explicitly factored into the calculations. As an aside, and perversely, where particular international standards have to be adhered too, and the costs of doing so are high, the BCR associated with a statistical domain may be low. GDP is an example of this, where the costs of provision are high relative to usage and implied value.
- **Wider effects** or externalities beyond the use and re-use by known organisations are not included in the core estimates. Examples of this include the net benefits accruing from

policy decisions made, or accountability leading to better government. This means the core estimates provided are narrow, and are likely to understate benefits to some degree.

- As noted by stakeholders, the aggregate value from all statistics in the UK – as **a functioning and independent evidence-base** rather than a collection of domains – is likely to exceed the sum of its parts. By definition, any attempt to consider statistics by type or domain will not take into account such synergies.

## 9.2 Potential improvements

We are of the opinion that improved management information from ONS on usage and users, as well as specific primary research into user preferences would improve the estimates included in this report greatly.

The analysis is suited to providing a relatively low-cost estimate of the value that each domain generates through use (and to a lesser extent) re-use. These estimates are best suited for use where an average value is required. They do not provide particularly robust measures for use in assessing the marginal value of changes in statistical provision.

To that end we make suggestions as to three ways in which the evidence base on the value of statistics might be improved in future.

**A. Management Information**

As part of our analysis of usage, we were provided with a wide range of useful information. The bulk of this information, however was provided across a range of Excel files with non-consistent metadata and naming conventions. This necessitated a large scale exercise in SQL to aggregate and cleanse data pertaining to usage by specific users.

We understand that the reason for the lack of consistency in the management information provided is that we required a full year of data (2014), and this data is associated with the old website. Since then a new, more user-friendly website has been developed and we understand that the management information associated with the new website is of better quality/consistency.

Nonetheless, it may be worth investigating how the information and key metrics from the new website can be 'sweated' to deliver better insight on usage and how users are using data.

In doing so, this may be useful for ONS more broadly in terms of the monitoring and evaluation of performance in terms of usage and other factors that feed into valuations as a by-product.

This links to the second suggestion below.

**B. User Survey**

The preferred option in our analysis was a user survey to reveal preferences and ultimately place a monetary value on different characteristics of data. Such a willingness to pay survey would allow marginal benefits to be considered in a business case setting where ONS is able to understand how investment in statistics might deliver marginal changes in those desired characteristics (e.g. timeliness, or reliability).

It was agreed that now was not the time to conduct such a study given the costs involved – both financially to ONS and in terms of the time/compliance costs for participating organisations. This is especially resonant given the large scale consultation with users in the wake of the Bean Review.

In our view there is merit in revisiting the type of 'soft-factor' willingness-to-pay analysis we outlined earlier in this report at some point. Perhaps this is best in two-to-three years' time – by then improvements in statistical provision will take effect. This will also allow the survey

to consider, ex-post, how users feel the quality of statistics, and characteristics thereof, have improved over the intervening period.

### C. Econometric Analysis

Alternatively, if there is a widely held view that the quality of statistics has improved over the next few years and that the above user survey is not best use of resource, it may be worth using econometric analysis to consider the welfare gains arising from those improvements. Using the macro production function approach previously harnessed by Haskel & Woodbridge to estimate the contribution of big data is one possibility.

This would give a composite picture of the value contribution of UK statistics across the economy as a whole. Analysis could be conducted to see how this value has changed over a two-to-three year period. It might also yield an estimate of the elasticity of GDP (and employment) to statistical provision and use.

# Annex 1: Bibliography


ACIL Tasman, 2009. Spatial Information in the New Zealand Economy. Wellington: Land Information New Zealand.

Beagrie, C., 2012. Economic Impact Evaluation of the Economic and Social Data Service. New Zealand: The Centre for Strategic Economic Studies (CSES).

Beagrie,N. and Houghton, J., 2016. The Value and Impact of the European Bioinformatics Institute. Salisbury: Charles Beagrie ltd.

Bean, C., 2016. Independent Review of UK Economic Statistics. London.

Bonina, C., 2013. New business models and the value of open data: definitions, challenges and opportunities. London: London School of Economics and Political Science.

Carpenter, J. and Watts, P., 2013. Assessing the value of OS Open Data to the Economy of Great Britain- Synopsis. London: Ordnance Survey.

Capgemini Consulting, 2015. Creating Value through Data. London: European Data Portal.

Capgemini Consulting, 2013. The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data. London: Capgemini Consulting.

Chebli, L., Goodridge, P and Haskel, J., 2015. Measuring activity in big data: new estimates of big data employment in the UK market sector. London: Imperial College Business School Discussion Paper 2015/04.

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.

Deloitte, 2012. Open Growth: Stimulating Demand for Open Data in the UK. London: Deloitte Analytics.

Department for Transport., 2010. Role of Soft Factors in Patronage Growth and Modal Split in the bus market in England.

Fox, M., 2013. City Data: Big, Open and Linked. Enterprise Integration Laboratory.

Goodridge, P. and Haskel, J., 2015. How much is UK business investing in big data? London: Imperial College Business School Discussion Paper 2015/05.

Goodridge, P. and Haskel, J., 2015. How does big data affect GDP? Theory and evidence for the UK. London: Imperial College Business School Discussion Paper Discussion Paper 2015/06.

Granickas, K., 2013. Understanding the Impact of Releasing and Re-using Open Government Data. European Public Sector Information Platform.

Grueber, M. and Tripp S., 2011. Economic Impact of the Human Genome Project. Colombus: Battelle Memorial Institute.

Gruen, N., Houghton, J. and Tooth, R., 2014. Open for Business: How Open Data can help achieve the G20 Growth Target. Lateral Economics.

Gurin, J., 2014. Open Data Now: The Secret to Hot Startups, Smart Investing, Savvy Marketing, and Fast Innovation. New York: McGraw-Hill.

Hensher, D., 1994. Stated preference analysis of travel choices: the state of practise. Netherlands: Kluwer Academic Publishers.

HM Treasury., 2011. The Green Book Appraisal and Evaluation in Central Government. London.

Holder, A., 2006. Cost-benefit analysis of monetary and financial statistics, London: Quarterly Bulletin Article Bank of England.

Johnson, P., 2015. UK Consumer Statistics: A Review. London: UK Consumer Statistics Authority.

Keen, J., Paige, R. and Rooksby. J., 2013. Big data + politics = open data: The case of health care data in England. Policy and Internet.

Keisler, J., 2014. Value of information: Facilitating targeted information acquisition in decision processes. Boston: Management Science and Information Systems Faculty Publication Series.

London Economics, 2015. Met Office – General Review, Economic Analysis –Final Report.

Mayo, E. and Steinberg, T., 2007. The Power of Information: An Independent Review.

McKinsey Global Institute., 2013. Open Data: Unlocking innovation and performance with liquid information. McKinsey & Company.

Nilsen, K, 2007. Economic Theory as it applies to Statistics Canada: A review of the literature. Toronto: Statistics Canada.

Pearce, D. and Ozdemirogly, E., 2002. Economic Valuation with Stated Preference Techniques Summary Guide. London: Department for Transport, Local Government and the Regions.

Pollock, R., Stepan, P. and Valimaki, M., 2010. The Value of the EU Public Domain. Cambridge: Cambridge Working Papers in Economics 1047, Faculty of Economics, University of Cambridge.

PwC., 2015. Nesta and the Open Data Institute, Open Data Challenge Series Final Report. London: Nesta.

Rhind, D., 2014. What is the Value of Open Data? London: Advisory Panel on Public Sector Information.

Royal Statistical Society., 2016. Response from the Statistical Society to the Bean Review of Economic Statistics. London.

Ubaldi, B., 2013. Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. Paris: Organisation for Economic Co-operation and Development.