

MEASURING THE ECONOMIC VALUE OF DATA AND CROSS- BORDER DATA FLOWS

A BUSINESS PERSPECTIVE

OECD DIGITAL ECONOMY
PAPERS

August 2020 No. 297

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors. Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to Directorate for Science, Technology and Innovation, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

Note to Delegations:

This document is also available on O.N.E. under the reference code:

DSTI/CDEP/MADE(2019)4/FINAL

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

© OECD (2020)

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.

Foreword

This paper investigates how the economic value of data can be conceptualised and measured from a business perspective. It first discusses data monetisation as a strategy for developing new business models, as well as enhancing “traditional” business models. Secondly, taxonomies for data are reviewed and a new taxonomy proposed with a specific focus on measuring the business value of data. Here discussion is centred on four stylised ‘data monetisation strategies’ that are commonly used by companies to generate new streams of revenue, or to improve existing business processes or products. The role of cross-border data flows as a key enabler of our global economy is also looked at, leading to the concept of a ‘global data value chain’ based on the idea that digitalisation enables the physical detachment of data collection, analysis, storage, and monetisation. Finally, the most promising avenues for measuring the economic value of data are summarised and discussed.

The report was initially prepared to inform an expert workshop on measuring data and data flows convened by the OECD and ESCoE in London in November 2018. It benefits from the detailed discussions which took place at that workshop as well as further input and feedback from delegates to the OECD Working Party on Measuring and Analysing the Digital Economy and members of the OECD Secretariat.

Measuring the Economic Value of Data and Cross-Border Data Flows: A Business Perspective

*David Nguyen and Marta Paczos
UK Economic Statistics Centre of Excellence*

Abstract

The process of collecting, aggregating and analysing data for the purpose of successful operation is nothing new for companies. However, the amount and variety of data they use has increased dramatically in recent years. In fact, data have often become a central element in business models, posing fresh challenges to researchers and policymakers alike. In this paper, we investigate how the economic value of data can be conceptualised and measured from a business perspective. We first discuss data monetisation as a strategy for developing new business models, as well as enhancing “traditional” business models. Secondly, we review taxonomies for data and propose a new taxonomy with a specific focus on measuring the business value of data. Here our discussion is centred on four stylised ‘data monetisation strategies’ that are commonly used by companies to generate new streams of revenue, or to improve current business processes or products. We also discuss how different data characteristics and types affect economic value. Next, we examine the role of cross-border data flows as a key enabler of our global economy. We discuss how and why businesses transfer data across borders, as well as the broad scale and value of cross-border data flows. To do so we present the concept of a ‘global data value chain’, based on the idea that digitalisation enables the physical detachment of data collection, analysis, storage and monetisation. Finally, we summarise and discuss the most promising avenues for measuring the economic value of data and consider their feasibility in the short and long-term.

Keywords: digital, science and technology

Acknowledgements

This research paper has been funded by the Organisation for Economic Co-operation and Development (OECD) within the *Going Digital* project and as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE). Our work has benefitted greatly from discussions at the joint OECD-ESCoE workshop “Data on Data and Data Flows” which took place in November 2018 at the National Institute of Economic and Social Research (NIESR) in London. We are grateful for comments and feedback received from participants at the WPMADe meeting in May 2019 at the OECD in Paris and the AEA/ASSA 2020 Annual Meeting in San Diego. Further, we thank a number of individuals for very helpful comments and suggestions on earlier drafts of the paper. This includes Nadim Ahmad, Sarah Box, Alessandra Colecchia, Carol Corrado, Diane Coyle, David Gierten, Daniel Ker, Wendy Li, Michael Mandel, Christian Reimsbach-Kounatze, Rebecca Riley, Rachel Soloveichik, Peter van de Ven and Andrew Wyckoff. All remaining errors are our own.

Executive summary

The process of collecting, aggregating and analysing data for the purpose of successful operation is nothing new for companies. However, the amount and variety of data they use has increased dramatically in recent years. In this paper, we investigate how the economic value of business data can be conceptualised and measured.

In the first step towards developing a taxonomy with which to categorise data and conceptualise and measure its economic value, it is helpful to discuss the concepts of **data-enabled** and **data-enhanced businesses**. This distinction focuses on the core function of data within each business model: data-enabled businesses are companies that have developed revenue generation strategies fully reliant on data and that would not exist without access to large amounts of data and advanced data analytics. On the other hand, data-enhanced businesses exploit data to better coordinate pre-existing business operations, facilitate decision-making and to introduce new goods and services; data does not alter or determine their core business models.

However, the distinction between both is not dichotomous and we discuss four related revenue generation strategies: data-enabled businesses will likely generate revenue from (1) selling or licensing data, or (2) selling entirely new data-related products, whereas data-enhanced businesses are more likely to use data to (3) improve existing products, or (4) overall productive capacities and business efficiency. Examples of such companies and relevant data monetisation strategies are discussed in detail. These four strategies are then put together in a framework of data monetisation strategies across different business models, which can be used to assess the degree to which businesses are relying on data to generate revenue (and how this changes over time).

Next, to further address the data value-generation process we elaborate on the concept of the ‘data value chain’. It is composed of four stages: i) data collection, ii) data aggregation, iii) data analysis, and iv) data use and monetisation, all of which are underpinned by data storage and (cross border) data flows. However, the economic value of data is a combination of many factors, including the information content that it carries, demand for data, and its actual or intended use. Despite obvious empirical challenges in measuring data characteristics in a standardised way, we review key properties that are associated with economic value from the perspective of businesses. These include linkability, accessibility, timeliness, trustworthiness, and scarcity.

To provide a broader context we review most commonly used data taxonomies and typologies. It should be noted that a *universal* taxonomy covering all types of data does not exist. Rather, different policy questions and considerations around data use by businesses call for different classifications and typologies. We review several ways to distinguish data types, based on the funding for data generation or collection, data usage rights, data subjects, method of data generation, and data source.

We extend our discussion to include the international nature of many business models, by looking at the role of cross-border data flows and barriers to these. While one might argue that these flows do not influence the value of data *per se*, they are important as they critically enable the value creation processes within the ‘*global data value cycle*’. We discuss different types of cross-border data flows and conclude that the volume of data transferred is not necessarily helpful when trying to establish their value in an economic sense. At the same time, the formation of barriers to cross-border data flows can have a material impact on the ability of companies to monetise their data as the creation of economic value is often very much dependent on the ability to move and aggregate data across a number of locations scattered around the globe. Different types of barriers are discussed, in addition to studies estimating the magnitude of their

economic impact, and indicators of restrictiveness such as the Digital Services Trade Restrictiveness Index (DSTRI).

The paper concludes with an overview of the most promising avenues for measuring the economic value of data and a discussion on their feasibility in both the short- and long-term. Here we also discuss what aspects of data and data flows are already measured (e.g. use of cloud computing services, data management & processing and e-commerce). We assess different measurement approaches, including the valuation of data based on market prices (e.g. based on data brokers, data insurance, or data breaches), valuation of data as a knowledge-based or intangible asset (e.g. using cost-based methods), valuation of data based on the business model (i.e. income-based approach as in the key framework developed in this paper), or input-output tables of the economy.

Table of Contents

Foreword	3
Measuring the Economic Value of Data and Cross-Border Data Flows: A Business Perspective	4
Abstract	4
Acknowledgements	4
Executive summary	5
1. Introduction	9
2. The use of data in new business models: towards a taxonomy	10
2.3. The use of data: a taxonomy of business models and business types	14
3. A taxonomy of data types for the purpose of economic measurement	19
3.1. Review of data taxonomies or classifications	19
3.2. Data characteristics	20
3.3. Data types	21
4. Cross-border data flows	24
4.1. Creating value by transferring data across borders	24
4.2. The scale and scope of cross-border data flows	26
4.3. Barriers to cross-border data flows	28
4.4. Measurement challenges arising from cross-border data flows	29
5. Approaches to measuring the economic value of data	31
5.1. What aspects of data are already measured?	31
5.2. Review of approaches to measuring the economic value of data	32
6. Discussion and implications for measurement agenda	37
References	39
Notes	44

Tables

Table 3.1. Overview of different data types	21
---	----

Figures

Figure 2.1. The Global Data Value Cycle	13
Figure 2.2. Businesses and data	15
Figure 2.3. The use of data: data-enabled vs. data-enhanced business models	16
Figure 2.4. Data monetisation across business models and sectors	17
Figure 3.1. Data categories as set out in ISO/IEC standard 19944 “Cloud services and devices: Data flow, data categories and data use”	20
Figure 4.1. Total used cross-border bandwidth, 2005-2017	25
Figure 4.2. Global Submarine Cable Map, 2018	27

Figure 4.3. Global data centre traffic, by type and Consumer Internet Protocol (IP) traffic, by sub-segment, 2015-22 30

Boxes

Box 2.1. How data enable business functions across different business models 10
Box 2.2. Valuing data in business models based on providing ‘free’ digital services 12
Box 4.1. Types of barriers to cross-border data flows..... 28
Box 5.1. Bankruptcy case studies highlighting the value of business data 34

1. Introduction

In their most basic form, data can be described as the unordered and unprocessed representation of any types of observations that are quantified and stored in symbols (OECD, 2013a). As part of the digital transformation, the use of data has become increasingly valuable to businesses and has spurred the creation of entirely new business models. Data, or more precisely the information data hold, often function as a critical input into the production processes of goods and services. For some firms, data are the most valuable asset they own. Businesses leverage data as business intelligence, use it for the purpose of process optimisation, improvement of products and services, and in research and development (R&D) activities (Magalhaes and Rosiera, 2017).

The economic literature looking specifically into the economic value of data for business operations and business productivity is limited.¹ This also means that neither economists nor statisticians have, as of yet, developed a consensus on the best way to conceptualise, classify, and value different types of data and data inputs into the production process. At the same time, data are becoming ubiquitous in business operations and ever more actions are leaving digital “footprints”. Nevertheless, its economic “value” is often ambiguous, making conceptualising and measuring it both theoretically and practically difficult.

These measurement issues are further exacerbated by the international orientation of many business models, which involve considerable cross-border data flows within and between companies. However, without the proper measurement and valuation of ‘data’ it is difficult to assess its role in terms of firm performance or product market structures. Conceptually and empirically, measurement problems arise at the company, industry and country levels. They potentially undermine the accuracy of current economic statistics and, as a result, the development of effective policies aimed at fostering growth and inclusive development in the digital era.

The aim of this paper is to investigate how to measure and value data with a focus on business data. We first discuss the role of data monetisation as a basis for new and traditional business models (Section 2). We then review existing data taxonomies or classifications and propose a new framework focused specifically on measuring the economic value of data used by businesses (Section 3). Next, we examine the increasing importance of cross-border data flows for international business in a global economy (Section 4). Here we also discuss different drivers of cross-border data transfers and barriers to them. Finally, we describe the most promising avenues for measuring the economic value of data (Section 5) and discuss their feasibility in the short- and long-term and implications for the economic measurement research agenda (Section 6).

Naturally, there are aspects related to the value and measurement of data that are beyond the scope of this paper. This includes the development of detailed ‘ready-to-go’ options to empirically measure different types of business data. While this is a necessary and desirable outcome of future research in the area, the main aim here was to conceptually place the business model related to the monetisation of data at the centre when thinking about how to measure the value of data. Nevertheless, the second part of the paper discusses a number of empirical measures that we could find based on existing studies or borrowed from related fields. Finally, while our discussion focusses on the private returns generated from data, we are aware that this has clear implications for discussions on under-priced or under-served goods or services. For example, businesses deciding to invest in (or acquire) data

likely do not “price-in” all potential externalities of that data being shared more widely (e.g. with researchers or the government). This could be due to a lack of awareness or business incentives, when data monetisation is not feasible or profitable.²

2. The use of data in new business models: towards a taxonomy

2.1. Data-enhanced and data-enabled business models

Companies have long been involved in the process of collecting, aggregating and analysing data for the purpose of running their businesses. However, it is only in recent years that the scale and scope of data used by companies has changed and that data have moved to the core of many business models.³ Some even suggest that “*so much of the world is instrumented that it is difficult to actually avoid generating data*” (OECD, 2016b: p.52). On the one hand, data help to better coordinate existing business operations (e.g. supply chains), facilitate decision-making and enable the improvement or introduction of new goods and services. In **Box 2.1** we provide a number of practical examples of the type of processes concerned. Crucially, this ‘incremental’ digitalisation does not alter or determine the core business models of these companies. We refer to these types of companies as “*data-enhanced businesses*”.⁴

For data-enhanced businesses, data facilitate the creation of new value within an established business model. This is closely related to the concept of “data-driven innovation (DDI)”, as outlined in OECD (2015).⁵ This intuition is confirmed by an empirical study based on 18,000 manufacturing firms in the US that shows how establishments that rely on data-driven decisions exhibit higher productivity and higher output in terms of value-added (Brynjolfsson and McElheran, 2016). More recently, Hughes-Cromwick and Coronado (2019) discuss in detail how the automobile, energy and financial services sectors rely on data for their short-term and long-term business decisions. By contrast, some businesses could be thought of as fully digital, or ‘data-native’ companies. Data are the lifeblood of their operations and the key enabler of their core revenue-generating activities. This includes online platforms that rely on data and data analytics to match users and providers of goods or services. We refer to these as types of companies as “*data-enabled businesses*”.⁶

Box 2.1. How data enable business functions across different business models

Potential uses of data flows to improve business functions:

- Automation of supply chains
- Consolidation of back-office operations
- Scalability of software via the cloud
- M2M communication (e.g. sensors, Internet of Things)
- Digital collaboration between teams (e.g. in R&D, sales, procurement)
- Online purchases of goods and services
- Use of mobile apps to deliver products and services
- Use of online platforms as intermediaries
- Analysis of big data

Examples reported by the Information Technology & Innovation Foundation, ITIF (2015):

1. Mining: To enhance efficiency and cut costs, Rio Tinto aggregates data from its laboratories, control systems, maintenance logs and surveillance cameras, located in different mining sites globally. It is reported that every day around 30GB of data are

transferred between its global sites and a data processing centre in Brisbane, where the information is analysed.

2. **Manufacturing:** Volvo and Scania can aggregate real-time vehicle information such as location and diagnostics data to improve driver safety, environmental impact and supply chain management. Volvo aggregates data from all countries in a centre in Sweden. Similarly, Unilever is operating two global data centres with a total of 4 000 servers, which enables it to use global consumer analytics based on a very large dataset.

3. **Oil & Gas:** Royal Dutch Shell aggregates data from around 10,000 sensors placed in oil wells using three cloud-enabled global data centres. These sensors have been co-developed with HP and enable the company to find new resources based on high-resolution seismic data.

4. **Aerospace:** Boeing aeroplanes transmit information in-flight which is analysed in near real time to identify problems early on and enable maintenance crews on the ground to be prepared for when the plane arrives. This reduces turn-around times and delays. Reportedly, a single engine of a Boeing 737 produces 20 TB of data per in-flight hour (though it is not clear what proportion of this data is actually transmitted).

5. **Healthcare:** Hospitals increasingly send medical images to doctors located in other countries for the purpose of diagnosis. This can reduce waiting times for patients. For example, the National Health Service (NHS) in England outsourced the processing of MRI scans using the company Alliance Medical which has around 200 imaging sites across Europe. The Swedish company Hermes Medical Solutions offers cloud-based software applications to share medical images across 30 countries, though 95% of patient data are stored in Sweden.

Sources: ITIF (2015), Business Roundtable (2015), US Department of Commerce (2016).

Data-enabled businesses are companies that have developed fully digital business models, that would not exist without access to large amounts of data and advanced data analytics (e.g. Amazon, Uber, Twitter, Booking.com and Airbnb).⁷ In addition, many start-ups are founded without a clear initial revenue model but with the expectation of future data-related revenue streams. For example, TravelPerk is a start-up using a freemium⁸ business model based on taking affiliate commissions on bookings. However, “*down the road, it also has its eye on generating a data-based revenue stream via paid-tier trip analytics*” (TechCrunch, October 2018). A more mature transport and mobility app, CityMapper, monetises some of its data on users’ travel patterns by providing a transport network analysis tool. It has also attempted to run its own bus service to compete directly with other forms of transport and is developing new software for buses to transmit real-time data to scheduling systems.

For both data-enabled and data-enhanced businesses, the value-added related to data is often a by-product of data analytics and data itself. Moreover, as highlighted by the examples above, the ways in which firms monetise the data-data analytics nexus often depend on the specific business model they adopt (Li et al., 2019).⁹ Therefore, when discussing the value of data, it is useful to consider a data value chain or data value cycle (OECD, 2013b; Li et al., 2019).

The recent literature on measuring ‘free digital services’ holds some further insights for how businesses turn data in economic value. Some studies are reviewed in **Box 2.2**.

Box 2.2. Valuing data in business models based on providing ‘free’ digital services

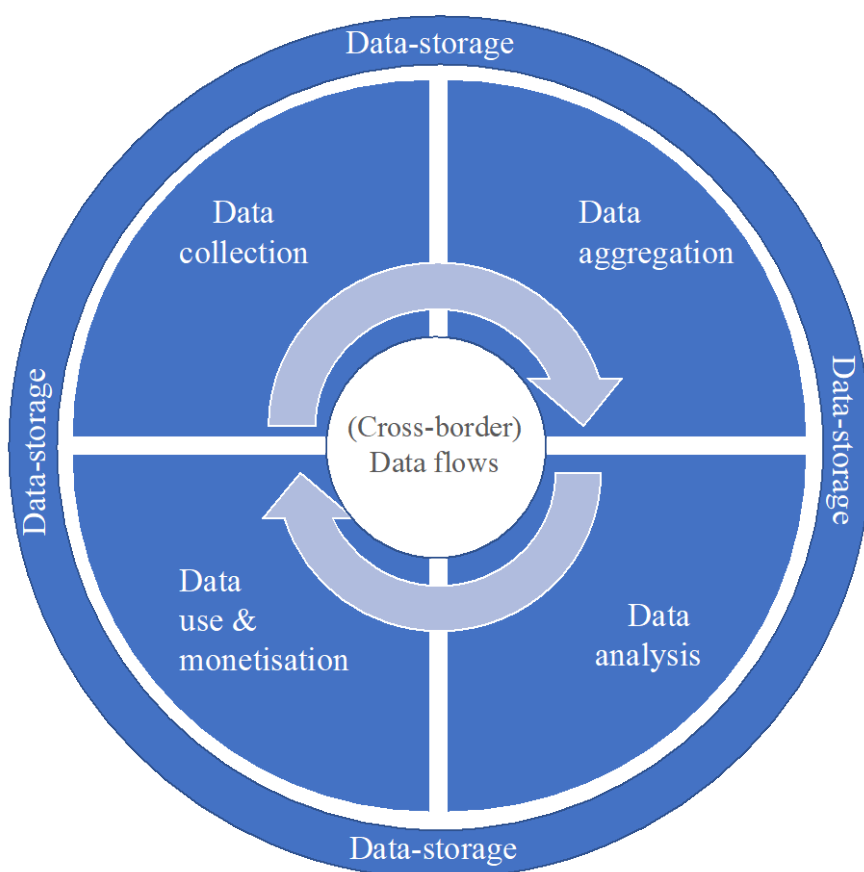
- Digitalisation now affects almost all aspects of the economy and new digital goods pose great challenges to our current measures of consumer welfare and GDP (Coyle, 2017). One challenge is that new digital services are often free at the point of use, especially for data-enabled businesses that rely on network effects for gathering data (i.e. social networks). However, this does not mean that the data have no value. For example, the market capitalisation of Facebook, which also owns WhatsApp and Instagram, exceeded USD 600 billion in mid-2018 illustrating the considerable value the market believes that revenue streams based on data collected via “implicit” or “barter” transactions can generate.
- The “professional network” LinkedIn which is operating a data-enabled business model, was acquired by Microsoft in 2016 for a total of USD 26 billion, which makes it possible to link data on individuals with other Microsoft products such as Office and Outlook and in turn increases their value. Shapiro and Aneja (2019) estimate that revenues associated with monetising Americans’ personal data collected by the major Internet search engines, social media platforms, data brokers, credit card companies and healthcare data businesses amounted to more than USD 78 billion in 2018.
- One attempt to derive the value of ‘free’ digital services, e.g. via massive online choice experiments as done by Brynjolfsson, Collis and Eggers (2019). Similarly, Nakamura et al. (2018), treat free content as a barter transaction where consumers and businesses receive content in exchange for exposure to advertising or marketing, and ultimately, *“households are treated as active producers of viewership services that they barter for consumer entertainment.”* (Nakamura et al., 2018; p.2). Since consumers provide their data for free in exchange for free services this approach is directly related to the value of data. Ahmad and Schreyer (2016) describe these triangular transactions between the users of digital products, advertisers (as data users) and the service providers, but they also highlight the conceptual issues that these types of valuation approaches entail. For example, while it is difficult to interpret the magnitudes of valuations, Coyle and Nguyen (2020) show that changes and rankings in valuations of free digital products can be very insightful.
- ~~Another method is to use massive online choice experiments as done by Brynjolfsson, Collis and Eggers (2019), who estimate that online services such as Wikipedia or WhatsApp generate annual consumer surplus in the range of billions of US dollars. Since people are not directly paid for their data by social media platforms, it is difficult to say how valuable it is. However, considering the large amounts of welfare that people seem to get from these free services can serve as an indication. Future extensions of these types of studies might consider asking people more directly about how they would value their data in monetary terms.~~

2.2. The global data value chain

To gain some further insights into how exactly data can generate value for businesses we build on the concept of a ‘data value chain’ (see e.g. Rayport and Sviokla, 1995; Visconti et al., 2017; OECD, 2015). We distinguish four stages in the process of deriving value from data: i) data collection, ii) data aggregation, iii) data analysis, and iv) data use and monetisation. All four stages are underpinned by data storage and (cross border) data flows which occur throughout (see Figure 2.1).

The collection of raw data (stage 1) can take place in a single place or across many places and countries simultaneously. However, the additional value creation is limited when businesses are faced with barriers to data transfers and aggregation (stage 2), particularly when these involve cross-border transactions. The next stage is data analysis of aggregated data or multiple combined datasets. This can take place in yet another location and hence involve (further) data transfers across borders.¹⁰ Finally, the monetisation stage generates additional data that can feed into the data value chain to make it a global ‘data value cycle’. When it comes to the storage and computing that underpins data collection, aggregation, analysis and monetisation, businesses can benefit from economies of scale by centralising all of their data in one or a few locations. Others might find it beneficial to store copies in data centres across the globe, e.g. to protect it from disasters and to reduce access times (latency). Cloud computing is a key enabling technology of this process.

Figure 2.1. The Global Data Value Cycle



Source: Authors' own elaboration.

The concept of the ‘data value cycle’ can also pose challenges to the standard distinction between process and product innovation. For instance, for many automobile companies, data serve as an input in the production process, constituting a process innovation. In a similar vein, gathering more data can also fuel the improvement of the algorithms used for data analytics. Examples include Facebook’s process of ranking of all available posts to be

displayed on a user's news feed (McGee, August 2013) or Amazon's A9 algorithm used to match users' search queries to products they are most likely to purchase (Wikipedia, s.d.).

On the other hand, the business models that involve data collection coupled with advanced data analytics could eventually lead to the introduction of new data-based services. Here we are thinking of manufacturing firms that extend their product range as well "servitise" their business model – making a gradual transformation from a traditional goods manufacturer into a (data-driven) services provider. One example is the car manufacturer BMW that offers "CarData" – a set of individually tailored customer services based on car-generated telematics (BMW, s.d.). Similarly, John Deere, a manufacturer of agricultural machinery, uses IoT solutions to offer data-fuelled and AI-driven agritech platforms (John Deere, s.d.). It should be noted that these data-driven improvements often allow for collecting further data and, in turn, broaden the potential for even more new or improved products. The examples above also relate closely to economies of scale and scope experienced by both data-enabled and data-enhanced businesses, and their implications for economic measurement. Manufacturing firms that add services to their portfolio are not new, and earlier approaches may indeed offer insights into how to integrate data-enhanced business models into economic statistics. The key difficulty here is to separate the value-added from the physical product from the data-enabled service.

Demand economies of scale as demonstrated by constant improvement of Facebook's algorithm thanks to the increasing number of its users, or data-driven solutions for manufacturing companies where accuracy improves with more data inputs, further complicate the assessment of the value of underlying data that is exploited by the algorithm. Similarly, measurement is difficult because data-reliant companies can offer different-but-similar customised varieties of new products at a very low marginal cost (economies of scope).¹¹

Considerations of economies of scale versus scope, product versus process innovation, and servitisation of manufacturing, demonstrate that answering the question of "*What determines the value of data?*" is a non-trivial task. It is fair to state that there is no standardised approach for measuring the 'value' of data and its importance from a business perspective. However, there seems to be a consensus that simple measures of data volumes are not good enough to approximate the economic value of data, or data flows (OECD, 2019a; HMT, 2018).

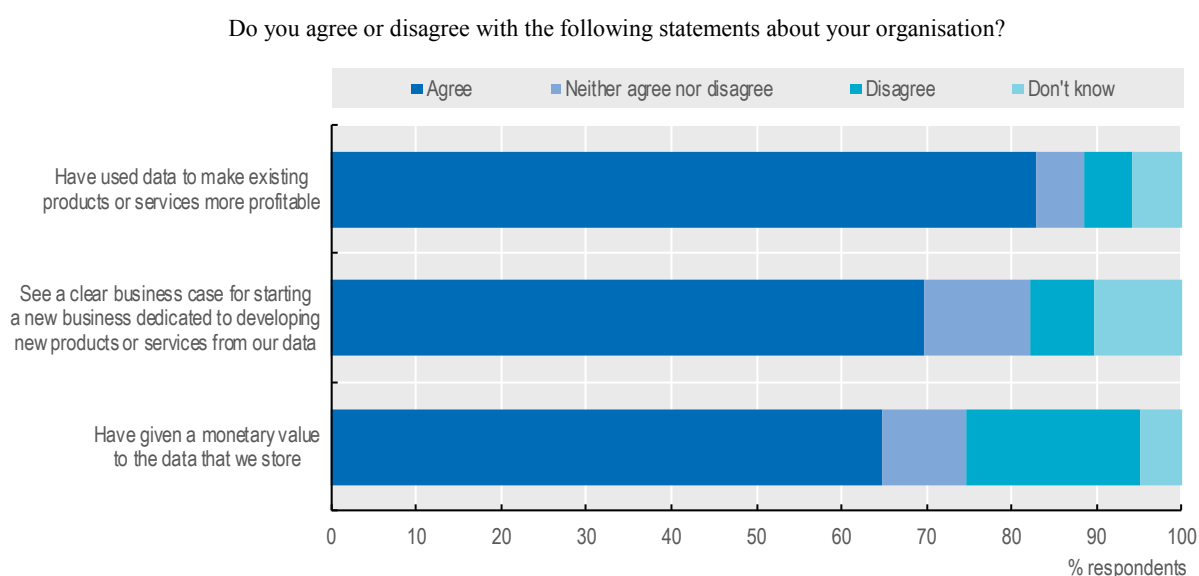
Straightforward data quantity measures, such as bits and bytes, have little connection with the information contained within each data 'unit'. As an illustration, the transfer of a new car design intuitively carries a different value potential than an individual's purchase history or clickstream. Cisco (2018) show that video accounted for 75% of all Internet Protocol (IP) traffic in 2017, the greatest single category of online data flow. Considering data volume versus value is further complicated by the use of ever-improving compression techniques widely applied to flows of data. Understanding the context in which data is used, combined with a categorisation of different data types and characteristics, is essential for the purposes of economic measurement of the value related to data.

2.3. The use of data: a taxonomy of business models and business types

The economic value of data is a combination of many factors, including the information content that it carries, demand for the data, and its actual or intended use.¹² This is closely related to the idea of *relevance* – the ways in which data can be exploited within a particular business model. In a survey conducted by the Economist Intelligence Unit in 2015, close

to 90% of almost 500 executives stated that data were now used by “the majority or all parts of the business”. (The Economist Intelligence Unit, 2015). This serves as an indication of how relevant data have become across constituent parts of many businesses. Interestingly, 65% of respondents stated that their company had assigned a monetary value to the data they store (see Figure 2.2). Furthermore, 83% reported that their companies made use of data to increase profitability of existing products and 70% saw a clear business case to develop new goods or services based on their data. Finally, more than half saw a business case for selling data owned by their company.

Figure 2.2. Businesses and data



Note: Agree summarises “strongly agree” and “somewhat agree” responses. Disagree summarises “somewhat disagree” and “strongly disagree” responses. Based on a survey of 476 senior executives worldwide.

Source: OECD based on “The Business of Data”, The Economist Intelligence Unit report (2015).

Some companies have developed internal data valuation methods and prepare internal measures of performance (or key performance indicators, KPIs) based on such data characteristics. For instance, the asset valuation models summarised in Laney (2011) include methods to approximate the *market value of information* by trying to capture the income that can be generated by selling, renting or bartering data and the *loss value of information*, which aims to measure the cost to replace the data and the financial impact to the business if the data were lost. Gartner (2016) propose two sets of firm KPIs related to the level of (i) firm digitalisation, as well as the (ii) direct improvement of current business functions based on data and new business models.

Based on the previous two paragraphs it becomes clear that the monetisation of data can come in different shapes and forms. This includes selling data directly, improve operations by sharing data, use data to develop new products and services, and improve existing products and services. However, what they have in common is that they are directly related to a particular business model.

The taxonomy developed and presented here focuses on the role of data across different business models and within businesses that combine multiple business models. A *key factor* is the *actual or intended use of data to generate revenue now or in the future*. Below we

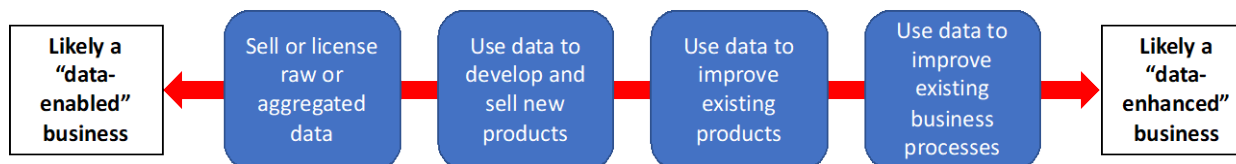
also discuss several data characteristics associated with the value of data for businesses, though we acknowledge the empirical challenge in measuring those in a standardised way. Taken together we believe this can provide the conceptual basis for the economic measurement and valuation of data since data only become valuable to businesses if it is (or is intended to be) used in a profit-generating way.

In our taxonomy, the ‘business model’ takes centre stage as it describes the ways in which a company generates and captures economic value, i.e. new or improved revenue streams. In our understanding, the adopted business model seeks to answer central questions, including what to offer to which customer group as well as how to deliver economic value and for what price. Hence, the business model seems an appropriate starting point when thinking about how businesses rely on data – directly or indirectly – in order to create value, develop a competitive edge and, ultimately, to generate streams of revenue. Certainly, a company can adopt multiple models at the same time.

In **Figure 2.3**, we illustrate how data could be used in different business models within a company. We distinguish four categories of data-related business models and illustrate how they relate to the distinction between data-enhanced or data-enabled businesses:

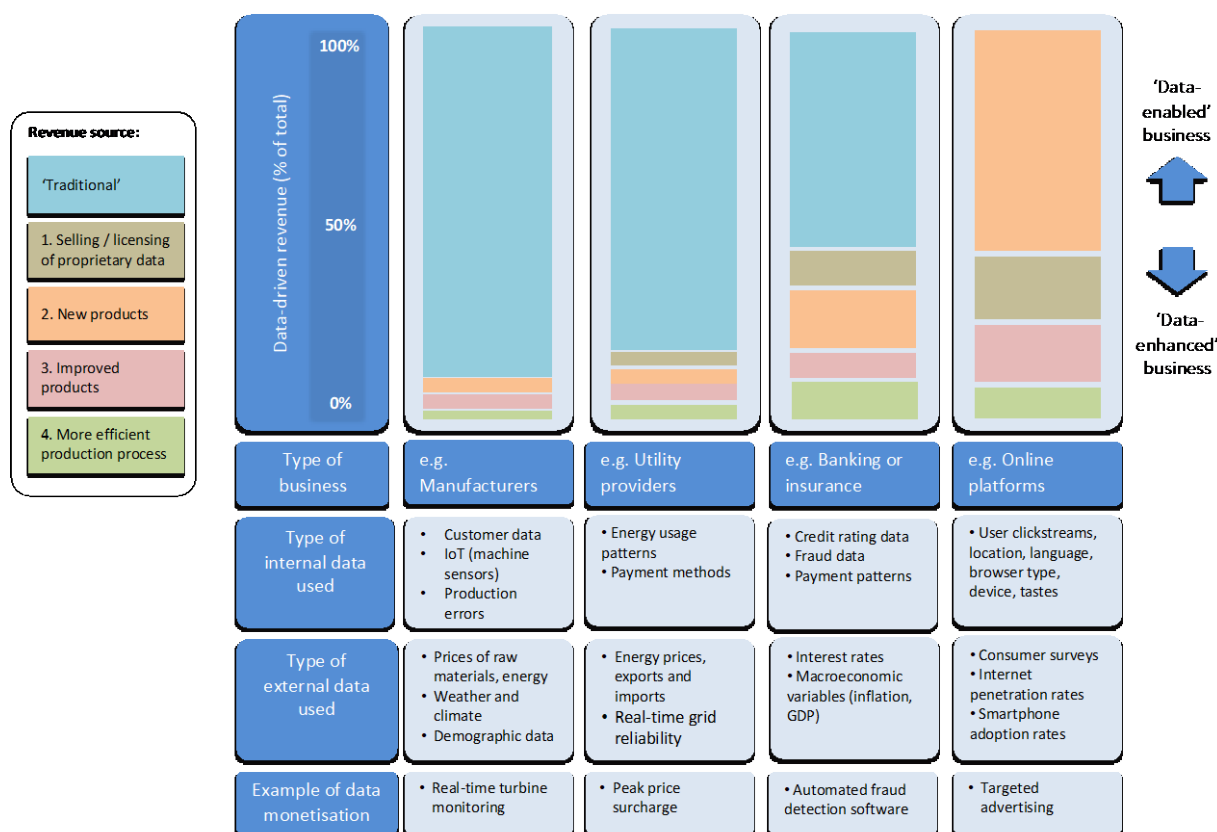
- Category 1. Selling or licensing raw or aggregated data;
- Category 2. Developing and selling new data-related products;
- Category 3. Use data to improve existing products;
- Category 4. Use data to improve production processes or business efficiency.

Figure 2.3. The use of data: data-enabled vs. data-enhanced business models



Source: Authors' own elaboration.

Figure 2.4 provides a schematic illustration of our taxonomy that relates stylised business types to the four different data-related business models outlined in **Figure 2.3** based on the use of internal and external data. The key metric we propose to measure the economic ‘value’ of data for a business is the share of total revenue that is derived from the monetisation of data (in some form or another).¹³ For the distinction of data-enabled and data-enhanced businesses, we refer to their ‘core’ business model, defined here as the ‘main’ source of revenue generation (i.e. more than 50% of total revenue). This share is illustrative as estimates do not appear to be readily available. Nevertheless, developing such a metric in the future seems highly relevant. For example, findings based on the McKinsey Global Survey (2018) suggest that data monetisation is an increasingly important driver of revenue growth. It is reported that the monetisation of data contributes to 10% or more of the total revenue for 32% of high-performing businesses and 9% of all other businesses.¹⁴

Figure 2.4. Data monetisation across business models and sectors


Source: Authors' own elaboration

The coloured bars show – in a highly stylised way – how much the different revenue streams of a business potentially contribute to its overall revenue. In addition to the four data-driven business models (Figure 2.3) we separate ‘traditional’ sources of revenue, which in simplified terms relates to ‘non data-driven’. Whenever data-driven revenue contributes to more than half of total revenue (the dashed line), we speak of a ‘data-enabled’ business (as opposed to a data-enhanced business). In addition, for each type of business we illustrate what types of internal and external data it may have access to for monetisation purposes. Again, this is a highly stylised representation for the simple purpose of classifying different types of businesses with regards to how they can generate economic value from data. By using internal or external sources of data, or a likely a combination of both, businesses can find new ways to create economic value for themselves and their customers. Note that external data, despite possibly being free (or non-rival), carries value in the production process when integrated with the overall business model. The last row in Figure 2.4 provides examples of such data monetisation strategies.

The first column refers to the example of a manufacturer of transport goods (e.g. Airbus, Boeing); while the insights would be analogous for other manufacturing businesses. For instance, the bulk of total revenue is likely driven by the sales of aircraft and parts as well as providing maintenance and training services. However, based on sensors that can be installed in those aircraft, the business is able to reduce operational interruptions, maximise aircraft utilisation and flight operations (data monetisation category 3) as well as offer a

new additional service of real time remote access to aircraft data (category 1) (Airbus, October 2018) and (Boeing).

Similarly, based on the second column of **Figure 2.4**, the main stream of revenue for an energy company would be the provision of electricity. However, using customer energy usage data combined with freely available external data on energy prices, it can design new pricing schedules, e.g. peak usage surcharge (category 4). Alternatively, it may decide to monetise those data by directly selling them, if compatible with data privacy regulations (category 1).

On the other end of the spectrum, data-enabled businesses would only be able to offer a minimal fraction of their goods or services if they could not collect the data (or rely on externally acquired data and data analytics) to improve their services, or to license the data access. In this sense, their revenue streams are completely data-dependant. However, there exists a degree of differentiation among those businesses in terms of their reliance on a specific business model. Li et al. (2019) show how this is the case for different online platforms (column 4 in **Figure 2.4**). For instance, the revenue of the online platform Amazon Marketplace predominantly comes from the supply of a data-enabled service – a buyer-seller matching platform (category 2). However, Amazon also uses data to offer sellers the opportunity to promote their products to some individuals (category 3)¹⁵, and licenses access to internally collected customer behaviour data (category 1). The company also constantly uses data to improve its algorithms (category 4).

In addition, based on a data-driven understanding of customer needs, Amazon also offers its own products that directly compete with independent sellers on its platform. This includes ‘Amazon Basics’, a private-label for home goods, office supplies and tech accessories, launched by Amazon in 2009. Based on the vast amounts of data it can access, these products can be designed to suit the taste and price range of specific groups of consumers.¹⁶ At the extreme, almost all of the revenue of data brokering companies such as Experian or LexisNexis is derived from licensing the access to data (category 1) and the delivery of analytical data-based services (category 2).

With advancements of artificial intelligence (AI) new business models will emerge and firms will undergo further transformations. It is possible to imagine that, in the near future, data-enhanced businesses such as car manufacturers will see major changes in their core business models as they shift from offering ‘traditional’ goods (i.e. sale of vehicles) to ‘digital’ services (e.g. sale of ‘mobility solutions’ like on-demand or subscription-based rental of vehicles via the Internet). Some examples of firms that already undertook this shift include German car manufacturer BMW, which started to offer mobility services, and Finland-based Konecranes – a manufacturer of cranes that uses data analytics to offer a new service called “TrueConnect”. The service allows continuous monitoring of cranes and their operators (e.g. to predict wear and tear or maintenance requirements) with the final goal of selling crane “*uptime or number of lifts*” rather than just cranes.¹⁷

3. A taxonomy of data types for the purpose of economic measurement

3.1. Review of data taxonomies or classifications

A broad taxonomy that distinguishes economic and social dimensions (i.e. personal vs. non-personal data, open vs. closed data) from the technical dimensions of data (i.e. user- vs. machine-generated data) is developed in OECD (2013a). These are mainly discussed within the context of ‘big data’, which *prima facie* refers to data volume, making it difficult to link it to its economic value. Various classifications of personal data are reviewed in OECD (2013b), including a discussion of methods for assigning monetary values to personal data. It should be noted that a *universal* taxonomy covering all types of data does not exist. Rather, different policy questions and considerations around data use by businesses call for different classifications and typologies.

The Swedish National Board of Trade (2015) provides a definition and taxonomy of personal data in order to exemplify what kind of data is used by companies and what categories of personal data are relevant for international data transfers. Their report differentiates between *volunteered*, *observed* and *inferred* personal data. Based on a small survey of Swedish firms, the report relates different categories of data transfers to the specific business functions they enable or complement (e.g. technical data to upgrade software). Beyond personal data, it also proposes a classification of different types of data based on how data are used:

- Corporate data;
- End-customer data (business-to-consumer; B2C);
- Human resources data;
- Merchant data (business-to-business; B2B); and
- Technical data.

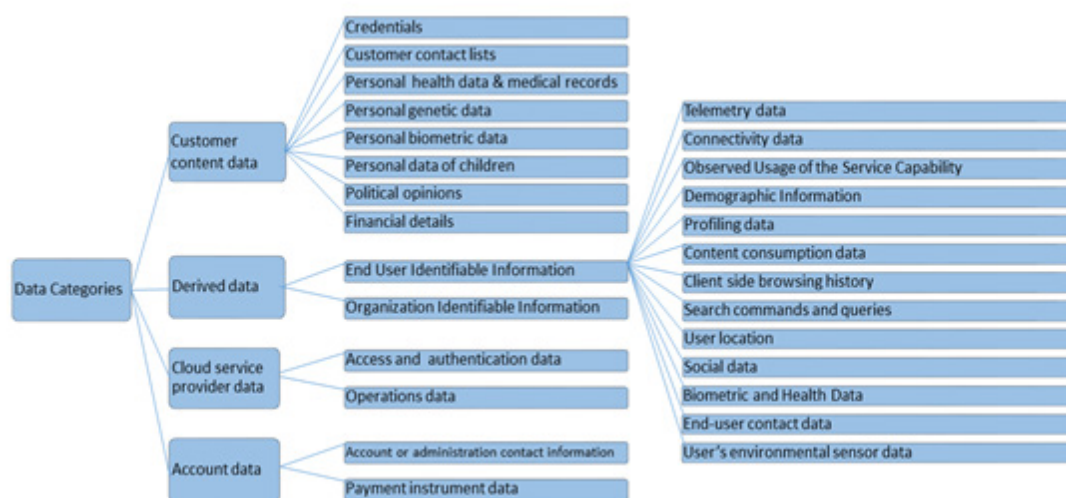
The U.S. Department of Commerce (2016) provides a categorisation of four types of data flows for the purposes of discussing how their cross-border transfer may be captured by economic statistics. The aim of their taxonomy is to scope the potential for measuring the economic impact of cross-border data flows and to put forward recommendations to improve available economic metrics for them. The four types are:

- Purely non-commercial data traffic (e.g. government and military communications);
- Transaction data flows between buyers and sellers at a market price (e.g. online banking or advertising);
- Commercial data and services exchanged between or within businesses or other related parties at zero market price (e.g. design information);
- Digital data and services delivered to and from end-users at zero market price (e.g. free email, free maps and navigation, social media).

In the advent of the EU’s General Data Protection Regulation (GDPR) enforcement date in May 2018, many businesses searched for support in classifying their data with the goal of coming up with GDPR compliant data use statements. Many have relied on the existing ISO/IEC standard 19944 on “Data flow, data categories and data use”¹⁸, which, most

importantly for our purposes, contains a well-developed section discussing different data categories (**Figure 3.1**). The four main categories are customer content data (incl. biometric and contact information), derived data (identifying either end-users or organisations), data by cloud service providers, and account data (payment or admin data). This is attractive from the point of view that companies are likely to both understand and use this classification. However, the primary focus of this typology is data privacy in the context of individual and personal data rather than its value in an economic sense.

Figure 3.1. Data categories as set out in ISO/IEC standard 19944 “Cloud services and devices: Data flow, data categories and data use”



Source: Reproduced from ISO/IEC 19944:2017. Copyright ISO/IEC. All rights reserved.

3.2. Data characteristics

Data have a number of distinctive features, which contribute to their unique nature as an intangible asset. These are an important determinant of usability within different business models. Most importantly, data is non-rivalrous in nature (OECD, 2013a; Mandel, 2017) and it is not subject to the standard wear and tear of tangible assets. This means that data can be used multiple times without inherently diminishing their value. In principle, data can be exploited and re-exploited infinitely at low marginal cost. Usually, it is data infrastructure and analytics that mainly determine the costs of data re-use.¹⁹

A simple Google search for ‘data monetisation’ returns almost 18 million results discussing countless data characteristics that are potentially associated with value. In an early contribution on this topic, Laney (2011) discusses a number of factors that affect data ‘utility’ and, ultimately, economic value from a business perspective - providing a list of ‘objective’ and ‘subjective’ data characteristics. Although they could prove difficult to measure in a standardised way, those data features are fundamental to how data can be utilised to generate value for both data-enabled and data-enhanced businesses. Data are most likely to provide a basis for monetisation and value-creation if they are:

- **Linkable**²⁰ – can be merged with the other datasets;
- **Accessible** – easily retrieved and/or integrated into business processes;

- **Disaggregated** – at the desired level;
- **Timely** – updated with sufficient frequency to meet the business requirements, e.g. annually, daily or in real time;
- **Trustworthy** – deemed credible by those using it; data are unbiased and impartial, and do not depend on the judgment, interpretation, or evaluation of individuals (see Open Data Watch, 2018);
- **Representative** - records do not contain missing fields, data are representative enough to meet business requirements;
- **Scarce** – proprietary or secret, difficult to come by.

There are also some data characteristics that are often discussed in the context of big data and its features – known as the “3 V’s of big data”: volume, variety, velocity (Gartner 2011). For example, data volume, understood as data being a collection of a sufficient number of observations (closely related to the statistical power of data), could impact the data’s value-generating potential. However, as noted above, volume alone will not be a sufficient characteristic that determines economic value. When considering the costs of storing and processing large amounts of data (though those have decreased drastically in recent years), ‘hoarding’ large a volume of irrelevant data could even be detrimental to business performance.²¹

3.3. Data types

By building on the OECD (2013a) classification, we can consider data ‘types’ based on ownership or right to use, data source, funding of data collection and maintenance, data access, and methods of collection. Note that the categories, listed in **Table 3.1**, are not mutually exclusive as some datasets (acquired or generated) will fall into more than one category. Crucially, it is often the case that there is a continuous mapping of data types within a given criterion. For example, the distinction between what is personal and what is non-personal data has become increasingly difficult with new data mining techniques or data analytics solutions. It could be the case that new (or newly combined) data sources enable the identification of an initially anonymised individual (e.g. comparing data on IP address, cookies, geo-location, social media profiles, and clickstreams), thus making non-personal data, personal. This process of “browser fingerprinting” is discussed in great detail by Cyphers and Gebhart (2019) and often based on triangulating information on screen resolution, language setting, time zone, and software version.²²

Table 3.1. Overview of different data types

Key aspect	Data Type	Description
<i>Funding of data collection and maintenance</i>	<i>Private sector data</i>	Data that is funded, created, maintained and held by the private sector, e.g. car company in-house generated data on their production processes, or a database of purchases from an online grocery shop.
	<i>Public sector data</i>	Data that is funded, created, maintained and held by the public sector; Example: data on health

		records of patients or individual tax records, or data originating from the US GPS satellites ²³ .
Ownership or right to use	<i>Proprietary data</i>	Data with clearly defined ownership that is protected by Intellectual Property Rights or any other rights with a similar effect (OECD, 2019a); this could include <i>individual</i> data as well as <i>organisational</i> data ²⁴ .
	<i>Open data (or public domain data)</i>	Data that is publicly available (as opposed to proprietary data), free to use by anyone for any purpose without any legal restrictions. It is not protected by Intellectual Property Rights, copyrights or any other similar legal rights; Example: <i>data.gov.uk</i> , which is a repository maintained by the UK Government to make non-personal UK government data available to the public.
Data subject	<i>Personal data</i>	Personal data is any data that allows for the identification of an individual data subject (OECD, 2013b). It can cover public and private sector data, e.g. user-generated content (e.g. blogs, photos, tweets) or geo-location data from mobiles as well as public sector data (e.g. police records, social security numbers).
	<i>Organisational data</i>	Organisational data describes data that allows for the identification of organisations. This data is usually controlled by organisations themselves, either legally or for contractual reasons. It can also be held by public bodies such as tax authorities. It is often commercially sensitive data.
Data generation	<i>User created data</i>	User created data is data that has been made available by an individual (e.g. telemetry tracking data, consumer behaviour data collected through mobile apps or social media posts). This can be volunteered data (i.e. “active”), observed data (i.e. “passive” or “implicit”), or derived data about a user (see also OECD, 2019c).
	<i>Machine generated data</i>	Machine generated data, e.g. machine-to-machine communication (M2M); Internet of Things (IoT), i.e. data collected from sensors.
Data source	<i>Internal data</i>	Internal data is data that collected and consolidated from different branches within a business. For example, lists of purchase orders from the sales department, transactions from accounting or any other internal source which is responsible for recording information about a business’ commercial interactions.

	<i>External data</i>	External data is data not collected internally, but rather obtained from a source outside of company - for instance, by purchasing access to a proprietary database. This could be <i>acquired</i> data as well.
--	----------------------	--

Source: Authors' own elaboration.

In the previous sections of this paper, we have discussed the difficulties related to measuring the value of data from the perspective of businesses. We have developed a conceptual framework to approach the issue based on the business types, alongside discussing some key data characteristics and types that are relevant for measuring the value of data. It should be clear by now that this is no trivial exercise as it is related to the longstanding question of the 'value of information'. The following section adds another dimension to this challenge, discussing the global nature of some data, including the use of cross-border data flows that underpin many modern businesses operations.

While one might argue that these flows do not influence the value of data *per se*, we think they are important as they critically enable the value creation processes highlighted in the global data value cycle in Section 2. At the same time, the formation of barriers to cross-border data flows can have a material impact on the ability of companies to monetise their data. We first discuss the motivation of businesses to transfer data across borders, followed by a brief discussion on both the scale and scope of data flows and the main barriers to cross-border data flows.

4. Cross-border data flows

4.1. Creating value by transferring data across borders

The increasing digitalisation of the global economy is not only driving data flows within countries but also across borders (European Commission, 2017). The main reason is that digitalisation enables the physical detachment of data collection, aggregation or analysis, storage, and use or monetisation, as outlined in Section 2. Each of these steps can take place in multiple countries, raising various issues for the measurement and valuation of data and data flows. This trend is likely going to accelerate due to the decreasing costs, as well as increasing sophistication and ease of use of digital innovations such as cloud computing, the Internet of Things (IoT), edge computing, artificial intelligence and machine learning tools, robotics and automation, and digital distributed ledger-based transactions (i.e. blockchains).

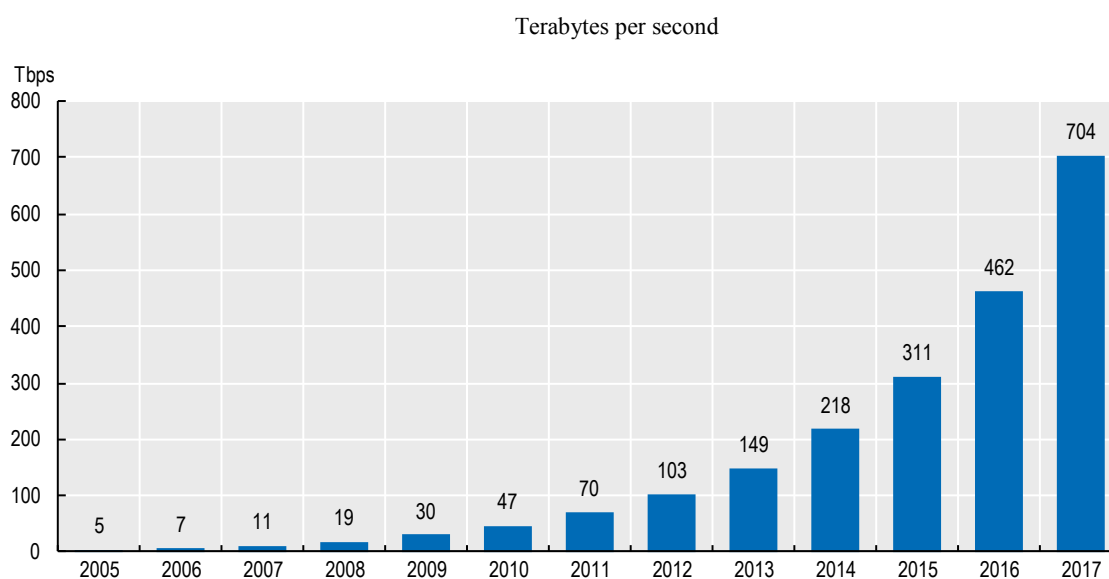
Cross-border data transfers enable businesses to build and maintain complex global value chains. In other words, the creation of economic value is often very much dependent on the ability to move and aggregate data across a number of locations scattered around the globe. The ability to transfer data internationally enables firms to effectively coordinate their research and development, supply chains, production, sales, and post-sales processes (US Department of Commerce, 2016; OECD, 2018a). While the rationale behind introducing barriers to cross-border data flows can be grounded in national security or data privacy concerns, or the desire to protect domestic markets, it is clear that in many cases impediments to international data transfers can have severe negative economic impacts on businesses and ultimately on complex value chains and trade.

In terms of *volume*, global cross-border data flows were estimated to have exceeded 700 terabytes per second (Tbps) in 2017 (see **Figure 4.1**), which means that since 2007 they increased by a factor of 64 (McKinsey Global Institute, 2019).²⁵ Of course, volume (e.g. measured in bytes) is not necessarily a good indicator of the value of data, i.e. the information that it encodes, which is related to the potential for current or future monetisation (see also Section 2.2). A previous study estimates that since 2015 the value of cross-border data flows has exceeded the value of cross-border merchandise trade (McKinsey Global Institute, 2016). However, the challenge remains in defining what metrics are appropriate for assessing the actual ‘value-added’ of such flows (e.g. per GB of data transferred) as this can differ vastly across firms and sectors depending on both the informational content of the data itself and the analytics applied to it (among other factors). The location of economic value creation is also the basis for national-level indicators such as GDP, and estimates of productivity, meaning there is a clear need to analyse data value chains in detail for policy-making processes.

It is estimated that in 2014 the international flow of data added USD 2.8 trillion to the global economy and this is expected to grow to USD 11 trillion by 2025 (McKinsey Global Institute, 2016).²⁶ The key question for measuring the *value* of cross-border data flows is related to how businesses actually derive value from data. As discussed in Section 2, data can either be used to enhance or augment existing business functions, or to enable entirely new businesses models and revenue streams. This distinction also has a cross-border dimension that can further complicate the way data flows can be measured as data can be transferred across national jurisdictions.

In many cases it is possible that businesses aggregate data in locations (e.g. headquarters or data analytics centre), which are not located in the same country where the original data is collected (see Global Data Value Chain, **Figure 2.1**). For example, consider a set of data points that are volunteered by, and collected from, the users of an online social network free of charge and, hence, they do not generate any financial transactions in the country where the user is based. However, once those data points are transferred and aggregated with millions of other data items from across the globe, they become the basis for data analytics and thus for value creation. Eventually, they are monetised by the provision of data-based services (e.g. in the form of targeted advertising) or by licensing access to the database. This potentially can have large implications for taxation of multinational enterprises.²⁷

Figure 4.1. Total used cross-border bandwidth, 2005-2017



Source: Telegeography; McKinsey Global Institute analysis, 2019

Nowadays, cross-border transfers of data underpin virtually all business relations in international trade, international investment and global supply chains. Data transfers are also fundamental in enabling business operations of multinational companies. This includes coordinating human resource and R&D operations across foreign affiliates, often using cloud-based solutions (Swedish National Board of Trade, 2014). In addition, the initiation and completion of cross-border transactions between buyers and sellers typically involves transferring some data, e.g. regarding contracts or specifications of orders, packaging and delivery.

Similar to intermediate inputs in the 'analogue' value chain, the data value chain may require data to be transferred across borders multiple times, or even on an on-going (real-time) basis. For example, multinational enterprises owning plants in various countries are often integrated in the same production process and need to constantly monitor and communicate their production volumes.

4.2. The scale and scope of cross-border data flows

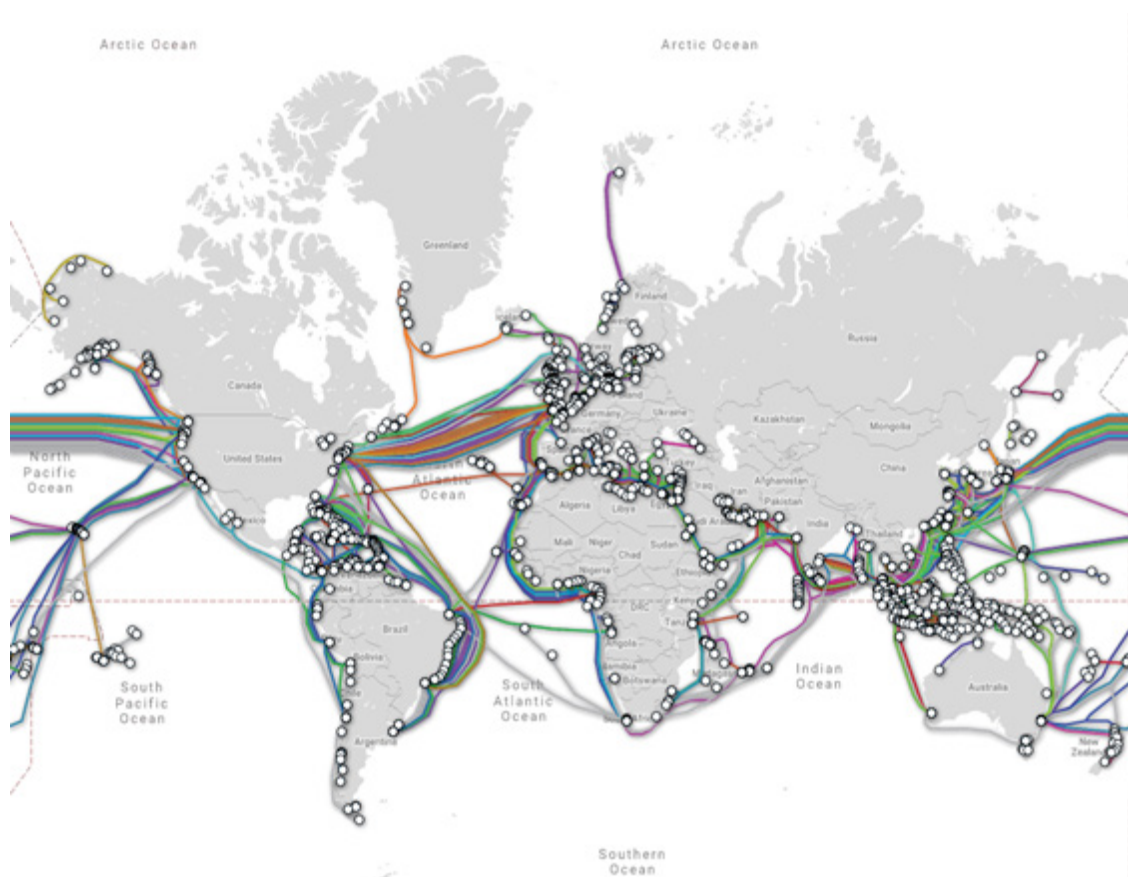
Data transferred across borders are not fundamentally different from data transferred within countries. However, latency rates can be higher for data packages sent across longer distances, which can affect their value by making data less accessible, less timely, less frequent, less reliable, etc.²⁸ Latency can also increase costs of data analytics if data have to be locally stored in multiple locations. Data flows across borders can take place within businesses, between businesses (B2B), between businesses and consumers (B2C), and between machines (M2M). The latter is driving increasing amounts of global Internet traffic based in the IoT. Cisco (2017) estimates that by 2021 there will be 11.6 billion connected devices, including M2M devices, and that the traffic they generate will increase by 70% annually between 2016 and 2021. In **Box 2.1**, several examples were given of businesses making use of cross-border data flows in their day-to-day activities.

Many manufacturing companies monitor the status, performance and condition of their machines in different locations. For example, Volkswagen and AWS (Amazon Web Services) announced the co-development of the ‘industrial cloud’ in March 2019 with the aim of connecting “*data from all machines, plants and systems in all factories*” (Volkswagen, s.d.). This is enabled by sensors using cellular or satellite connectivity to send signals. Real-time data is then aggregated at a global level and with the potential to be monetised via a new service. In another example, reported in Pepper and Garrity (2016), a General Electric jet plane turbine generates around 400 GB of data per day, that could potentially be used to diagnose problems or optimise efficiency in real-time.

Similarly, many online platforms offer their services globally, relying on cross-border data flows to deliver digital matching services (e.g. Uber, Airbnb, eBay). At the same time, they also collect transaction and consumer behaviour data in various locations, which further need to be transferred across borders in order to be stored, aggregated and analysed. Finally, insights based on aggregated global data serve as the basis for commercial services that can be delivered in multiple locations (e.g. targeted advertising, or demand forecasting, price elasticities of consumers).

At least at an inter-continental level, the bulk of data is transferred via submarine cables, making their use a useful indicator for the volume of cross-border data flows. A second indicator of the scale of global data infrastructure are large data centres that enable the storage of data, as well as remote computing via the Internet (cloud computing).

The installed capacity of submarine cables can provide a global view on which markets are most integrated in terms of data connectivity. In 2017, there were around 428 submarine cables in service, with a total length of around 1.1 million kilometres (TeleGeography Blog, February 2017). **Figure 4.2** reveals that some parts of the globe are much more connected than others. Not surprisingly, the trans-Atlantic route between the East Coast of the US and Europe, and trans-Pacific from the West Coast of the US to East Asia are especially well-served with capacity. However, there is significant and increasing capacity in South-East Asia as well.²⁹

Figure 4.2. Global Submarine Cable Map, 2018

Note: White dots represent cable landing points

Source: Submarine Cable Map, TeleGeography (TeleGeography), www.submarinecablemap.com, licensed under CC BY-NC-SA 3.0 (creative commons).

Data centres are another important component of the global data infrastructure. They consist of servers that can be used exclusively by a firm (often in the form of private cloud or co-location services) or rented on demand from cloud service providers such as AWS or Microsoft Azure (public cloud), or anything in-between (hybrid cloud). By 2020, global IT infrastructure spending on traditional on-premises data centres is predicted to be surpassed by spending on off-premises private and public cloud (IDC, s.d.). By some estimates there were around 8.6 million data centres globally in 2017, though this number is likely going to decline in the future due to ongoing centralisation in larger data centres (Cushman and Wakefield, 2016).

International bandwidth usage is increasingly shifting towards content providers such as Amazon, Google, Facebook, and Microsoft. In the past few years, their share of international bandwidth usage has risen sharply to reach 54% in 2018, a similar share to traditional Internet backbone providers (Mauldin, 2017). To ensure they can meet increasing demand for their services, content providers themselves have become large players in the development of global data infrastructure, including the construction of submarine cables and data centres (Mauldin, 2017).

4.3. Barriers to cross-border data flows

Governments across the world face various challenges to their existing legal doctrines brought about by digitalisation and have various reasons for imposing restrictions to cross-border data flows. These are often related to data protection and privacy considerations (see **Box 4.1** for a more detailed overview) and their existence can offer ways of implicitly valuing the targeted data and data flows. In simple terms, lower economic costs to restricting them would suggest a lower value of the data flows. Note that while these costs could be considerable, there could be political reasons that would justify incurring them.

Research shows that the number of regulations on data flows across national borders has been increasing rapidly in recent years (Casalini and Lopez Gonzalez, 2019; Information Technology and Innovation Foundation, 2017). Ferracane (2017) reports that in 2017 the total number of restrictions reached 87 across 64 economies³⁰, of which half were put in place in countries located in the Asia-Pacific region. Casalini and Lopez Gonzales (2019) estimate that as of 2019 the total count of data regulations, including different types of regulations relating to data transfers and local storage requirements, is above 200.

Studies have estimated that the economic costs associated with restricted international data flows can be large. This suggests that they are highly valuable.³¹ For example, the US International Trade Commission (2014) estimates that US GDP would be 0.1 to 0.3% higher if foreign digital trade barriers were removed. Similarly, for the EU, barriers to data flows are estimated to reduce GDP by 0.4 to 1.1%, depending on the strength of data localisation requirements (ECIPE, 2014). Another study finds that data regulations lead to a 0.48% reduction of real GDP in the EU (Bauer, Ferracane, and van der Marel, 2016).

Box 4.1. Types of barriers to cross-border data flows

- Local storage and local processing regulations (i.e. the requirement to keep and/or process data on servers located within a given country).
- Data protection regulation (i.e. laws governing the collection, use and transfer of personal data. The most comprehensive example is GDPR in the European Union, which has been in force since May 2018).
- Competition and antitrust law adapted to digital markets (i.e. a set of economic policies that are designed to favour the exporting conditions faced by digital, data-enabled enterprises of a particular nationality, e.g. EU Parliament voting for the legal breaking up of Google operations in the EU).
- Cybersecurity (i.e. a collection of technologies, processes and controls designed to protect systems, networks and data from an unauthorised exploitation, e.g. EU working towards the introduction of a certification process for IoT devices to increase their cybersecurity).
- Intellectual property rights (e.g. on digital content such as music, movies and books).
- Restrictions on Internet use, censorship and blocks against data transfers.

Sources: ECIPE (2014), Swedish National Board of Trade (2014), US Department of Commerce (2016), Bauer, Ferracane, and van der Marel (2016), ITIF (2017), Meltzer and Lovelock (2018).

Others study cloud computing services, as a key advantage of the cloud is that services can be accessed remotely from anywhere as long as users can connect to the Internet. This means restrictions to cross-border data flows potentially have severe negative consequences for access to these services, which can be used to store or process data, access software, or host platforms – usually at much cheaper costs than on-premises IT equipment. Berry and Reisman (2012) conclude that in 2010 cloud services contributed around USD 1.5 billion to US exports and USD 1.4 billion to the sales of US foreign affiliates. In a study on the UK, Coyle and Nguyen (2018) calculate that cloud computing potentially added USD 2.8 billion to UK exports in 2017.

4.4. Measurement challenges arising from cross-border data flows

The US Department of Commerce (2016) quotes a number of challenges that impede the measurement of cross-border data flows. This includes the fact that there is limited evidence on how firms actually use cross-border data flows to generate value, but also the lack of a standard nomenclature as well as official data on cross-border data flows. The reports stresses that existing studies are infrequent, not transparent enough on their methods and often designed to capture only ‘tech’ sectors. It is clear that these fundamental roadblocks need to be resolved as a matter of urgency to enable robust and coherent measurement of the economic value of cross-border data flows.

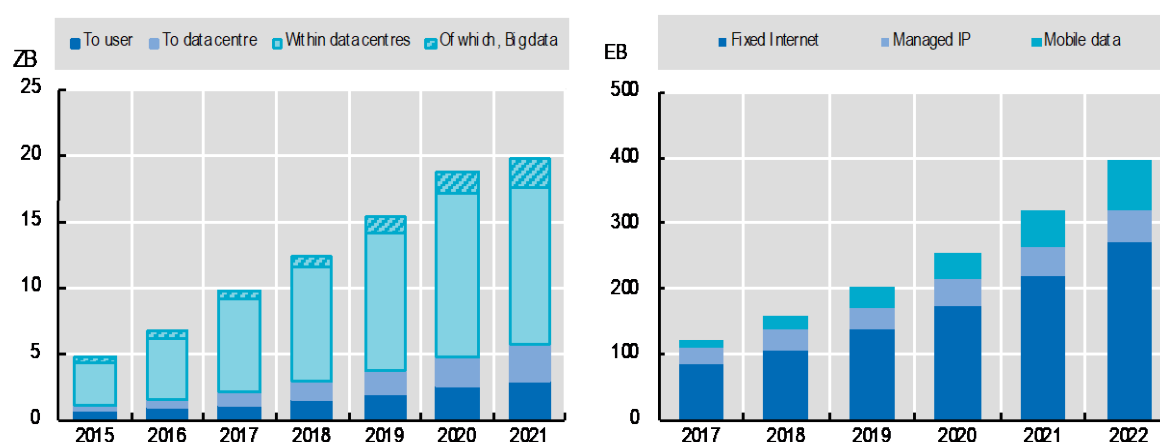
Some aspects of cross-border data flows are already measured. For instance, Ferencz (2019) introduces the OECD Digital Services Trade Restrictiveness Index (DSTRI), which builds on the OECD Services Trade Restrictiveness Index and provides information on the barriers companies face while supplying services using electronic networks. DSTRI indices are available for 44 countries and quantify the impediments that affect any services traded digitally in the main five categories: (1) Infrastructure and connectivity³², (2) Electronic transactions, (3) Payment systems, (4) Intellectual property rights, and (5) Other barriers affecting trade in digitally enabled services. In a similar vein, Ferracane et al. (2018) have developed the Digital Trade Restrictiveness Index (DTRI) that measures the restrictiveness of digital trade for 64 economies. A sub-category of DTRI captures the intensity of restrictions to cross-border data transfers, which can be applied within standard gravity models in order to quantify the impact on trade flows. This information is of high relevance as international trade agreements are starting to include clauses governing data protection and data flows (e.g. the EU – Japan Economic Partnership)³³.

The drawback of this type of study is that it is more difficult to account for heterogeneity of various data types and data monetisation strategies that are needed to map value to industry sectors.

It is clear that approximating the value of data flows between countries by the volume of data has significant limitations.³⁴ This follows directly from our description of different data monetisation strategies and business models (see Section 2). The value of data depends on the information content it carries as well as on how it is used now or intended to be used in the future.³⁵ Another reason is that the bulk of data traffic is expected to take place within data centres (see **Figure 4.3**; Cisco, 2018), which is further complicated in the case of cross-border flows. For example, two businesses might exchange data within a data centre in a third country and hence no cross-border flow might be recorded, even though a database might have moved across borders in the legal sense. Therefore, it seems that looking at sheer data volumes or installed bandwidth (as shown in **Figure 4.2**) can be a misleading when analysing the associated value of these flows.³⁶

Another approach would be to look in more detail into what types of data are being transferred, and how their distribution differs across industries and sectors. For example, Cisco (2018) report global data centre traffic broadly divided into ‘consumer’ and ‘business’, with further sub-divisions by type of application (see **Figure 4.3**).³⁷ Such detailed data could be analysed based on the data taxonomy developed in Section 3, and help to classify which types of data flows are high- or low value-added. If broken down by sector and related to types of barriers this could prove to be highly valuable for policymakers. For example, it could be that both automotive manufacturing and motion picture production are characterised by large volumes of data traffic, though the value added in films might be comparably low.

Figure 4.3. Global data centre traffic, by type and Consumer Internet Protocol (IP) traffic, by sub-segment, 2015-22



Note: Zettabytes per year (left-hand panel) and Exabytes per month (right-hand panel). “To data centre” refers to traffic flowing from one data centre to another, for example, moving data between clouds, or copying content to multiple data centres as part of a content distribution network. “To user” refers to traffic that flows from the data centre to end users through, for example, streaming video to a mobile device or PC. “Within data centres” refers to traffic that remains within a data centre, for example, moving data from a development environment to a production environment within a data centre, or writing data to a storage array.

Source: OECD (2019a), *Measuring the Digital Transformation*, calculations based on Cisco Global Cloud Index 2016-2021 and Cisco Visual Networking Index 2017-2022, January 2019.

One important challenge is posed by the transfer of data that takes place within multinational enterprises (van der Marel, 2015). In some sense, this is similar to measurement issues related to firms’ internal transactions of intangible inputs such as R&D (UNECE, 2015, Rugman & Eden, 1985). A review of that literature might offer additional insights. In terms of measurement, it would be relevant to look into networks of affiliates within the same business group, as cross-border coordination is often driven by offshoring motives.

Lastly, researchers should make use of new and experimental data sources. The Internet itself can provide new opportunities for measurement of some aspects of global data flows that are beyond statistical surveys. This includes methods of web scraping, IoT search engines, and sensor-based data generated by “smart-meters”.

5. Approaches to measuring the economic value of data

5.1. What aspects of data are already measured?

Although the measurement of the value of data still remains a challenge, we are able to measure some aspects of data and data flows already. For example, some national statistical organisations³⁸ conduct surveys covering the use of the Internet and digital technology (e.g. the Community survey on ICT usage and e-commerce in enterprises conducted by EU member states, or the Survey of digital technology and Internet use in Canada). These surveys are designed to measure the use of digital technologies by businesses. However, they do not directly measure the intensity of use of data or digital technologies by businesses, which could be added to existing surveys relatively easily. For example, regarding the use of cloud computing services, Eurostat currently only requires a “Yes/No” question. It should be noted however that such data still provide valuable insights. The OECD has coordinated experimental analysis using these data for factor and cluster analysis on the complete set of the digital technology-related questions in ICT surveys from four countries (including Italy, Poland, Sweden, and the United Kingdom) with the aim of the measuring of the digital maturity of businesses (OECD, 2019a). Furthermore, this analysis will be extended to look at the adoption and impact of data-intensive technologies including cloud computing services and data analytics.

Regarding other statistical sources, Ker (forthcoming, [DSTI/CDEP/MADE(2019)6]) reviews what relevant breakdowns are prescribed in standard classifications of both products and industries, finding various products related to data storage, management, and processing as well as identifying several industries where firms primarily engaged in compiling and monetising databases are classified. However, data availability for such detailed product and industry classes is found to be very limited in practice.

Data centres, being an important component of the global data infrastructure, serve also as a potential source for informative statistics. Cisco (2018) presents some forecasts on data transfers by origin and destination, distinguishing the following flows:

- From data centre to user (i.e. traffic that flows from the data centre to end users through the Internet or IP Wide Area Network);
- Flows of data from data centre to data centre;
- Flows within data centres (i.e. traffic that remains within the data centre, though excluding traffic within the same rack of servers).

As shown in **Figure 4.3**, it is estimated that by 2021 over 70% data flows will take place within individual data centres. Interestingly, the report states that the traffic between data centres is growing faster than either traffic to end-users or traffic within data centres. The reason given is “*the increasing prevalence of content distribution networks, the proliferation of cloud services and the need to transfer data between clouds, and the growing volume of data that needs to be replicated across data centres*” (Cisco 2018). We think that working with the providers of data infrastructure represents a unique opportunity to gain vital insights into the use of data by businesses.

5.2. Review of approaches to measuring the economic value of data

This section provides an overview of various approaches to measuring the value of data with a focus on how businesses use data as an asset in their production process to generate revenue. Most data are not transacted on traditional markets where we can observe, and measure associated prices. As a consequence, the use of market (or: market-equivalent) prices is often not possible and alternative approaches need to be found. In this section we review some of general existing valuation methods that were developed to e.g. measure intangible assets but also propose more experimental methods with the aim to stimulate further discussion. Where possible we comment on feasibility and strength or weaknesses of approaches. We should also note that there are no off-the-shelf methods to measure the value of data and much more research on this topic is needed.

5.2.1. Data valuation based on market prices

The value of a good or service is often identified with the market equilibrium price, i.e. the meeting point between demand and supply. However, there are various challenges that make this approach difficult to apply to data. First, we need the existence of a well-defined market. While this might exist for some data (e.g. on marketing preferences of individuals), it could prove to be difficult to find for many other types of data, often for good reasons. In addition, as databases are traded relatively infrequently, transaction-based valuation may often rely on obsolete information (similar to housing stock).

Secondly, since the value of data is highly context-dependent, the same dataset could be valued differently across different data suppliers, users and regulators. As noted by OECD (2013a): “[...] *economic experiments and surveys in the United States indicate that individuals are willing to reveal their social security numbers for USD 240 on average, the same data sets can be obtained for less than USD 10 from U.S. data brokers such as Pallorium and LexisNexis*”. In addition, the ‘true’ value of data may not be known to the seller and hence the market price could be misleading (i.e. due to asymmetric information).

Nevertheless, studying data brokers offers an interesting avenue for data valuation based on market prices, as they typically value their own datasets by quoting detailed, itemised price lists for compiled databases. Moreover, this approach is in line with 2008 System of National Accounts as “*databases for sale should be valued at their market price, which includes the value of the information content*” (OECD, 2009: p.120). However, the identification of data brokering firms themselves may be problematic, as they are not classified to a single class in standard industry classifications (OECD, 2013b).

Data brokerage companies typically collect information from publicly available personal records and then aggregate, store and sell it to different customers (described in detail in Gellman and Dixon, 2013). Their business models range from specialised business-to-business services offering information such as background criminal record checks of individuals or detailed information on potential business partners or competitors (e.g. Experian, LexisNexis) to other business models such individual localisation services (e.g. Locate Plus, LocatePeople.org).³⁹

The insurance market could offer another potential avenue for deriving the value of data held by businesses, since insurance companies need to assess the value of data held in-house or in data centres. Insurance against data breaches is becoming more important given the strong reliance on data by an increasing number of businesses. Therefore, the pricing of (data) insurance as well as (possibly) negotiations around insurance claims in case of

data breaches could both potentially be informative (although the latter could also include litigation payments or legal costs and hence go beyond the value of data per se).

Related to the insurance of data itself is the notion of insuring continuous service delivery against failures of enabling IT infrastructure, including data centres. Interruptions of service provision to the end-user can entail significant costs to businesses (loss of revenues as well as reputation). Since under certain circumstances, the infrastructure providers might be liable for such losses, they take out insurance to protect themselves against such eventualities (Data Center Knowledge, July 2016). There are a number of factors shaping these estimations, including third party liability and first part losses.⁴⁰

In a recent study sponsored by IBM (IBM, 2019), the Ponemon Institute (2018) conducted interviews with more than 2,200 IT, data protection, and compliance professionals from 477 companies that have experienced a data breach between July 2017 and July 2018. They found that the average cost of data breaches in this period was USD 3.86 million. While data breaches entails all kinds of costs this can at least serve as an indication. According to another survey of 1,800 global business decision makers conducted by NTT Security (NTT, 2018), data breaches entail significant costs to businesses, including:

- Loss in customer confidence, damage to reputation and brand, financial losses;
- Expected revenue drops in excess of 10% on average;
- Costs of recovery, estimated to be around USD 1.5 million on average;
- Costs of cybersecurity insurance, which 38% of organisations have in place.

This section discussed a number of approaches that should be explored to measure the business value of data at market prices. Specifically, we proposed to look at data brokers and insurance for data breaches related to the data themselves or the data infrastructure. The exact application of these experimental sources of information to the valuation of data in production process would necessarily have to be further developed and its limitations would need to be addressed (e.g. separation of ‘data value’ from the costs of legal process or disentangling the contribution of data to output valuation in case of insurance information). Nevertheless, we find these approaches as interesting options. The following sections look at other approaches that may be useful when considering the value of data and data flows.

5.2.2. Data as a knowledge-based asset

The intangible assets literature holds a number of insights when thinking about measuring the value of data, as databases and software are an existing category of intangible assets.⁴¹ Corrado et al. (2005) outline a ‘best practice’ method for measuring intangible investment in the National Accounts framework, distinguishing digitised information, innovative property, and economic competences/organisational capital. This can be useful when thinking about data as an knowledge-based intangible asset that contributes to the production process of firms and previous reports have called for measuring the value of digitised data as an intangible asset (OECD, 2014).

To measure intangible capital in the national accounts, surveys distributed by statistical offices ask companies about their purchases, disposals and own-account investment in databases and software. For example, the UK Office for National Statistics uses the Quarterly Acquisitions and Disposals of Capital Assets Survey (QCAS) to collect information on the value of capital assets bought and sold by businesses in the private sector. However, the measurement of the value of data in these surveys suffers from a

number of problems. Firstly, response rates on databases tend to be relatively low.⁴² Secondly, own-account databases are typically recorded as part of own-account software, making it difficult to single out the value of data itself.⁴³ Thirdly, at the industry-level it is particularly difficult to gauge expenditures on the creation of databases, as data are to some degree produced in virtually all sectors of the economy.⁴⁴

As a consequence, direct measures of investment in databases are often inadequate, and therefore, statistical offices often rely on a cost-based approach. The cost-based method uses labour market data, for the number of computer software developers (or other data-related occupations such as economic researchers), their average hourly costs, combined with assumptions on time-use, non-wage labour costs, and overheads, in order to estimate the values of databases.⁴⁵ This approach has been explored recently by Statistics Canada (2019) based on 8 occupational groups. Their calculations assume that some occupational groups spend a higher share of their time producing data. For example, while ‘data entry clerks’ are assumed to spend 100% of their time on producing data, it is only 20-30% for ‘economists and economic policy researchers and analysts’. Following this experimental method, Statistics Canada estimates that the total investment in data amounted to CA\$9.4 billion to CA\$14.2 billion 2018.

It is also important to note that although databases qualify as an asset under the 2008 System of National Accounts (SNA 2008)⁴⁶, the data they hold are explicitly excluded.⁴⁷ Ahmad and van de Ven (2018: p.4) discuss in more detail how based on SNA 2008, “databases should reflect only the value of the underlying database management systems and the costs associated with the digitisation of data”⁴⁸. This recommendation reflected the view that the underlying value (information content) associated with the data itself was *de facto* a non-produced asset, with outright purchases of databases that included the intrinsic value of the underlying data recognised in the accounts as goodwill”. However, this approach can pose challenges in practice. When databases (and the underlying value of data) are acquired for example in taking over another company, the transfer is captured as ‘goodwill’, and, so in-line with the implicit treatment that data are non-produced.⁴⁹ However, in practice, when databases are acquired separately the transactions may be recorded as Gross Fixed Capital Formation at the market price of the acquisition (including the value of the underlying data). Discussions are on-going in the context of the update of the 2008 SNA to assess whether this needs to be addressed and, in particular, how to make data more visible in the accounts.

Under some conditions, data also qualify as an asset under the International Financial Reporting Standards rules⁵⁰, yet it should be stressed that from an individual company perspective intangible investment and data assets are likely to be largely underreported (Tax Adviser, May 2018).

Box 5.1. Bankruptcy case studies highlighting the value of business data

Case 1: The bankruptcy of American Caesars Entertainment Operating Corp. Inc, 2015

During “chapter 11” bankruptcy proceedings, some creditors argued that Caesars’ customer loyalty program “Total Rewards” held data worth USD 1 billion. In support of this assessment, a report by a bankruptcy court examiner did note instances where sold-off Caesars properties suffered a decline in earnings due to the loss of access to customer analytics based on the Total Rewards database. However, it was also observed that it might prove difficult to sell the Total Rewards data, e.g. for incorporation into another company’s

loyalty program. As Short and Todd (2017) note: “Although the Total Rewards system was Caesars’ most valuable asset, its value to an outside party was an open question”.

Case 2: Bankruptcy of RadioShack, 2015

When the American company RadioShack filed for bankruptcy protection in 2015, the most valuable asset that was put forward for bidding by the company’s creditors was a database with records of purchases of roughly 67 million RadioShack customers, along with their addresses.

Source: Short and Todd (2017).

A related approach based on the market valuation of data is to exploit information revealed in mergers and acquisitions (M&A), which can potentially be motivated by access to data (e.g. “*data-seeking foreign direct investment*”). This is feasible since the legal process requires the acquired company to assign a value to its assets, which includes intangible assets such as databases and software. Anecdotal evidence suggests that the number of M&A transactions that are motivated by data access has been increasing (HuffPost, February 2017). Some deals that were at least partially motivated by data access include Microsoft’s acquisition of LinkedIn in 2016 and purchase of GitHub in 2018. Some Central Banks or National Statistics Offices keep track of such activities and could add new questions into their existing forms or surveys relatively easily.

Bankruptcy cases provide another opportunity to explore the value of business data as negotiated in court. We provide some examples of high-profile court cases in **Box 5.1**. Future research could look into what valuation methods are used by courts and whether they could be used in other settings as well.

5.2.3. Revenue-based valuations and the ‘data value cycle’

In Section 2 we proposed a stylised taxonomy of different business models that are either data-enhanced or data-enabled. Based on Figure 2.4 we suggested to explore the share of revenue that is driven by data monetisation across different types of firms (e.g. manufacturers, utility providers, banks, or online platforms). In practice this would involve building up a repertoire of case studies across a number of businesses in different sectors that in turn could be generalised to a sector. This should be supported by business surveys and periodic revisions to keep shares up to date. In terms of theory this approach has could be integrated with the notion of treating data as a knowledge-based asset as discussed in Section 5.2.2. We believe this approach is feasible in practice.

Digital platforms are a key example of data-enabled businesses (see also Section 2.3) and are disrupting a large number of industries, including transport, accommodation, retail, and travel services. Li et al. (2019) propose a classification of online platforms⁵¹ based on eight major types, related to their underlying business models. For their empirical analysis, they select key companies for each platform type to estimate a range for the economic value of their data. Focusing on the ‘data value chain’, they show how online platforms create value from data in stages, and how the value derived at each stage varies. Their approach recognises two sources of data monetisation inferred from market transactions as well as Selling, General and Administrative Expenses⁵² as a proxy for a firm’s investment into a data-driven business model. Then, an approach similar to R&D capitalisation is applied to infer the value of the underlying data.

Li et al. (2019) find that online platforms can vary in their degree of vertical integration in the data value chain, and that this variation can determine the means and effectiveness of their data monetisation. For example, Amazon's Marketplace value of data is estimated to be USD 125 billion which accounts for 16% of the company's current market value. Overall, their study suggests large benefits from vertical integration within the data value chain, and online platforms that are also data holders are best placed to exploit these.

One interesting example of a data-enabled business is Citymapper, which has been experimenting with different ways to monetise the travel and movement data it has on individuals using its app. While it discontinued a dedicated bus service in London, it started offering the *Citymapper Pass* at prices which undercut the official Transport for London prices for weekly travelcards by £4.10 in 2019. A crude calculation would suggest that the business hence values annual travel and movement data for an individual at around $52 * £4.10 = £213.20$ (approximately 270 USD). It would be interesting and relatively straightforward to seek similar examples for different types of data.

5.2.4. Other approaches to trace value of data flows

To better understand and trace the value associated with data flows it could be worth to exploit variation found in input-output tables. As the OECD's "*Measuring the Digital Transformation*" measurement roadmap suggests, "*Superimposing Input-Output tables with data-flow tables [could be used] to assess whether flows of value added are accompanied by flows of data.*" (OECD, 2019a). Research by the European Commission supports this idea as most data are "*intermediary goods that are used in production processes by other parties*" (Duch-Brown et al., 2017: p. 28). It is fair to note that the feasibility of this approach depends on solving some of the conceptual and measurement challenges related to data as mentioned above. For example, one would need at least to have a clear idea about the value-added by sector that is enabled by data assets in order to track how this value is flowing across sectors. Hence, we see this as approach as less feasible in the short-run.

This approach would capture the reliance of different industry sectors on inputs from data-intensive industries; both domestic and foreign (see van der Marel, 2015). As the level of reliance differs across sectors, input-output relations would provide a picture of the relative importance of data inputs. In practical terms, the difficulty will be to actually assess which sectors are providing data inputs to others, where, and to what degree. While it appears easier to assign the output of some specific sectors to these activities (e.g. 63.1 – "Data processing, hosting and related activities; web portals"), it would require considerable research to figure out the appropriate coefficients other sectors. We note that this is similar to some approaches and ideas discussed in the context of digital trade (e.g. López González and Jouanjean, 2017).

6. Discussion and implications for measurement agenda

What is the purpose of measuring the value of data? In 1987 Robert Solow famously stated that “*You can see the computer age everywhere but in the productivity statistics*” (Solow, July 1987) – those words could be paraphrased today in the context of the seeing data everywhere but in business balance sheets. It is clear that a better understanding of how data contributes to adding value and raising productivity, in order to inform policymakers on a wide range of areas where the impact of data-driven innovations is inevitable. But, crucially, this requires consistent and reliable statistics that are able to capture the complexity of data uses in the modern economy.

The development of methods for measuring the value of data is not an easy task and much more research on the topic is vital. Also, both research and statistical communities would benefit from more systematic thinking about this measurement challenge and available options. As noted by ESCoE-OECD November 2018 workshop participants, some of the issues raised in this paper might resemble the measurement challenges related to R&D activities, software and intangible assets more generally. This also means we do not have to start from scratch but can use methods and concepts tested and applied before. However, it is important to stress that the challenges we are faced with when discussing the value of data often go well beyond those surmounted when measuring R&D and software.

However, some characteristics of data could make it a special sort of intangible asset: some data does not always become obsolete quickly⁵³ and it can be continuously combined with other data to create new value. Other types of data could be ‘worthless’ in a matter of seconds. This is further complicated by the fact that the value of data is highly related to context and dependent on its use within businesses. This paper discussed some of the data characteristics that one would need to consider when looking into this issue further.

When it comes to valuing data and data flows, various avenues for further investigation have been highlighted. We foresee the approach of valuing data based on a business model perspective as the most comprehensive and therefore as the most promising (e.g. Li at al., 2019). Nevertheless, more effort could be made towards a search for settings where market valuations of data can be observed, with an aim to identify businesses of interest and aggregate this information across industries. Here, information on data pricing from data brokering agencies could prove helpful. Also, in the short term it seems feasible to derive the value of data from data breaches, data insurance and the data-motivated Mergers & Acquisitions – at least for some sectors.

It is certainly helpful (and feasible) to introduce new questions to existing business surveys, which would allow both for a more precise measurement within currently adopted methods in the national accounts and for a better understanding of different areas of data-driven businesses activity across sectors. This could include the measurement of data licenses that are in use across sectors, or how the demand for people doing data-related work within a firm and sector changes over time. In the long term, an update of the Standard Occupational Classification (SOC) to account for “new” data-related occupations should be explored. More systematic work on supplementary National Accounts estimates (e.g. in the context of Digital Supply-Use Tables [[SDD/CSSP/WPNA\(2019\)1](#)]) need to be explored further. Lastly, also a more systematic approach to analyse specific business models and data monetisation strategies will be unavoidable in the longer term.

This paper has highlighted a number of conceptual and empirical challenges regarding the measurement of the value of data and data flows and proposed a number of approaches that should be explored further to start accounting for this increasingly important resource in modern economies. While the challenges seem daunting, we argue to put the business model at the centre of any method of valuation.

References

- Ahmad, N. and Schreyer, P. (2016), “*Measuring GDP in a Digitalised Economy*”. OECD Statistics Working Paper 73, OECD Publishing: Paris.
- Ahmad, N. and van de Ven, P. (2018), “Recording and measuring data in the System of National Accounts”. Paper prepared for the Meeting of the Informal Advisory Group on measuring GDP in a digitalised economy, 9 November 2018, OECD.
- Bauer, M., Lee-Makiyama, H., van der Marel, E., and Vershelde, B. (2014), “The costs of data localisation: Friendly fire on economic recovery”, ECIPE Occasional Paper No. 3/2014.
- Bauer, M., Ferracane, M.F., van der Marel, E. (2016), “Tracing the Economic Impact of Regulations on the Free Flow of Data and Data Localization”, Global Commission on Internet Governance (GCIP) Series paper No. 30, May 2016.
- Bloom, N., Schankerman, M., and Van Reenen, J. (2013), “Identifying Technology Spillovers and Product Market Rivalry”, *Econometrica*, 81(4): 1347-1393.
- Brynjolfsson, E., Collis, A. and Eggers, F. (2019), “Using massive online choice experiments to measure changes in well-being”, *Proceedings of the National Academy of Sciences*, 116(15): 7250-7255.
- Brynjolfsson, E. and K. McElheran (2016), "Data in Action: Data-Driven Decision Making in U.S. Manufacturing", Working Paper 16-06, Center for Economic Studies, U.S. Census Bureau.
- Business Roundtable (2015), “Putting Data to Work: Maximizing the Value of Information in an Interconnected World”, Business Roundtable, January 2015.
- Casalini, F. and J. López González (2019), “Trade and Cross-Border Data Flows”, OECD Trade Policy Papers, No. 220, OECD Publishing, Paris.
- CIGI (2018), “*A National Data Strategy for Canada: Key Elements and Policy Considerations*”, CIGI Paper No. 160, Centre for International Governance Innovation, February 2018.
- Cisco (2017), “*Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021*”, White Paper, Cisco, February 2017.
- Cisco (2018), “*Cisco Global Cloud Index: Forecast and Methodology, 2016–2021.*” White Paper, Cisco, November 2018.
- Cisco (2019), “*Cisco Visual Networking Index: Forecast and Trends, 2017–2022*”, White Paper, Cisco, February 2019
- Corrado, C., Hulten, C. and D. Sichel (2005): "*Measuring Capital and Technology: An Expanded Framework*", NBER Chapters, in: *Measuring Capital in the New Economy*, pages 11-46, National Bureau of Economic Research, Inc.
- Coyle, D. (2017): “*Do-it-yourself digital: the production boundary and the productivity puzzle*”, ESCoE Discussion Paper 2017-01, Economic Statistics Centre of Excellence, January 2017.

- Coyle, D. and Nguyen, D. (2020), “Valuing Goods Online and Offline: The Impact of Covid-19”, CEPR Covid Economics, Issue 33, 30 June 2020.
- Coyle, D. and Nguyen, D. (2018), “*Cloud Computing and National Accounting*”, ESCoE Discussion Paper 2018-19, Economic Statistics Centre of Excellence, December 2018.
- Cushman & Wakefield (2016), “*Data Centre Risk Index 2016: Informing Global Location Strategies in a Digital World Expanding at a Phenomenal Pace*”, Cushman & Wakefield, July 2016.
- Cyphers, B. and Gebhart, G. (2019): “*Behind the One-Way Mirror: A Deep Dive Into the Technology of Corporate Surveillance*”, Electronic Frontier Foundation, December 2019.
- Deloitte, (2017): “*Assessing the value of TfL's open data and digital partnerships*”, Deloitte, July 2017.
- Duch-Brown, N., Martins, B. and Mueller-Langer, F. (2017). “*The economics of ownership, access and trade in digital data*”, JRC Technical Reports: JRC Digital Economy Working Paper 2017-01, European Commission.
- European Commission (2015). Public Consultation on the regulatory environment for platforms, online intermediaries, data and cloud computing and the collaborative economy, Brussels.
- European Commission (2017). Exchanging and Protecting Personal Data in a Globalised World. Communication from the Commission to Parliament and the Council, Brussels.
- Ferencz, J. (2019), “The OECD Digital Services Trade Restrictiveness Index”, *OECD Trade Policy Papers*, No. 221, OECD Publishing, Paris.
- Ferracane, M.F. (2017). Restrictions to Cross-Border data flows: a Taxonomy, ECIPE Working Papers No. 1/2017.
- Ferracane, M.F., Lee-Makiyama, H. and van der Marel, E. (2018). *Digital Trade Restrictiveness Index*. ECIPE Report April 2018.
- Gartner (2011): “Gartner Says Solving “Big Data” Challenge Involves More than Just Managing Volumes of Data”, Press release.
- Gellman, R. and Dixon, G. (2013), “*Data Brokers and the Federal Government: A New Front in the Battle for Privacy Opens*”, World Privacy Forum Report, October 2013.
- Hartmann, P.M., M. Zaki, N. Feldmann, and A. Neely (2014): “Big Data for Big Businesses? A Taxonomy of Data-driven business models used by start-up firms”, Cambridge Service Alliance working paper.
- Hofheinz, P. and Mandel, M. (2015). Uncovering the Hidden Value of Digital Trade: Towards a 21st Century Agenda of Transatlantic Prosperity. Interactive Policy Brief Issue 19/2015. The Lisbon Council and Progressive Policy Institute.
- Hofheinz, P. and Osimo, D. (2017). Making Europe a Data Economy: A New Framework for Free Movement of Data in the Digital Age. Policy Brief, the Lisbon Council.

- Hughes-Cromwick, Ellen, and Julia Coronado (2019): "The Value of US Government Data to US Business Decisions." *Journal of Economic Perspectives*, 33 (1): 131-46.
- ITIF (2015). *Cross-Border Data Flows Enable Growth in All Industries*. Information Technology & Innovation Foundation. February 2015.
- ITIF (2017). *Cross-Border Data Flows: Where Are the Barriers, and What Do They Cost?*. May 2017.
- Ker, D.(forthcoming). "Perspectives on the value of data and data flows", [DSTI/CDEP/MADE\(2019\)6](#), OECD, Paris.
- Kommerskollegium (2014). No transfer, No Trade: the Importance of Cross-Border Data Transfers for Companies Based in Sweden. National Board of Trade, Sweden. January 2014
- Laney D. (2011): "Infonomics: The Economics of Information and Principles of Information Asset Management", The Fifth MIT Information Quality Industry Symposium, July 13-15, 2011.
- Li, W.C.Y., Nirei, M., & Yamana, K. (2019), „Value of Data: There’s No Such Thing As A Free Lunch in the Digital Economy”, *U.S. Bureau of Economic Analysis Working Paper*, Washington, DC.
- López González, J. and M. Jouanjean (2017), “Digital Trade: Developing a Framework for Analysis”, OECD Trade Policy Papers, No. 205, OECD Publishing, Paris.
- López González, J. and J. Ferencz (2018), “Digital Trade and Market Openness”, OECD Trade Policy Papers, No. 217, OECD Publishing, Paris.
- Magalhaes, G., and C. Roseira (2017): “Open Government Data and the Private Sector: An Empirical View on Business Models and Value Creation.” *Government Information Quarterly*.
- Mandel, M. (2017). “The Economic Impact of Data: Why Data is Not Like Oil.” Progressive Policy Institute.
- Mauldin, A. (2019). “A Complete List of Content Providers' Submarine Cable Holdings”. (TeleGeography Blog, November 2017). Accessed February 2020.
- McKinsey (2018): “Achieving business impact with data: A comprehensive perspective on the insights value chain.” Digital/McKinsey, April 2018.
- McKinsey Global Institute (2016a): “Digital Globalization: The New Era of Global Flows”. McKinsey & Company, February 2016.
- McKinsey Global Institute (2016b): “The Age of Analytics, competing in a data-driven world”. McKinsey & Company, December 2016.
- McKinsey Global Institute (2019): “Globalization in Transition: The Future of Trade and Value Chains”. McKinsey & Company, January 2019.
- Meltzer, J.P. and P. Lovelock (2018). *Regulating for a global economy: Understanding the importance of cross-border data flows in Asia*. Brookings Global Economy & Development Working Paper No. 113, March 2018.

- Nakamura, Leonard, Jon Samuels, and Rachel Soloveichik (2018): “‘Free’ Internet Content: Web 1.0, Web 2.0 and the Sources of Economic Growth”; Paper prepared for the 35th IARIW General Conference; Copenhagen, Denmark, August 20-25, 2018.
- OECD (2009): “*Handbook on Deriving Capital Measures of Intellectual Property Products*”, OECD Publishing, Paris.
- OECD (2013a): “Introduction to Data and Analytics (Module 1): Taxonomy, Data Governance Issues, and Implications for further Work”, Paper circulated for consultation; OECD, [DSTI/ICCP\(2013\)13](#).
- OECD (2013b), “*Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value*”, OECD Digital Economy Papers, No. 220, OECD Publishing, Paris.
- OECD (2014), “*Measuring the Digital Economy: A New Perspective*”, OECD Publishing, Paris.
- OECD (2015), “*Data Driven Innovation: Big Data for Growth and Well-being*”, OECD Publishing, Paris.
- OECD (2018a), “Trade and cross-border data flows”, Preliminary Draft, TAD/TC, OECD Publishing, Paris.
- OECD (2018b). *Working Party on International Trade in Goods and Trade in Services Statistics: Result of the 2018 WPTGS Stocktaking Questionnaire*. [STD/CSSP/WPTGS\(2018\)3](#), 28th February 2018.
- OECD (2019a), *Measuring Digital Transformation: A Roadmap for the Future*, OECD Publishing, Paris.
- OECD (2019b), *Digital Innovation: Seizing Policy Opportunities*, OECD Publishing, Paris.
- OECD (2019c), *Understanding Enhanced Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing Paris.
- Pepper, R. and Garrity, J. (2014). The Internet of everything: How the network unleashes the benefits of big data. *Global Information Technology Report 2014*.
- Rayport, J. F. and Sviokla, J.J. (1995), “Exploiting the Virtual Value Chain”, *Harvard Business Review*, Vol. 73, No. 6, pp. 75–85.
- Rugman, A.B., and L. Eden (1985). *Multinationals and Transfer Pricing*. London: Routledge.
- Shapiro, R. and S. Aneja (2019) “*Who Owns Americans’ Personal Information and What Is It Worth?*” Future Majority report.
- Short, J.E. and S. Todd (2017): “What is your data worth?”, *MIT Sloan Management Review*, Vol. 58, No.3.
- Statistics Canada (2019): “*The value of data in Canada: Experimental estimates*”, Latest Developments in the Canadian Economic Accounts, Statistics Canada, 10th July 2019.
- System of National Accounts 2008. New York: United Nations, 2009.

Swedish National Board of Trade (2015): No Transfer, No Production: A Report on Cross-border Data Transfers, Global Value Chains, and the Production of Goods”.

The Economist Intelligence Unit (2015): “The Business of Data”, The Economist Intelligence Unit report.

United Nations Economic Commission for Europe (2015): “Guide on Measuring Global Production”, December 2015.

US International Trade Commission (2014). *Digital Trade in the U.S. and Global Economies, Part 2*. Publication Number 85, August 2014.

US Department of Commerce (2016). Measuring the Value of Cross-Border Data Flows. September 2016.

van der Marel, E. (2015). Disentangling the Flows of Data: Inside or Outside the Multinational Company?, ECIPE Occasional Paper 07/2015.

World Bank (2016): “World Development Report 2016: Digital Dividends”.

Notes

¹ However, we acknowledge that the literature on organisational change, innovation, and knowledge-based capital holds relevant insights for studying the economic value of data as well.

² This is related to the literature on R&D spillovers, where social returns to R&D investment are shown to be higher than private returns, meaning that companies on their own are under-investing in R&D (Bloom et al., 2013).

³ (The Economist, 2017) headlined that “*The world’s most valuable resource is no longer oil, but data*”. This is underpinned by the fact that the most valuable companies nowadays include a number of technology giants that critically rely on intangible assets, including vast databases, to generate large shares of their revenue (e.g. Apple, Microsoft, Amazon, Alphabet, Facebook, and Alibaba).

⁴ Related to this, the OECD (2019b) distinguishes primarily physical (e.g. agriculture), mixed digital & physical (e.g. retail), and primarily digital end products (e.g. media).

⁵ DDI is defined as “significant improvement of existing, or the development of new, products, processes, organisational methods and markets” (OECD, 2015: 17).

⁶ Certainly, businesses might also collect some data aimed neither for sale, nor for use in a production process, e.g. data could be gathered for the purpose of compliance with a government regulation. For simplicity, we abstract from those considerations (they can be thought of as a form of taxation).

⁷ Amazon started operating as an online bookstore in 1994, and it now also sells consumer electronics, produces movies, operates a supermarket chain, and provides cloud services, among others. We describe Amazon as a data-enabled business since its core business model is based on offering ‘infinite shelf space’. The company achieves this by providing large quantity of products at low prices delivered fast. It can do so only with predictive data analytics and search algorithms at its heart. See also blog (Hunter, March 2019).

⁸ Freemium is a pricing strategy by which a product or service is provided free of charge, but a premium is charged for additional features or services. Another example is Spotify, that charges a subscription fee for ad-free access to its vast music library.

⁹ The business model may change over time as new opportunities to monetise data arise which were not necessarily conceived initially.

¹⁰ Theoretically the nominal price (and hence value) of some data can be higher in the case of prohibitively high barriers on cross-border flows. However, we are mainly concerned with data that only becomes valuable when aggregated at a global level.

¹¹ Also, as highlighted by participants from the business community at OECD-ESCoE workshop in November 2018, a non-negligible share of value-added is generated by data analytics, which itself critically depends on the quantity and quality of the underlying data.

¹² Another important aspect to mention is that it matters whether data have been collected or acquired legally and, hence, whether it can be legally commercialised. This is illustrated by the recent fine for the parenting club *Bounty* which sold 34m personal data records to 39 different companies (incl. Axicom, Equifax and Sky) without asking for the users’ permission. The British Information Commissioner’s Office, which led the investigation case described that “*data sharing was an integral part of their business model at the time*”; see (The Guardian, April 2019).

¹³ We are aware that not all companies (especially some start-ups) have positive revenue streams. This taxonomy is highly stylised and revenue can be replaced by company ‘valuation’, which is naturally based on future revenue streams.

¹⁴ High performers are defined as companies that had annual growth rate of 10% or more over the past 3 years.

¹⁵ Shapiro and Aneja (2019) provide estimates of the value of American personal data of based on the digital advertising revenue of the major online platforms. They show, based on these companies’ financial statements, that in 2018 they earned USD 111.1 billion from U.S. advertisers targeting American consumers. Moreover, “*Google and Facebook dominated this area in 2018, accounting respectively for 37.1 percent (\$41.3 billion) and 20.6 percent (\$22.9 billion) of total digital advertising revenues.*” (p. 9)

¹⁶ According to TJI Research, Amazon sells products using 146 private label brands and 640 Amazon exclusive brands (last updated November 2019), across different product categories including clothing, electronics, food, furniture, household goods, and healthcare (TJI Research, s.d.).

¹⁷ Covered in Harvard Business School “*Managing the Future of Work*” podcast, 25th October 2018. Konecranes reports that it already connected 18,000 cranes globally to TrueConnect and employs 600 software engineers to develop control systems (Harvard Business School, October 2018).

¹⁸ ISO (the International Organisation for Standardisation) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by these organisations to address particular fields of technical activity.

¹⁹ This is closely related to the concept of ‘data velocity’, which is used to refer to the speed of data flows. In particular, it describes the efficiency of data analytics and network infrastructure enabling smooth processing of data flows.

²⁰ Data link-ability is directly linked to data variety. Data variety relates to “*the capacity to analyse a variety of mostly unstructured data sets from sources as diverse as web logs, social media, mobile communications, sensors and financial transactions. This capacity is often associated with the capacity to link these diverse data sets (linked data).*” (OECD, 2013b)

²¹ Though beyond the scope of this paper, we also note the potential costs on the environment due to the vast amounts of energy needed to power data centers.

²² We do not aim to provide an exhaustive list, but rather focus on the most frequently used data types that are relevant in determining the economic value of data. For example, we do not discuss “big data” separately as we do not see it as a data type. We would rather relate it to a set of data characteristics - such as volume - discussed in the previous section.

²³ In some instances, data can be generated by the private sector and held by the public sector (e.g. tax records, employment records). Businesses spend considerable resources to file returns or fill-in business surveys on an ongoing basis.

²⁴ Data can also be ‘protected’ by being held as trade secrets, as is often the case for recipes (NOLO, s.d.)

²⁵ However, as already mentioned, Cisco (2018) reports that as much as 75% of all IP traffic in 2017 was video streaming. <This is the second time this is mentioned and I think you really need to make explicit what you are suggesting their nature (video) means for their value>

²⁶ The MGI model estimates the contribution of various flows - including data - to approximate their impact on real GDP. Data flows are approximated by cross-border used bandwidth from TeleGeography (sum of capacity for Internet backbones, private networks and switched voice networks). MGI run the model for 97 countries for the years 1995 - 2013 and find that a 10% increase in cross-border data flows raises GDP by 0.2.

²⁷ For example, The Guardian (2018) reports that in 2017, Facebook only paid £15.8m in tax in the UK, though it recorded GBP 1.3 billion in British sales. The article notes that taxable profits had been reduced by ‘administrative expenses’ of GBP 444m. In addition, the UK tax-to-sales ratios of Amazon, Google, and Apple were similarly low. According to Facebook’s Annual Report 2017, the company recorded a profit margin of 50% on its total global sales of USD 40 billion. This highlights the need for further research into the potential role of data transfers that are internal to the firm and to what degree they are related to the location of value creation.

²⁸ Latency refers to the time delay in sending data packages across physical distance, e.g. via fibre optics.

²⁹ Similarly, the Global Internet Map can give some insights into the data volume in terms of installed bandwidth (TeleGeography, 2018).

³⁰ This includes all EU-28 and OECD member countries.

³¹ Various studies are summarised in ITIF (2017).

³² Among other categories, this area covers policies that affect connectivity such as measures on cross-border data flows and data localisation.

³³ The partnership was signed in July 2018 and is the largest trade agreement ever negotiated by the EU, covering around a third of global GDP. It is complemented with reciprocal adequacy agreement on data protection, meaning both countries will recognise respective data protection systems. Once ratified, this will create the largest area of safe cross-border data flows. Source: European Commission, 17th July 2018 (European Commission, July 2018)

³⁴ A study by McKinsey Global Institute (2016) relates data volume (or installed bandwidth) to its economic value using regression analysis. This makes sense insofar as countries with more data flows (in terms of volume) are more likely to have more data flows with higher values.

³⁵ It also depends on whether data can be used legally, which might differ across countries.

³⁶ This can be underpinned when considering that building and maintaining the physical infrastructure to support cross-border (as well as within border) data flows is costly.

³⁷ Consumer traffic is defined as traffic originating with or destined for consumer end users; business traffic is traffic originating with or destined for business end users. Furthermore, consumer segment encompasses Other Consumer Apps, Social Networking, Video Streaming and Search whereas business segment is composed of Computation, Database/Analytics/IoT, Collaboration and Enterprise Resource Planning and Other Business Applications (Cisco, 2018).

³⁸ In some cases it could be other organisations such as universities or research institutes that conduct these types of surveys.

³⁹ It should be noted that data brokers often are not collecting their own datasets from scratch. Instead, they might rely on various public and commercial datasets, to organise or aggregate the data for easier use. Therefore, the market value of a dataset from the perspective of data brokers could be much lower (or higher) than its original production cost.

⁴⁰ This includes: Damage to the facility, premises and equipment; Employee health and safety; Business and service interruption; Data security and privacy; Expenses resume operations after a loss event; Regulations and compliance risks; and Human error.

⁴¹ It should be noted that the intangible assets literature offers many insights into the value of data, but its focus is on long-lived assets. We acknowledge that many datasets are not capital assets in this sense and that their rate of depreciation can be very high. However, in some sense we abstract from some of the implications for measurement in this paper.

⁴² For example, for the ONS QCAS the response rates vary from 67% to 85%; see (Office for National Statistics, 2014).

⁴³ Moreover, the System of National Accounts 2008 guidelines suggest valuing own-account databases based on the cost of organising and compiling data in an easy to use form, rather than the underlying data (Section 10.113). On the other hand, the value of the underlying data is included in purchased databases (10.114). It is likely that most of the databases are however developed on an own-account basis and rarely transacted on the market. Hence, surveys are likely to significantly underestimate of the total value of data.

⁴⁴ Moreover, the Central Product Classification until recently has not provided an adequate set of categories that cover databases without including too many other items. Central Product Classification Ver. 2 introduced in 2008 new single category “Original compilations of facts/information” (83940) relates to databases.

⁴⁵ It is also worth noting that surveys often use some sort of extrapolation from larger businesses to the smaller ones. This is because large companies are typically asked to fill much more detailed forms, which later serve as a basis for imputation of, for instance, the asset breakdown in a given industry. This implicitly puts fairly strong assumptions on the asset mix distribution across the companies’ universe.

⁴⁶ “All databases holding data with a useful life of more than one year should be recorded as fixed assets providing they meet the general definition of an asset (i.e. are expected provide benefits to their owners and over which ownership rights are exercised).” (OECD, 2009; p. 102)

⁴⁷ For example, the ONS questionnaire for the Quarterly Acquisitions and Disposals of Capital Assets Survey asks businesses to report databases that “are files of data that are organised to enable effective use of, and access to, the data”, instead of the data itself.

⁴⁸ In the National Accounts framework, both the database management software and on-going running costs (e.g. for data storage services) are excluded from the capitalised value of databases. Further detail on this can be found in Ker (2019) [[DSTI/CDEP/MADE\(2019\)6](#)].

⁴⁹ The important point to note here is that the exclusion of data from the production boundary (and so do not add to GDP when they are included in a database) is not the same thing as saying that data do not have value, nor that they are not important sources of output and growth, as the SNA fully records revenue streams generated from the use of data.

Data could be classified as an intangible asset under IFRS if it is “*identifiable, non-monetary asset without physical substance*”; see (IAS Plus, s.d.).

⁵¹ The online platform is defined by the [European Commission \(2015\)](#) as “an undertaking operating in two (or multi)-sided markets, which uses the Internet to enable interactions between two or more distinct but interdependent groups of users so as to generate value for at least one of the groups” (p.5) (European Commission, January 2016).

⁵² SG&A (alternatively SGA, SAG or SGNA) are Selling, General and Administrative Expenses which are a major non-production cost reported in the balance sheets of companies.

⁵³ A separate discussion on the depreciation rates appropriate for data and different data types is welcome, however beyond the scope of this paper (see also note 22).