Organisation for Economic Co-operation and Development

DSTI/CDEP/GD(2022)1/FINAL

Unclassified | English - Or. English

2 February 2023

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INNOVATION
COMMITTEE ON DIGITAL ECONOMY POLICY

Measuring the value of data and data flows

JT03511489

The report was approved and declassified by the Committee on Digital Economy Policy on 27 September 2022 and prepared for publication by the OECD Secretariat.

This publication is a contribution to Phase III of the OECD Going Digital project, which aims to provide policy makers with the tools they need to design and implement better data policies to promote growth and well-being.

For more information, visit www.oecd.org/going-digital.

#GoingDigital

Note to Delegations:

This document is also available on ilibrary as OECD (2022), "Measuring the value of data and data flows", *OECD Digital Economy Papers*, No. 345, OECD Publishing, Paris, https://doi.org/10.1787/923230a6-en.

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.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Note by the Republic of Türkiye:

The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Türkiye recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Türkiye shall preserve its position concerning the “Cyprus issue”.

Note by all the European Union Member States of the OECD and the European Union:

The Republic of Cyprus is recognised by all members of the United Nations with the exception of Türkiye. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Foreword

This report discusses different approaches to data valuation, provides estimates of the value of data and data flows, and puts forward a measurement agenda for the future.

The report was drafted by Simon Lange, John Mitchell, Vincenzo Spiezia and Jorrit Zwijnenburg. It was prepared under the supervision of Andy Wyckoff, Director of the OECD Directorate for Science, Technology and Innovation (STI), and Audrey Plonk, Head of the Digital Economy Policy Division in STI. The report benefits from comments and suggestions from staff of the OECD's Directorate for Science, Technology and Innovation (Angela Attrey, Gallia Daor, Christian Reimsbach-Kounatze and others) as well as from participants at the virtual OECD Workshop "Measuring the Value of Data and Data Flows", held on 7 April 2022. Angela Gosmann and Mark Foss provided editorial support. This publication is a contribution to IOR 1.3.1.2.3 of the 2021-2022 Programme of Work and Budget of the Committee on Digital Economy Policy.

Table of contents

FIGURES

Executive Summary

Overview

There is a widely shared notion that data have become an increasingly important input into the production of many goods and services. But just how important? What is the value of data – their contribution to economic growth and well-being? A better understanding of the contribution of data to growth can help inform important academic and policy debates.

This report discusses different approaches to the valuation of data, their advantages and shortcomings and their applicability in different contexts. In addition, it provides estimates of the value of data and data flows based on these approaches. It is mainly focused on the monetary valuation of data produced by private economic actors and their recording in economic statistics. Other important aspects of the value of data are beyond the scope of this report. Issues such as the value of making government data openly available or the value to individuals of keeping personal data private will be explored in future work.

Findings

Data have a specific combination of economic characteristics that distinguish them from other inputs into economic production

Data are non-rival but excludable. They exhibit economies of scale, which, coupled with information asymmetries and weak ownership regimes, hinder the emergence of multilateral data markets. Hence, the value of data will be affected by the governance framework that determines how they can be created, shared and used.

As most data are not traded, only a small portion of their value can be measured based on market statistics

Only a small portion of the value of data can be measured based on market statistics. This portion includes firms' revenues, international trade and market valuation. In the United States, the revenues from direct sales of data were estimated at USD 33.3 billion in 2019. In the same year, exports of data services from the European Union (EU27) and the United States were equal to USD 18.6 billion and USD 6.7 billion, respectively. Venture capital investments in "big data" firms, which reflect the investors' evaluation of the future revenues of these firms, reached USD 35.6 billion in 2021.

The sum-of-cost approach appears the most promising to estimate the value of own-account data

The sum-of-cost approach appears the most promising to estimate the value of own-account data i.e. data produced by a firm for its own use rather than for sale. This approach derives the value of data based on the costs incurred to produce it. As such, it is consistent with the valuation of other own-account intellectual property products, e.g. software, and research and development. International guidance on the implementation of this approach in the System of National Accounts is being developed.

Experimental estimates based on the sum-of-cost approach suggest a sizeable value of investment in data in OECD countries

Estimates by national statistical institutes (NSIs) suggest that investment in total data assets was 2.2-2.9% of total value added in Australia (2016), 1.4-1.9% in Canada (2018), 2.4-3.0% in the Netherlands (2017) and 0.8% in the United States (2020). Estimates by academia, based on a broader definition of data assets than used by NSIs, range between 3.8-6.6% of the market sector's value added in selected EU countries.

Recommendations

Develop product and industry classifications to help measure the value of data

Although the bulk of data is not traded on the market, this report has shown that market statistics, such as revenues, exports and expected revenue streams – as reflected by venture capital investments – are a key tool for measuring the value of data. However, current product and industry classifications are not suited to delineate data. The United States is the only economy where statistical nomenclatures make it possible to measure – although imperfectly – revenues from the sales of data. Developing product and industry classifications to better capture data products and data-related activities remains a priority for measuring the value of data.

Develop international statistical guidelines to measure data investment and assets

A consensus has emerged on the approach for measuring data assets within the SNA with respect to data produced for firms for their own use. However, the implementation of this approach is still in its infancy and will require considerable further work. Therefore, developing international statistical guidelines for the measurement of data investment and assets will be a major task in the years ahead.

Develop dedicated survey tools and econometric approaches to estimate value of cross-border data flows

Within the scope of economic statistics, the measurement of the value of cross-border data flows remains elusive, as discussed in this report. This situation calls for a good deal of invention: in using information not collected for this purpose; in developing dedicated survey tools, including Internet-based ones; and in applying new econometric approaches to estimate the value of cross-border data flows.

Engage policy makers from diverse disciplines to help measure value of data

The above considerations apply to the measurement of the economic value of data within the SNA framework. However, data have value for consumers, firms, governments, research activities and society at large that goes beyond the scope of macroeconomic statistics. Data may also generate negative value when their use is detrimental to some individuals or organisations. Developing concepts to think about these channels of value creation and statistical frameworks to measure them should be part of any measurement agenda for data. Given the multidisciplinary nature of this task, the engagement of different policy and technical communities within the OECD is an indispensable prerequisite for its success.

1 Introduction

There is a widely shared notion that data have become an increasingly important input into the production of many goods and services. But just how important? What is the *value of data* – their contribution to economic growth and well-being? A better understanding of the contribution of data to growth can help inform important academic and policy debates. This section introduces key concepts associated with data and outlines the structure of the report.

1.1 Why measure the value of data?

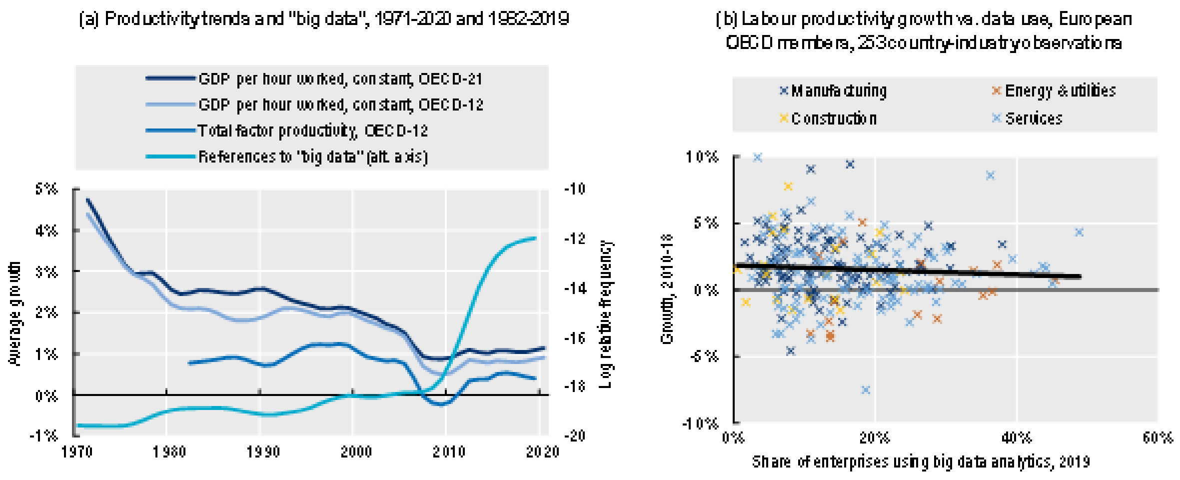
There is a strong notion that the vast amounts of data generated by search engines, online social media and other digital technologies are having profound impacts on the economy and people's lives. Data have been called the "new oil"1 and "the world's most valuable resource" (The Economist, 2017[1]), suggesting they are increasingly important in the economy. Yet putting precise numbers on the costs and benefits of data in the economy and for society at large has thus far remained elusive. How much does the private sector invest in the collection, processing and analysis of data? How does it change in response to, say, the introduction of new data protection regulations? Do increasing stocks of data held by private firms and their uptake of data-driven business models contribute to overall economic growth and well-being?

Measuring the value of data in the economy is critical to inform three key policy debates. These relate to productivity growth, the potential risks associated with use of personal data and cross-border data flows.

First, "big data" have raised hopes to significantly accelerate productivity growth. However, although data and digital technologies have proliferated in recent decades, their effects are not apparent in productivity statistics (Figure 1.1) (Goldin et al., 2021[2]). Some have argued that the contribution of new technologies, including data-driven machine learning and big data analytics, might not be adequately captured in productivity statistics (the *mismeasurement hypothesis*) (Brynjolfsson, Rock and Syverson, 2021[3]).

Others have pointed to other channels, including the potential adverse effects of data on competition and, thus, allocative efficiency (Furman et al., 2019[4]; Prüfer and Schottmüller, 2021[5]) or to slowing technology diffusion (Andrews, Criscuolo and Gal, 2016[6]; Corrado et al., 2021[7]). OECD work favours the latter hypothesis (OECD, 2022[8]), but better measurement of data and their value would help resolve such debates and better target policy attention (Ahmad, Ribarsky and Reinsdorf, 2017[9]).

Figure 1.1. Productivity growth and "big data"



Note: See endnote[2] for a full list of countries and industries included. Left panel: Average growth rates and log relative frequency based on a local polynomial regression with a bandwidth of 1.5 years. Right panel: The thick black line is a least-squares regression line with slope parameter -0.017 (*p*-value = 0.300). One extreme observation for manufacturing – growth in labour productivity of 14.0%, uptake of big data analytics by 8.7% of enterprises – has been omitted from the figure.
Source: Left panel: OECD, based on data from the OECD (2022[10]) and Google Ngram (Lin et al., 2012[11]). The latter are licensed on a Creative Commons Attribution 4.0 International License. Right panel: OECD, based on data from Eurostat (2022[12]) and OECD (2022[10]).

Second, the use of some types of data (e.g. personal data) can have positive economic and societal benefits. However, it may also entail costs in the form of reduced privacy (OECD, 2022[13]), personal liberty and agency (Zuboff, 2019[14]; Zuboff, 2019[15]). The downside risks associated with access to and monetisation of personal information have been a long-running concern but have received renewed attention in recent times (OECD, 2019[16]).

Finally, the debate about cross-border data flows encompasses concerns about the protection of privacy and intellectual property rights, along with ensuring access to information for regulation, national security and support of domestic industries (López González, Casalini and Porras, 2022[17]; OECD, 2022[18]). There has been a notable increase in measures addressing cross-border data flows. This has either placed conditions on the cross-border transfer of data or required that data be stored locally, potentially affecting economic activity (Casalini and López González, 2019[19]; OECD, 2022[18]).

The implications of data use and governance for the economy and society are the subject of sibling reports to this publication (OECD, 2022[13]; OECD, 2022[8]; OECD, 2022[18]). These debates, just like the measurement debate, point to the need for a better and more nuanced understanding of the value of investment in data and their contribution to growth and well-being.

1.2 Outline

This report discusses several complementary approaches to the valuation of data, their advantages and shortcomings and their applicability in different contexts. In addition, it estimates the value of data and data flows based on these approaches.

The report is mainly focused on the monetary valuation of data produced by private economic actors and their recording in economic statistics. Other important aspects of the value of data are beyond its scope. The value of data for consumers, firms, governments, research and society at large, for example, will be explored in future work.

The report is organised as follows:

- **Section 2** sets the scene by defining data, charting the rise of the data economy and delineating the role of data in economic production. It discusses the economic characteristics of data and their implications for measurement.
- **Section 3** explores the feasibility of the market-based approach to measuring the value of data and provides some examples of statistics based on this approach.
- **Section 4** reports progress towards recording data in macroeconomic statistics and presents some experimental estimates of investment in data in a sample of OECD countries.
- **Section 5** summarises the main findings of the report and proposes a draft measurement agenda.

2 Data and the economy

What are data and how are data different from other economic goods? This section first discusses the drivers of the increasing importance of data in economic activity: falling costs of data processing and storing; increasing activity and uptake of digital technologies; and rapid advances in artificial intelligence. It then analyses the economic characteristics of data, which comprise non-rivalry and excludability, spillovers and externalities, economies of scale, increasing returns to scale, synergies and low specificity. Finally, the section examines the implications of these characteristics for measuring the value of data.

2.1 What are data?

In common parlance, the term "data" is interchangeable with terms such as "information," "evidence" or even "knowledge". According to OECD (2021[20]), the term data "refers to recorded information in structured or unstructured formats, including text, images, sound and video". In other instances, the term "data" is used to refer to Internet Protocol (IP) traffic or the volume of digitised information stored on servers and other hardware.

While both notions are relevant in a variety of policy and technology debates, they are either too broad or ill-suited for measuring the value of data. Instead, this report builds on the definition of data recently proposed in the context of the System of National Accounts (SNA):[3] "Information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities" (ISWGNA, 2022[21]).

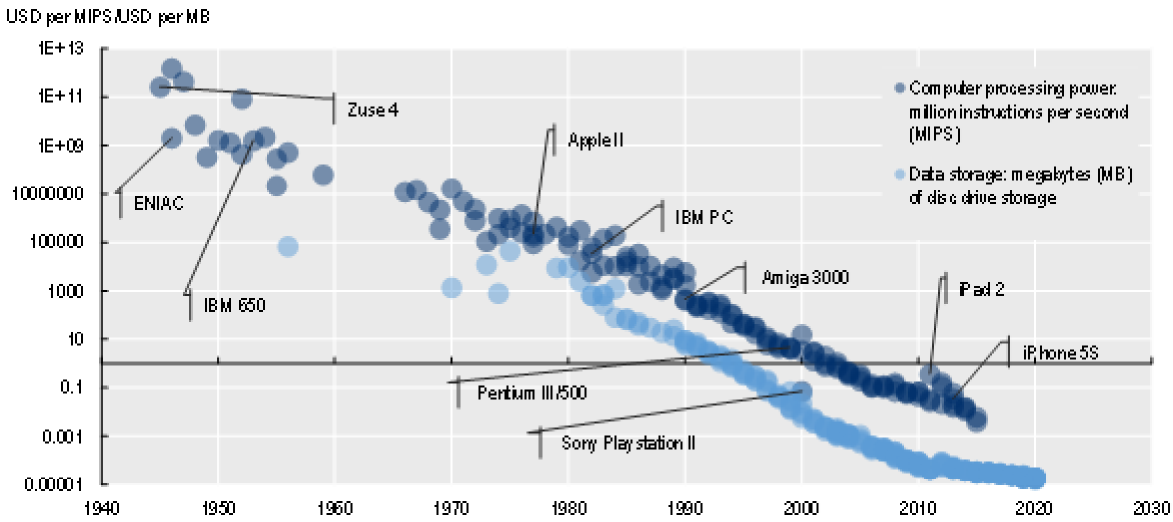
2.2 The rise of the data economy

Data have long been important to economic activity. Yet three technological developments have led to a rapid increase in the importance of data in economic activity (Carriere-Swallow and Haksar, 2019[22]): falling costs of data processing and storing; increasing activity and uptake of digital technologies; and rapid advances in artificial intelligence.

Falling costs of data processing and storing

Advances in digital hardware have reduced the costs of collecting, processing and storing data (Figure 2.1). The real costs of both computer processing power and data storage have halved every 15 months over the second half of the last century. Meanwhile, the price of software started falling at a much more moderate rate of about 2% per year from the early 1980s (US Bureau of Economic Analysis, 2022[23]).

Figure 2.1. Cost of computer processing power and data storage



Notes: MB = megabyte; MIPS = million instructions per second. Costs are deflated using the consumer price index (US Bureau of Economic Analysis, 2022[24]).
Source: OECD, based on data collected by Moravec (2022[25]) and McCallum (2021[26]).

Increasing connectivity and uptake of digital technologies

While falling hardware prices are long-term trends, two more recent developments have created vast amounts of easy-to-collect digital data: increasing connectivity and an increase in uptake of digital services.

The number of Internet users has increased substantially. In 2000, on average, only one-quarter of the population of OECD countries was using the Internet; by 2020, the share had increased to close to 90%.

Not only has the number of users increased but also the frequency and intensity of Internet use. Connectivity first shifted from personal computers in offices and homes to smartphones. More recently, it has moved to other connected devices embedded with sensors – the Internet of Things (IoT). By one estimate, there were 20 billion connected devices globally by 2019, up from 8.8 billion in 2010. Of this amount, IoT devices accounted for half (AIOTI, 2020[27]).

Rapid advances in artificial intelligence

Finally, increasing availability of digital data has coincided with rapid advances in artificial intelligence (AI). This is especially true for machine learning, an applied subfield of AI premised on finding patterns in large datasets as a basis for predictions (Agrawal, Goldfarb and Gans, 2018[28]; OECD, 2019[29]). AI – a term coined in the 1950s – has been through ups and downs in terms of interest and funding ("AI winters"). However, substantial progress has been made using multi-layered neural networks ("deep learning") in diverse areas such as image and speech recognition and machine translation (Varian, 2019[30]).

The increasing uptake of AI by private firms is widely seen as having the potential to drive productivity growth by spurring innovation and increasing automation. The number of patents granted by the United States Patent and Trademark Office each year in the area of machine learning grew from only 28 in 2007 to nearly 2 650 in 2021, a growth rate of nearly 40% per year (Baruffaldi et al., 2020[31]; United States Patent and Trademark Office, 2021[32]).

2.3 Economic characteristics of data

Data have been compared variously to oil, sunlight and infrastructure. They also share some characteristics with ideas or knowledge. However, while there are parallels in each case, data have specific characteristics that distinguish them from all of the above.

Non-rivalry and excludability

Data are non-rival: they can be used over and over again, by different actors and at the same time without being used up (Jones and Tonetti, 2020[33]). At the same time, data are generally excludable: in most cases, organisations and firms can prevent other people or institutions from accessing their data. While non-rivalry is inherent to data, excludability depends on the legal framework regulating access to and ownership of data (Ostrom, 2010[34]).

Spillovers and externalities

Excludability is closely related to the idea of spillovers, which occur when entities other than the one collecting the data can benefit from them or incur costs. If these benefits (costs) are not transferred to prices, they generate positive (negative) externalities. While data are, in principle, excludable, spillovers can still occur, such as in the case of data breaches.

Coupled with non-rivalry, positive spillovers provide a strong rationale, from a social welfare standpoint, to share data widely, e.g. to improve production processes and enhance productivity. In the case of personal data, however, negative externalities are possible. For example, sharing one's personal data can reveal information about others (Acemoglu et al., 2019[35]).

Economies of scale

The production of data exhibits economies of scale, i.e. the cost of producing a dataset is high relative to the cost of producing additional copies, which is negligible. Economies of scale have implications for pricing: when marginal costs are near zero, the price is determined entirely by the demand, i.e. the value that users ascribe to data. As consumers can have widely different valuations of the set of data, value-based pricing naturally leads to differential pricing (Shapiro and Varian, 1999[36]).

Increasing returns to scale

Data can give rise to increasing returns to scale in production, i.e. more economic activity generates more data, which in turn generate more economic activities, and so on (Jones and Tonetti, 2020[33]). There is an important potential interaction between this positive feedback loop and non-rivalry: if data are shared among firms, increasing returns to scale might also operate at the level of the economy. Data might also exhibit increasing returns to scale as an input into machine learning (Posner and Weyl, 2018[37]).

Synergies

Data exhibit synergies (or complementarities) in three ways. First, the value of data increases in the presence of other data of the same type (Coyle et al., 2020[38]). For instance, a data point about any one person may have limited value. However, combined with data about the same person over a long period or about other people, a data point can reveal trends or patterns. Second, the value of data increases in the presence of other data with complementary attributes. Official economic statistics, for example, are representative at the level of the economy yet published with a considerable time lag. When these statistics are combined with high-frequency private-sector data from digital platforms, they can reveal patterns (Chetty et al., 2020[39]). Third, the value of data increases in the presence of other, non-data inputs, especially other intangibles. Examples include specific technologies, e.g. information communication technology hardware and software, sensors and skills. Also, as machine learning depends critically on data as an input, an increase in the uptake of machine learning might significantly increase the value of data in the future (Agrawal, Goldfarb and Gans, 2018[28]; OECD, 2022[8]).

Low specificity

Compared to other products, data exhibit a lower degree of specificity, i.e. they can be used in a larger range of production activities. Often, data are re-used for purposes different from the one intended. For example, anonymised mobile call data records of telecommunications services providers have been re-used to monitor and control the spread of COVID-19 (Reimsbach-Kounatze, 2021[40]).

2.4 Implications for the measurement of the value of data

What do non-rivalry, excludability, synergies, scale effects, externalities and low specificity imply for the valuation of data? Non-rivalry of data and their potential excludability imply that value creation from data depends on the extent to which data are shared, used and re-used throughout the economy, i.e. the set of rules and institutions that govern data openness and sharing.

Low specificity makes it hard to evaluate the value of data as the same dataset can be re-used for new, unforeseen purposes.

The notion that data sharing can enhance value generation is underpinned by the synergies of data. Additional value from the same data can be generated if one firm shares its data with another firm that has access to complementary skills, technologies or ideas.

The more data are shared, the more value can potentially be created from them. At the same time, the cost of producing data might not adequately reflect their contribution to value creation. This would occur if the benefits generated through sharing and re-using data with more firms or individuals are uncompensated.

Scalability, economies of scale and increasing returns to scale, have implications for data markets that are similar to those of other intangible assets and especially information goods (Shapiro and Varian, 1999[36]).

Data markets will tend towards monopolistic competition, which might imply challenges for valuations of data based on market prices and quantities. These issues will be discussed in more detail in Section 3.

Non-rivalry implies that not enough data are being shared. However, negative externalities associated with personal data could lead to too much data sharing at too low a price. This might happen if individuals do not factor in privacy costs to others (Acemoglu et al., 2019[35]). Hence, even if data markets would exist on a large scale, information on prices and volumes might not adequately reflect the valuation of their data by individuals.

Overall, the characteristics of data imply their costs and prices may underestimate or overestimate their economic value. For positive externalities or non-decreasing returns, value would be underestimated. For negative externalities or monopolistic competition, value would be overestimated.

These issues are not specific to data. Other assets, such as research and development, may also exhibit externalities and non-decreasing returns. These effects, however, cannot be measured but only estimated through further economic or econometric analysis.[4] Therefore, while they may not measure the whole value of data, costs and prices are the necessary prerequisites to arrive at reliable and comparable estimates.

Key take-aways

- Falling costs of data processing and storing, increasing connectivity and uptake of digital technologies, as well as recent advances in AI, have led to a rapid increase in the importance of data in economic activity.
- Data have a specific combination of economic characteristics that distinguish them from other production inputs and have implications for the measurement of their value.
- Non-rivalry of data and their potential excludability imply that value creation from data depends on the extent to which data are shared, used and re-used throughout the economy, i.e. the set of policies and institutions that govern data openness and sharing.
- Negative externalities associated with personal data could lead to too much data sharing at too low a price. Hence, information on prices and volumes might not adequately reflect users' valuation of their data.
- Overall, the characteristics of data imply that costs and prices may not provide a fully accurate measure of their economic value. Yet they are essential to arrive at reliable and comparable estimates, which may be further enhanced through economic or econometric analysis.

3 Sales of data and data-intensive services

How are data shared among economic agents and what are the implications for the prospect of measuring their value based on observed quantities and prices? While specific examples of data markets exist, the properties of data significantly limit the extent to which they are traded under standardised terms. Own-production and own-use and sales of data-intensive services are far more common. Hence, market-based valuations of data face significant challenges in practice. This section explores such questions by examining types of data markets, lack of economies of scale for multilateral data markets, sales of data and data-intensive services, international trade in data services and venture capital to big data firms.

3.1 Data markets

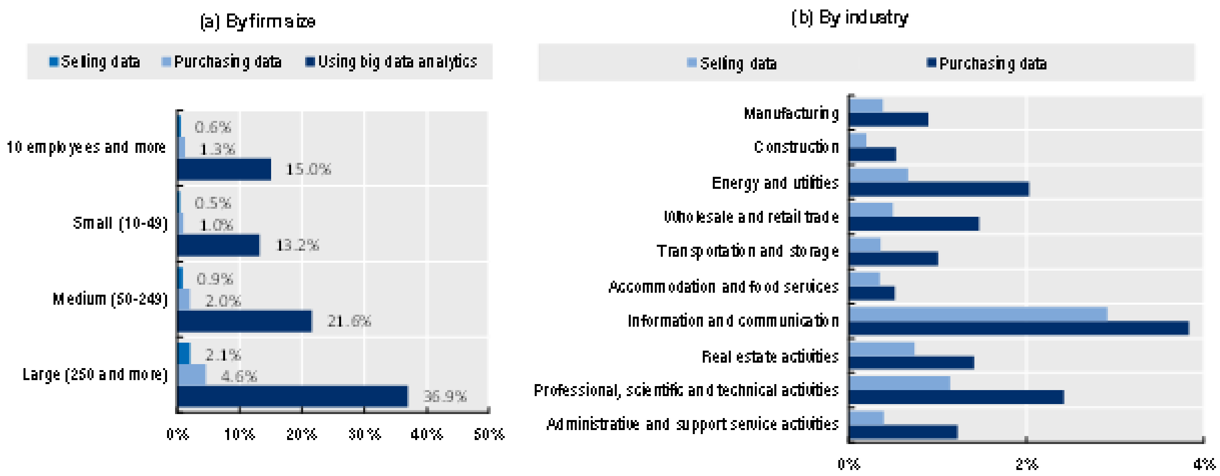
Sales of data

Data can be valuable to many different users. Due to their non-rivalry (Section 2), data may be employed in applications or contexts far removed from the locus of their production (World Bank, 2021[41]). Therefore, markets for data that facilitate efficient matching between buyers and sellers might significantly increase their economic value. If data were primarily traded in markets – if the economic unit producing the data was not the same as the unit using them – the size of the market would provide a straightforward measure of the value of the data. In fact, the System of National Accounts (SNA) recommends valuing assets at the price observed in markets or estimated from observed market prices to the extent possible (Section 4).

However, significant amounts of data collected by private entities are not traded in markets. For instance, 15% of enterprises in Europe on average analysed big data in 2019. However, only 1.3% purchased data and only 0.6% sold data to other economic units (Figure 2.1a). Although larger firms tend to trade data more, the shares of those purchasing and selling data was, on average, less than 5% and 2%, respectively. This held true even among enterprises with 250 employees or more.

Figure 2.1. Share of enterprises using, purchasing and selling data

Europe, 2019



Notes: The data cover the business economy but exclude financial services. Purchases and sales here include licensing of access to data. Countries included are listed in endnote.[5]
Source: OECD, based on data from Eurostat (2022[12]), *Digital Economy and Society* (database), https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database (accessed 27 January 2022) .

These statistics also suggest that data trading is unevenly distributed across different industries (Figure 2.1b). The share of firms purchasing and selling data is higher in the information and communication sector and lower in manufacturing and construction, as well as in blue-collar services such as the hospitality sector.

Types of data markets

Koutroumpis, Leiponen and Thomas (2020[42]) distinguish four types of data markets:

- **Bilateral markets (one-to-one)** are markets in which bilateral trading is sustained by relational contracts that can involve high transaction costs, such as search costs and costs associated with reputation-building. The “data brokers” are an example of such markets.
- **Dispersal markets (one-to-many)** are markets in which a single seller transacts with many buyers for the same data under standardised terms of exchange. An example are markets in which data are distributed through application programming interfaces (APIs) such as the Twitter “firehose”. Transaction costs are unlikely to be an issue in dispersal markets. However, it can be difficult to prevent buyers from diluting the value of the purchased data by making them available to others.
- **Harvesting markets (many-to-one)** are markets in which data are generated through the use of digital services and sellers make data available to a single service provider, often in exchange for free access to the service. Again, transaction costs are low, but the availability of data depend on the nature and popularity of the underlying service.

- **Multilateral markets (many-to-many)** are often referred to as "data marketplaces": platforms through which large numbers of sellers and buyers can potentially transact under standardised licensing models and regulations regarding data access and usage.

There have been numerous attempts to create multilateral data markets since 2010 (Spiekermann, 2019[43]). However, they have often been challenging and have experienced elevated exit rates. Prominent examples of data marketplaces launched in the early 2010s that subsequently failed include Microsoft's Azure Data Marketplace, Datacoup, Handshake, Kasabi and xDayta. Other marketplaces such as InfoChimps or Data Market deviated from their original business plan.

Why are multilateral data markets not emerging at scale?

The specific characteristics of data pose challenges to the emergence of large, multilateral data markets.

Economies of scale

As noted in Section 2, the production of a given dataset is subject to economies of scale – high fixed costs and low, near-zero marginal costs: digital data are "copyable" at nearly no cost. As for other products that share this characteristic, pricing of data cannot be cost-based: setting the price at the level of marginal costs times a mark-up does not work if marginal costs are zero (Shapiro and Varian, 1999[36]). Instead, information goods are typically priced according to consumer value. As buyers of data can have widely different valuations of the same data, value-based pricing naturally leads to differential pricing, e.g. versioning.

Economies of scale per se do not prevent markets from developing, as shown by many examples of information goods for which markets exist (e.g. software, movies, books, news and so on). However, they do help to explain why multilateral data markets are unlikely to emerge. First, as data are valued differently by different buyers, bilateral exchanges and price discrimination will be more frequent than standard licensing models and regulations characteristic of multilateral markets. Second, the copyability of data, coupled with a lack of traceability and weak ownership rights (see below), implies that it is easy to sell data on. This reduces the prospect for further sales for the original producer of the data.

Lock-in and holdup risks

A firm whose activity depends on data from another firm would become locked into this transactional relationship if switching to another supplier is unfeasible or costly. This leaves the data seller in a position to appropriate the profits of the data buyer. For instance, it could raise the price of its data, a situation known as a "holdup" (Shapiro and Varian, 1999[36]).

If such holdup problems are anticipated and cannot be dealt with adequately through a contract, firms might be reluctant to depend on an external data supplier. Therefore, they may decide to internalise the production of data (own account) or to acquire the data supplier (vertical integration). In either case, data would not be traded on the market.

Data as a strategic asset

Data-collecting firms might be reluctant to offer their data for sale if it lowers barriers to market entry and exposes them to competition (Jones and Tonetti, 2020[33]). Data generated by an incumbent firm are informative about the strengths and weaknesses of its operations, an insight that new entrants could use to challenge it. If the expected benefits from selling or licensing data are lower than the costs associated with higher competition, incumbent firms have a disincentive to sell their data.

Information asymmetries from data as an experience good

Data are an "experience good", a good whose quality cannot be determined by a prospective buyer other than through purchase and consumption (Nelson, 1970[44]; Vining and Weimer, 1988[45]). If prospective buyers are only able to observe average quality, they will be willing to pay a price commensurate with average quality. This will lead sellers with high-quality data to withdraw from the market, thus lowering average quality and driving down prices further. By the iterative withdrawal of the highest-quality data on offer, markets might unravel entirely – an issue known as "market for lemons" (Akerlof, 1970[46]).

Limited appropriation possibilities for ownership of data

Appropriation of the returns to intangible goods can be pursued through legal instruments that facilitate and protect control rights (Teece, 1980[47]; Levin, Cohen and Mowery, 1985[48]). Intellectual property rights, such as patents, copyrights, and trademarks, are available to protect an idea, a technology or an expression (Gans and Stern, 2010[49]). By contrast, the legal instruments available to protect data are less well defined. Databases are theoretically protected under copyright in most jurisdictions. However, copyright typically only protects the structure and organisation of the database, not its individual observations (Koutroumpis, Leiponen and Thomas, 2020[42]). Courts often take the position that the actual data are facts that cannot be subject to copyright.[6] Because data are non-rival, weak appropriation regimes pose challenges for the development of data markets. In particular, potential sellers might not be able to prevent buyers from re-selling data or otherwise passing them on, which might depress prices in the primary market. Additional complications arise in the case of personal data.[7]

Externalities associated with personal data

Data about one consumer may be informative about other consumers, which can lead to too-low prices (Acemoglu et al., 2019[35]; Bergemann and Bonatti, 2019[50]) and too-large collection of personal data (Choi, Jeon and Kim, 2019[51]). This negative externality might help explain why consumers are rarely paid for sharing their personal data and are only compensated with free access to digital services (Acquisti, John and Loewenstein, 2013[52]). In addition**,** a data-selling firm may face a reputational externality, e.g. if the buyer experiences a data breach or uses the data in fraudulent ways[8] (de Cornièr and Taylor, 2020[53]).

In summary, data markets require an extraordinary amount of trust among participants to achieve scale. Buyers need to trust the quality of the data, their legitimacy and the reliability of their supply. Meanwhile, sellers need to trust that trading partners will not use the data in harmful ways. Achieving the required level of trust is easier in long-running, bilateral relationships than in multilateral markets.

Sales of data-intensive services

Faced with these challenges, firms that collect significant amounts of data have often opted for alternative ways to monetise their data. For example, they sell services whose value is largely accounted for by the underlying data. The market for sponsored search advertising is one example of this kind of bundled transaction (Bergemann and Bonatti, 2019[50]; OECD, 2020[54]). Rather than selling information about users directly, search engines typically use it to improve the quality of their advertising products. Financial intelligence is another example of data-intensive services: firms like Bloomberg L.P. and Thomson Reuters combine financial data with computer hardware and software (the Bloomberg Terminal and the Reuters 3000 Xtra system) to provide financial intelligence services.

3.2 Sales of data and data-intensive services

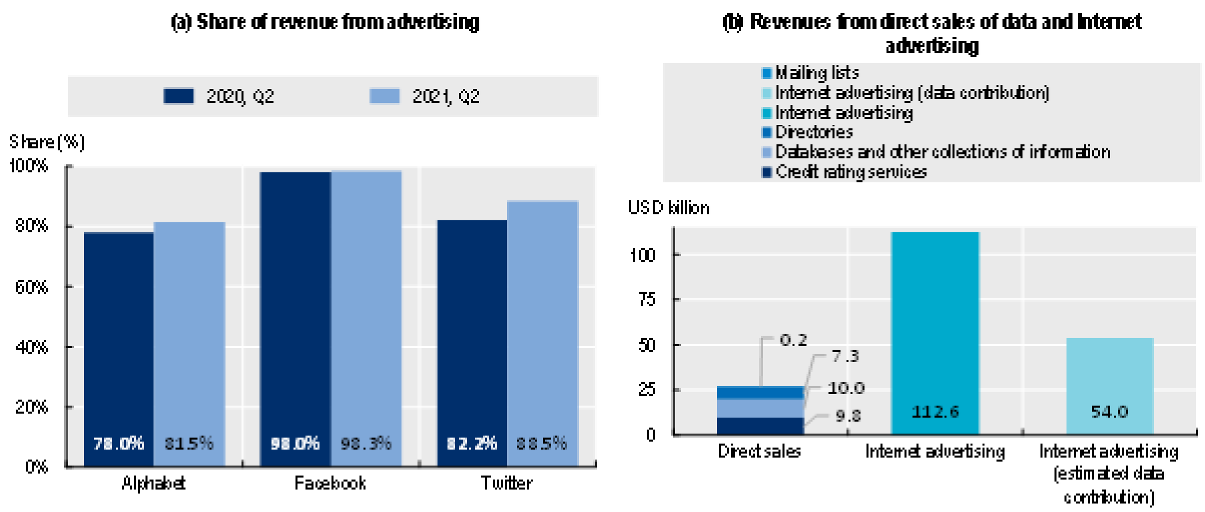
Revenues from sales of data and data-intensive services

This subsection provides illustrative estimates of the size of the market for data. Given the challenges associated with trade in data discussed above, prices and quantities observed in data markets only capture a small part of the value of data. The United States is used as the primary example here, chiefly because the US Census Bureau (2017[55]) provides detailed information about sales at the level of products that does not seem to be available for other countries (Ker and Mazzini, 2019[56]).

The first step is to decide which products should be considered relevant data. Products were classified in the 2017 US Economic Census based on the North American Product Classification System (NAPCS).[9] Four of 3 534 valid broad-line NAPCS codes were identified as capturing the sale of "data" as defined in this report. They are databases and other collections of information, directories, mailing lists and credit rating services.[10]

As noted above, Internet advertising is the leading example of a data-intensive service, accounting for a large share of revenues of companies that collect large quantities of data (Figure 3.3a). Therefore, revenues from Internet advertising are reported here as a benchmark.

Figure 3.3. Revenues from sales of personal data and Internet advertising



Source: Left panel: OECD, based on income statements from Alphabet (2021[57]), Meta/Facebook (2021[58]) and Twitter (2021[59]). Right panel: OECD, based on US Census Bureau (2017[55]).

As data are only one input into targeted Internet advertising, total revenues overstate the value of personal data in its production. However, several studies show the contribution of the use of personal data to Internet advertising revenue is substantial.[11] For instance, a 2020 study that analysed data from AdChoices,[12] the Internet industry's self-regulatory programme, concludes that the price of ads uninformed by people's personal data was 52% less than the cost of comparable ads informed by those data (Johnson, Shriver and Du, 2020[60]).

Figure 3.3b provides estimates of the size of the markets for data and Internet advertising in the US economy. Taken together, sales of data generated USD 27.3 billion in revenues in 2017. Of this amount, credit rating services and databases and other collections of information each accounted for 36-37%

(USD 9.8 billion and USD 10.0 billion, respectively) and directories for 27% (USD 7.3 billion). Mailing lists accounted for only 0.6% of the total.

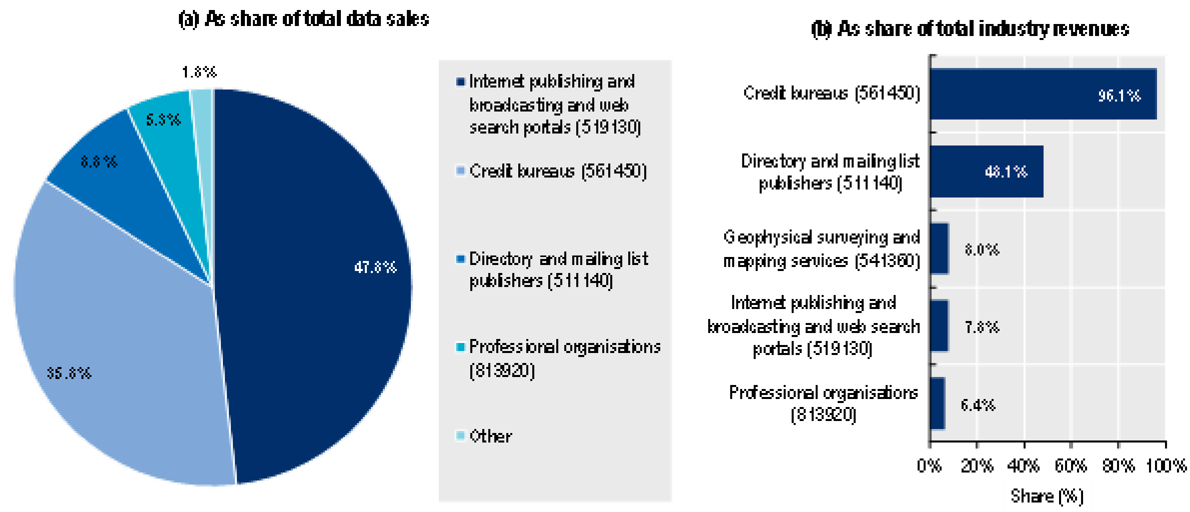
By contrast, Internet advertising generated USD 112.6 billion in revenues in 2017, more than four times the revenue generated through direct sale of data. Applying the above-mentioned estimate by Johnson, Shriver and Du (2020[60]), the contribution of data to revenues from Internet advertising would amount to USD 54.0 billion, i.e. twice as much as revenues from the sale of data.

Industry profile

What sectors do firms selling data hail from and how important are data sales within these industries? Internet publishing, broadcasting and web search portals accounted for 48% of data sales revenues in the United States in 2017. These were followed by credit bureaus (36%), directory and mailing list publishers (8.8%), and professional organisations (5.3%) (Figure 3.4a). Data sales accounted for most of the revenues of credit bureaus (96%) and nearly half of the revenues of directory and mailing list publishers (48%) (Figure 3.4b). For other industries, including Internet publishing and broadcasting and web search portals, data sales accounted for less than 10% of industry revenues.

Figure 3.4. Revenues from sales of data by industry

United States, NAICS classifications, 2017



Source: OECD, based on US Census Bureau (2017[55]).

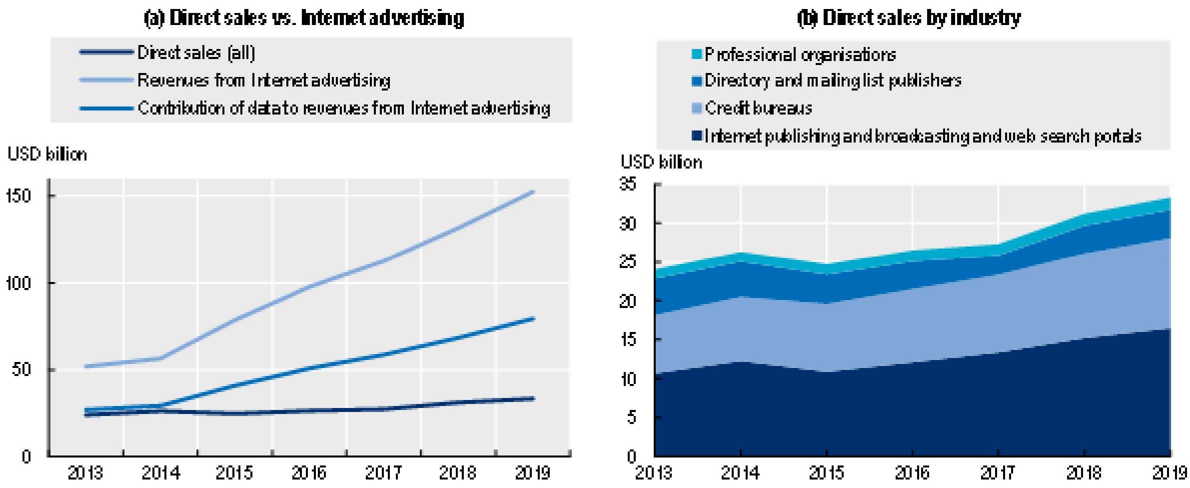
Trends in sales of data and data-intensive services

How did revenues from direct data sales and Internet advertising change over time? The 2017 Economic Census used above, which provides a high degree of detail on revenues by product, occurs only once every five years. Statistics on revenues by source from the 2019 Service Annual Survey were used to obtain a first-order approximation of trends over time.

Results reported in Figure 3.5 indicate that revenues from Internet advertising increased at a rate of 16.0% per year. This resulted in a near tripling of revenues between 2013 and 2019 from USD 51.9 billion to USD 152.5 billion. By contrast, the market for direct sales of data grew at a more moderate rate of 5.6% per year from USD 24 billion to USD 33.3 billion during the same period.[13]

Figure 3.5. Estimated trends in sales of data and data-intensive services

United States, 2013-19



Source: OECD, based on US Census Bureau (2017[55]; 2019[61]).

3.3 International trade in data services

Trade statistics provide information on the value of exports and imports of data services. In particular, the component "Other information services" (OECD, 2022[62]) includes:

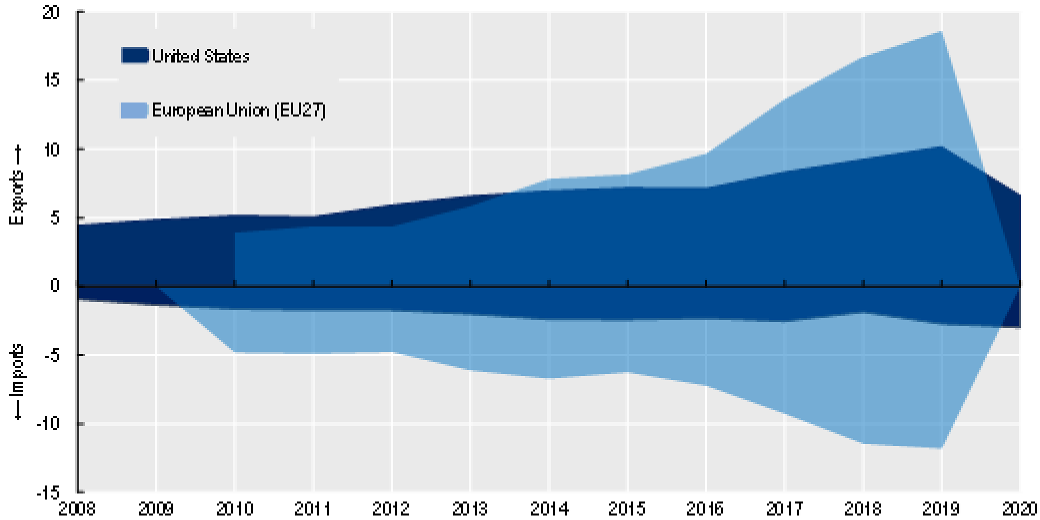
- database services, e.g. database conception, data storage and the dissemination of data and databases
- web search portals, e.g. search engine services that find Internet addresses for clients who input keyword queries.

Statistics on the above services measure the value of international trade in data services where data are the main object of the transaction. However, they do not include the value of data that are traded as part of an international transaction in goods or other services. As such, statistics based on the above definition are likely to underestimate the actual value of trade flows in data.

In 2019, exports of data services to the rest of the world amounted to USD 18.6 billion in the European Union (EU27) and USD 10.2 billion in the United States (Figure 3.6). Import of data services from the rest of the world in the EU27 and the United States amounted to USD 11.8 billion and USD 2.7 billion, respectively.

Between 2010 and 2019, exports of data services to the rest of the world increased by 4.7 times in the EU27 and doubled in the United States. Over the same period, imports of data services from the rest of the world increased as well but at a lower rate: by 2.5 times in the EU27 and by 1.7 times in the United States. The 2020 statistics, available for the United States only, show a sharp contraction in data services exports (-35%) following the onset of the COVID-19 pandemic.

Figure 3.6. Exports and import of data services to/from the rest of the world



Source: OECD, (2022[62]), https://stats.oecd.org/Index.aspx?DataSetCode=TISP_EBOPS2010.

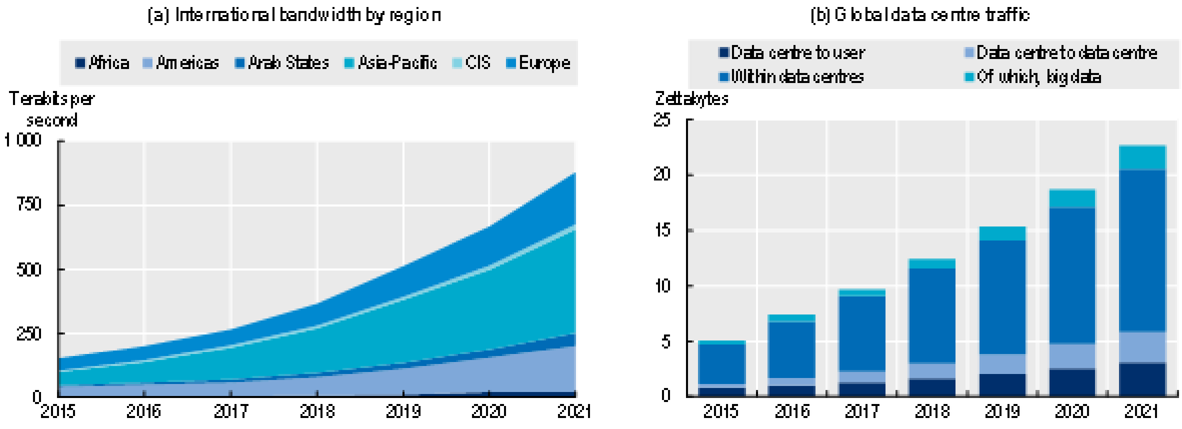
3.4 Cross-border data flows underpinning international trade

International trade flows supported by the use of digital technologies generate corresponding flows of digital data. These data may include information about products, prices, quantities and payment methods but also names and other personal information of the parties involved. The transmission of any data between two firms or individuals in different jurisdictions constitutes a cross-border data flow.

Unlike trade in data services discussed above, the transfer of data is not the main object of these transactions. Therefore, the associated data flows are not recorded in international trade statistics. More broadly, there is no statistical classification or official statistics on cross-border data flows. These features make their measurement particularly challenging.

A few proxies have been suggested to measure the volume of cross-border data flows. A common measure is international bandwidth i.e. the average traffic load of international fibre-optic cables and radio links for carrying Internet traffic (Figure 3.7a). In a similar vein, some reports (World Bank, 2021[41]) have pointed to global IP traffic as an indicator of cross-border data flows (Figure 3.8, right-hand axis).

Figure 3.7. International bandwidth and global data centre traffic



Note: Left panel: All data are ITU estimates. Right panel: ZB = Zettabytes.
Source: Left panel: OECD, based on ITU (2021[63]). Right panel: OECD (2019[64]). Calculations are based on the Cisco Global Cloud Index 2016-2021 and the Cisco Visual Networking Index 2017-2022, January 2019.

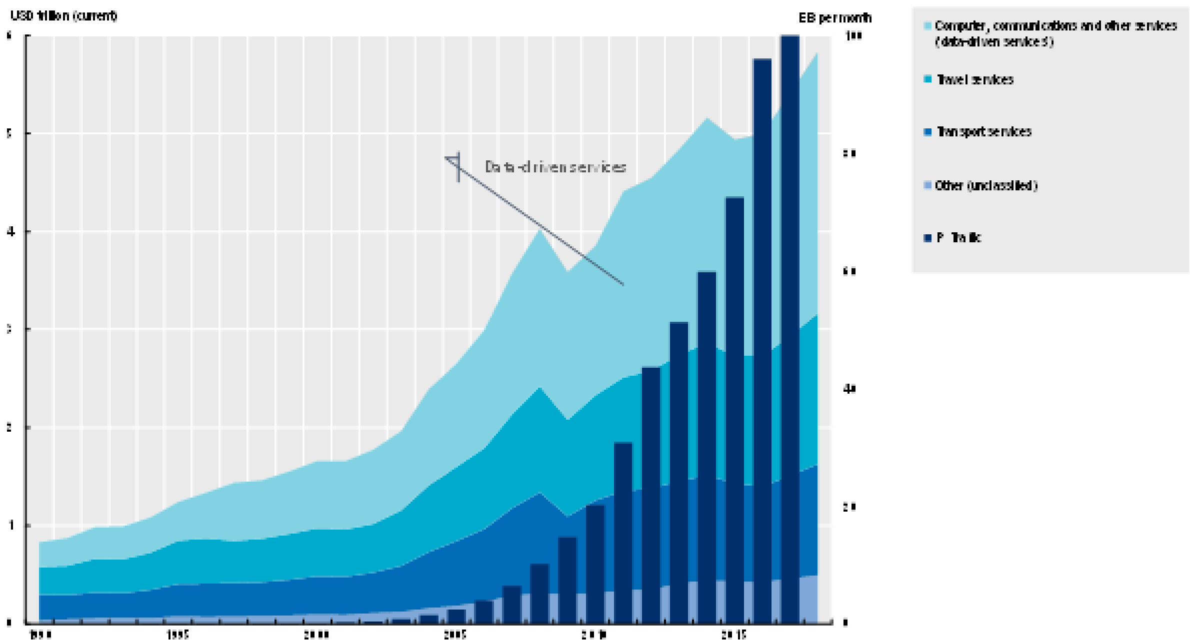
Measuring the value of data flows by some proxy for the volume of data has significant limitations. The value of data depends on the information it carries, as well as on its actual or potential use. Indeed, a large fraction of IP traffic is not generated by data as defined in this report. For instance, by one estimate, video accounted for 60% of the total downstream volume of Internet traffic in 2019 – with Netflix alone accounting for 12% (Cullin, 2019[65]).

Another shortcoming of international bandwidth and IP traffic measures is that, according to some estimates (OECD, 2019[64]), the bulk of data traffic seems to occur within data centres (Figure 3.7b). This implies that, if two businesses exchange data within a data centre in a third country, international bandwidth and IP traffic metrics would record no cross-border flow (Nguyen and Paczos, 2020[66]).

A further limitation of IP-based measures is their reliance on some non-representative sample of IP traffic. As the amount of data transferred over the Internet is just too large to be fully measured or to design a representative sample, IP traffic metrics rely on a limited number of observation standpoints. For example, they rely on an Internet Exchange Point, at very few points in time, under the incorrect assumption that these data points are representative of the whole IP traffic.

A different type of measure that has been used as a proxy for cross-border data flows is the value of trade in data-driven services (Figure 3.8) or potentially services enabled by information communication technology (ICT) (UNCTAD, 2015[67]). This includes all service transactions delivered remotely over ICT networks. These measures run into two major difficulties. On the one hand, it is difficult to assess to which degree data-driven or potentially ICT-enabled services are actually delivered via ICT networks. On the other, they do not capture cross-border flows of data that are likely to underpin trade transactions not delivered over ICT networks, including merchandise trade.

Figure 3.8. Global trade in data-driven services and IP traffic



Note: IP = Internet Protocol; EB = exabyte.
Source: OECD, based on World Bank (2021[41]).

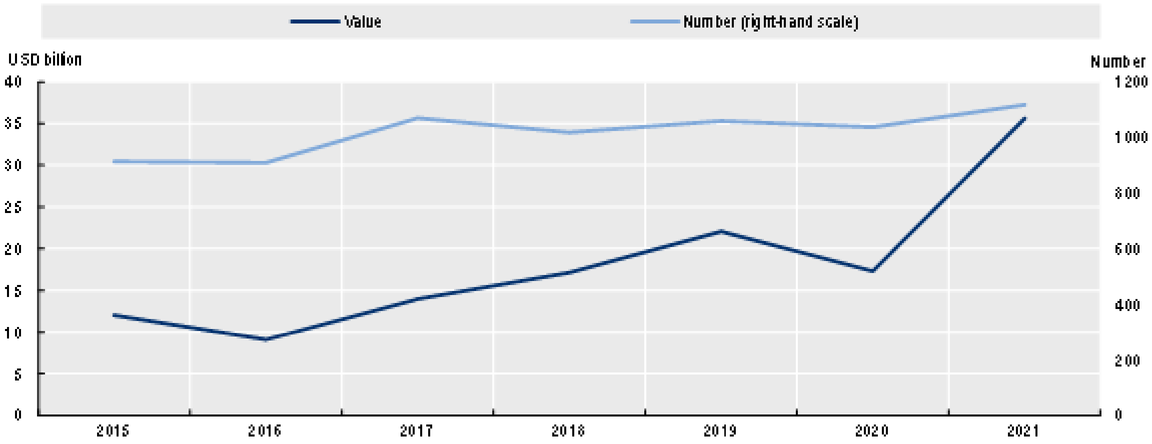
3.5 Venture capital to big data firms

Venture capital (VC) is a form of private equity financing. Specifically, it is equity capital provided to enterprises not quoted on a stock market. This is particularly relevant for young companies with innovation and growth potential but untested business models and no track record. Venture capitalists take on the risk of financing new or growing businesses with perceived long-term growth potential with the expectation that some supported firms will become successful. Therefore, VC investment in big data firms reflects the investors' evaluation of the future revenues from the data owned by these firms. In this sense, it provides a proxy for the value of their data.

Figure 3.9 shows the growth in VC investments in big data firms worldwide in 2015-21. The number of deals increased by 22%, from 913 to 1 117 between 2015 and 2021. The value of VC investment increased even more, nearly tripling from USD 12.1 billion to USD 35.6 billion over the same period.

Figure 3.9. Number and values of venture capital deals in big data firms worldwide

2015-21

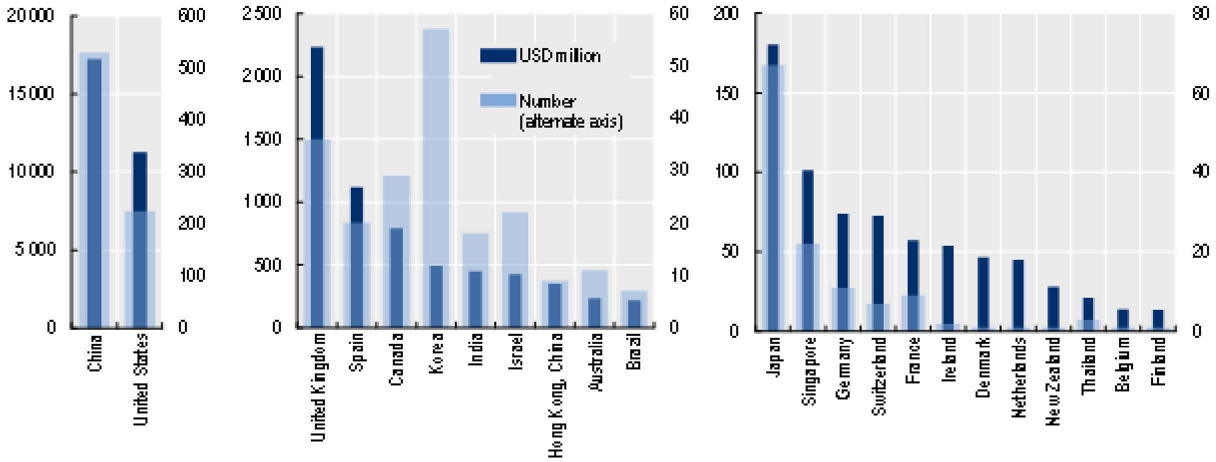


Source: OECD, based on Preqin Pro, www.pro.preqin.com (last accessed on 14 February 2022).

Figure 3.10 shows the number and value of VC deals in big data firms by country in 2021 (the latest available year). The People's Republic of China (hereafter "China") and the United States are well ahead, for both the number of deals (530 and 225, respectively) and their value (USD 17.3 billion and USD 11.3 billion, respectively). The other countries are at quite some distance. In 2021, there were 36 VC deals in big data firms for a value of USD 2.2 billion in the United Kingdom; 20 deals for a value of USD 1.2 million in Spain; and 29 deals for a value of USD 793 million in Canada.

Figure 3.10. Number and values of venture capital deals in big data firms by country

2021



Note: To keep the figure readable, countries with big data VC investment below USD 10 million are not shown.
Source: OECD, based on Preqin Pro, www.pro.preqin.com (last accessed on 14 February 2022).

Key take-aways

- Due to the specific characteristics of data, data markets are difficult to establish and sustain. Indeed, significant amounts of data collected by private entities are not traded in markets.
- As most data are not traded, only a small portion of their value can be measured based on market statistics. The latter include firms' revenues, international trade and market valuation.
- In the United States, the revenues from sales of data are estimated at USD 33.3 billion in 2019. In the same year, exports of data services from the European Union (EU27) and the United States were equal to USD 18.6 billion and USD 6.7 billion, respectively.
- Venture capital investments in "big data" firms, which reflect the investors' evaluation of the future revenues of these firms, reached USD 35.6 billion in 2021.
- There is no statistical classification or official statistics on cross-border data flows. In addition, measures based on international bandwidth, IP traffic or data-driven services seem a poor proxy for the value of these flows.

4 Recording data in macroeconomic statistics

As data have become an important input in production, their value should be adequately captured in macroeconomic statistics. Explicitly incorporating data into the System of National Accounts would ensure comparable measurement across countries and allow for a range of policy and analysis, including on productivity. This section discusses recent progress and outstanding challenges in recording data in macroeconomic statistics and presents experimental estimates of the value of investment in data assets.

4.1 Data in the System of National Accounts

Data are a crucial input into productive activities across almost all facets of the economy. Data are both used one-off (as intermediate consumption) and repeatedly over time (in the form of an asset) as an input into the production of goods and services. However, data are not explicitly identified as an input to production or as a standalone asset in macroeconomic statistics, according to the System of National Accounts (SNA) 2008 (UNSD et al., 2009[68]). While a database asset exists in the SNA2008, "the cost of acquiring or producing the data" is explicitly excluded when calculating its value.[14] Furthermore, the way expenditure on data assets is captured limits the interpretability of the outputs or may even misrepresent data's contribution to production (Ahmad and van de Ven, 2018[69]). This has resulted in concerns about whether the full economic role of data is appropriately represented in macroeconomic statistics, and in particular in gross domestic product (GDP). Explicitly incorporating data into the SNA would ensure a comparable approach to measuring of data across countries. All countries would compile estimates of investment in, and stock of, data in a consistent way. This could then be used for a variety of policy and analysis, including on productivity.

The SNA2008 framework is being updated, with a planned release in 2025. The update will allow outputs to better reflect developments in the economy since the last time the framework was codified.[15] Explicitly incorporating businesses' production and use of data into the national accounts is a key priority in this update.

A harmonised methodology to record data in macroeconomic statistics is also being developed, although this may not be finalised by 2025.[16] In this regard, the recognition of data as output and investment will need to address several measurement issues, which are noted in this section.

4.2 Valuation of data

The SNA2008 explains that assets should be recorded "at the value at which they might be bought in markets at the time the valuation is required" and that "[i]deally, values observed in markets or estimated from observed market values should be used" (SNA2008 §2.60). However, as discussed in Section 2, several factors limit the scope for data to be traded in a traditional market. As a result, most data used in production are produced on an own-account basis, i.e. by the same firm who uses them (Figure 2.1a). Hence, alternative asset valuation techniques are often required. When market prices are not available, the SNA2008 provides two methods for estimating the value of an asset: either net present value of the asset or sum-of-costs of production.

Net present value

Net present value (NPV) uses potential future income that may be derived from an asset as an alternative approach to estimating the current value of an asset. The NPV approach is already used in some areas of the SNA, for example, to value natural resources.

In theory, NPV may provide an accurate measure of the value of data assets. In practice, the re-usability of data (Section 2) may be higher than for other assets due to the absence of physical deterioration. This may make it difficult to estimate the exact future income streams from data assets. In this regard, the NPV approach requires numerous assumptions that may be difficult to justify. National statistical institutes (NSIs) are likely to encounter significant difficulties sourcing the required information from businesses that either produce data assets or make products using data assets. Reinsdorf and Ribarsky (2019[70]) point out that any calculation using NPV would involve assumptions that are "unacceptable for national accounts purposes".

Sum-of-cost approach

The sum-of-cost approach consists in measuring the value of output by summing the costs of production: intermediate consumption, compensation of employees, consumption of fixed capital used in production,[17] a net return to fixed capital used in production (also known as "mark-up") and taxes (minus subsidies) on production.

As discussed in Section 2, the characteristics of data are such that their production costs may not provide a fully accurate measure of their economic value. This is because it may be difficult to factor in the impact of non-decreasing returns, externalities or monopolistic competition. This issue is not specific to data. Other assets within the boundaries of the SNA2008 (e.g. research and development) also tend to exhibit the above features. These effects cannot be measured statistically but only estimated through economic or econometric analysis. Therefore, while they may not measure the whole value of data, the costs of production and price information are the necessary prerequisites to estimate them.

Within the SNA2008, the sum-of-cost approach is used in several scenarios. For example, it is used to measure the output of goods and services provided for free (or at insignificant prices) by government or non-profit organisations (e.g. education and health services), collective non-market output (defence, social security), or the production of goods and services retained by the producer for its own final consumption or capital formation.

The SNA2008 provides only generic guidance on which costs to include in the sum-of-cost approach. However, the planned revision will outline specific guidance for the measurement of data (ISWGNA, 2022[21]). It will note costs associated with the following:

- planning, preparing and developing a data production strategy
- accessing, recording and storing information embedded in observable phenomena (OP),[18] which may include, but are not limited to, the explicit purchases of OP or already produced data
- designing, organising, testing and analysing data to draw information and conclusions from them.

These costs include staff time and costs of items used as intermediate consumption in the production of data. In addition, they would include an estimate for consumption of the fixed capital used in the own-account production of data and a mark-up for net operating surplus for market producers.[19] The latter should reflect the future profitability of the newly created data, ensuring the estimate resembles the market value to the extent possible. This is critical for the use of this information in analyses (e.g. appropriately accounting for any changes in productivity).

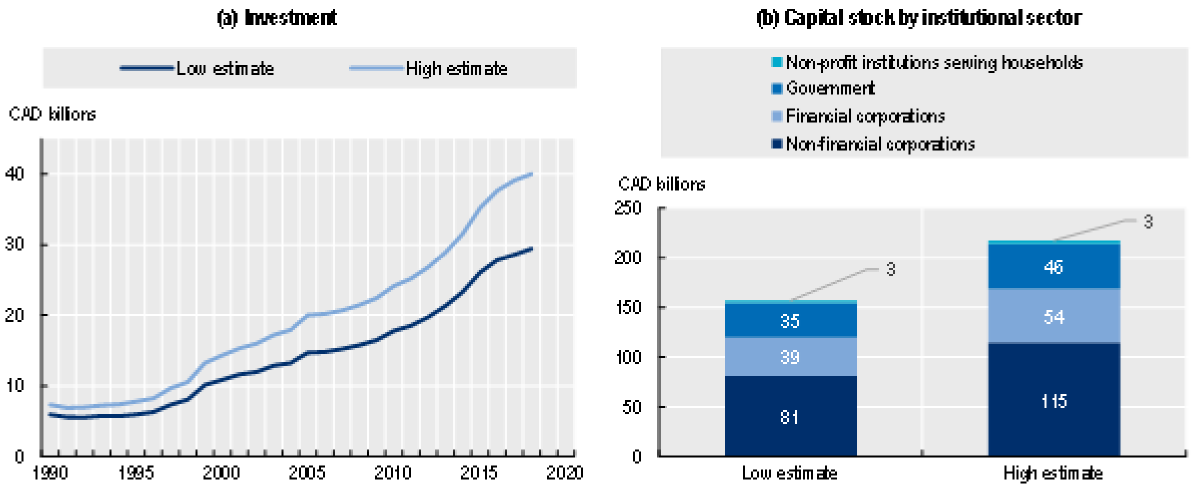
The mark-up will differ across companies (and industries). It will also be influenced by the level of data governance in countries, possible uses of the data and the competitiveness of the markets. While difficult to estimate, research and the sharing of methods and learnings across compilers should provide a foundation for best practices. Additionally, NSIs already use the sum-of-cost approach, including a mark-up, to value other own-account intangible assets. Therefore, applying this approach to data would be methodologically consistent to current practices.

4.3 Evidence from experimental studies

Estimates of data investment and stocks

Several experimental estimates of investment in data assets are available at this point, both from NSIs and academia. All of them rely on the sum-of-cost approach. Statistics Canada (2019[71]), for example, used wage and employment information as the basis for a preliminary set of estimates of investment in and stock of total data assets.[20]

Figure 4.1. Experimental estimates of the value of data investment and capital stocks in Canada



Note: CAD = Canadian dollars. Lower and upper range values result from alternative assumptions about the time devoted by different occupations groups to the creation of data assets.
Source: Statistics Canada (2019[71]).

Estimates of annual investment range between CAD 29.5 billion and CAD 40.1 billion in 2018, growing consistently since 1990 in nominal terms (Figure 4.1a). These estimates of investment in data assets were then combined with specific assumptions to generate estimates of the capital stock of data using the perpetual inventory method.[21] The stock of data assets in the Canadian economy in 2018 was estimated to be worth between CAD 157 billion and CAD 218 billion (Figure 4.1b).

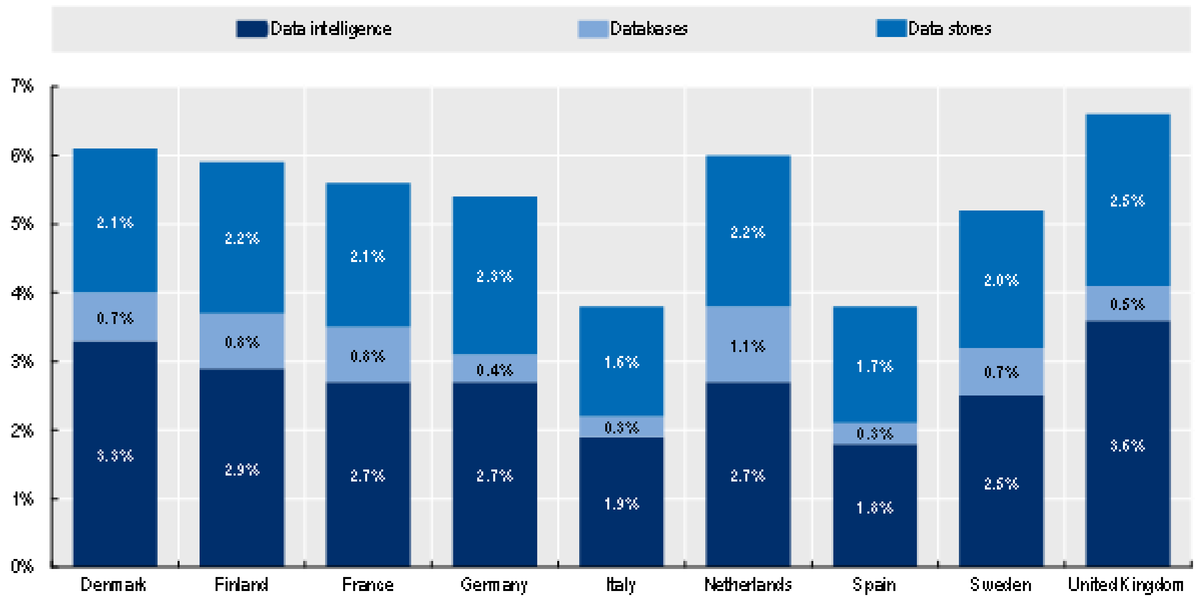
Importantly, these estimates cannot be added to current investment statistics as they overlap to a degree with some assets already measured in the SNA, notably software and databases, as well as research and development (R&D). Therefore, further work is required to calculate the overlap and refine the estimates of investment in and stock of total data assets (Statistics Canada, 2019[71]).

The work by Statistics Canada was followed by initiatives at Statistics Netherlands (de Bondt and Mushkudiani, 2021[72]), the Australian Bureau of Statistics (Smedes, Nguyen and Tenburren, 2022[73]) and the US Bureau of Economic Analysis (Calderón and Rassier, 2022[74]). Investment in total data assets relative to value added was estimated at 2.2-2.9% – depending on alternative hypotheses – in Australia (2016), 1.4-1.9% in Canada (2018), 2.4-3.0% in the Netherlands (2017) and 0.8% in the United States (2020). Several other countries continue to advance this work, although they do not publish final aggregate estimates yet (Mitchell, Ker and Lesher, 2021[75]).

Goodridge and Haskel (2015[76]) and Goodridge, Haskel and Edquist (2021[77]) also created estimates for several economies. More recently, Corrado et al. (2022[78]) estimated investment in data assets in nine European countries along the three stages of the data value chain: data stores, database and data intelligence (Figure 4.2). Overall, market-sector investment in data assets over 2010-18 ranged between 3.8% and 6.6% of gross value added, with data intelligence. In other words, the calculation integrated data with advanced analytic tools, accounting for about half across countries in the sample. These figures are based on a broader definition of data assets than the NSI estimates reported above.

Figure 4.2. Estimated average annual investment in data assets in selected European countries

As a percentage of gross value added of the market sector, 2010-18



Note: Estimates are sum-of-cost estimates for market-sector industries. Data stores are defined as raw records not yet cleaned, formatted or transformed for analysis; databases as transformed raw data suitable for some form of data analytics or visualisation; and data intelligence as the further integration of data with advanced analytical tools.
Source: Corrado et al. (2022[78]).

Estimated contribution of data assets to productivity growth

Is there any evidence to support the notion that capitalising data would significantly affect measured productivity growth? Two empirical studies suggest a negative answer to this question. Goodridge and Haskel (2015[79]) for the United Kingdom and Goodridge, Haskel and Edquist (Goodridge, Haskel and Edquist, 2021[77]) for 13 European countries estimate data investment and capital stocks based on the sum-of-cost approach. They then use a growth accounting framework to investigate the potential effects of capitalising data.

Their results suggest that recording data investments in the United Kingdom would have increased GDP growth over 2005-12 by only 0.02% per year. Similarly, in the 13 European countries considered in the study, expanding the asset boundary to include data would have increased labour productivity growth in 2011-16 by 0.04% per year only, from 0.79% to 0.83%.[22]

In explaining these modest effects, the authors note that much of the investment in data – 74% in the United Kingdom and 62% in Europe – seems to be already captured in the SNA as part of fixed investment in software and databases, and R&D.

4.4 Practical challenges for the sum-of-cost approach

While the empirical studies discussed above provide estimates of the value of data investment and stocks, their authors acknowledge the estimates are experimental in nature and likely require refinement. Therefore, additional guidance is needed. This should ensure the sum-of-cost approach is implemented consistently across countries and over time and that estimates meet the standards of statistical rigour.

The approach should also be implemented in accordance with accounting principles. In several ways, this may be considered straightforward for data as some characteristics clearly conform to definitions and classifications in the SNA. For example, the SNA defines a fixed asset as a produced economic good that is “used repeatedly or continuously in production processes for more than one year” (SNA2008 §10.11). Since data are often used repeatedly in the production process, they would appear to fit this category. Data could thus be considered alongside other fixed assets in the accounts, e.g. machinery and equipment, buildings and computer hardware.

However, some of the unique characteristics of data, including some outlined in Section 2, push established conceptual boundaries of the SNA2008 or create practical complications. These challenges include:

- ambiguity regarding the extent to which the value of data is the result of production activities
- how to record the expenditure associated with the regular addition of new information to existing data assets
- the dual purpose of many assets used in the production of both data assets and other goods and services.

These challenges, and how the national account community is attempting to overcome them, will be briefly discussed in the remainder of this section. The aforementioned revision to the SNA2008 is ongoing and the guidance has not yet been finalised. However, ongoing theoretical research combined with empirical evidence, such as that presented above, have allowed for the sharing of possible solutions with statistical compilers.

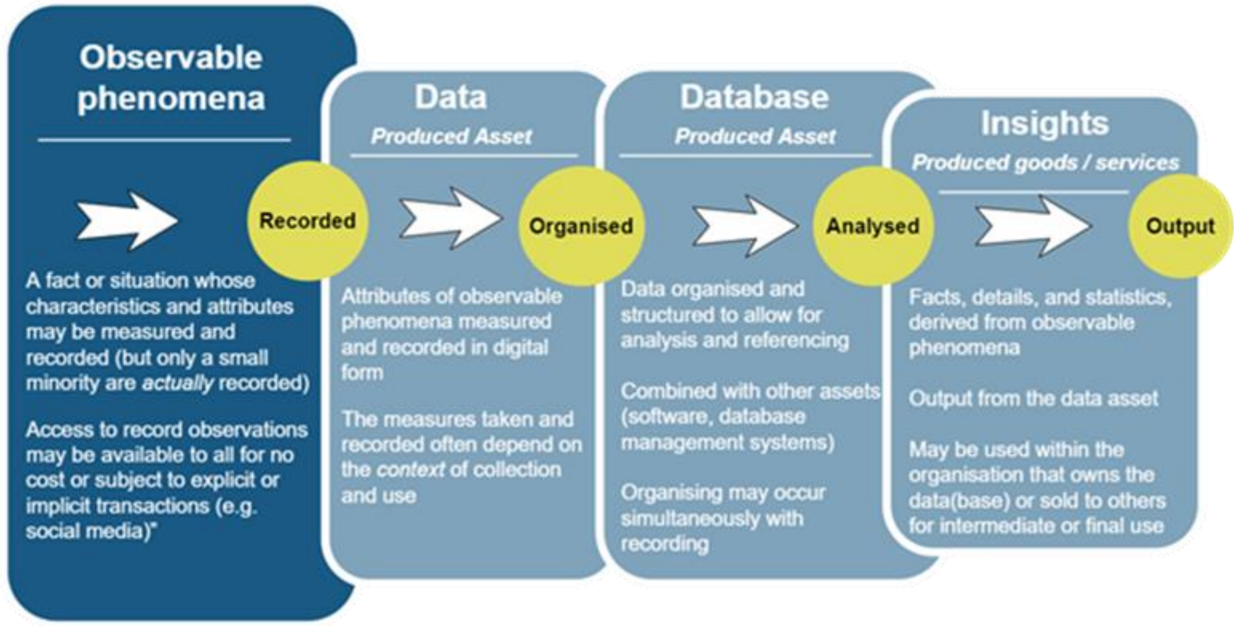
Ultimately, the nuances of how data are produced and consumed in the economy will require a range of assumptions or compromises by compilers. These will need to ensure that measurement is both feasible and consistent across countries. These assumptions and compromises will also be discussed in the remainder of the section.

Is the value of data entirely the result of productive activities?

Within the SNA framework, assets are considered as either produced or non-produced. Only the value added generated by production is recorded as part of GDP. Data raise a challenge to this clear-cut delineation. On the one hand, some of their value stems from the underlying OP, which are not produced. On the other, some of their value – likely the largest part of it – is produced, e.g. the creation of a database or the insights gained from the analysis of data (Figure 4.3).

Accounting for the value stemming from the information embedded in OP would be a significant departure from the current national accounts framework. The introduction of a hybrid asset class, accounting for both produced and non-produced value, would have considerable implications for other areas of the accounts, e.g. depreciation, estimates of capital stock and capital services, as well as productivity measurement. Furthermore, the valuation of the non-produced part of data would be problematic as the same data may be deployed to alternative users, who will place a different economic value on them (Section 2).[23] From the perspective of the national accounts, this would simply result in a revaluation of the asset. Therefore, the value of data should be entirely accounted as the result of production.

Figure 4.3. Data-information chain from an SNA perspective



Source: Based on Mitchell, Ker and Lesher (2021[75]).

Accounting for the addition of new information to existing data assets

Most assets, after being produced, receive maintenance, a current operating expense that is not capitalised. Alternatively, they may also undergo “major improvements, additions or extensions” (SNA2008 §1.56), which are counted as investment and therefore add to the overall capital stock. Unlike tangible assets, data do not deteriorate physically and their production is often a work in progress: expenditure is undertaken to continuously add new data points to an existing data asset. As a result, the distinction between maintenance and improvement expenses, which is challenging at the best of times, is even more difficult in the case of data.[24]

The expected service life of data varies greatly depending on the content of the information embedded and its use. However, through obsolescence, the vast majority of data assets will begin to decline in value if no additional expenditure is undertaken and no new information content is added. Therefore, each addition of an individual information element may not seem to constitute a “major improvement, addition or extension” (SNA2008 §1.56). However, it still “improves the assets performance, increase[s] their capacity or prolong their expected working lives”. Therefore, the expenditure associated should be recorded as new capital investment.

Accounting for the dual purpose of assets used in production

As discussed previously, the sum-of-cost evaluation of databases in the national accounts does not include the costs associated with “acquiring or producing the data” (SNA2008 § 10.1130). The inclusion of these costs would allow for the value of data to be captured. However, the process of “acquiring data” and the associated costs are not always clear (Box 4.1).

For instance, OP used in the production of data are often accessed via the provision of free digital services. What proportion of the costs of providing these services should be accounted as a cost of acquiring and producing data?

Box 4.1. How firms gain access to observable phenomena

An idiosyncrasy of data production in the economy, which is relevant for its incorporation into the SNA, concerns the different ways that organisations and data producers obtain access to observable phenomena (OP) to collect the information they convey. OP, which are considered “facts or situations whose characteristics or attributes may be recorded” (Mitchell, Ker and Lesher, 2021[75]), are fundamental to the ongoing creation of data. Importantly, obtaining this access may include an explicit monetary payment between two units. It thus requires explicit recording in the SNA.

Firms obtain access to OP not available publicly through one of three methods (ISWGNA, 2020[80]):

- **Exchanged for a free service.** Firms, especially digital ones, often provide consumable services for free (or at low cost) to gain access to OP generated by the users (Section 2). Services offered for free include social networks, search engines and free applications providing specific services.
- **As a by-product of the normal production process.** OP are often generated as a by-product during the production of market goods and services. The information content embedded in these OP can often be used to improve efficiencies across the business. Examples may include inventory controls, transaction records and customer purchase histories.
- **By explicitly purchasing access.** Although limited, there are clear examples where data producers make explicit payments in exchange for obtaining access to OP. Examples include firms that make cash payments for purchase receipts or for product surveys.

Of the three options, the first two are by far the most common for accessing OP (see also Section 3). However, neither of them (consuming a free digital service or information being created as a by-product of production) constitute an economic flow in the sense of the SNA. As such, the event is not considered an act of production and will not be recorded in the SNA. Conceptually, gaining access to the information elements embedded in the OP is fundamental to the creation of data. However, from the SNA perspective, the first stage of data production is the recording of these information elements.

Conversely, if access to OP is explicitly purchased, a monetary transaction is occurring. The transaction must thus be recorded in the national accounts. The nature of this transaction (capital or current expenditure) and the product/asset being purchased is just one example where additional recommendations may be required to ensure consistent treatment in the measurement of data.

This complicated question is not unique to the production of data; other assets in the economy are often used to produce more than a single output. However, data are omnipresent, often created as a by-product of conventional production. When data are combined with the lack of market prices, it heightens the challenge. Therefore, statistical compilers must make some assumptions and test them empirically to develop guidance on the production and use of data in the SNA.

Other considerations

Further issues need to be addressed to finalise the guidance on the production and use of data in the SNA. These include how data are sold, both on an exclusive and non-exclusive basis; the presentation of data in the accounts, i.e. separately or combined with other intangible assets; and the assumptions concerning service lives and retirement distributions used when calculating the capital stock of data.

Overall, the unique characteristics of data, including the way they are produced and used, do not fundamentally exclude data from being classified alongside other fixed assets already recorded in the SNA. Rather, they imply that statistical compilers will need to reach a broad consensus on certain topics to ensure that data are measured in a consistent and comparable way. These decisions will continue to be further explored as part of the update of the macroeconomic statistical standards.

Furthermore, practical guidance will be developed to help countries implement the relevant recommendations in order to arrive at high-quality and comparable results across countries. This will leverage off the experience of NSIs compiling first experimental estimates (as explained above). The guidance will focus on issues such as the collection of relevant input data; the derivation of mark-ups for different companies and industries; and asset lives and depreciation schedules for different types of data. In so doing, it will carefully consider the different levels of statistical infrastructure across countries.

Key take-aways

- The explicit inclusion of data into macroeconomic statistics as an asset used repeatedly in production is a priority in the current revision of the SNA.
- The sum-of-cost approach is regarded as the most feasible approach to arrive at high-quality and comparable estimates of investment in data assets. This is consistent with the valuation of other own-account intellectual property products (e.g. software and R&D).
- Data should be considered entirely the result of production, despite the significant and fundamental contribution that non-produced information make to the value of data.
- While data assets are likely be included in the forthcoming revision of the SNA, the development of guidelines for measuring the production and transactions of data will continue even after the SNA is updated.
- Experimental estimates for selected OECD countries suggest the value of data assets is sizeable. However, their inclusion in macroeconomic statistics seems to have a modest effect on growth and productivity as a significant share of data are likely already recorded in the SNA under other intangible assets (e.g. software and databases, and R&D).

5 A measurement agenda for the value of data

In this section, the report provides selected estimates of the value of data, using the United States as an example. It explores how such estimates are wide ranging and depend on the type of data considered and the method employed to estimate their value. It delineates the main axes of a measurement agenda for the value of data, addressing the need to better capture data products and data-related activities, to develop international statistical guidelines and dedicated survey tools, and, finally, to engage different policy and technical communities within the OECD.

5.1 Estimating the value of data

This report has contributed to advance the measurement of the economic value of data. It has done so within the System of National Accounts (SNA), by looking at the value of data sold and purchased on the market as well as the value of data produced by firms for their own use. Table 5.1 summarises the estimates provided in the report, using the United States as an example. The table shows there will generally be a wide range of estimates of the value of data, depending on the type of data considered and the method employed to estimate their value.

Table 5.1. Select estimates of the value of data in the US economy

	USD (billions)	Yearly growth rate
Revenues from direct sales of data, 2019	33	5.6% (nominal) in 2013-19
Data contribution to Internet advertising revenues, 2019	79	16% (nominal) in 2013-19
Yearly investment in own-account data assets, business sector, 2021 (sum-of-cost estimate)	186	4.2% (real) in 2002-21
Net stock of data assets, business sector, 2021 (sum-of-cost estimate)	421	3.8% (real) in 2002-21

Source: Own calculations (see Section 3) and Calderón and Rassier (2022[74]).

5.2 Recommendations

The report has provided several useful findings but has also pointed out areas for further methodological and statistical work. Taken together, these areas delineate the main axes of a measurement agenda for the value of data.

Develop product and industry classifications to help measure the value of data

Although the bulk of data is not traded on the market, this report has shown that market statistics, such as revenues, exports and expected revenue streams – as reflected by venture capital investments – are a key tool for measuring the value of data. However, current product and industry classifications are not suited to delineate data. The United States is the only economy where statistical nomenclatures make it possible to measure – although imperfectly – revenues from the sales of data. Developing product and industry classifications to better capture data products and data-related activities remains a priority for measuring the value of data.

Develop international statistical guidelines to measure data investment and assets

A consensus has emerged on the approach for measuring data assets within the SNA with respect to data produced for firms for their own use. However, the implementation of this approach is still in its infancy and will require considerable further work. Therefore, developing international statistical guidelines for the measurement of data investment and assets will be a major task in the years ahead.

Develop dedicated survey tools and econometric approaches to estimate value of cross-border data flows

Within the scope of economic statistics, the measurement of the value of cross-border data flows remains elusive, as discussed in this report. This situation calls for a good deal of invention: in using information not collected for this purpose; in developing dedicated survey tools, including Internet-based ones; and in applying new econometric approaches to estimate the value of cross-border data flows.

Engage policy makers from diverse disciplines to help measure value of data

The above considerations apply to the measurement of the economic value of data within the SNA framework. However, data have value for consumers, firms, governments, research activities and society at large that goes beyond the scope of macroeconomic statistics. Data may also generate negative value when their use is detrimental to some individuals or organisations. Developing concepts to think about these channels of value creation and statistical frameworks to measure them should be part of any measurement agenda for data. Given the multidisciplinary nature of this task, the engagement of different policy and technical communities within the OECD is an indispensable prerequisite for its success.

References

Acemoglu, D. et al. (2019), "Too much data: Prices and inefficiencies in data markets", *NBER Working Paper*, No. 26296, National Bureau of Economic Research, Cambridge, MA, https://doi.org/10.3386/w26296. [35]

Acquisti, A., L. John and G. Loewenstein (2013), "What Is Privacy Worth?", *The Journal of Legal Studies*, Vol. 42/2, https://doi.org/10.1086/671754. [52]

Agrawal, A., A. Goldfarb and J. Gans (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, Harvard Business Review Press. [28]

Ahmad, N. (2005), "Follow-up to the measurement of databases in the National Accounts", *Canberra II Issues Paper*, No. SNA/M1.06/19.1, OECD, Paris, https://unstats.un.org/unsd/nationalaccount/aeg/papers/m3Databases.PDF. [81]

Ahmad, N., J. Ribarsky and M. Reinsdorf (2017), "Can potential mismeasurement of the digital economy explain the post-crisis slowdown in GDP and productivity growth?", *OECD Statistics Working Papers*, No. 2017/09, OECD Publishing, Paris, https://doi.org/10.1787/a8e751b7-en. [9]

Ahmad, N. and P. van de Ven (2018), "Recording and measuring data in the system of National Accounts", paper presented at the 12th meeting of the Advisory Expert Group (AEG) on National Accounts, Luxembourg, 27-29 November, https://unstats.un.org/unsd/nationalaccount/aeg/2018/M12.asp. [69]

AIOTI (2020), "IoT and edge computing convergence", No. 19 May 2020, AIOTI blog, https://aioti.eu/aioti-strategic-foresight-through-digital-leadership-iot-and-edge-computing-convergence/. [27]

Akerlof, G. (1970), "The market for "lemons": Quality uncertainty and the market mechanism", *The Quarterly Journal of Economics*, Vol. 84/3, p. 488, https://doi.org/10.2307/1879431. [46]

Alcobendas, M., S. Kobayashi and M. Shum (2021), "The impact of privacy measures on online advertising markets", *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.3782889. [82]

Alphabet (2021), "Alphabet announces second quarter 2021 results", 27 July, Press Release, Alphabet, Mountain View, CA, https://abc.xyz/investor/static/pdf/2021Q2_alphabet_earnings_release.pdf?cache=4db52a1. [57]

Andrews, D., C. Criscuolo and P. Gal (2016), "The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy", *OECD Productivity Working Papers*, No. 5, OECD Publishing, Paris, https://doi.org/10.1787/63629cc9-en. [6]

Arthur, C. (2013), “Tech giants may be huge, but nothing matches big data”, 23 August, The Guardian, https://www.theguardian.com/technology/2013/aug/23/tech-giants-data. [83]

Baruffaldi, S. et al. (2020), “Identifying and Measuring Developments in Artificial Intelligence: Making the Impossible Possible”, *OECD Science, Technology and Industry Working Papers*, No. 2020/05, OECD Publishing, Paris, https://doi.org/10.1787/5f65ff7e-en. [31]

Bergemann, D. and A. Bonatti (2019), “Markets for information: An introduction”, *Annual Review of Economics*, Vol. 11/1, pp. 85-107, https://doi.org/10.1146/annurev-economics-080315-015439. [50]

Brynjolfsson, E., D. Rock and C. Syverson (2021), “The productivity J-curve: How intangibles complement general purpose technologies”, *American Economic Journal: Macroeconomics*, Vol. 13/1, pp. 333-372, https://doi.org/10.1257/mac.20180386. [3]

Calderón, J. and D. Rassier (2022), “Valuing the US data economy using machine learning and online job postings”, paper prepared for the 37th IARIW General Conference, https://iariw.org/wp-content/uploads/2022/08/Calderon-Rassier-IARIW-2022.pdf. [74]

Carriere-Swallow, Y. and V. Haksar (2019), “The economics and implications of data: An integrated perspective”, *SPR Departmental Paper*, No. 19/16, International Monetary Fund, Washington, DC. [22]

Casalini, F. and J. López González (2019), “Trade and Cross-Border Data Flows”, *OECD Trade Policy Papers*, No. 220, OECD Publishing, Paris, https://doi.org/10.1787/b2023a47-en. [19]

Chetty, R. et al. (2020), “The economic impacts of COVID-19: Evidence from a new public database built using private sector data”, Opportunity Insights, https://opportunityinsights.org/wp-content/uploads/2020/05/tracker_paper.pdf. [39]

Choi, J., D. Jeon and B. Kim (2019), “Privacy and personal data collection with information externalities”, *Journal of Public Economics*, Vol. 173, pp. 113-124, https://doi.org/10.1016/j.jpubeco.2019.02.001. [51]

Corrado et al., C. (2021), “New evidence on intangibles, diffusion and productivity”, *OECD Science, Technology and Industry Working Papers*, No. 2021/10, OECD Publishing, Paris, https://doi.org/10.1787/de0378f3-en. [7]

Corrado, C. et al. (2022), “Data, digitization, and productivity”, in *Technology, Productivity and Economic Growth*, University of Chicago Press, https://www.nber.org/books-and-chapters/technology-productivity-and-economic-growth/data-digitization-and-productivity. [78]

Coyle, D. et al. (2020), “The Value of Data – Policy Implications”, Bennett Institute and Open Data Institute, Cambridge MA, https://www.bennettinstitute.cam.ac.uk/publications/value-datapolicy-implications/. [38]

Cullin, C. (2019), “2019 Global Internet Phenomena Report”, 10 September, Press Release, Sandvine, San Jose, CA, https://www.sandvine.com/press-releases/sandvine-releases-2019-global-internet-phenomena-report. [65]

de Bondt, H. and N. Mushkudiani (2021), “Estimating the value of data in the Netherlands”, paper prepared for the IARIW-ESCoE Conference, 11-12 November, https://iariw.org/wp-content/uploads/2021/10/bondt_paper.pdf. [72]

de Cornièr, A. and G. Taylor (2020), "Data and competition: A simple framework, with applications to mergers and market structure", *Working Paper*, No. 1076, Toulouse School of Economics, https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/wp/2020/wp_tse_1076.pdf. [53]

Eurostat (2022), *ICT Usage in Enterprises*, (database), https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database (accessed on 27 January 2022). [12]

Furman, J. et al. (2019), *Unlocking Digital Competition*, Report of the Digital Competition Expert Panel, United Kingdom, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf. [4]

Gans, J. and S. Stern (2010), "Is there a market for ideas?", *Industrial and Corporate Change*, Vol. 19/3, pp. 805-837, https://doi.org/10.1093/icc/dtq023. [49]

Gibson, J. (2004), "Re-reifying data", *Notre Dame Law Review*, Vol. 163, pp. 164-242, https://scholarship.richmond.edu/cgi/viewcontent.cgi?article=1113&context=law-faculty-publications. [94]

Goldfarb, A. and C. Tucker (2011), "Privacy regulation and online advertising", *Management Science*, Vol. 57/1, https://doi.org/10.1287/mnsc.1100.1246. [84]

Goldin, I. et al. (2021), "Why is productivity slowing down?", *Oxford Martin Working Paper Series on Economic and Technological Change*, No. 2021-6, Oxford Martin School, University of Oxford, https://www.oxfordmartin.ox.ac.uk/publications/why-is-productivity-slowing-down/. [2]

Goodridge, P. and J. Haskel (2015), "How does big data affect GDP? Theory and evidence for the UK", *Discussion Paper*, No. 2015/06, Imperial College London Business School, https://spiral.imperial.ac.uk/bitstream/10044/1/25156/2/Goodridge_2015_06.pdf. [79]

Goodridge, P. and J. Haskel (2015), "How much is UK business investing in big data?", *Discussion Paper*, No. 2015/05, July, Imperial College London Business School, https://spiral.imperial.ac.uk/bitstream/10044/1/25159/2/Goodridge%202015-05.pdf. [76]

Goodridge, P., J. Haskel and H. Edquist (2021), "We see data everywhere except in the productivity statistics", *The Review of Income and Wealth*, https://doi.org/10.1111/roiw.12542. [77]

ISWGNA (2022), *Recording of Data in the National Accounts*, Intersecretariat Working Group on National Accounts, https://unstats.un.org/unsd/nationalaccount/RAdocs/DZ6_GN_Recording_of_Data_in_NA.pdf. [21]

ISWGNA (2020), *Recording and Valuation of Data in National Accounts*, prepared for the 14th meeting of the Advisory Expert Group on National Accounts, Intersecretariat Working Group on National Accounts, https://unstats.un.org/unsd/nationalaccount/aeg/2020/M14_5_3_1_Recording_of_Data_in_SNA.pdf. [80]

ITU (2021), "Measuring Digital Development. Facts and Figures 2021", web page, https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx (accessed on 1 May 2022). [63]

Johnson, G. (2013), "The impact of privacy policy on the auction market for online display advertising", *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.2333193. [85]

Johnson, G., S. Shriver and S. Du (2020), “Consumer privacy choice in online advertising: Who opts out and at what cost to industry?”, *Marketing Science*, Vol. 39/1, pp. 33-51, https://doi.org/10.1287/mksc.2019.1198. [60]

Jones, B. and L. Summers (2020), “A calculation of the social returns to innovation”*, Working Paper Series*, No. 27863, National Bureau of Economic Research, Cambridge, MA, http://www.nber.org/papers/w27863. [86]

Jones, C. and C. Tonetti (2020), “Nonrivalry and the economics of data”, *American Economic Review*, Vol. 110/9, https://doi.org/10.1257/aer.20191330. [33]

Ker, D. and E. Mazzini (2019), “Perspectives on the value of data and data flows”*, OECD Digital Economy Papers*, No. 299, OECD Publishing, Paris, https://doi.org/10.1787/a2216bc1-en. [56]

Kerry, C. (2019), “Should consumers be able to sell their own personal data? No: It would do little to help consumers, and could leave them worse off than now”, 13 October, Wall Street Journal, https://www.wsj.com/articles/should-consumers-be-able-to-sell-their-own-personal-data-11570971600 (accessed on 16 December 2021). [88]

Koutroumpis, P., A. Leiponen and L. Thomas (2020), “Markets for data”, *Industrial and Corporate Change*, Vol. 29/3, https://doi.org/10.1093/icc/dtaa002. [42]

Levin, R., W. Cohen and D. Mowery (1985), “R & D appropriability, opportunity, and market structure: New evidence on some Schumpeterian hypotheses”, *The American Economic Review*, Vol. 75/2, pp. 20-24, http://www.jstor.org/stable/1805564. [48]

Lin, Y. et al. (2012), “Syntactic annotations for the Google Books Ngram Corpus”, *July, Proceedings of the ACL 2012 System Demonstrations*, pp. 169-174, https://dl.acm.org/doi/10.5555/2390470.2390499. [11]

López González, J., F. Casalini and J. Porras (2022), “A Preliminary Mapping of Data Localisation Measures”*, OECD Trade Policy Papers*, No. 262, OECD Publishing, Paris, https://doi.org/10.1787/c5ca3fed-en. [17]

McCallum, J. (2021), “Memory Prices 1957+”, webpage, https://jcmit.net/memoryprice.htm (accessed on 10 January 2022). [26]

Meta (2021), *Q2 2021 Earnings*, 28 July, Meta, https://investor.fb.com/investor-events/event-details/2021/Facebook-Q2-2021-Earnings-/default.aspx. [58]

Mitchell, J., D. Ker and M. Lesher (2021), “Measuring the economic value of data”*, OECD Going Digital Toolkit Notes*, No. 20, OECD Publishing, Paris, https://doi.org/10.1787/f46b3691-en. [75]

Moravec, H. (2022), “Processor List”, webpage, https://frc.ri.cmu.edu/~hpm/book97/ch3/processor.list.txt (accessed on 5 January 2022). [25]

Nelson, P. (1970), “Information and consumer behavior”, *Journal of Political Economy*, Vol. 78/2, pp. 311-329, http://www.jstor.org/stable/1830691. [44]

Nguyen, D. and A. Paczos (2020), “Measuring the economic value of data and cross-border data flows: A business perspective”*, OECD Digital Economy Papers*, No. 297, OECD Publishing, Paris, https://doi.org/10.1787/6345995e-en. [66]

OECD (2022), “Data shaping firms and markets”, *OECD Digital Economy Papers*, No. 344, OECD Publishing, Paris, https://doi.org/10.1787/7b1a2d70-en. [8]

OECD (2022), *EBOPS 2010 - Trade in Services by Partner Economy (database)*, http://dotstat.oecd.org/Index.aspx?DataSetCode=TISP_EBOPS2010 (accessed on 6 April 2022). [62]

OECD (2022), “Fostering cross-border data flows with trust”, *OECD Digital Economy Papers*, No. 343, OECD Publishing, Paris, https://doi.org/10.1787/139b32ad-en. [18]

OECD (2022), “Responding to societal challenges with data: Access, sharing, stewardship and control”, *OECD Digital Economy Papers*, No. 342, OECD Publishing, Paris, https://doi.org/10.1787/2182ce9f-en. [13]

OECD (2022), *STAN STructural ANalysis Database (SNA08, ISIC REV.4 version)*, (database), http://oe.cd/stan (accessed on 18 February 2022). [10]

OECD (2021), *Recommendation of the Council on Enhancing Access to and Sharing of Data*, OECD/LEGAL/0463, OECD, Paris, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463. [20]

OECD (2020), *Competition in digital advertising markets*, https://www.oecd.org/daf/competition/competition-in-digital-advertising-markets-2020.pdf. [54]

OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en. [29]

OECD (2019), *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD Publishing, Paris, https://doi.org/10.1787/276aaca8-en. [16]

OECD (2019), *Measuring the Digital Transformation: A Roadmap for the Future*, OECD Publishing, Paris, https://doi.org/10.1787/9789264311992-en. [64]

OECD (2015), *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris, https://doi.org/10.1787/9789264229358-en. [92]

OECD (2009), *Measuring Capital – OECD Manual 2009: Second edition*, OECD Publishing, Paris, https://doi.org/10.1787/9789264068476-en. [91]

Ostrom, E. (2010), “Beyond markets and states: Polycentric governance of complex economic systems”, *American Economic Review*, Vol. 100/3, pp. 641-672, http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.3.641. [34]

Pilat, D. (2004), “The economic impacts of ICT – What have we learned thus far”, paper prepared for the 4th ZEW Conference on the economics of information and communication technologies, Manheim, 2-3 July. [89]

Posner, E. and G. Weyl (2018), *Radical Markets*, Princeton University Press, Princeton & Oxford. [37]

Prüfer, J. and C. Schottmüller (2021), “Competing with big data”, *The Journal of Industrial Economics*, Vol. 69/4, pp. 967-1008, https://doi.org/10.1111/joie.12259. [5]

Reichman, J. and P. Samuelson (1997), “Intellectual property rights in data?”, *Vanderbilt Law Review*, Vol. 50, pp. 52-166, https://scholarship.law.duke.edu/faculty_scholarship/455. [93]

Reimsbach-Kounatze, C. (2021), "Enhancing access to and sharing of data: Striking the balance between openness and control over data", in *Data Access, Consumer Interests and Public Welfare*, Nomos, Baden-Baden, https://doi.org/10.5771/9783748924999-25. [40]

Reinsdorf, M. and J. Ribarsky (2019), "Measuring the digital economy in macroeconomic statistics: The role of data". [70]

Shapiro, C. and H. Varian (1999), *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business Review Press, Boston, MA. [36]

Smedes, M., T. Nguyen and B. Tenburren (2022), "Valuing data as an asset, implications for economic measurement", paper prepared for Economic Implications of the Digital Economy Conference, 9-10 March, https://www.abs.gov.au/system/files/documents/7bfccb4ddb8aded818330bebe6b76b14/Smedes%20-%20Valuing%20data%20as%20an%20asset.pdf. [73]

Spiekermann, M. (2019), "Data marketplaces: Trends and monetisation of data goods", *Intereconomics*, Vol. 54/4, pp. 208-216, https://doi.org/10.1007/s10272-019-0826-z. [43]

Statistics Canada (2019), "The value of data in Canada: Experimental estimates", *Latest Developments in the Canadian Economic Accounts*, 19 July, Statistics Canada, Ottawa, https://www150.statcan.gc.ca/n1/en/pub/13-605-x/2019001/article/00009-eng.pdf?st=ifEOEPUK (accessed on 3 February 2022). [71]

Teece, D. (1980), "Economies of scope and the scope of the enterprise", *Journal of Economic Behavior & Organization*, Vol. 1/3, pp. 223-247, https://doi.org/10.1016/0167-2681(80)90002-5. [47]

The Economist (2017), "The world's most valuable resource is no longer oil but data", 6 May, The Economist, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data. [1]

Tonetti, C. (2019), "Should consumers be able to sell their own personal data? Yes: It would encourage sharing of data – a plus for consumers and society", 13 October, Wall Street Journal, https://www.wsj.com/articles/should-consumers-be-able-to-sell-their-own-personal-data-11570971600 (accessed on 16 December 2021). [90]

Twitter (2021), "Twitter announces second quarter 2021 results", 22 July, Press Release, Twitter, San Francisco, CA, https://s22.q4cdn.com/826641620/files/doc_financials/2021/q2/Q2'21-Earnings-Release.pdf (accessed on 13 January 2022). [59]

UNCTAD (2015), "International Trade in ICT Services and ICT-enabled Services: Proposed Indicators from the Partnership on Measuring ICT for Development", *Technical Note*, No. 2, United Nations Conference on Trade and Development, Geneva, https://unctad.org/system/files/official-document/tn_unctad_ict4d03_en.pdf. [67]

United States Patent and Trademark Office (2021), *PatentsView*, website, https://patentsview.org/ (accessed on 27 January 2022). [32]

UNSD, W. et al. (2009), *2008 System of National Accounts*, United Nations Statistcs Division, Eurostat, International Monetary Fund, OECD, World Bank. [68]

US Bureau of Economic Analysis (2022), “Nonresidential: Intellectual property products: Software”, *Private Fixed Investment, Chain-type Price Index*, Federal Reserve Bank of St. Louis (FRED), https://fred.stlouisfed.org/series/B985RG3A086NBEA (accessed on 11 January 2022). [23]

US Bureau of Economic Analysis (2022), “Personal consumption expenditures”, *Chain-type Price Index*, Federal Reserve Bank of St. Louis (FRED), https://fred.stlouisfed.org/series/DPCERG3A086NBEA (accessed on 11 January 2022). [24]

US Census Bureau (2019), *2019 Service Annual Survey*, US Census Bureau, https://www.census.gov/programs-surveys/sas.html (accessed on 4 February 2022). [61]

US Census Bureau (2017), *2017 Economic Census*, https://data.census.gov/cedsci/advanced (accessed on 13 January 2022). [55]

US Supreme Court (1991), *Feist Publications, Inc., v. Rural Telephone Service Co.*, https://supreme.justia.com/cases/federal/us/499/340/. [87]

Varian, H. (2019), “Artificial intelligence, economics, and industrial organization”, in Agrawal, A., J. Gans and A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press. [30]

Vining, A. and D. Weimer (1988), “Information asymmetry favoring sellers: A policy framework”, *Policy Sciences*, Vol. 21/4, pp. 281-303, http://www.jstor.org/stable/4532147. [45]

World Bank (2021), *World Development Report 2021: Data for Better Lives*, World Bank, Washington, DC, http://data https://datacatalog.worldbank.org/search/dataset/0037991 (accessed on 19 June 2020). [41]

Yu, P. (2019), “Data producer’s right and the protection of machine-generated data”, *Tulane Law Review*, Vol. 93, pp. 859-929, https://scholarship.law.tamu.edu/facscholar/1318/ (accessed on 5 January 2022). [95]

Zuboff, S. (2019), “Surveillance capitalism and the challenge of collective action”, *New Labor Forum*, Vol. 28/1, https://doi.org/10.1177/1095796018819461. [14]

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Public Affairs, New York. [15]

Notes

[1] The phrase “data is the new oil” is commonly attributed to British mathematician and entrepreneur Clive Humby (Arthur, 2013[83]).

[2] Left panel: OECD-21 include Australia, Belgium, Canada, Chile, Germany, Denmark, Spain, Finland, France, Ireland, Italy, Japan, Korea, Luxembourg, the Netherlands, Norway, New Zealand, Portugal, Sweden, the United Kingdom and the United States. OECD-12 include Australia, Belgium, Canada, Denmark, Finland, France, Germany, Italy, the Netherlands, Sweden, the United Kingdom and the United States. Right panel: countries included are Austria, Belgium, Denmark, Estonia, Finland, France, Greece, Hungary, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden and the United Kingdom. Industries included pertain to the business economy with the exception of financial services and real estate activities. ISIC-rev.4 codes C10-12, C13-15, C16-18, C19-23, C24-25, C26, C27-28, C29-30, C31-33 (all manufacturing), D-E (energy and utilities), F (construction), G45, G46, G47, I, J, M and N (all services).

[3] There are other definitions of data, including in work by the OECD (2015[92]).

[4] See Jones and Summers (2020[86]) for an example of the large literature.

5 Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden and the United Kingdom.

6 For instance, in *Feist Publications, Inc., v. Rural Telephone Services Co.*, the United States Supreme Court established that information alone cannot be protected by copyright (US Supreme Court, 1991[87]). However, other jurisdictions, notably the European Union, recognise "database rights" as a *sui generis* right distinct from copyright as the latter is not available for databases that aim to be complete. For legal perspectives on intellectual property rights in data, see Reichman and Samuelson (1997[93]), Gibson (2004[94]) and Yu (2019[95]).

7 Legal frameworks such as the European Union's General Data Protection Regulation allocate specific rights to the data subject (e.g. a Facebook user), implicitly leaving residual rights to the data collector (Meta) as a purely de facto property right. In addition, there is uncertainty as the legal and regulatory environment for data markets – and especially for markets for personal data – is a matter of considerable public debate and might evolve. For instance, while some commentators and scholars have argued for giving consumers ownership rights over their data (Tonetti, 2019[90]), others have argued that data ownership is not feasible or conceptually flawed (Kerry, 2019[88]).

8 An example in case is the reputational cost incurred by Facebook following the fraudulent use of its data by Cambridge Analytica.

9 Neither the NAPCS nor the NAICS are designed to track revenues from data sales. However, it cannot be ruled out that significant data sales are recorded under nondescript code such as "database design and development services" or "other information services".

10 While the inclusion of databases and other collections of information, directories and mailing lists is straightforward, credit rating services are included because of the activities that firms in the associated industry, credit bureaus, engage in: "compiling information, such as credit and employment histories on individuals and credit histories on businesses, and providing the information to financial institutions, retailers, and others who have a need to evaluate the creditworthiness of these persons and businesses". Also, credit rating agencies are often mentioned as examples of so-called data brokers.

11 Other studies are provided by Alcobendas, Kobayashi and Shum (2021[82]), Johnson (2013[85]), Goldfarb and Tucker (2011[84]). In these studies, estimates of the contribution of personal information to Internet advertising range from 38.5% to 65%.

12 The online display advertising industry opted to implement the AdChoices programme as a means of self-regulation in 2010. AdChoices provides notice by superimposing the AdChoices logo over the corner of online display ads and enables consumers to opt out of behaviourally targeted advertising. Consumers who opt out still see ads, just not ads that are targeted based on their previous browsing behaviour.

13 Again, it is important to note that the contribution to data to revenues from Internet advertising might be considerably less than total revenues – the trends are indicative, not the levels.

14 The reasons for this are set out in Ahmad (2005[81]), a document presented in discussions leading up to the 2008 SNA revision. In summary, there was a concern this inclusion would lead to the inadvertent capitalisation of knowledge.

[15] Updates of the SNA have taken place at irregular intervals, the last dating back to 2008 and the one before that to 1993. The updates are usually predicated on a concern that certain elements of the economy are no longer appropriately recorded in a manner required by users or consistent with other economic statistical standards.

[16] In this way, the introduction of data into the SNA is likely similar to that of R&D in the SNA2008. In that revision, the capitalisation of R&D came with the following important caveat: “R&D should be recognized as part of capital formation. In order to achieve this, several issues have to be addressed. These include deriving measures of R&D and development, price indices and service lives. Specific guidelines, together with handbooks on methodology and practice, will provide a useful way of working towards solutions that give the appropriate level of confidence in the resulting measures” (SNA2008 §10.104).

[17] Consumption of fixed capital is an SNA term synonymous with the accounting concept of depreciation.

[18] Observable phenomena are “facts or situations whose characteristics or attributes may be recorded” (Mitchell, Ker and Lesher, 2021[75]).

[19] While this lists of costs is relatively straightforward, certain assumptions, such as the occupations chosen, the share of time certain occupations spend producing data, or the exact operating surplus mark-up, will certainly require more research to converge at broadly accepted assumptions.

[20] Total data assets consists of data, database and data science.

[21] Estimates of the capital stock can be obtained from investment data based on the perpetual inventory method. For more information, see OECD (2009[91]).

[22] By comparison, similar studies conducted for ICT investment more broadly typically find that they accounted for 0.3%-0.9% of growth in GDP per capita over 1995-2002 (Pilat, 2004[89]).

[23] This is an instance of the more fundamental challenge of adjusting for differences in quality when output is measured based on the sum-of-cost approach. In the case of data, differences in quality may be due to the intrinsic information embedded in the underlying observable phenomena rather than to production activities.

[24] While this is not a unique problem – other intangible assets such as software and R&D face similar measurement questions – the systematic addition of data points to many data assets magnifies the importance of the decision. This is in clear contrast to most existing fixed assets within the SNA, which are produced once and used repeatedly as an input into production, from which they steadily decline in value over time.