Data Use Externalities

Report to Department for Digital, Culture, Media and Sport by Belmana with the University of the West of England

August 2021

Glossary

Term	Description
Data ecosystem	A data ecosystem is a collection of infrastructure, analytics, and applications used to capture and analyse data. The term ecosystem is used rather than 'environment' because data ecosystems are intended to evolve over time and the term refers to interactions between multiple data users.
Data trusts	A data trust is a mechanism for individuals to take the data rights that are set out in law and pool these into an organisation - a trust - in which trustees make decisions about data use on their behalf.
Data value chain	Viewing data use as a process with stages increasing the value of data, a sequence also called the 'data lifecycle'.
Economies of scale	Economies of scale (also referred to as supply side economies of scale) occur where production of a good on increasingly large scales tends to lead to increasingly low costs per unit made.
Economies of scope	An economy of scope means that the production of one good reduces the cost of producing another related good.
Externality	A secondary or unintended consequence that affects a person, organisation or group without their involvement in the underlying transaction. An externality can be either beneficial ('positive') or costly ('negative').
Friction costs	Costs incurred executing a transaction, such as in product information gathering.
GDPR	General Data Protection Regulation is regulation on the handling of citizens' personal data by public and private sector organisations.
Non-rivalrous	Non-rivalry means that one person's enjoyment of a good does not diminish the ability of other people to enjoy the same good.
Non-excludable	Non-excludability means that people cannot be prevented from enjoying the good.
Network externalities	A network effect (also called demand-side economies of scale) is where the value a user derives from a good or service depends on the number of users of compatible goods or services. They are typically positive.
Open data	Open data is a dataset that: ● Meets certain standards of accessibility (usually meaning published online). ● Structures in a way that is machine-readable; and ● Can be used by anyone for any purpose (because of the licence that it is published under).
Personal data	Data from which a person can be identified is personal data, including data that can be combined with other information to identify a person.
Personal data store	A Personal Data Store helps gather, store, manage, use and share personal information.
Public good	A commodity or service that is made available to all members of society.

Term	Description
R&D externalities	Indirect, external effects of research and development, such as the general increase in knowledge for society through R&D.
Spillover	Effects on parties not involved in an underlying transaction - similar to externalities - but ones that could be ‘internalised’. Internalised means the benefits or costs are somehow recouped by the parties involved in the underlying transaction.

Executive Summary

1. This study was commissioned by the Department for Digital, Culture, Media and Sport to:

   - Identify the likely positive and negative externalities of data use (that is, the wider social impacts of data use) and

   - Provide an assessment of the viability of methods that could be used to value them.

2. The study is primarily based on a literature review. This was substantiated by interviews with a range of relevant policy, academic and practitioner experts.

Data use externalities

3. An externality is a secondary or unintended consequence that affects a person, organisation or group without their involvement in an underlying transaction or process. A classic example is the impact on property prices caused by a desirable or undesirable new development nearby (such as a golf course, or airport)

4. Markets are typically less efficient when externalities occur, so it is important to understand when, where and to what extent they happen so proportionate responses can be implemented.

5. The characteristics of data and data use mean externalities occur in a wide range of activities and sectors that use data.

Typology of Data Use Externalities

6. The research for this report has developed a **typology of externalities of data use:**

   - **Supply-side externalities**, associated with producing data products (these are divided in this report into economies of scale, and economies of scope

   - **Demand-side externalities, including** economic efficiencies as services and products are consumed or enter markets (divided into friction costs and network effects)

   - **Legal externalities** are the indirect effects on rights such as those associated with personal privacy or intellectual property

   - **Wider external effects** are all other externalities associated with data use, including **environmental changes from greater efficiency, the cost of the unpaid labour needed to create some data, and other social externalities**

Locating Data Use Externalities

7. The research also looked into what kinds of data use activities might lead to which types of externalities.

Measuring Externalities in Data Use

8. Three **approaches were considered in this report,** modelling approaches, willingness to pay/ willingness to accept surveys, and other types of econometric approaches (such as regression analysis).

9. The findings of this report are that each approach:

    - has been used already within data policy to further our understanding of the contribution of data use and the impact of externalities

    - has strengths and weaknesses, and could benefit from being used together in some way; this has typically not been done so far in the literature analysed.

Recommendations and next steps

10. This report highlights how important the design of any valuation study is, with this often requiring some tailoring of the research method to the issues being considered.

11. Research on the cost to businesses of cyber security and data breaches appears to be the most developed area of data use externalities. While this is promising, it also indicates more work is needed on the valuation of data use externalities.

12. Policies improving data sharing are a stage later in the value chain than the generation of data, at the aggregation/analysis stage. The focus is enabling controlled data sharing, balancing privacy and security negative externalities with the potential benefits from re-use. The policies are often sector or use specific and enable the wider ecosystem to access data. These areas may be ones where data use externality valuation techniques can be used tailoring to the specific sectors, uses and the stage in the data value chain.

1. Introduction

1. The National Data Strategy sets out the opportunities associated with data use.[1] Data and its use can impact economic and social outcomes in positive and negative ways. Firms use large volumes of data to spot patterns and gain insights, monetising these in various ways.[2] They also typically put considerable effort into protecting the security of their data to prevent misuse. At an individual level, sharing information with an app might facilitate use of a helpful service like tailored travel alerts. If there's a cyber attack however, you may then get nuisance calls or spam.

2. The concept of an externality helps understand these wider social and economic harms and benefits. An externality is a secondary or unintended consequence, positive or negative, that affects a person, organisation, or group without their involvement in an underlying transaction or process.

3. Any externality of data use can lead to inefficient behaviours, as businesses, consumers and other economic actors use data without regard for these external effects. These inefficient behaviours are sometimes called "market failures".

4. This study was commissioned by the Department for Digital, Culture, Media and Sport to:

   - Identify the likely positive and negative externalities of data use
     - Understand whether these externalities arise at a particular point of the data 'life cycle'
     - Review the existing estimates of the scale or scope of these externalities
   - Provide an assessment of the viability of methods that could be used to value externalities:
     - What are the types of methods that have been or could be used to capture the value of externalities arising from data use
     - Assess the viability of these methods, to inform future research.

5. The findings come mostly from a literature review. Alongside this, interviews were conducted with policy makers and academics researching the economic and social impacts of data use.

This report

6. Chapter 2 expands on the definition of externalities, and then considers the types of externality identified through our research:

   - 'Supply-side' externalities that are associated with producing data products (or other products)
   - 'Demand-side' or economic efficiency externalities, that are likely to appear as services and products are consumed or enter markets

[1] DCMS (2020)

[2] Ker and Mazzini (2020)

- ‘Legal’ externalities, which cover effects on the individual, entity or system affected such as effects on privacy or intellectual property

- ’Wider environmental and societal’ externalities, effects beyond those relating directly to data use such as reduced carbon emissions from improved efficiency.

7. Chapter 3 discusses the ways data is used within organisations and within the data ecosystem. For both of these, it explores when these types of externality are likely to occur. The chapter highlights the importance of “platforms”.

8. Chapter 4 considers how data use externalities are associated with specific policies, covering open data standards and the use of settings that enable data sharing while retaining controls that mitigate harm.

9. Chapter 5 looks at approaches that estimate values of these externalities,including:

- ‘Modelling’ approaches, which has been used to estimate the value of externalities (sometimes called “shadow prices”)

- ‘Willingness to pay/ willingness to accept’ surveys, which collect data about what individuals are prepared to pay or have paid in relation to externalities

- ‘Econometric approaches’, where the effects of policies – with a focus on tackling market failures faced by businesses – are assessed using firm-level or similar detailed data.

2. Defining data use externalities

Chapter summary

- This chapter covers:
    - How characteristics of digitised data[3] and its use make understanding externalities an important facet of the 'life cycle' of data; and
    - Different types of 'externalities'.
- People and businesses[4] can benefit or suffer from data and data use even when they did not directly spend time or money in that data's creation, curation or analysis. These benefits or costs are called 'positive' and 'negative' externalities.[5]
- This is because of the particular characteristics of data, namely that:
    - Data is 'non-rival' **in nature.** That means that it can be used by more than one actor simultaneously without impacting other users or uses, for approaching zero extra cost and hence moving, sharing or publishing data is easy
    - Data typically represents information about people, things or events, in a condensed (numerical) format and hence mathematical or statistical summary are easy
- Markets are typically more efficient when the benefits of a product or service benefit the buyer exclusively, and the profits go to the seller exclusively. Understanding externalities and building them into transactions is therefore an important part making markets more efficient.
- Externalities take different forms. Those that relate to data use identified in this research are
        - **Su**pply-side
        - Demand-side
        - Legal externalities, and
        - A wider set o**f indirect effects**

[3] for the rest of this report, 'data' is used to mean 'digitised data' unless otherwise explicitly mentioned

[4] in economic terminology the term 'actor' or 'party' is used to describe people and businesses involved in a behaviour or transaction

[5] an example of positive externalities is when Transport for London published real-time operational data, which was then used by the public and software developers

Introduction

1. For conventional *private goods*, which can only be used by one person at a time (they are *rival*)[6] and others' use can be prevented (they are *excludable*), markets typically work well (see Figure 2.1 below).

2. Where goods or services don't meet these criteria, issues such as the free rider problem arise. The costs and benefits of a transaction increasingly affect people beyond the transaction itself, and the incentive for individuals to sell or buy them diminishes.

3. A positive or negative externality can, over time or through regulation of some kind, become internalised. This is when the externalities are factored into the transaction price. This process of internalising externalities is a key part of making markets work efficiently.[7]

4. Data is not a conventional private good. Data can be replicated at approaching zero cost, or accessed by multiple people from a single source. In these ways data can be non-rival.[8] And any structures to exclude others is in this context of data being easily replicated and re-used multiple times.

Figure 2.1: Rival and Excludable Goods

Rivalry and excludability are spectrums rather than binaries. With some goods and services, such as an apple, one party consuming all of it prevents another from consuming any of it. With other goods, such as a beach or a road, many parties can consume it but after certain thresholds there are diminishing returns, i.e. the beach is full or the road is congested.

Data can be located anywhere on this grid.

		Rivalry at the point of consumption	
		High	Low
Excludability at the point of consumption	High	**Private Goods** For example, an apple	**Club Goods** For example, paid data standards
	Low	**Common Goods** For example, coal fields	**Public Goods** For example, National Defence

[6] Bolded terms are defined in the glossary

[7] Inefficiency might look like, for example, underinvestment in R&D given that others might benefit from your ideas.

[8] Haskell & Westlake, 2018, p. 65

5. The potential for externalities to occur when data is used is high.[9] Non-rivalry enables the combining and re-use of data for different purposes. For example, a person's activity tracker created by a smartwatch - a datafile created to understand fitness - can be combined with an individual's credit file - a dataset to understand credit worthiness - and used by a marketing firm to increase sales and profits, by advertising the right trainers at just the right price point.

6. This data use may have externalities. The marketing firm could use data to establish that fitness inclined people are prepared to pay a higher price. The right price for another person may become higher, a negative externality of the original data use. Equally, an unrelated person may benefit if the marketing firm’s insight sets in motion a change so that the right trainers are produced more cheaply. People may feel their rights are being compromised through uses of data to which they did not consent or understand.

7. The types of externalities that arise can be summarised as:

   - Supply-side externalities: **Economies of scale** and **economies of scope** in the use of data and the innovation aspects.

   - Demand-side externalities: Reduced **friction costs** in markets and network **spillovers** from the use of data.

   - Legal externalities: Privacy and data security externalities.

   - Wider external impacts, recognising that data use can meet other policy objectives such as environmental or societal impacts.

   The following sections provide more detail on each of these types of externality.

Data use externalities

Supply-side externalities

8. ‘Supply-side’ here refers to externalities that indirectly act on the companies that provide the products or services. They can be further subdivided into Economies of Scope and Economies of Scale.[10]

   - Economies of Scope can occur by re-using and combining existing data for a new purpose. This occurs when data is shared, as the same data need not be recreated multiple times, and when data created for one purpose is used for a new purpose.[11]

   - Economies of Scale are the reduction in per unit costs of a product or service as the total quantity of production increases.[12]

[9] London Economics (2020) maps the economic effects.

[10] Ctrl-Shift, 2018 and Martens, 2021

[11] For example, businesses provide details about their products online, which can then be aggregated with other business product information and turned into a product or service itself, which benefits consumers as a whole (Jones and Tonetti, 2020).

[12] For example, as the number of data points grows about consumers' actions, research done more accurately represents reality and hence can lead to more efficient product development, more tailored products or marketing (Aswani, 2017; Cameron, 2015; Leroux, 2016).

9. Some of these supply-side externalities are *internalised,*[13] which can lead to some companies getting and staying ahead of their competitors (‘*winner takes all’* scenarios)[14] (see demand side externalities below, and Chapter 3).

Demand-side externalities

10. Demand-side externalities are factors that relate to those who purchase products or services. They can be subdivided into ***Friction Costs*** **and** *Network Externalities*

- Friction Costs are all non-direct costs.[15]

- Network externalities occur where products or services interact and reinforce one another, such that the value changes with changes in numbers of users.[16]

Legal externalities

11. Legal externalities relate to both Intellectual Property Rights and Personal Privacy and Protection.

- *Intellectual property rights* stem from data creation having a cost, while data can be copied and disseminated quite easily. This means the property rights framework is associated with externalities.[17]

- *Personal Privacy and Protection externalities* relate to data that can be used to understand aspects of an identifiable individual’s private life.[18]

Wider external effects of data use

12. These are more diffuse or less directly traceable to data use. They include:

- Environmental externalities. There are both positive and negative impacts on the environment resulting from data use, emissions caused by the infrastructure on which data

[13] That is, the costs or benefits eventually get factored into the price of the services or products.

[14] Jones and Tonetti (2020), CMA (2020)

[15] For example, technological changes mean a product is sold online, rather than needing retail space, (Almunia, 2012; Goldfarb and Tucker, 2019, p.3; Shapiro and Varian, 1999)or digital verification makes it easier to certify trustworthiness or reputation (Brynjolfsson and Smith, 2000; Brynjolfsson et al., 2006)

[16] For example, data driven businesses which act as an intermediary like Facebook or Uber become more useful with more users; or transport authorities may see wider network impacts as better-informed consumers means lower demand at peak times (London Economics, 2020; CTRL-SHIFT, 2018; Deloitte, 2017).

[17] Martens, 2021.

[18] For example, when an individual shares data with a company but doesn’t know the company sells it on in a way that compromises individual’s privacy (Jones and Tonetti, 2020; Acemoglu et al., 2019, p. 3 ;Nissenbaum, 2011) or a company pays for cyber security to help prevent a breach (Bandhyopadhyay et al., 2009)

relies (which come with an environmental cost), as well as reductions to emissions following e.g. R&D and streamlining of business processes, or moving services online.[19]

- (unpaid or underpaid) labour costs of data creation. Much of this is paid, but some unpaid individuals (for example leaving reviews on websites, providing data about themselves as they transact) do not get paid for the benefits companies then derive.[20]
- Social externalities

Concluding points

13. These externalities occur in a wide range of activities and sectors, because data use is so ubiquitous. The challenge is to locate where the externalities might arise in a manner to facilitate analysis and evidence gathering.

14. A further challenge is differentiating effects that are internalised or should be integrated into the value generation of the data owner from the effects that are indirect. There is a dynamic to this, with externalities becoming internalised due to changes in market structure or policies. For example, the incentive to bring complementary datasets together can incentivise mergers or data-related barriers to competition.

[19] There have been studies on the carbon footprint of the cloud and – in the wake of the pandemic – studies looking at the natural experiment of switching activities that did not use data (e.g. conferencing) to online data-using services (e.g. Burtscher, 2020).

[20] Diepeveen and Wdowin, 2020; Savona, 2019. A distinction is made by Statistics Canada (2019) between data that is produced by businesses and governments for their own use but not sold in the marketplace, and data that is supplied by households to businesses and governments as payment-in-kind in exchange for other services, as is the case for Facebook, Google and many other online service providers.

3. Locating externalities

Chapter summary

- This chapter covers:
    - Data value chains (that is, the different stages of data use and where costs and benefits accrue)
    - The wider data-use ecosystem
    - How the different types of externality described in chapter 2 fit within both the value chain and the data-use ecosystem
- The data value chain involves the processes of:
    - Data generation
    - Data collection or aggregation
    - Data analysis
    - Data exchange
- Along the data value chain:
    - Supply side externalities mostly occur at early stages (planning, curation and analysis)
    - Demand side externalities mostly occur at later stages (sharing/ publishing)
    - Legal externalities occur along the entire value chain (as different actors are involved)
    - Wider external effects of data use mostly occur at later stages
- This report considers the actors in the data-use ecosystem to include:
    - Researchers and academics
    - Regulators for data privacy and legal issues
    - Standardisation bodies
    - Investors, venture capitalists, and incubators
    - Organisations providing resources and services to develop the commercial potential of the ecosystem.

Introduction

1. This chapter looks at evidence around where externalities might occur. There are two ways to look at this.

   - The first characterises data use as a linear process, a sequence commonly called the **'data value chain'** or 'data lifecycle'.

   - While externalities will derive from activities in the data value chain, their indirect effect can fall outside the actors involved in the chain. Therefore this report considers approaches that articulate the wider set of stakeholders that are involved or affected by data use, a **'data ecosystem'**.

Data value chains

Data use and adding value

2. The way data is used by organisations can be broken into stages that add value, see Figure 3.1.[21] There is no single, widely-agreed approach, but the value chain is typically broken into:

   - Initial stages of data identification, generation

   - ‘Collection’ stages that involve capture, collection, storage, cleaning etc,

   - ‘Analysis’ stages of data processing, statistical analysis, linkage etc[22]

   - ‘Exchange’ stages involve data sharing, publication, deriving insight, repurposing and or utilising data for financial or public benefit.[23]

3. Data's value increases as it moves through the value chain, with raw individual data carrying the lowest value. The aggregation of data, and its treatment, add or give it value.[24]

[21] Diepeveen and Wdowin (2020), Mawer (2015) and Corrado (2019)

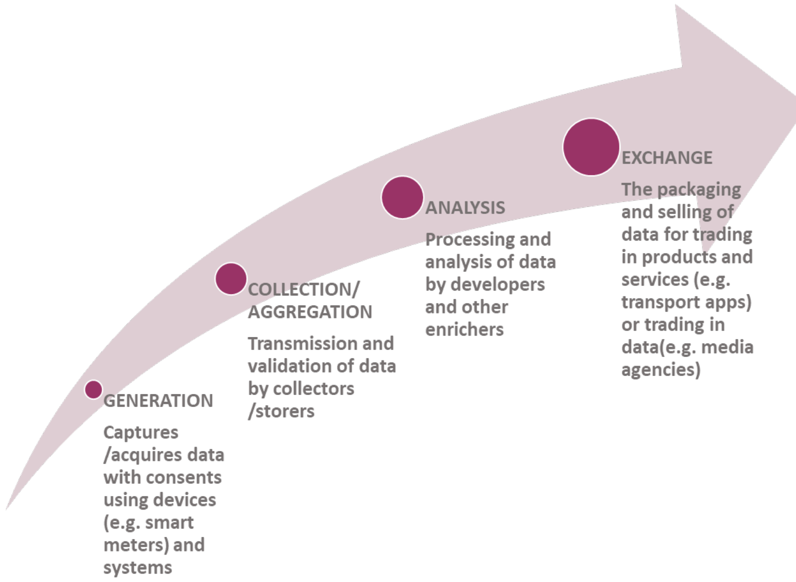
[22] Mawer, 2015; Corrado, 2019; Savona, 2019

[23] there are many other models, see ODI (2018b)

[24] OECD, 2015; Schimmelpfennig and Ebel, 2016; Brynjolfsson and McElheran, 2019

Figure 3.1: Developing the value chain looking at underpinning activities

The simplified data value chain is derived from GSMA (2018) and OECD work (Nguyen and Paczos, 2020). The GSMA work explores data value with an industry structure perspective. The nature of data results in a tightly integrated value chain where the organisation that collects the data is very likely to keep control and ownership of that data through all steps towards developing the final output. Nguten and Paczos look specifically at international data flows from a business perspective.



Externalities and the data value chain

4. Much of any value added is captured by the organisation or other actors in the chain. It is internalised in business profitability, remuneration for employees, price changes and quality improvements for consumers enhancing market shares for producers. Where the overall outputs are for a public good, the value added may be in terms of services delivered.

5. However, along the data value chain, it is possible to locate where externalities might occur, summarised in Table 3.2 below.[25]

[25] This table summarises insights from a range of case studies (Lammerant et al, 2015, Wdowin and Diepeveen, 2020)

Table 3.2: Externalities emerging from activities in the value chain

	Generation	Collection/ Aggregation	Analysis	Exchange
Supply-side: Economics of scale	●	●	●	
Supply-side: Economics of scope	●	●	●	
Demand-side: friction costs				●
Demand-side: network effects				●
Legal externalities: Intellectual property	●	●	●	●
Legal externalities: Personal Privacy and Protection	●	●	●	●
Wider external effects: Environmental			●	●
Wider external effects: (unpaid) labour costs		●		
Wider external effects: Social externalities			●	●

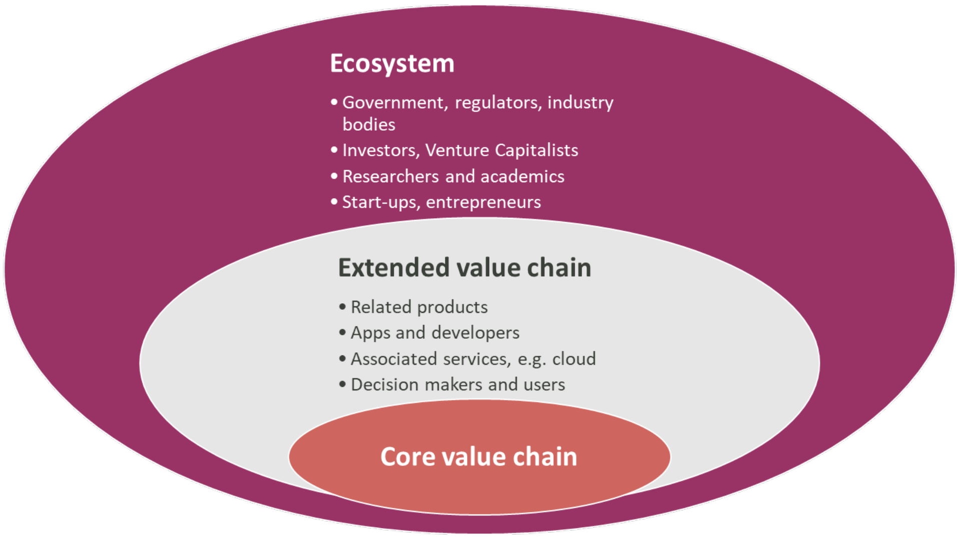
Understanding the ecosystem and locating externalities

6. The ODI has defined a data ecosystem as consisting of :

- Data infrastructure – such as data assets, standards, technologies, policies
- People, communities and organisations that are affected by the value created by it[26]
- Market structure in which the data value chain operates
- Actors not integrated into the data value chain but affected as data is used.[27]

[26] ODI, 2018b
[27] GSMA, 2018

Figure 3.2: Developing the value chain through an ecosystem approach



Adapted from Curry et al. 2016

7. An ecosystem approach can be used to locate what kinds of actors are affected, and in what ways, by data use indirectly.[28] This report found that the existing evidence base is not currently broad enough to complete such an analysis.

Internalising effects within the ecosystem

8. An increasing number of mergers in the digital sector involve data, with potential reductions in competition and increasing barriers to other market entrants.[29] Businesses holding data may refuse to provide access to other firms, or the dominant firm may collect excessive amounts of data.[30]

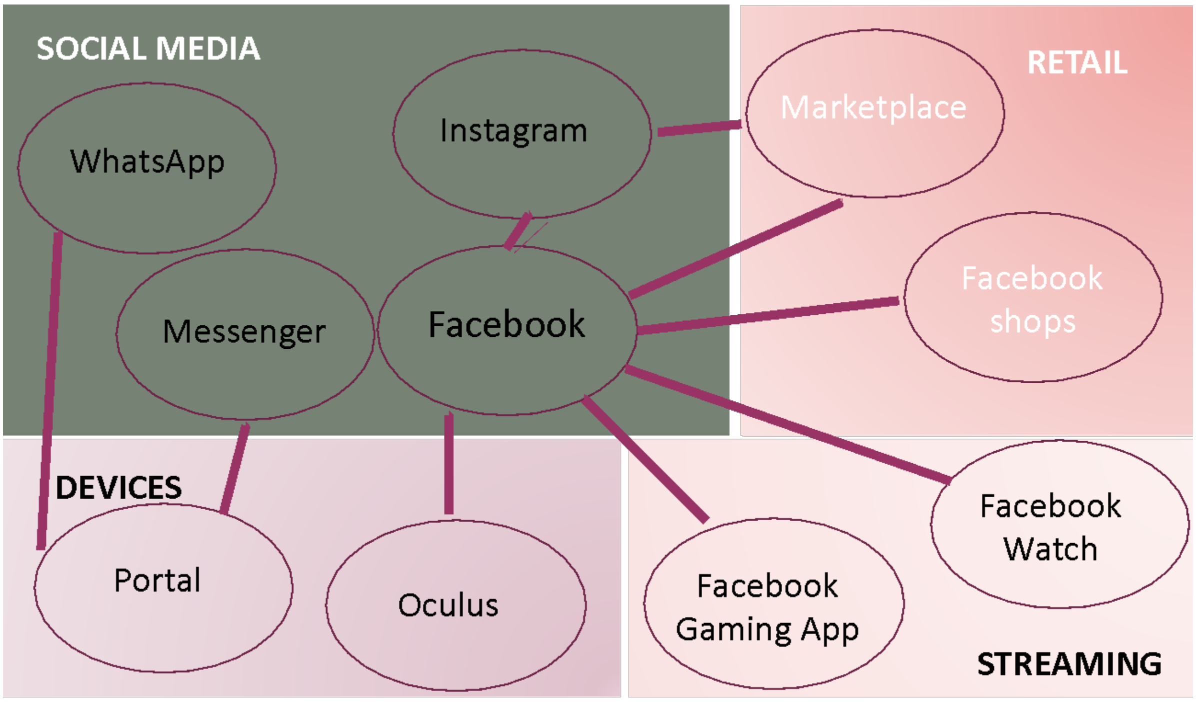
9. A feature of many internet and data driven businesses is that they are platforms. They act as intermediaries that bring together different players in the data value chain. There may be scope to provide multiple services, with companies expanding into and operating in adjacent or even unrelated areas, either launching new services themselves or acquiring other companies.

[28] For example, the publication of TfL transport data led to app developers being able to create citymapper, which in turn benefitted the wider public. ODI, 2018b; Deloitte, 2017

[29] Argentesi et al. 2019; Grunes and Stucke, 2016; Furman et al., 2019.

[30] Autorite de la Concurrence and Bundeskartellamt, 2016; Scott Morton et al., 2019

Figure 3.4: CMA Illustration of the Facebook ecosystem



The Competition and Markets Authority (CMA, 2020, p57) presents an illustration of both Facebook and Google's ecosystem. Here the Facebook example illustrates:

- successful digital companies in recent years have built large ecosystems of complementary products and services around their core service.
- the channelling of online activity into these ecosystems.

10. Figure 3.4 indicates a use of an ecosystem analysis. A set of consequences identified are that entry would be deterred and that market control insulates the most profitable aspects of the platform such as advertising.

11. Beyond these costs to the ecosystem as a whole, there are benefits including integration of a wide range of products and services which can deliver efficiency savings, potentially reducing prices. It can also improve the consumer experience overall, by increasing the ease with which a range of different services are accessed.

12. Recent initiatives such as Open Banking has focused on data portability as a means to reduce the friction costs borne by users on moving to other suppliers of services. However, these do not change data driven network effects. [31]

Concluding points

13. This chapter is primarily offering tools to locate the externalities that occur as data is used. It explores where a search for externalities might focus. The ecosystem approaches provide a

[31] Martens et al 2020; ODI and Fingleton, 2019

toolkit that could be used to locate who may benefit or lose from data use externalities. While case studies using this approach exist - namely CMA digital advertising study, and TfL Open Data Evaluation by Deloitte - further research could develop this.

4. Externalities and data interventions

Chapter summary

- This chapter covers three policy case studies:
  - Open data standards
  - Secure research environments
  - Personal data stores
- The themes covered include
  - Opening data and data standards
  - The use of settings and standards that enable data sharing which have the controls to mitigate the negative externalities that may arise

Introduction

1. This section looks in detail at the externalities of three policies, open data standards, secure research environments, and personal data stores.
2. Information for this section was gathered from
   - evaluative literature on each policy
   - Interviews with relevant policy makers and researchers at bodies from ODI, OECD, Nesta, DCMS, TfL, Mydex and ONS.

Open data standards

3. Open data standards are reusable agreements that make it easier for people and organisations to publish, access, share and use data. Their development has been a feature of the opening of (public) data and this has been integrated into government with the creation in March 2020 of a UK Data Standards Authority. This section reviews two evaluations of case studies of this kind of intervention.
   - One case study evaluated standards for the open data about leisure activities.[32] It centres on the operational data held by leisure centres about sports and other leisure activities that can be booked by members of the public. The study highlights that the standards were developed to internalise externalities, with a first focus being the search costs for those that would like to use leisure facilities. Data about leisure activities was made accessible so that developers could create apps embedding the data and enabling users to easily view what activities were on offer and book the right activities for them. The direct effects were reduced friction costs as search for leisure activities became easier.
   - The evaluation explores other outcomes and links the data use to increased physical activity. The study models impact including a reduction in premature deaths, health cost savings and increased productivity. While the model can be criticised due to its reliance on

[32] Frontier Economics 2019

the assumptions that underpin their estimates and the simplified approach taken to link new users to the impacts, it develops evidence around indirect benefits of data use.

4. The other case study is the well known evaluation of the opening of Transport for London data using a standard.[33]

  - Users can make better decisions on when and how to travel, which can improve transport system efficiency. There may be spillovers, in that the better decision making at traveller level could then affect all travellers if – for example – it leads to lower fares or reduced crowding.

  - Opening data enables those with appropriate skills to develop applications Interviews with those that worked on this policy suggest that

4.o.1. Capturing indirect benefits were an explicit part of the reason for the policy.

4.o.2. The incentives and instructions to put data online did not then guarantee the data would be used, and it could remain under-analysed or unused.

4.o.3. Returns could be lower in open data applications due to entry costs being low. This then did highlight the need to both open data and have parallel open data standards to facilitate further re-use.

Data access enabling research and development

ONS Secure Research Service

5. The Office for National Statistics has developed a Secure Research Service. This is a secure setting, providing access to data that has been made as safe as possible but within technical and procedural constraints that remove the chance of a disclosure of sensitive data. This was designed to balance the trade-off between analysis of rich data sources and the risk of compromising privacy. [34]

6. In the environment, a combination of measures enables researchers to use the data, provide analytical outputs, link administrative and survey datasets together. Recent outputs include analytical efforts to understand the health surveys and records during the Covid crisis.

7. ONS are developing an evaluation framework for the impacts of the access they provide.

[33] Deloitte, 2017

[34] Desai et al., 2016 describe the “Five Safes” framework.

Smart Data initiatives

8. Smart Data policies include Open Banking.[35] Beyond primary goals of improving banking experiences for customers, these facilitate secure accessing and sharing of data.[36]

9. Infrastructure for this has been rolled out, reducing barriers to entry through easing customer signing up, reducing the friction costs of transacting and setting up accounts.[37] This touches on portability of data, where individuals have a means to transfer their data to a third party securely.

10. Innovation facilities that afford access have been developed. Alongside access to data, there is a significant level of business incubation, with open banking projects providing start-ups and small businesses with support through funding, technical, legal, marketing, and commercial assistance. A focus is the legal and privacy aspects of a route to market for any applications developed, given the sensitivity of the data, such as the European Commission sponsored Data Pitch.

Discussions about enabling R&D using data

11. Interviews with those that worked on these policies highlighted the following themes:

    - Given the sensitivity of the data, the actions to mitigate the negative externalities (data protection of personal data, vetting of those accessing the data) helps ensure positive outcomes.

    - There was a recognition of open data increasingly being complemented with data sharing that was more closed.[38] It was felt that some bodies which held data are moving to more collaborative research projects with businesses and not-for-profit organisations to help develop new products.

    - The balance between widening data use and the privacy of data collected in operational settings was highlighted. The example of transport data was given, which can be disclosive even if all the normal personal identifiers are removed, because they can still demonstrate patterns of travel.

    - Any data sharing draws in the skills and capabilities of the wider ecosystem, especially the “creators” (i.e. app developers), delivering new services to segments that otherwise would not benefit.

---

[35] The focus for Open Banking specifically, is banking data of SMEs and one million consumers. It enables mechanisms by which the consumer gives consent to a third party to initiate payments, submit confirmation of funds requests or access account information held at their bank.

[36] DCMS, 2020

[37] ODI and Fingleton, 2019

[38] in the terms of the ODI Data Spectrum

- The value of the indirect effects of data use are difficult to measure and include crucial measurement gaps, such as locating the costs and benefits in nations when activities and uses were global.
- Benefits have been substantial but there are costs associated with mitigating the negative externalities.
- Allowing researchers access through a safe setting involved technological, legal and operational investments.

Personal data stores

12. When organisations collect and hold data about people for their business needs, they will typically hold this on their own systems. Any legal permissions needed (e.g. under GDPR) will be obtained at the point of collection. In this way, if an individual wants to check or change these permissions, they will need to go to the organisations themselves. A personal data store is an alternative model for storing data, with data about an individual going into their own data store. The store then allows individuals to share chosen portions of this data with service providers under their control.

Mydex CIC

13. The Mydex CIC technology uses a system of 'tokens' that verify facts about citizens (such as proofs of address, age, disability or educational qualification), held safely in the citizen's own personal data store. In this way, data from multiple sources about an individual is aggregated in their personal data store.
14. If someone is eligible to receive a service or a benefit they *are* aware of, sometimes they are automatically eligible for other services or benefits which they *aren't* aware of. Data passporting is a solution to this problem. Providers can match anonymised profiles of data held in the individual's data store to the requirements for accessing other services. Data passporting is a process already used or enabled through other schemes, such as Open Banking.

Discussion about data stores and data use externalities

15. Citizens incur friction costs due to the time, effort and money they spend finding the information they need to access services, filling in forms and providing information about themselves
16. A data aggregator reduces these friction costs in theory. For example, Mydex notes that citizen financial advice providers previously would have invested considerable time with a client to understand the individual's financial position, but this activity was essentially already completed.
17. As data stores can facilitate multiple uses of the same data (by providing a rich, secure, pre-verified personal data store) there is potential to capture economies of scale and scope.
18. There are investments needed to create a personal data store, such as setting up the technological and organisational processes to make data accessible to the store.
19. The extent to which the potential benefits of personal data stores are realised in practice have not yet been evaluated. For example, there is the potential that privacy benefits are

exchanged for additional environmental costs, or reduced quality of research leading to no net changes to utility overall.

Concluding points

20. From the supply side, the interventions - particularly personal data stores - facilitate internalising many of the externalities of multiple uses of the same data (economies of scope) and then putting in place the structure across multiple individuals and organisations (economies of scale). There are trade-offs between benefits from sharing data (such as new products, better insights, more efficient delivery of services), and costs (investment in infrastructure and processes).

21. Evaluations examining all the externalities discussed in Chapter 4 were not found.

5. Measuring externalities

Chapter summary

- This chapter covers approaches to valuing the externalities associated with data use:
  - ‘Modelling’ approaches, which assess the overall value of data.
  - ‘Willingness to pay’ surveys, which collect people’s views on the value of an externality in the absence of a market price.
  - ‘Econometric’ approaches, using survey data and modelling to understand an externalities effect on productivity and firm-performance.

Introduction

1. As externalities occur beyond a transaction itself, there is no market price to value them. Valuing externalities is useful for:

   1.1. Helping to determine their importance

   1.2. Setting a budget for mitigating the risk of harms

   1.3. Promoting policies to enable positive externalities.

2. The overall value of an external effect is often estimated using the value of some unit of an externality and multiplying by the quantity of an external impact.

Modelling

3. In this context, a model is a range of data and assumptions pulled together for estimation purposes, in this instance models are used to estimate externality impacts. They may be used to forecast the future outcomes or consider the impacts of a past policy.

Applications in data use externalities

4. Several studies have looked at the value of open data using a model. For example, TfL’s open data and digital partnerships has been valued, estimating the time savings for network passengers to be between £70 and £90 million per annum, which increases to £130 million per annum when wider effects are accounted for.[39] The basis for estimates is the modelled levels of time savings monetised using values for time.

5. Key to the approach is to estimate a set of outcomes that can be monetised. Data sharing could lead to new mobility solutions reducing congestion, improving freight efficiency, and causing fewer accidents. The different steps from data usage to the outcomes are modelled. Further, many of the impacts assessed are regarded as externalities, in that impacts such as reduced congestion would not be internalised.

6. The Open Data Institute (ODI) commissioned evaluations to estimate the economic and social impact of its R&D programme. London Economics (2019 and 2020) produced a framework to assess the economic contribution. The analysis brought together quantitative and qualitative

[39] Deloitte, 2017

evidence. It provides high level estimates of the economic impact and considers the reasonableness of findings using sensitivity tests.

7. A theory of change underpins the modelling and is used to understand externalities. The method allows attribution, by breaking down the overall change into testable steps towards this overall effect including any that are not internalised. The study quantifies the economic value of a particular open standard.

8. A case study approach is used to evaluate projects within the ODI's R&D programme.[40] One case study examined the impact of banks using a common standard to share bank account data. The research estimates the economic impact on the increase in new signups of third-party overdraft services since the introduction of the standard, an outcome that can be monetised. They estimate the potential savings to individuals in aggregate of between £1.3 and £2.2 million.

9. The dynamics after data use R&D can also lead to externalities. A J-curve over time is observed, observing early downward effects of R&D because of mismeasurement, missing impact and the up-front investments affecting outcomes adversely.[41] There are then positive effects. The analysis then uses a model calibrated on R&D outcomes applying this to data using activity, artificial intelligence. The model estimates the spillovers and network type externalities.

10. Industry studies have also looked at data use and the benefits of adopting digitalisation. The 2017 Review into Industrial Digitisation was accompanied by a benefits model.[42] The review captures benefits by looking at barriers removed following adoption of certain technologies,then modelling variations in:

    10.1. The uptake of different technologies (e.g., artificial intelligence, additive technologies)

    10.2. The uptake between different sectors

Lessons from other policy areas

11. Using models to quantify externalities is common; a stocktake in 2012 suggests DfT, DWP, DH and MOD each have over 50 business-critical models that are used for policy simulation, including estimating externalities.[43] There is also guidance on proportionate quality assurance of models used in policy making.[44]

12. Looking at the example of models built by DfT, there is guidance on modelling:

    12.1. Time and impact of improved travel time

    12.2. Preventing a fatality/accident (both used in transport appraisal)

    12.3. Carbon dioxide and other pollutants

[40] Frontier Economics, 2020

[41] Brynjolfsson and Syverson, 2020

[42] Accenture, 2017

[43] MacPherson, 2012

[44] HM Treasury, 2015, The Aqua Book

12.4. Wider social policies, such as expenditure to government due to transition into employment

13. Where values cannot be based on guidance, the Green Book suggests that survey-based approaches may be considered to provide estimates.[45]

Valuations based on surveys

14. Surveys can be used to ask large numbers of people or organisations about a given issue, and in that way get a representative view of scope and scale. For establishing the value of externalities (where there is no obvious market price) a traditional survey has limited use.[46]

15. There are however types of surveys that estimate prices of goods and services that do not have a market price, broadly falling into two types:

15.1. **'Stated preference' approaches**. Simply put, these ask respondents their willingness to pay (WTP) or willingness to accept (WTA) compensation for a hypothetical change in the level of provision of a non-market good.[47]

15.2. One of the strengths of stated preference approaches is their flexibility. They are applicable to a wide range of externalities and can use scenarios and carefully designed surveys to engage with past and future changes. The method is most robust when the focus is an item with which survey respondents are familiar. Often, this then means either using materials to enhance realism or placing into the survey design specific questions or alternative wordings of questions to test robustness.

15.3. However, results of this type are often very different from prices elicited through more complex methods[48].

15.4. **'Revealed preference' approaches**. These use observed behaviour to estimate the value of indirect effects. For example, by analysing house prices (observed behaviour) and data on the characteristics of region and economy, you can 'reveal' how much people value living near e.g. green spaces.

15.5. Revealed preference approaches are most robust when there are natural experiments, allowing data to be collected over different scenarios.

[45] HMT, 2020

[46] There are direct costs of data use which obviously don't fall into this category. Notably, there are direct losses to business following a cyber breach. DCMS conduct an annual Cyber Security Breaches Survey, which includes questions on impact (including estimated costs) of cyber security breaches. Other surveys are used to collect similar information in other countries, such as SAS OpRosk Global Data, which compiles information on publicly reported operational losses in excess of US$100,000 (SAS Corporation, 2015). The sensitivity of collecting data about security breaches however remains high.

[47] Mitchell and Carson, 1989

[48] Benndorf and Normann, 2018

Applications in data use externalities

16. Stated preference approaches are already used to explore the trade offs between security, and ease of use and detail of personal information..[49][50] This research looked at different options to mitigate privacy concerns, not values; consumers might:

    16.1. Accept targeted advertising using their data (i.e. some privacy loss) in exchange for free web services.[51]

    16.2. Pay a premium to purchase from firms with better privacy protection.[52]

17. Many studies find a pronounced discrepancy between participants WTP and WTA.[53] Individuals expect to be paid more money to share their data than they would be willing to pay to regain or retain their privacy.

18. To some extent, this is unsurprising given the huge uncertainties that are involved in these decisions. It is near impossible for individuals to foresee what their data may be used for, and what the consequences for them personally might be.[54] When answering survey questions, respondents might consider the potential benefits from sharing their data which they have come to expect, such as a more personalised offering, even if this is not explicitly part of the scenario. Indeed, individuals are willing to trade-in their personal information for targeted recommendations as in-kind payment.[55]

Lessons from other policy areas

19. Contingent valuation studies are common in a number of public policy areas. Transport Value of Time studies are commissioned by the government. The surveys establish a 'value' for reducing an hours delay experienced by one individual for different modes of transport.[56]

20. This approach has benefited from verification through other research approaches, notably 'natural experiments' including when congestion charges were introduced. As this changed real-life patterns of travel, it provided evidence to check stated preference survey methods .[57]

21. In other policy areas, the WTA:WTP ratio for a stated preference study has also been researched. Studies observe much larger ratios for environmental goods, such as protection of endangered species(sometimes on the order of 10:1, compared to similar studies on data protection referenced above of ~2:1) .[58] The magnitude of this disparity remains to be explained, but a plausible account is that in the environmental context, a high figure for WTA reflects a kind of moral outrage.

---

[49] e.g. direct financial rewards (Culnan and Armstrong, 1999; Laufer and Wolfe, 1977) and improved quality of service such as personalisation and social benefits (Chellappa and Sin, 2005; Smith et al., 2011)

[50] Savage and Waldman, 2015

[51] Schumann et al., 2014

[52] Tsai et al., 2011

[53] Winegar and Sunstein, 2019; Grossklags and Acquisti, 2007

[54] Acquisti and Grossklags, 2005

[55] Li and Unger, 2012

[56] See DFT, 2016

[57] Fezzi et al., 2014

[58] Cummings et al. 1986

Econometric approaches to external effects

22. Econometric approaches are where data is put into a statistical model which incorporates economic theory, to better understand the relationship between economic variables. Econometric approaches are quite varied, and a range have already been used within and beyond data policy.

23. When used to explore externalities, the challenge for these approaches is to identify what portion of effects seen on firms or the wider economy are due to externalities. For example, there are sectors of the economy that currently do not invest in artificial intelligence approaches that are likely to benefit but have been seen to under-invest. This suggests benefits are possible but would not be fully internalised, with the main mechanisms being the network externalities and reduced friction costs associated with data use.

Applications in data use externalities

24. There have been a number of studies that look at the impact of externalities using an econometric approach within data policy. Most notably:

    24.1. The OECD finds that firms that use data exhibit faster labour productivity growth than those that do not, by approximately 5% to 10%.[59]

    24.2. Research from the United States, being at the frontier of data-driven decisions in manufacturing, is linked with improvements in revenue-based productivity of 4% to 8%.[60] The authors show that timing, however, is essential. Leading adopters of data analytics are receiving the biggest gains, while laggards that reach the frontier later tend to have lower net benefits or none at all.

    24.3. Analysis of German firm-level data finds evidence that use of data and analytics increases the likelihood of a firm becoming a product innovator, as well as for the market success of product innovations.[61] These results hold for both manufacturing and service sectors but are contingent on firms' investment in IT-specific skills. Others have documented similar findings.[62]

    24.4. A survey of 500 UK firms which are commercially active online found using data that online activity is associated with higher productivity. [63] The authors found that the type of data use that had the greatest impact on productivity was 'data analysis and reporting of data insights' , whereas amassing data has little or no effect on its own.

Lessons from other policy areas

25. Econometric approaches that assess wider economic impacts of data use rely on firm-level data. This allows the impact of, for example, benefits from R&D to be assessed and tracked at firm-level.

---

[59] OECD, 2020

[60] Brynjolfsson and McElheran, 2019

[61] Niebel, Rasel and Viete, 2018

[62] Bajari et al., 2019; Wamba et al., 2017

[63] Bakhshi, Bravo-Biosca and Mateos-Garcia, 2014. In this study, 'using data' about online activities included the collection, analysis and deployment of online customer data.

26. The spillover effects of multiple similar companies operating in a geographically similar area (referred to as 'agglomeration effects') have been studied in this way. Research indicates that firms benefit from economies of scale (which are internal to the firm) and network effects (which exist between firms) through:

26.1. Network drivers, such as a location providing pools of workers who have a variety of skills

26.2. Increasing returns to scale (within firms) in intermediate inputs and/or

26.3. Relative ease of communication (between firms) and obtaining supplies, workers, and innovative ideas due to the proximity among firms.

27. Through these mechanisms, when firms in related industries cluster together the costs of production may decline due to competing multiple suppliers, greater specialisation, and division of labour.The disadvantages of competitors taking customers are therefore sometimes outweighed by the advantages of that cluster attracting more suppliers and customers than a single firm could have done alone.[64]

28. These externality effects have been measured and valued for integration into value for money modelling. Underpinning transport guidance are estimates derived from estimates of the spillovers, so that transport interventions can consider these impacts as schemes are assessed.[65]

29. This type of analysis has been carried out on digital clusters.[66] This evaluation of the impact of the Tech City programme in London found that the policy increased cluster size and density, especially for 'digital tech' plants, where revenue/worker and high-growth firm activity also rose. But for a larger set of incumbents 'digital content businesses, the policy also led to de-concentration and lower revenue productivity.

Concluding points

30. There are methods with quite mature applications in other sectors that can inform our understanding of the scale and scope of externalities within data policy; indeed, this is already happening.

31. Each set of approaches considered here have costs and benefits, which implies deploying these methods in a coordinated way for triangulation.

32. All these methods, modelling approaches especially, rely on ultimately subjective decisions about what impacts are included and accounted for, studies such as the 2017 Review into Industrial Digitisation highlight the benefits of:

32.1. Expert groups to calibrate and validate model assumptions

32.2. Assuring quality by comparing results from other models and running sensitivity analysis.

---

[64] Glaesar and Mare, 2001

[65] Venables et al., 2014; DFT, WebTAG, 2018

[66] Nathan, 2019

6. Conclusions

1. Some themes emerge looking across the literature reviewed, discussions with experts and placing the evidence about data use externalities within a framework.

2. The externalities associated with data use are often analysed in terms of opening data. There has been a drive to release data held by public bodies. This can be understood in terms of externalities, a means to capture indirect external effects further down the value chain. Further, the early action has relatively low indirect and direct costs. Privacy externalities for open data are low as they are usually anonymised. Data costs are minimal as data is the by-product of other activities.

3. However, there is a need to consider how wider data use can be enabled where data cannot be opened. The spectrum of data types means that not all data is suitable for openness. This study has sought to understand where external effects will be important, in terms of the data value chain and the data use ecosystem. This proves useful because it provides a more granular, though somewhat generic, framework to locate externalities. Further work could usefully test this framework empirically on historic case studies.

4. A key feature of the case studies considered in this literature review - Open Banking, personal data stores etc - are that they work across an ecosystem. Data is 'opened' by the structure of the policy that minimises legal and security risks, and hence allows value to be added. These ecosystem effects**,** the wider value captured in the ecosystem of innovative businesses, researchers, data aggregators, regulators beyond the value chain of the core data use, highlights that different sectors and types of businesses and organisations act in capturing or being affected by externalities.

5. This forms the basis for looking at valuing externalities. The methods employed range from those grounded in currently used methods and values. To estimate the indirect effects of data use on transport planning, values for externalities are available. However, the methods available for valuing externalities that are specific to data use, such as privacy, need maturing. The techniques used are well-understood but, when applied to data use externalities, are less accurate than in other policy areas because the methods work best when interviewees understand the range of potential impacts of that use.

6. These themes suggest some areas where next steps on analysing data use externalities may focus. Policies are improving data sharing at a stage later in the value chain than the generation of data, at the aggregation/analysis stage. The focus is enabling controlled data sharing, balancing privacy and security negative externalities with the potential benefits from re-use. The policies are often sector or use specific.

7. These areas may be ones where data use externality valuation techniques can be used, tailoring to the specific sectors and uses. In tying the evidence gathering more directly to examples, evidence gathering can be more specific and better specified, defining the direct and indirect costs and benefits in terms of the policies, products and wider evidence about the sector or use. This may then improve the validity of any additional evidence collected about data use externalities.

References

Acemoglu, D., Makhdoumi, a., Malekian, A & Ozdaglar, A (2019), *Too Much Data: Prices and Inefficiencies in Data Markets*, Working Paper No. 26296, Cambridge, Massachusetts, National Bureau of Economic Research.

Acquisti, A. and Grossklags, J. (2005), *Uncertainty, Ambiguity and Privacy,* Paper submitted to the 4th Annual Workshop on Economics and Information Security.

Almunia, J. (2012), *Competition and personal data protection*, Speech to the European Commission, Brussels.

Argentesi, Elena, Paolo Buccirossi, Emilio Calvano, Tomaso Duso, Alessia Marrazzio, and Salvatore Nava (2019). "Ex-post Assessment of Merger Control Decisions in Digital Markets." Document prepared by Lear for the Competition and Markets Authority.Aswani, S. (2017),3 Ways Consumer Goods Companies Should Use Consumer Data.from Clarabridge: https://www.clarabridge.com/blog/3-ways-consumer-goodscompanies-use-consumer-data/

Autorite de la Concurrence and Bundeskartellamt (2016). "Competition Law and Data".

Bajar, P., Chernozhukov, V., Hortacsu, A. and Suzuki, J. (2019), 'The impact of Big Data on Firm Performance: An Empirical Investigation, AEA Papers and Proceedings, 109, pp. 33-37

Bakhshi, H., Bravo-Biosca, A. and Mateo-Garcia, J. (2014), The analytical firm: Estimating the Effect of data and online analytics on firm performance, Nesta Working Paper no. 14/05, London, Nesta.

Bandyopadhyay, T., Mookerjee, V. S., & Rao, R. C. (2009). Why it managers don't go for cyber-insurance products. Communications of the ACM-Scratch Programming for All, 52(11), 68. https://doi.org/10.1145/1592761.1592780

Benndorf, V. and Normann, H-T. (2018), 'The willingness to sell personal data', *The Scandinavian Journal of Economics*, 120, 4, pp. 1260-1278.

Beresford, A. R., Kubler, D. and Preibush, S. (2012), 'Unwillingness to pay for privacy: A field experiment', *Economics Letters*, 117, pp. 25-27.

Brynjolfsson, E. and McElheran, K. (2019), *Data in Action: Data-Driven Decision Making and Predictive* Analytics *in U.S. Manufacturing*, Working Paper Number 3422397, Rotman School of Management.

Brynjolfsson, E. and Smith, M, D. (2000), 'Frictionless Commerce? A Comparison of Internet and Conventional Retailers', *Management Science*, 46, 4, pp. 563-585.

Brynjolfsson, E., Rock, D. and Syverson, C. (2020), *The Productivity J-Curve: How intangibles complement general purpose technologies*, Working Paper 25148, Cambridge, MA, National Bureau of Economic Research.

Cameron, N. (2015), Optus: Improving product innovation with a data-driven customer view. from CMO: https://www.cmo.com.au/article/574445/optus-improving-productinnovation-via-data-driven-customer-view/

Chellappa, R. K. and Sin, R. G. (2005), 'Personalization versus privacy: An empirical examination of the online consumer's dilemma', Information Technology and Management, Vol. 6, No. 2-3, pp. 181-202.

CMA (2020), *Online platforms and digital advertising*, Market study final report, London, Competition and Markets Authority.

Castro, L., Lo, A. W., Reynolds, T., Susan, F., Vaikuntanathan, V., Weitzner, D., & Zhang, N. (2020). SCRAM: A Platform for Securely Measuring Cyber Risk . Harvard Data Science Review. https://doi.org/10.1162/99608f92.b4bb506a

Corrado, C (2019), *Data as an Asset: Expanding the Intangible Framework*. Presented at the EMAEE 2019 Conference on the Economics, Governance and Management of AI, Robots and Digital Transformation – Held at SPRU, University of Sussex, 3-5 June 2019.

Ctrl-Shift (2018) *Data Mobility: The personal data portability growth opportunity for the UK economy*, Report, London, The Department of Digital, Culture, Media and Sport.

Culnan, M. J. and Armstrong, P. K. (1999) 'Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation', Organization Science, Vol. 10, No. 1, pp. 104-115

Cummings, R. G., Brookshire, D. S., Schulze, W. D., Bishop, R. C., & Arrow, K. J. (1986). Valuing environmental goods: an assessment of the contingent valuation method. Totowa, N.J: Rowman & Allanheld.

Curry, E (2016) 'Chapter 3: The Big Data Value Chain: Definitions, Concepts and Theoretical Approaches'. In J.M. Cavanillas et al. (eds.), New Horizons for a Data-Driven Economy, pp. 29-37.

DCMS (2020) *National Data Strategy*. London: Department for Digital, Culture, Media and Sport.

Deloitte (2017) Assessing the value of TfL's open data and digital partnerships, slides, London, Deloitte.

DFT (2016) Understanding and Valuing Impacts of Transport Investment Value of Travel Time Savings, London: Department for Transport.

DFT, WebTAG, 2019 TAG UNIT A2.1: Wider Economic Impacts Appraisal, May. Department for Transport.

Diepeveen, S. and Wdowin, J. (2020) The Value of Data, Cambridge, Bennett Institute for Public Policy.

Fezzi, C., I. Bateman and S. Ferrini (2014) "Using revealed preferences to estimate the value of travel time to recreation sites". *Journal of Environmental Economics and Management*, 67 (2014): 58-70.

Frontier Economics (2019) Evaluating the economic and social returns. Frontier Economics Project Report for ODI.

Frontier Economics (2020) *Geospatial Data Market Study,* Great Britain, London, Geospatial Commission.

Furman, Jason, Dianne Coyle, Amelia Fletcher, Derek McAuley, and Philip Marsden (2019). Unlocking Digital Competition. Report of the Digital Competition Expert Panel. HM Government.

Goldfarb, A and Tucker, C. (2019) 'Digital Economics', *Journal of Economic Literature*, 57, pp. 3-43.

Graham, D. and Gibbons, S (2019) 'Quantifying eider economic impacts of agglomeration for transport appraisal: Existing Evidence and Future Directions', *Economics of Transportation*, 19, 100121.

Grossklags, J. and Acquisti, A. (2007), When 25 Cents is Too Much: An experiment on Willingness to Sell and Willingness to Protect Personal Information, The Workshop on the Economics of Information Security.

Grunes, Alan and Maurice Stucke (2016). Big data and competition policy. Oxford University Press.

GSMA (2018), *The Value Chain*, Executive Summary, London, GSM Association.

Hartman, P. M., Zaki, M., Feldmann, N., Neely, A. (2014), A Taxonomy of Data-driven Business Models used by Start-up Firms, Working Paper, Cambridge, Cambridge Service Alliance.

Haskel, J, and Westlake, S. (2017), Capitalism without Capital: The Rise of the Intangible Economy. Princeton University Press.

HMT (2015), The Aqua Book: Guidance on producing quality analysis for government, London, HM Treasury.

HMT (2019), The economic value of data: discussion paper, Great Britain, HM Treasury.

HMT (2020), The Green Book: Central Government Guidance on Appraisal and Evaluation, London, HM Treasury.

Jones, C. I. and Tonetti, C. (2020), 'Nonrivalry and the Economics of Data', American Economic Review, 110, 9, pp, 2819-2858.

Ker, D. and E. Mazzini (2020), "Perspectives on the value of data and data flows", OECD Digital Economy Papers, No. 299, OECD Publishing, Paris, https://dx.doi.org/10.1787/a2216bc1-en.

Kim, T., Barasz, K. and John, L. K. (2019), 'Why am I seeing this ad? The effect of ad transparency on ad effectiveness', Journal of Consumer Research, 45, 5, pp. 906-932.

Laufer, R. S. and Wolfe, M. (1977) 'Privacy as a concept and a social issue: A multidimensional developmental theory', Journal of Social Issues, 33, 3, pp. 22-42.

Leroux, P. (2016) Get More Value From Your Data And Deliver Tailored Customer Experiences. from D!gitalist Magazine: http://www.digitalistmag. com/customer-experience/2016/01/27/get-more-value-from-data-deliver-tailored-customerexperiences-03959890

Li, T. and Unger, T. (2012), 'Willing to pay quality personalization? Trade- off between quality and privacy', European Journal of Information Systems, 21, 6, 621-642

Lin, T. (2020), Valuing intrinsic and Instrumental preferences for privacy, Paper presented at the Boston Cyber Alliance Talk, Boston University.

Liu, Brynjolfsson, Dowlatabadi (2018) Do digital platforms reduce moral hazard? The case of uber and taxis, Working Paper 25015, Cambridge, Massachusetts, National Bureau of Economic Research.

London Economics (2019) Study of the economic and social returns from the ODI's innovation programme. London

London Economics (2020) Independent assessment of the Open Data Institute's work on data trusts and on the concept of data trusts, Report to the Open Data Institute, DTE-005, London, London Economics.

Macpherson, N. (2013) Review of quality assurance of government models. London: HM Treasury.

Martens, B., A de Streel, I. Graef, T. Tombal and N. Duch-Brown (2020) "Business-to-Business data sharing: An economic and legal analysis", JRC Digital Economy Working Paper 2020-05. Brussels: European Commission.

Martens, B. (2021) "An economic perspective on data and platform market power", JRC Digital Economy Working Paper 2020-09. Brussels: European Commission.

Mawer, C (2015) Valuing Data is Hard, report for Data Valuation Methods, Silicon Valley, Silicon Valley Data Science.

Mitchell, R. C. and R. T. Carson (1989), Using Surveys to Value Public Goods: The Contingent Valuation Method. Baltimore: John Hopkins University Press.

Morton, Scott, Theodore Nierenberg, Pascal Bouvier, Ariel Ezrachi, Bruno Jullien, Roberta Katz, Gene Kimmelman, A Douglas Melamed, and Jamie Morgenstern (2019). "Report: Committee for the Study of Digital Platforms-Market Structure and Antitrust Subcommittee". George J. Stigler Center for the Study of the Economy and the State, The University of Chicago Booth School of Business.

Nathan, M. (2019). 'Does Light Touch Cluster Policy Work? Evaluation the Tech City Programme', CEP Discussion Paper No. 1648.

Niebel, T., Rasel, F. and Viete, S. (2018), 'BIG data – BIG gains? Understanding the link between big data analytics and innovation, Economics of Innovation and New Technology, 28, 7, pp. 1-21.

Nissenbaum, H. (2011) A contextual approach to privacy online. Daedalus 140(4): 17

ODI (2017), The case for government *involvement to incentivise data sharing in the UK intelligent mobility sector*, Briefing Paper, London, Open Data Institute.

ODI (2018a), *The role of data in AI business models*, Report, Open Data Institute.

ODI (2018b), Personal data in transport: exploring a framework for the future, Report, Open Data Institute.

ODI and Fingleton (2019) Open Banking, Preparing for Lift Off, Report, Open Data Institute and Fingleton.

OECD (2020), *Digital Economy Outlook*, Paris: OECD Publishing.

OFT (2006) 'The commercial use of public information (CUPI)', Office for Fair Trading report, http://www.opsi.gov.uk/advice/poi/oft-cupi.pdf

SAS Corporation. (2015). OpRisk Global Data. https://www.sas.com/content/dam/SAS/en_us/doc/productbrief/sas-oprisk-global-data-101187.pdf

Savage, S. J. and Waldman, D. M. (2015) 'Privacy tradeoffs in smatphone applications', Economics Letters, 137, pp. 171-175.

Savona, M. (2019). “The Value of Data: Towards a Framework to Redistribute It”, SPRU Working Paper Series 2019-21 (October).

Schimmelpfennig, D. and R. Ebel (2016), ‘Sequential adoption and cost savings from precision agriculture’, Journal of Agricultural and Resource Economics, 41, 1, pp. 97-115.

Schumann, J. H., Wangenheim, F. V. and Groene, N. (2014), ‘Targeted Online Advertising: Using Reciprocity Appeals to Increase Acceptance among Users of Free Web Services’, Journal of Marketing, 78, 1, pp. 59-75.

Shapiro, C. & Varian, A. (1999), *Information Rules – A Strategic Guide to the Network Economy.* Boston, Mass.: Harvard Business School Press.

Sheehan, K, B. and Hoy, M, G. (2000), Dimensions of Privacy Concern among Online Consumers. Journal of Public Policy & Marketing, 19, 1, pp. 62-73.

Smith, H. J., T. Dinev, and H. Xu (2011). Information privacy research: an interdisciplinary review. MIS quarterly 35 (4), 989{1016.

Statistics Canada (2019), The value of data in Canada: Experimental Estimates, ISSN Number 1705-9658, Canada, Statistics Canada.

Tsai, J. Y., Egelman, S., Cranor, L. F. and A. Acquisti (2011), ‘The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study’, Information Systems Research, 22, 2, pp.254–268.

Venables, A., J. Laird and H. Overman (2016) “Transport investment and economic performance: Implications for project appraisal”, https://www.gov.uk/government/ publications/transport-investment-and-economic-performance-tiep-report

Wamba, S. F., S. Akter, A. Edwards, G. Chopin, and D. Gnanzou. (2015), ‘How ‘Big Data’ Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study’, International Journal of Production Economics, 165, pp. 234–246.

Winegar, A. and Sunstein, C. (2019) ‘How much is data privacy worth? A preliminary investigation’, Journal of Consumer Policy, 42, pp. 425-440.

Annex A: Research Aims and Scope

These are the research aims and scope set out in the invitation to tender for this research.

1. The overall objective of the project is to identify the likely positive and negative externalities of data use (that is, the wider social impacts of data use) and then provide an assessment of the viability of methods that could be used to value them. The primary aims of this research are:

   - a list of potential externalities directly related to the use of data, at each stage of the data 'life cycle'.
   - a greater understanding of which methods are used to capture, quantify and value the externalities of data use.
   - an assessment of methods that are used in capturing externalities in other fields that might be used or adapted for use in understanding data externalities; and
   - an assessment of the viability of these methods.

2. This might entail reviewing methods for valuing social and economic externalities used in other disciplines beyond those used in the area of data. The key topics that we wish to understand are:

   - A much more tightly defined typology of perceived externalities of data use, and to understand the tangible and intangible externalities associated with data use and creation.
   - What are the broad challenges to valuing externalities? E.g. are they in identifying them, or in the quantification, or something else?
   - Are these externalities different from other social and economic externalities being valued in other contexts/disciplines (in terms of assessing their valuation)? And if so, how?
   - What are the different methodologies/models used to value similar externalities?
   - Is there any evidence about what the different levels of awareness and attitudes are about both positive and negative externalities among different sectors? To what extent does the inability to measure externalities restrict companies from appropriate and effective use and sharing of data?
   - What is the practicality and feasibility of these methodologies? That is, out of the methodologies identified that could be used for valuing the externalities of data assets, what are the trade-offs between feasibility of carrying out the methodology, and the insights gained?
   - Are there differences between which externalities are of most importance to different types of organisation (that is, those operating in public, private and third sector, and within different sectors of the economy)?
   - Can we derive 'proof of concept' values for the externalities and methods identified?

Annex B: Literature review approach

1. The literature review looked initially at key studies (such as those cited in the invitation to tender and recent review references on the value of data use to design a literature search strategy. The strategy was developed to papers that are wide-ranging and comprehensive, as the topic was quickly established as wide ranging.

2. Standard databases (Google scholar) and licenced access to library resources (DeepDyve) were used. Search strategies used search terms to gather relevant evidence. These are in table A2.1 and including so-called “bag of words” searches. A search was followed by sifting out those studies that appear least relevant. Key to the approach was that searches were then followed up, using citations of the studies identified to “snowball” beyond initial lists. Such citation follow up used filtering, such as focusing on recent studies or identifying key authors or research groups.

3. The search terms were also used in more focused searches as appropriate. The terms would be used with search restricted to key research groups or web domains (OECD, Nesta, ODI, www.gov.uk). These were used differently for different aspects of search, e.g. government websites were a focus for policy model searches and follow ups.

4. Reviewing of documents was undertaken in stages. An initial template driven literature review of documents was undertaken at the first stage. This involved identifying findings and coding these to topics based on the research questions. In following stages, the template driven approach was found to be less necessary, as studies being considered could be reviewed into the topics that had been established.

Table A2.1: Search words

Ref	Search
0	ITT citations and citations from Diepeveen
1	"data externalities" spillovers
2	“data use” economics
3	"economic value of data"
4	externalities typology "data use"
5	"data value chain"
6	"value chains" "big data"
7	digital productivity
8	privacy economics data international
9	data privacy externalities valuation
10	big data information asymmetry
11	international data flows privacy

Ref	Search
12	“willingness to pay” privacy
13	“willingness to accept” privacy
14	“willingness to sell” privacy
15	UK government policy models
16	“open data” (policy area)
17	(policy area) data privacy economics
18	"open banking" externalities spillovers
19	"location data" externalities
20	"open transport data"
21	“business critical model”
22	“artificial intelligence” “open data”
23	“data driven decision making”
24	“data use” security externality
25	“big data” non-excludable
26	“security costs” survey
27	“data spectrum”
28	“five safes”
31	“data use” platform

Annex C: Summary of key valuation studies reviewed

Table C1: Studies modelling data use externalities

Externality	Study	What it does	Estimates
Economic efficiency, privacy	Deloitte (2017)	Estimates based on open data provided by Transport for London of the outcomes of open data on traffic, appraising economies of scale and scope, quantifying impacts in terms of travel outcomes, which can be valued using Green Books consistent values	Gross value added of £12 - 15 million per year for businesses which also directly created more than 500 jobs.
Spillovers	Brynjolfsson and Syverson (2020)	Develop a model for missing and miss-measuring general purpose technologies such as AI. Model generates a Productivity J-Curve and then used to analyse the historical roles of intangibles tied to R&D, software, and computer hardware.	Find substantial and ongoing Productivity benefits, following a 'J-Curve'
Economic efficiency	Pollock (2008); Carpenter and Watts (2013)	Modelling economic value of public sector information open data policies using behaviours surveys to understand economic value of new used of data after opening and modelling wider economic impacts.	Highlight difficulties in modelling
Economic/ societal impacts	Accenture (2017)	Economic and societal impact of digital technologies within UK manufacturing to support the recommendations of the Industrial Digitalisation Review. Industry level analysis of value at stake due to barriers primarily in adoption.	Estimates of value to industry through new revenues and cost reduction and also value to individuals (new, better products and services) and society (healthcare improvements, reduced carbon waste, safety at workplace)

Externality	Study	What it does	Estimates
Economic	Office of Fair Trading (2006[2])	The effects of changing the data-sharing arrangements employed by the public sector information (PSI) holders were surveyed. The study estimates the direct impact of PSI (i.e. the producer surplus generated by the PSI holders)	Direct impacts £66 million per annum and indirect impact (including the consumer surplus of PSI re-use) was around £518 million, also identified the distortion of downstream competition in the private sector through restricted access to raw data, transferring direct impacts to consumers.
Privacy	Winegar (2019)	Asking about consumers' willingness to pay (WTP) or willingness to accept (WTA)	The median participant is willing to pay relatively little ($5 per month) for privacy but demands much more ($80 per month) to give up privacy. This is an unusually large disparity between WTP and WTA.
Privacy	Lin (2020)	Empirically separate two motives for consumers to protect privacy: an intrinsic motive, which is a "taste" for privacy; and an instrumental motive, which reflects the expected economic losses from revealing one's private information to the firm. Combining a two-stage experiment and a structural model	Find that consumers' intrinsic preferences for privacy range from 0 to 5 dollars per demographic variable, exhibiting substantial heterogeneity across consumers and categories of personal data.
Privacy	Beresford et al. (2012)	Set up two fictitious online shops where experiment participants had to buy a DVD, with one of the two shops requiring more personal data to be revealed in the purchasing process than the other. In this experiment, real stakes were involved, as participants really bought a (subsidised) DVD.	When the shop that requires more information sells the DVD for EUR 1 cheaper, participants are more likely to shop there. When the price is the same, participants are equally likely to choose either store. Therefore, the authors conclude that participants are not willing to pay for privacy.
Friction	Liu, Brynjolfsson, Dowlatabadi (2018)	Compare driver choices at Uber with taxis by matching trips so they are subject to the same optimal route	Drivers in taxis detour on airport routes, with non-local passengers experiencing longer detours and these detours lead to longer travel times.

Externality	Study	What it does	Estimates
Friction	Brynjolfsson and McElheran (2019)	Provide empirical study of the diffusion of data driven decision-making using firm-level linked data.	Friction makes adoption of productivity enhancements costly; slowing spread and lowering productivity