



Department for
Business, Energy
& Industrial Strategy



Companies House

VALUING THE USER BENEFITS OF COMPANIES HOUSE DATA

Report 1: The Methodological Framework

BEIS Research Paper Number 2019/015

September 2019



OGL

© Crown copyright 2019

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3 or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to us at: enquiries@beis.gov.uk

Contents

1 Introduction	4
1.1 Objectives of the study	4
1.2 Background and context	4
1.3 User types	7
1.4 Structure of this report	7
2 Overview	9
2.1 Purpose of the valuation framework	9
2.2 Uses of CH data	13
2.3 Benefits of CH data	14
3 Valuation methodologies	16
3.1 Range of methods	16
3.2 Data requirements	21
4 Approach	23
4.1 Value chains for CH data user types	23
4.2 Summary	25
ANNEXES	26
Annex 1 References	27
Annex 2 Literature review	28

1 Introduction

This report has been prepared by ICF Consulting Services Limited (ICF) in collaboration with Economics for the Environment Consultancy (eftec). It sets out the background and methodology for the research commissioned by the Department for Business, Energy and Industrial Strategy (BEIS) to value the user benefits of Companies House (CH) data.

1.1 Objectives of the study

It is well understood that requirements on companies to provide data to CH imposes costs on those companies. These costs are captured in Impact Assessments that accompany relevant legislative proposals. Less well understood is the extent to which the consolidation of CH data benefits businesses, consumers and/or wider society.

It is not possible to use market valuation techniques to value the impact of CH data given that the vast majority of CH data are now available free of charge. This study, which draws heavily on non-market valuation techniques, is intended to fill this evidence gap. BEIS will use this study to inform Impact Assessments and policy development, including CH's contribution to the UK Industrial Strategy. The research has three inter-related aims:

1. to estimate the value of CH data for different types of user, e.g. companies, public sector, creditors, consumers and individuals;
2. to assess changes in those values over time; and
3. to identify the specific pieces of CH data that generate the greatest user value.

1.2 Background and context

Companies House (CH) is the registrar of companies in the UK. UK companies are required by the Companies Act 2006 to be incorporated and registered with CH, provide annual financial accounts and confirmation statements, and inform CH of changes to their companies and/or directors. CH is an Executive Agency of BEIS and is responsible for:

- incorporating and dissolving limited companies;
- examining and storing company information; and
- making information available to the public.¹

¹ <https://www.gov.uk/government/organisations/companies-house/about>

There are 4.2 million UK companies currently registered with CH and more than 600,000 new companies are incorporated each year². UK companies file more than ten million documents with CH every year³. This provides CH with a broad range of data, including:

- basic information on companies (including company type, status, registered office address, etc.);
- financial records and information; and
- information on directors/officers and People with Significant Control (PSC).

CH offers a range of different services for both filing and accessing these data. This study is focused on the use of CH data, which are used to inform the decisions of many individuals and businesses. CH offers the following services to help individuals and businesses to find company information:

- Companies House Service (CHS) – a free service, available via the CH website and an application program interface (API), which can be used to search for and view a range of company information or receive email alerts (via the ‘Follow’ function). Users do not need to register to use this service.
- WebCheck – a free search facility, through which users can also purchase documents or record images for £1 per document. Users purchasing documents will need to register using an email address and password and make payments using debit/credit cards or Paypal.
- A mobile app – a free search facility that can be used to access basic company information on Android and iOS devices.
- Information centres (at Cardiff, Belfast, Edinburgh and London) – that provide public search rooms that can be used by the general public to access company information via CH’s digital services. Again, basic information is available free of charge, but other services incur charges such as printing, viewing specific documents, inspecting original documents, receiving certified copies and certificates of incorporation /designation, and information provided on DVD (see below).
- Contact centres – that receive telephone or email requests for information, which are sent out to users by post or email, or made available for collection at one of the information centres. Charges are incurred for these services (similar to those stated above for information centres).
- A DVD directory – produced monthly, providing details of all live companies and all companies that have become dissolved in the previous month. The DVDs can be

² Companies House (2019) Companies register activities: 2018 to 2019. Available at:

<https://www.gov.uk/government/statistics/companies-register-activities-statistical-release-2018-to-2019>

³ Companies House (2018) Companies House Management Information 2017/18. Available at:

<https://www.gov.uk/government/statistical-data-sets/companies-house-management-information-tables-2017-18>

purchased for £30 per month (or £300 per year) and have a built-in search facility. Data cannot be exported, except through separate, one-off versions of the DVD with exportable files that cost £1,000.

- Additional data products, including:
 - Company data product – a free monthly snapshot that can be used to obtain bulk downloads of basic company details.
 - Accounts data product – a free, downloadable zip-file, containing data files of company accounts filed at CH (available for individual days or months).
 - People with Significant Control (PSC) data product – a free, downloadable snapshot of all PSCs listed in CH data (updated daily).
 - Uniform resource identifiers (URI) – a free service providing URIs (unique web addresses) for each company listed with CH. This enables businesses to include a link on their own websites to direct users to their own CH data webpage, or use the CH data in their own software.
 - XML gateway search service – a computer-to-computer search service that allows users to search CH data from within their own software. A monthly subscription of £4.70 is charged for this (plus additional charges for some specific documents or requests, as described above).

CH therefore provides both free and paid access to its data and services, although the costs for accessing data have been decreasing over time. Before April 2011, CH charged customers for all data services. Between April 2011 and December 2012, some services became free of charge (e.g. information relating to companies' directors), while others saw their fees reduced (e.g. CHD monthly subscriptions, image downloads, bulk 'Usual Residential Address' data). The CHS was then launched in July 2015, which made access to all public data free of charge. This coincided with a significant increase in the total number of searches and requests for CH data from 300 million in 2014/15 to 2.2 billion in 2017/18⁴, but a decline in the number and proportion of paid searches. Of the 2.2 billion searches undertaken on the CH website in 2017/2018, only 830,000 were paid for (0.04%). This represents a significant decline from the 6.3 million paid searches in 2014/2015⁵ (2.1% of all searches), before the introduction of the CHS.

⁴ Please note that these figures are based on searches and requests made via CHS, CHD, XML, other web services, DVDs and contact centres. They exclude downloads of company information and data via API, which will be included in figures for subsequent years.

⁵ Companies House (August 2018) Companies House management information 2017/18. Available at: <https://www.gov.uk/government/statistical-data-sets/companies-house-management-information-tables-2017-18>

1.3 User types

During the research, the study team identified three main user types. These are:

- Direct users, such as companies, who use CH data to carry out due diligence on suppliers or customers. This category also includes researchers who use the information in their own research.
- Other users, described as 'intermediaries', who use CH data as an input to their own data products and services. This category includes credit reference agencies and other providers of financial data and information.
- Providers of public goods, such as government departments or law enforcement organisations, who use the information during the course of policy or investigative work.

The research identified relatively few household users. A user profile survey was undertaken as part of this study, which found that the general public represented around 17% of all users. It also found that these were relatively infrequent users and only represented around 4% of all transactions/interactions involving CH data (see Report 2 for further details). Therefore it was decided not to include household users as a category in this research.

The methodological approach used to undertake research with each of these user types is described in Section 4.

1.4 Structure of this report

The remainder of this report is structured as follows:

- Section 2 provides an overview of the valuation framework and benefits associated with uses of CH data;
- Section 3 summarises the range of methods that can be used to value the user benefits of CH data, along with their high-level data requirements; and
- Section 4 outlines the approach used in this study.

There are two supporting annexes:

- Annex 1 lists the data and literature sources used in developing the valuation framework; and
- Annex 2 provides a short review of previous studies that have estimated the value of public sector information provision.

This report sets out the methodology used to estimate the user benefits of CH data. It is the first of a series of reports that present the findings of the study. Further reports are described below:

- Report 2 presents willingness to pay (WTP) estimates for all users of CH data and provides a policy example of how this analysis can be used.
- Report 3 presents findings for a subset of users: private sector businesses that use CH data as an input to their own commercial products (i.e. those defined as 'intermediaries' in section 1.3). This report presents WTP estimates alongside findings from a qualitative survey of the most significant intermediaries who use CH's bulk data products.
- Report 4 presents findings for another subset of users: the 'providers of public goods'. Findings are presented from the WTP survey and a separate qualitative survey of users that are considered to provide public goods (e.g. transparency organisations and law enforcement agencies).

Finally, all the results have been summarised in a policy summary which also draws conclusions about the value of CH data to users.

2 Overview

This section presents the valuation framework, which was developed to provide a transparent conceptual basis for the study, ensuring a coherent approach to estimating the different aspects of the value of CH data for different types of user.

2.1 Purpose of the valuation framework

The purpose of this valuation framework is to set out the underpinning concepts for valuing the user benefits of CH data. It is intended to provide a reference point that traces out how to estimate the value derived by different types of user, taking account of methodological and practical considerations, such as the availability of data needed to estimate the user benefits. It also provides a coherent structure for drawing the study's results together for use in subsequent policy analyses.

The starting point for the framework are the principles for economic analysis set out in the HM Treasury Green Book⁶. Accordingly, the user benefits of CH data are measured by the value that is associated with the various consumption and production activities to which the data is an input. In assessing the user benefits:

1. Value should be measured in terms of 'opportunity cost', which reflects the best alternative use of a good or service. Market prices – the prices at which goods are exchanged between buyers and sellers – reflect opportunity cost (the economic value of a good/service) if they are not subject to distortions such as taxes and subsidies. Where a good is provided free of charge (i.e. there is no market price) or in cases where the price paid by users of the data does not reflect the opportunity cost associated with its provision, 'shadow prices' need to be estimated⁷. A range of (non-market) economic valuation methods can be used to estimate user benefits in these instances (Section 3).
2. User benefits are measured by the economic surplus that users gain from the provision of a good or service. From a 'consumption' or demand perspective, it is measured by consumer surplus, which is the difference between the (maximum) amount the user is willing to pay for a good/service, and the price that is actually paid (see Box 2.1 below). From a production perspective, this is measured by the

⁶ HM Treasury, The Green Book - Central Government Guidance on Appraisal and Evaluation, 2018.

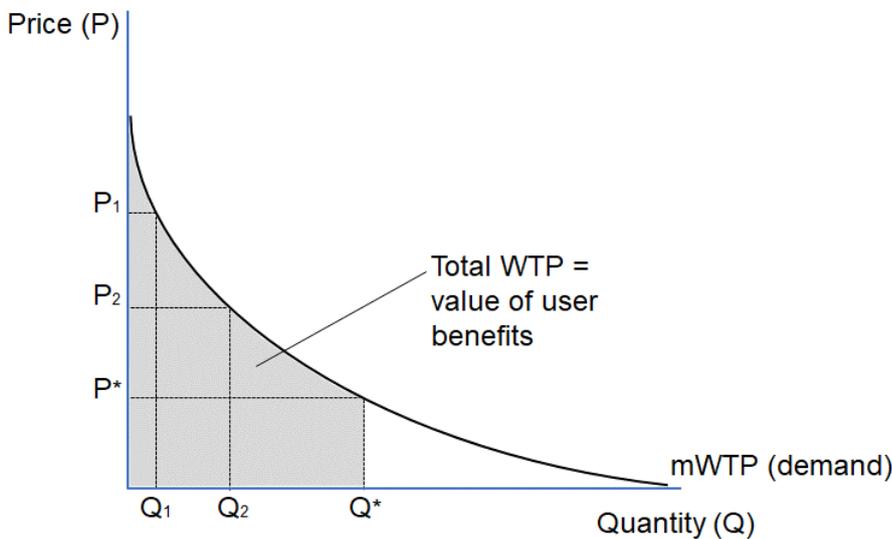
⁷ The term 'shadow price' simply refers to the estimated value of a good when a market price is not available. Shadow prices are commonly used for assessing environmental, social, and health impacts in policy appraisals. Examples outlined in the HM Treasury Green Book (see Annex 2; *ibid*) include values for changes in road, rail and aircraft noise exposure (see: <https://www.gov.uk/guidance/noise-pollution-economic-analysis>), or values for recreational uses of the natural environment (see: <https://www.leep.exeter.ac.uk/orval/>).

contribution to producer surplus, which is the difference between the price received for a good/service and the cost of providing it (see Box 2.2 below).

Box 2.1: Consumer surplus

Panel A below depicts a demand curve for a particular good (or service). The basic interpretation for an individual user is that it shows the quantity they will purchase (Q) as the price of the good (P) changes. Combining individual demand schedules across all users gives the overall (market) demand function for the good.

Panel A: demand schedule

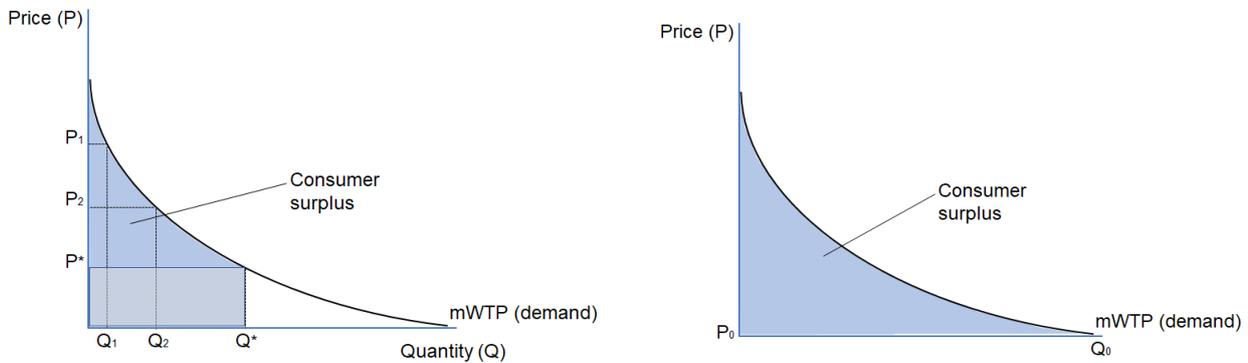


Interpreted another way, the demand schedule represents the user’s (marginal) willingness to pay (WTP) for successive quantities of the good. This is the maximum amount they would pay for each additional unit of consumption. For instance, for quantity Q_1 , the most the user would pay is P_1 , and for Q_2 the most the user would pay is P_2 . If the actual price that is charged is P^* , then the user would purchase up to Q^* . The maximum amount the user would pay declines as the quantity of the good increases due to diminishing marginal utility; that is, as the amount consumed increases, the additional utility (benefit) declines. Summing the maximum amount the user would pay for all quantities of the good up to the actual amount consumed (Q^*) gives their total willingness to pay. This represents the total value of the good to the user.

In the case of a market-priced good, the net welfare gain from users being able to buy Q^* at a price of P^* is the ‘surplus’ of value over the total purchase cost. This is known as consumer surplus and is the measure of welfare associated with provision of the good (see Panel B below; chart on the left). If the price of the good increased, consumer surplus would decrease (i.e. users are worse-off). If price decreases, consumer surplus increases and users are better off.

Box 2.1 cont.

Panel B: consumer surplus - market good (left) and non-market good (right)



For a non-market (un-priced) good that is provided free-of-charge – such as CH data – consumer surplus is equivalent to the entire area under the demand curve (Panel B; chart on the right). Effectively the price is zero (P_0) and the level of consumption is maximised (Q_0); i.e. no users are priced out of the market.

Since consumer surplus measures the wellbeing of individual users of a good/service, it can be summed across all users as the area under the aggregate demand function, and thereby establishes the total user benefit of a good.

In principle, therefore, the user benefits of CH data should be valued in terms of the economic surplus that is generated by the consumption and production activities that utilise it. However, the perspective that is adopted – consumer surplus or producer surplus – requires some consideration. Conventionally, consumer surplus is associated with the wellbeing of household consumers, whilst the principle users of CH data are businesses and organisations such as companies, creditors, investors, researchers, and public sector bodies. Nevertheless, these entities will have a demand schedule for CH data and their WTP should reflect how they use it and its value to their business.

The motivations that drive WTP will, though, vary across different types of organisation (see Section 2.2). For example direct users' WTP may be determined by how the availability and use of CH data allows them to make better informed business decisions. In contrast, the WTP of intermediaries – who incorporate CH data into their own data products and services – should reflect the value of the information as a productive input. For these users in particular, their 'producer' WTP should notionally be equivalent to the marginal revenue product of CH data as an input to production. This represents the most a producer would be willing to pay for one more 'unit' of data as an input, and traces out their (derived) demand for the input.

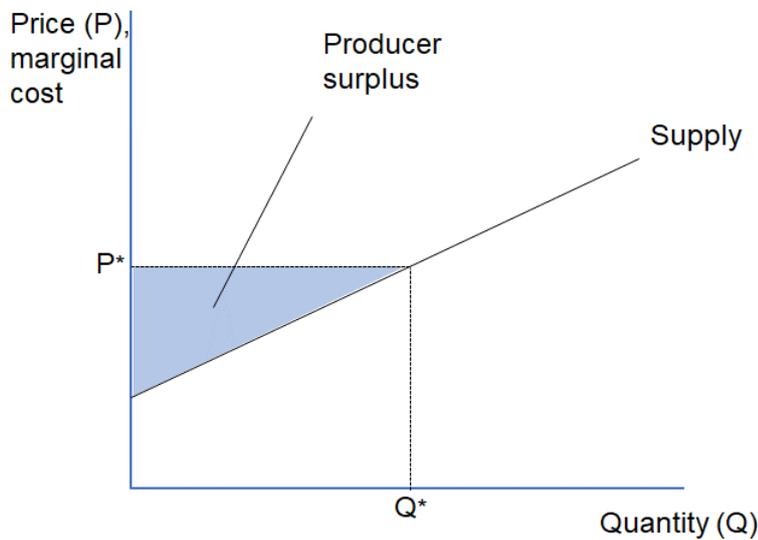
Moreover, a producer surplus perspective may be more appropriate for intermediaries, although what needs to be established in this context is the contribution of CH data to producer surplus (i.e. its economic rent) – this is likely to be best identified when CH data

is a significant input to production and therefore can be expected to also make a large contribution to producer surplus.

Box 2.2: Producer surplus

Producer surplus measures the welfare effect to a firm associated with the production of a particular good (or service). It is represented by the difference between the price the firm receives for its output and the minimum amount the firm is willing to accept for it (the 'supply price'). Since producer surplus measures the welfare gain to individual producers, it can be summed across all producers as the area above the aggregate market supply curve and below the price to determine the aggregate (total) producer surplus that accrues.

Panel A: producer surplus



In Panel A, the firm's willingness to accept (WTA) schedule (supply) is represented by its marginal cost curve (since profits are maximised when supply price = marginal cost). In order to supply Q^* , the firm's minimum WTA is P^* . The excess it earns over its minimum WTA for every unit of output up to Q^* is the producer surplus that is gained. This is equal to profit minus fixed costs and shows the net benefit to the firm from being able to sell at price P^* given variable costs consistent with marginal costs.

A key point to note, however, is that producer surplus measures the economic surplus generated by the combined factors of production for the supply of a good (e.g. labour, land, capital). To measure the return to any one productive input, its economic rent needs to be estimated. This is the excess of return over the minimum supply price for the productive input, which is determined by the next best alternative use of the input (i.e. its opportunity cost). Hence the value of CH data as a productive input is measured by the economic rent it generates.

2.2 Uses of CH data

CH data is a form of economic and business data, providing company registration and financial information about most companies and partnerships in the UK. It is made available to the public through the publication of the Company Register and the PSC Register. It can be accessed as 'raw' data files or as data images (e.g. pdfs), both of which are included in the definition of CH data used in this report. The information is provided at the level of an individual company, such as the list of beneficial owners, financial statements and annual filings, as well as in aggregate through statistical releases that could also be used for research and academic purposes (e.g. number of incorporated companies in the UK).

Company-level information is available in the form of the documents that are filed by companies via an online or mobile app search or an application programme interface (API). For the most part, these data are available free of charge to users via the Companies House Service⁸ and WebCHeck⁹; although, in the latter case, some documents/reports require payment of a nominal administration charge. For example, Companies House Direct¹⁰ is a paid-for subscription service that provides access to the same data.

Various commercial data products and services also include CH data, which is provided alongside other company information and value-added features such as credit ratings for companies (e.g. credit reference agency services), reports and analytics (e.g. FAME¹¹). Data are accessed in bulk from CH for these products by data services providers, who perform the role of 'intermediaries' in adding value to CH data and making it available to other users. Their value-added products are typically provided as subscription services to their users.

Users of the data are primarily concerned with checking the ownership of companies and their financial status, allowing businesses to gauge the credit-worthiness and reliability of previous, current and potential suppliers and customers. The data therefore helps inform decisions about whether to procure or provide services or products from/to certain suppliers. For direct users, an SME might use CH data to research a potential supplier to understand their financial position before deciding whether to contract their services. Alternatively, a financial services company might use CH data to access and assess company information to determine whether to provide credit to businesses, while a researcher in an academic institution might use the data as part of economic or business research.

⁸ See: <https://beta.companieshouse.gov.uk/>

⁹ See: <http://wck2.companieshouse.gov.uk/wcframe?name=accessCompanyInfo>

¹⁰ See: <http://direct.companieshouse.gov.uk/>

¹¹ See: <https://fame.bvdinfo.com/version-2018810/Home.serv?product=fameneo>

2.3 Benefits of CH data

There are different ways to look at the user benefits of CH data:

- **Asset value:** this approach treats CH data as an economic asset. It is a resource that generates a flow of value currently, with the expectation that it will continue to do so in the future. CH is effectively the steward of the resource, responsible for compiling the data and making it accessible for users. In doing so, CH incurs the costs of maintaining the asset for current and future uses. There are also ongoing costs incurred by the companies that provide the source information in line with regulatory reporting requirements.

The asset value perspective highlights that the value of data and information can depreciate over time. For instance, the more dated financial information becomes, the less value it has to most users, since it no longer reflects the current status of a company. Similarly, if the data is not maintained and kept up to date, its value as an economically productive asset will depreciate.

The conventional basis for valuing an asset of this nature is in terms of the discounted value of the future income stream (i.e. the revenue to CH net of the costs of maintaining the asset). For CH data, this would represent a partial valuation, since income is only generated from those users who pay fees to access CH data, including: (a) the purchase of some bulk data products (e.g. DVDs); (b) payment of a nominal fee (e.g. to access documents via the WebCheck service); or (c) payment of a subscription for the Companies House Direct (CHD) service. A more complete asset value can be estimated with the inclusion of the non-market component of the use value (i.e. from the use of free-of-charge services). As with the income stream component, the flow of value associated with the free-of-charge uses should be discounted over time to provide an estimate of the asset value.

- **Demand/consumption:** the broader view of the flow of value that is generated by CH data is provided by an assessment of the demand for it and the associated consumer surplus. As illustrated in Box 2.1, this captures both the priced uses of the data and the non-priced uses of the data, since consumption benefits (consumer surplus) are reflected by the difference between users' WTP to obtain it and the price paid (including cases where the price is zero).
- **Production input:** where CH data is used as a production input (i.e. for intermediaries), the user value can be measured in terms of the contribution to producer surplus. For direct users, there is also the potential to consider changes in producer surplus; for example through reduced operating costs as a result of the lower search effort that is needed to acquire the data and/or due to the lower purchase cost. This perspective implies a substitute source for the data that is costlier for the business or organisation – hence there is a gain in producer surplus since the business or organisation can achieve the same outcome (e.g. amount of revenue) at a lower cost; or for the same input/cost (e.g. staff time), achieve a higher level of output (e.g. increased productivity).

- **Value added:** this perspective is mainly relevant for intermediaries and the associated commercial data products and services that incorporate CH data. Here the data is effectively an intermediate input to a value chain and the net value of its use is assessed by considering differences between the value of a firm's output and the value of inputs, providing a measure of the value contributed by its production processes¹². Whereas estimates of producer surplus may be developed from bottom-up and firm-level assumptions, value added is ordinarily approximated using aggregate sector or regional economy level input-output models from national accounts. Expenditure by the final consumer – the users of the data service – is equal to the net outputs along the value chain minus the cost of the production input.

¹² Note that value added generated by a firm is not equivalent to the economic rent or marginal revenue product of the input since other production inputs (e.g. labour) are not treated as externally purchased inputs. Hence there is inherent bias in the value added approach which can lead to an over-estimate of the contribution of the production input of interest.

3 Valuation methodologies

3.1 Range of methods

There are two distinct approaches to estimating the user benefits of CH data. The first uses market-based data (e.g. transaction prices, revenues), whereas the second uses (non-market) economic valuation methods. The latter are concerned with the analysis of consumption, demand, production or supply relationships and involve estimating shadow prices and economic surplus values in cases where market prices are missing or distorted.

3.1.1 Market value-based analysis

Analyses that use market-based data essentially construct the demand curve for the good or service in question. This can involve analyses of varying degrees of sophistication¹³:

- **Point expansion:** estimation of the demand curve for a good or service using a market price and quantity. This takes an observation of a single point on the demand curve and applies a previously estimated or assumed elasticity¹⁴. This approach has previously been used in studies considering pricing regimes for the provision of public sector information (see Annex 2), where the point expansion from a revenue estimate (price and quantity) enables calculation of the consumer surplus associated with the provision of a good or service. It is a simplified approach – and potentially an over-simplification of the demand curve – and is reliant on the use of pre-determined elasticity estimates. However, it can be a pragmatic approach for providing order of magnitude estimates in situations where there is limited demand information available.
- **Residual imputation method:** this estimates values associated with intermediate goods and services. It attempts to approximate the economic rent or marginal revenue product that a non-priced input to production provides (e.g. open access data) to an output of known value. The value of the non-priced input is calculated by subtracting all other costs of production from the value of the output that is produced; hence the ‘residual’ value is assigned to the non-priced input. The approach is focused on a single point on the demand curve – but can be used in combination with other approaches that estimate the full demand curve for a good – and requires that all other production costs are known. The potential for error in the approach is significant, and hence is more appropriate where the non-priced input is

¹³ For technical summaries of the methods described in this section see: Young (2005) *Determining the Economic Value of Water – Concepts and Methods*, RFF Press; and Griffin (2006) *Water Resource Economics*, The MIT Press. Whilst these references focus on the estimating the economic value of water, there are parallels to assessments of the value of public sector data and information, such as non-market/priced provision, and both producer/intermediate and final consumption perspectives on the use value.

¹⁴ The elasticity estimate determines the slope of the demand curve. Combined with the price and quantity data this means that consumer surplus can be estimated.

a major input to production, since this implies a lower likelihood of over-estimating benefits.

- **Direct estimation:** if data are available, the demand curve for a good or service can be directly estimated by econometric analysis. This is reliant on the availability of consumer behaviour data for a range of price points that show how the associated consumption level for the good or service varies as the price changes. Conventionally, this should be the most accurate estimation approach and hence a more reliable basis for estimating consumer surplus, but it is reliant on the availability of suitable data.

3.1.2 Production function

The value of data as a production input can be assessed using a combination of data requirements:

- **Production function:** this analysis quantifies the relationship between an input to production and the output. Typically, production inputs are materials, labour and capital. Data inputs can be incorporated into the analysis in terms of the purchase costs (materials) or the search efforts in terms of staff time and costs (labour). The change in output that occurs when the data input changes – holding all other inputs constant – is the marginal product of data¹⁵. Multiplied by the output price for the product (good/service), this provides an estimate of the marginal value of the production input. The result provides a measure of a producer's (marginal) WTP for a production input. As with simpler residual imputation methods, the production function approach is more appropriate where the input in question is a key input to production (since a greater proportion of output is dependent on its availability).

3.1.3 Value added

As noted above, an approximate approach to assessing the value associated with use of a production input to a value chain is the value added approach:

- **Value added:** this utilises empirical results from (aggregate) economy-wide accounting systems (regional or national) in the form of input-output models. In basic terms it involves estimating the residual value of a production input by subtracting other input costs (purchased from other firms) from revenues. Notionally this appears to be similar to the residual imputation method, but there are, however, fundamental differences. For example, the value added perspective does not isolate the contribution of one input to production (e.g. raw data), rather it measures the productivity of all (purchased) production inputs (e.g. materials, energy). It also does not provide a 'complete' picture of the opportunity cost of other production inputs (e.g. wages paid to labour are treated as a part of the value added of a sector rather than a cost). Therefore, the value added of an input is invariably greater than an associated producer surplus measure that would be based on the value of marginal product (i.e. producers' WTP).

¹⁵ Strictly, this is the marginal physical product of the factor input.

Pragmatically, the value added approach provides an indication of the economic activity associated with a particular production input, but not a reliable measure of the benefit to producers. If data are available – or suitable assumptions are made – conceptually economic profit can be estimated by subtracting employment costs, while producer surplus could be estimated by subtracting fixed costs¹⁶.

3.1.4 Inferring the value of non-priced uses from market data

For non-priced uses of data, the emphasis is mainly on the application of non-market valuation methods to estimate demand (see below). However, an indicative valuation can be inferred based on market prices:

- **Inference from market values:** one approach noted in the HM Treasury Green Book in relation to estimating asset values for non-priced assets (e.g. discounted income stream) is to use a comparable income stream or sales value. Hence the marginal value associated with non-priced uses of CH data could be assessed by assuming that the price associated with ‘paid for’ uses of the data also represents the WTP of open access users. As with the point expansion, this approach is pragmatic in cases of limited data availability and may be suitable for providing order of magnitude estimates that can act as benchmarks and sense-checks for other approaches applying non-market valuation methods.

3.1.5 Non-market valuation methods

Non-market valuation methods are distinguished in terms of the revealed preference or stated preference approaches:

- **Revealed preference:** these methods focus on relationships that exist between the demand for a market-priced good (e.g. a commercial data services product) and preferences for a related non-priced good (e.g. CH data). The underpinning theoretical model for this analysis is the discrete choice model, which describes how consumers choose between two or more alternatives (e.g. different data products). By observing choices made between different products with different features (i.e. attributes such as the types of data and information that are provided, the analytical functions of the service, the subscription cost, etc.), the value of individual features to consumers – in terms of their (marginal) WTP – can be recovered. The approach is therefore dependent on suitable market choices being available to consumers, along with the associated data, which can be sourced at the aggregate market level or individual consumer level. Typically, individual consumer level data is captured through survey-based methods (e.g. household surveys), whilst market level data is collected from market data providers.
- **Stated preference:** these are survey-based approaches that present simulated markets for trading non-priced goods. The relevant approaches for estimating consumer WTP are the contingent valuation method and the discrete choice experiment. Generally, the contingent valuation method provides a value of a ‘bundled’ or ‘whole’ good or service in terms of a defined change from one level of

¹⁶ Producer surplus is the difference between total revenue and total variable cost (or marginal cost); economic profit is the difference between total revenue and total cost (fixed plus variable).

provision to another (e.g. a subscription to a data product). Discrete choice experiments are an attribute-based approach that break a good or service down into its characteristics or features and provide a value for (independent) changes in the provision of these attributes. It is based on the same underlying choice model as the attribute-based revealed preference approach, the difference being that it utilises simulated market choice data, rather than observed market choice to elicit the value of non-priced attributes. The choice of approach (contingent valuation vs. discrete choice experiment) largely depends on the requirements of the analysis and in practice the two methods can be complementary.

Overall, stated preference methods represent a more flexible non-market valuation approach as they can be used to value both previously experienced and future (i.e. not current or previously experienced) levels of provision of a good/service, subject to specifying a credible simulated market choice and context. In contrast, revealed preference methods have a more limited scope of application dependent on, for example, the presence of an attribute-based relationship between a market priced good/service and a non-priced characteristic. Moreover, because they are reliant on observed behaviour and data, they can only be used within the bounds of current or previous levels of provision of a good or service. Revealed preference would also not reflect public good and/or non-use benefits associated with the provision of a good/service. Where the objective is to estimate these values, then stated preference methods are typically the only approach that can be applied.

Data availability may also be a factor that constrains the ability to use revealed preference methods, in the same way that this can be a practical limitation for other market-based value analyses. In the context of data limitations, stated preference methods have an additional appeal in that these survey-based approaches can collect all the necessary data on consumer choices, characteristics, experiences and attitudes that are required to estimate demand for a good whilst controlling for a range of factors that are expected to influence WTP.

3.1.6 Cost-based approaches

A final approach to assessing user benefits is provided by cost-based approaches which can be used as a proxy for demand-based measures of value (i.e. a lower bound estimate of consumer WTP). In the context of CH data, the relevant cost-based proxy is the replacement cost – i.e. the cost of obtaining the same data and information inputs from an alternative source (at the least cost). Note that this could include search effort costs if the alternative option is for a user to compile the required information themselves. Overall, the approach is reliant on the assumptions that an alternative source is available, that the substitute represents the same level of provision (e.g. data quality), and that the user would be willing to incur the (higher) cost to obtain the alternative. As an approximate measure of WTP, the replacement cost does not reflect any consumer surplus for the use of a good/service over the cost of obtaining the substitute product.

3.1.7 Summary of methods

An overall summary of the range of methods described above is provided in Table 3.1.

Table 3.1: Summary of valuation approaches

Method	Description	Information Required	Limitations	Areas of application
Point expansion	Estimation of a demand curve using a demand elasticity and a known point on the demand curve.	A point on the demand curve (price and quantity) and exogenously determined elasticity.	Potential over-simplification of demand curve and results are sensitive to the assumed demand elasticity.	Can be applied generally, but dependent on demand elasticity estimate(s).
Residual imputation method	Estimation of the contribution and production benefits of a single input.	Costs of production and price and quantity of input of interest.	Provides one point on demand curve and significant potential for error in determining the costs of production.	Suited to intermediate goods that are a major factor of production for a good to cover potential inaccuracy.
Direct estimation of a demand curve	Applying econometric analysis to establish the demand curve for a good or service.	Observed market data showing how consumer demand varies with different price points.	Data availability, and may require shadow pricing techniques if market prices do not reflect opportunity costs.	Estimating consumer benefits of a good or service where the market exists.
Production function	Estimation of a continuous demand curve function for a production input (by estimating the form of the production function).	Observed physical relationships between quantity of input and production output. Requires econometric analysis.	Requires large amount of data to analyse that may be difficult to obtain. Results can be highly dependent on the assumed functional form for the production function.	Estimating the producer benefits of a good or service that is an input to production.
Value added	A measure of value based on the difference between a firm's output and the value of the inputs purchased from other firms.	Purchased input costs and the value of output, plus sector level input-output multipliers.	The approach over-estimates producer benefits since it does not isolate the value added associated with a single production input and treats some production inputs as value added activities (e.g. labour).	Suited to value chains from intermediate goods to final goods.
Inference from market values	Approach that assumes a closely associated market-based price is a suitable measure of WTP for a non-priced good.	A point on the demand curve (price and quantity) and exogenously determined elasticity (if point expansion is to be used).	This approach is effectively a form of 'value transfer' and assumes that consumer preferences in the market and non-market cases are consistent.	Cases where a suitable market-based equivalent good or service exists – note this could imply strong assumptions as to the relationship between 'price' and the level of provision (e.g. constant unit value vs. diminishing marginal value).
Revealed preference	Observing consumer behaviour in related markets to infer the	Observed market data and consumer choices between	Limited number of relationships between a consumed good and non-	Where a final consumer good

Method	Description	Information Required	Limitations	Areas of application
	value of a non-priced attribute of the good or service.	alternative products that feature the attribute of interest.	priced attribute that can be utilised.	incorporates the attribute of interest.
Stated preference	Use of simulated markets that present trade-offs for a non-market good or non-priced attribute with money.	Survey-based data.	Valid and robust application requires specification of a credible simulated market that can be administered in a survey-based format.	A range of contexts for the provision of a non-priced good, including future levels of provision.
Replacement cost	Use of the cost of an alternative good that provides the same/similar function as a measure of consumer WTP.	Market data for the price (cost) of the replacement good.	Can only be considered as a low-bound proxy estimate for consumer WTP since it does not reflect any consumer surplus above the price paid.	Cases where an appropriate substitute product is available.

3.2 Data requirements

A summary of the high-level data requirements for the range of valuation methods summarised in Section 3.1 is provided in Table 3.2.

Table 3.2: Summary of data requirements

Method \ Data requirement	Market data	Empirical data	Production data	Attribute data	Survey data
Point expansion	✓	✓	-	-	-
Residual imputation	✓	-	✓	-	-
Direct estimation	✓	✓	-	-	-
Production function	✓	-	✓	-	-
Inference from market values	✓	-	-	-	-
Revealed preference	✓	-	-	✓	✓
Stated preference	-	-	-	✓	✓
Replacement cost	✓	-	-	-	-

The main types of data requirement are:

- **Market data** – these are the price and quantity of data products in a functioning market. Different methods may require multiple data points, or a single price and quantity as summarised above.
- **Empirical data (results)** – these are results from previous studies, such as regression results, elasticity data or input-output multipliers.
- **Production data** – these require knowledge of the inputs into a production process. Certain methods require detailed information on the relationship between inputs and outputs.
- **Product (attribute) data** – information on the attributes (features) of data products and services.
- **Survey data** – where market prices for data and related market goods do not exist or are not sufficient, data on users' values are elicited through specially-designed surveys.

In practice, data requirements were a key consideration for the study since the available methods can only be used where data exist and the outputs of the analysis are only as good as the quality of the data. Obtaining firm-level production and cost data is a particular challenge, given its confidential nature, and is typically reliant on consultation or interviews with data users (i.e. intermediaries). Market data and empirical results can also be partial and reliant on supporting assumptions – hence the use of methods such as point expansion (see Annex 2).

4 Approach

4.1 Value chains for CH data user types

Figure 4.1 provides a set of outline value chains for the three types of user of CH data (as described in Section 1.3) specified for the study:

- direct users: includes businesses that use CH data to carry-out due diligence on suppliers or customers, and researchers who use the information in their own research;
- intermediaries: use CH data as an input to their own data products and services (including credit reference agencies and providers of financial data and information); and
- providers of public goods: government departments or law enforcement organisations that use CH data in policy or investigative work.

The value chains summarise how the concepts and different value perspectives can be applied to the different user types and uses of CH data. These are specified in generic terms for CH data, rather than detailing specific services or types of information (e.g. CH register vs PSC register). For the most part, these distinctions are not necessary at the conceptual level set out here.

One conclusion to draw is that a demand perspective is likely to provide the most consistent basis for estimating economic surplus across each of the user types. For direct users this is based on the assumption that WTP would be driven by how CH data contributes to better decision making, and how its ready availability implies lower search and operating costs for businesses. Conceptually, intermediaries' WTP should reflect the importance of CH data as an input to the services and products that they provide, while providers of public goods are likely to be more closely aligned to direct users in terms of the value of the data (based on better outcomes and lower operating costs).

In contrast, a producer surplus or value added approach is less tractable across the three user types. For direct users and providers of public goods, CH data is likely to be incidental to their main productive activities, hence determining the contribution to economic profit or surplus is expected to be challenging. Nevertheless, for intermediaries, a production-side analysis provides a useful 'sense-check' on surplus estimates derived via a demand-based approach – given the derived demand relationship for CH data – provided suitable data are available.

Figure 4.1: Outline value chains for the different types of user of CH data

Direct (mostly business) users of CH services

Company reporting and financial information	Final user
CH data 	Businesses and/or researchers 
Revenue to CH: none (zero price)	WTP for CH data Consumer surplus = max. WTP – price (over quantity purchased)
	Demand for CH data based on: <ul style="list-style-type: none"> - More informed decisions and better outcomes (e.g. higher revenue/utility) - Lower search effort (time/resource saving) - Lower business operating costs (vs. alternative sources)

Intermediaries

Company reporting and financial information	Intermediate user	Final user
CH data 	Value added data services providers 	Business users and/or researchers 
Revenue to Companies House: <ul style="list-style-type: none"> - none (for zero price data) - price x quantity (for fee-charging products/services) 	= Cost of purchased input Revenue to intermediary: price x quantity Producer surplus = revenue – variable costs of production	WTP for data services products Consumer surplus = max. WTP – price (over quantity purchased)
	WTP for CH data Demand for CH data based on: <ul style="list-style-type: none"> - Contribution of data to value of data service products (marginal revenue product) - Lower business operating costs (vs. alternative sources) 	Demand for data services products based on: <ul style="list-style-type: none"> - Value added features (e.g. analytics) - More informed business decisions and better outcomes (e.g. higher revenue)

Providers of public goods

Company reporting and financial information	Final user
Companies House data 	Public sector bodies and other providers of public goods 
Revenue to Companies House: <ul style="list-style-type: none"> - none (for zero price data) - price x quantity (for fee-charging products/services) 	Benefit of Companies House data / demand: <ul style="list-style-type: none"> - Cost savings and efficiency - Better outcomes (public goods and services)

4.2 Summary

Drawing on the preceding discussion, the study uses a survey-based approach to establish user demand for CH data, applying a stated preference method to estimate user WTP. Whilst a simulated market approach is not necessarily the ‘first choice’ from a valuation hierarchy perspective – even when recognising that the largely unpriced nature of CH data place the study within the domain of non-market valuation methods – there are various trade-offs to consider, including ensuring consistency of the analytical perspective across all user types and the availability of data.

As noted above, a stated preference approach offers considerable flexibility compared to alternative methods and is not reliant on third party data. All data requirements could therefore be met through the design of the survey and product information provided by CH. Furthermore, with the design of the survey, there was the potential to ‘build-in’ consistency checks for estimates of WTP. This included capturing data from direct users on marketed products that they purchase (substitutes) and also estimates of time/cost savings due to the ready availability of CH data. Further information on the design, testing, implementation and analysis of the stated preference research is provided in Report 2.

For intermediaries and providers of public goods, the stated preference approach has been supplemented by qualitative research that explored: their use of CH data products and services and data from other sources; the costs and revenues attributed to the use of CH data; and the availability of substitute data sources. This additional analysis for intermediaries and providers of public goods is largely qualitative in nature given the relatively small number of these users and interviews that were conducted.

As described above, the results of these research tasks are presented for the different types of user in the other reports in the series:

- Report 2 presents the findings of the stated preference research and resulting WTP estimates for all users of CH data, focusing in particular on the direct users;
- Report 3 presents specific findings for intermediaries, including WTP estimates from the stated preference research and analysis of the qualitative research;
- Report 4 presents a summary of findings from the stated preference research and qualitative research for providers of public goods; and
- a separate policy summary draws overall conclusions about the value of CH data to all users.

ANNEXES

Annex 1 References

Companies House (2019) Companies register activities: 2018 to 2019.

Companies House (2019) Business Plan 2019 to 2020.

Companies House (August 2018) Companies House management information 2017/18.

HM Treasury (2018) The Green Book - Central government guidance on appraisal and evaluation.

Koski (2011) Does Marginal Cost Pricing of Public Sector Information Spur Firm Growth? Helsinki: ETLA – The Research Institute of the Finnish Economy.

Pollock (2010) Welfare Gains from Opening up Public Sector Information in the UK. Cambridge: University of Cambridge.

PwC (2010) Economic Assessment of Spatial Data Pricing and Access. Canberra: ANZLIC – The Spatial Information Council.

COWI (2010) The Value of Danish Address Data. Danish Enterprise and Construction Authority, Copenhagen.

Pollock, Newbury and Bently (2008) Models of Public Sector Information Provision via Trading Funds. Cambridge: University of Cambridge.

Office of Fair Trading (2006) The commercial use of public information (CUPI).

Griffin (2006) Water Resource Economics, The MIT Press

Young (2005) Determining the Economic Value of Water – Concepts and Methods, RFF Press

Annex 2 Literature review

The value generated by access to and the provision of different types of public sector information has been considered by a number of previous studies (Table A.1).

Table A.1: Summary of previous studies

Author(s)	Title	Methodology	Results
Pollock et al. (2008)	Models of Public Sector Information provision via Trading Funds	Economic surplus (market values) approach comparing gains from moving to different cost pricing regimes.	Reducing the price of 'wholesale' or 'bulk' digital data has welfare-enhancing benefits. In the case of CH data, the benefits were estimated at £2.6 million per year.
Pollock (2010)	Welfare gains from opening up public sector information in the UK	Economic surplus (market values) comparing gains from moving to different cost pricing regimes.	Reducing cost of digital public sector information results in direct, indirect and efficiency gains in the range of £1.6 billion to £6 billion per year.
COWI (2010)	The value of Danish address data	Consumer surplus gains from making geographic information free using market price data.	Access to free geographic information resulted in direct benefits and savings to the value of €62 million over a 4-year period.
PwC (2010)	Economic Assessment of Spatial Data Pricing and Access	Economic surplus (market values) comparing gains from moving to different cost pricing regimes.	Making public information free delivers economic benefits that exceed alternative pricing models. The benefits are due to increased competition and innovation by downstream users.
Koski (2011)	Does marginal cost pricing of Public Sector Information spur firm growth?	Econometric analysis: <ul style="list-style-type: none"> • Random effects (RE) model captures relationship between pricing model and firm growth; and • Difference-in-difference (DiD) model estimates growth attributed to changes in pricing model. 	Reducing the price of data is positively related to the sales growth of firms, with the effect being more dramatic for Small and Medium Enterprises (SMEs). The growth effect becomes stronger following a two-year lag.

These studies have used a variety of approaches to provide empirical assessments, focusing particularly on alternative pricing regimes for public sector information and the impact on consumer and producer surplus. For instance:

- profit-maximising pricing – setting a price for the data that generates a profit over and above the cost involved in collecting and providing the dataset;
- cost-recovery pricing – setting the price for data at the same level as the costs related to its provision;

- marginal-cost pricing – charging a price that is equivalent to the cost of transmitting the data to a consumer (e.g. the cost of writing the data onto a DVD); and
- zero-cost pricing – free access and distribution of data to consumers.

Pollock et. al. (2008)¹⁷, Pollock (2010)¹⁸ and PwC (2010)¹⁹ all consider changes in economic surplus associated with different pricing regimes for public sector information. For instance, Pollock et. al. (2008) estimated the gain from a change in cost-recovery pricing to marginal cost pricing for the six largest public sector information holders (PSIHs) in the UK, including CH. The analysis focuses on the pricing of ‘bulk’ (wholesale) data and assumes other data product prices would remain unchanged. While information on the supply of data was readily available (e.g. type of data, number of downloads, and associated price), demand information was not. The analysis therefore applies results from the literature to estimate: (i) the elasticity of demand²⁰; and (ii) a value chain ‘multiplier’ that estimates the benefits from value-added products and services (e.g. software tools and consultancy services)²¹. For CH data, the analysis suggests that under the assumption of a ‘high’ demand elasticity and a ‘medium’ multiplier, a change from cost-recovery to a marginal-cost price policy would result in gross benefits of approximately £2.6 million per year with the Government incurring net costs of £681,000 annually.

Pollock (2010) applies the findings from the preceding Pollock et al. study to assess the benefits of a shift from a cost-recovery to a marginal-cost pricing regime for UK public sector information. The analysis uses an assessment by the Office of Fair Trading (OFT, 2006)²² that estimated the total income from sales of public sector information in the UK to be approximately £400 million. Based on this estimate of revenue and upper-end estimates for the elasticity of demand and multipliers, the gain in consumer surplus is estimated to be approximately £4.5 - £6 billion per year²³. The middle range estimate for the gain in consumer surplus is approximately £1.6 - £2 billion per year²⁴.

¹⁷ Pollock, Newbury and Bently (2008) *Models of Public Sector Information Provision via Trading Funds*. Cambridge: University of Cambridge.

¹⁸ Pollock (2010) *Welfare Gains from Opening up Public Sector Information in the UK*. Cambridge: University of Cambridge.

¹⁹ PwC (2010). *Economic Assessment of Spatial Data Pricing and Access*. Canberra: ANZLIC – The Spatial Information Council.

²⁰ The price elasticity of demand measures how the quantity of good or service that is consumed (demanded) responds to price (i.e. how much demand falls when price increases and vice versa).

²¹ The calculation that is applied is as follows: $Gains = 2/5F\lambda\epsilon$, where F is revenue under average cost pricing, λ is the value chain multiplier, and ϵ is the price elasticity of demand.

²² Office of Fair Trading (2006) *The commercial use of public information (CUPI)*. Available at: https://www.ec.europa.eu/newsroom/document.cfm?doc_id=1194

²³ Upper end estimates use: $\lambda = 8$, $\epsilon = 3.5$, resulting in estimated gains of approximately £4.5 - £6 billion per year.

²⁴ Middle range estimates use: $\lambda = 5$, $\epsilon = 2$ resulting in estimated gains of approximately £1.6 - £2 billion per year.

PwC (2010) provides an economic assessment of pricing regimes for topographic information, aerial photography, and cadastre and administrative boundaries for four Australian public sector agencies²⁵. Similar to Pollock et. al. (2008) and Pollock (2010) a comparative analysis is conducted to estimate the benefits of moving from a cost-recovery to zero-cost pricing regime. However, it is based on a dynamic effects model that attempts to factor-in potential changes in funding for the producer agency and the associated implications for data quality in terms of accuracy, currency and resolution. The model results indicate that in the first year following the change to zero-cost pricing, net benefits range from AUS \$1 million to AUS \$4.7 million per annum. Thereafter, the cumulative cost to the public sector agencies requires government funding to cover any shortfall, and by year 10 in the model, the outcome of zero-cost pricing ranges from a net loss of AUS \$0.4 million to AUS \$2.3 million annually.

A simplified approach to estimating the benefits of free access to geographic information (address data) in Denmark is provided by COWI (2010)²⁶. Between 2005 and 2009, the data had been delivered to approximately 1,200 public and private users. Using the price users paid for address data prior to free access, COWI calculated a gain in consumer surplus of €62 million for the study period. The associated cost to the public sector was calculated as €2 million, resulting in a net surplus of €60 million for the study period from 2005 to 2009. For the year 2010, gross benefits were estimated at €14 million with associated costs of €0.2 million.

An alternative perspective to the analysis of economic surplus is provided by Koski (2011)²⁷. The study by the Research Institute of the Finnish Economy examines firms' sales growth rather than market prices, to assess the impact of marginal cost pricing compared to other pricing regimes. The study uses data from various firms in Australia, Europe and the United States for the period from 2000 to 2007. Based on econometric analysis applying alternative model specifications²⁸, it is estimated that firms located in countries with marginal-cost or zero-cost pricing policies grew, on average, 15 per cent more per annum than firms located in countries where information is subject to cost-recovery pricing²⁹. Furthermore, a change in pricing policy is estimated to have an immediate economic benefit with firm growth of between 7 and 19 per cent higher in the two years after a change in pricing policy.

²⁵ i) Victoria's Department of Sustainability and environment topographic data, ii) Landgate topographic data, iii) Landgate aerial photography, and iv) Geoscience Australia topographic data.

²⁶ COWI (2010) The Value of Danish Address Data. Danish Enterprise and Construction Authority, Copenhagen.

²⁷ Koski (2011) Does Marginal Cost Pricing of Public Sector Information Spur Firm Growth? Helsinki: ETLA – The Research Institute of the Finnish Economy.

²⁸ A random effects model that linked firm growth to information pricing policy, and a difference-in-difference model that estimated the amount of firm growth attributable to a change in pricing policy.

²⁹ This is relative to the 2.9% average annual growth rate of firms located in countries that conduct cost-recovery or profit maximising pricing regimes.

This publication is available from: www.gov.uk/beis

If you need a version of this document in a more accessible format, please email enquiries@beis.gov.uk. Please tell us what format you need. It will help us if you say what assistive technology you use.