

# PERSPECTIVES ON THE VALUE OF DATA AND DATA FLOWS

---

OECD DIGITAL ECONOMY  
PAPERS

December 2020 **No. 299**

Note to Delegations:

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

DSTI/CDEP/MADE(2019)6/FINAL

This document, as well as any data and any 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 or third party. 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.

© OECD (2020)

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

## *Table of Contents*

<b>Executive Summary .....</b>	<b>6</b>
<b>Chapter 1. Introduction .....</b>	<b>8</b>
<b>Chapter 2. Data storage .....</b>	<b>9</b>
2.1. Introduction .....	9
2.2. Classification of data storage expenditures .....	11
2.3. Data and estimates .....	13
2.4. Concluding remarks.....	18
<b>Chapter 3. Output of firms compiling and selling databases .....</b>	<b>20</b>
3.1. Introduction .....	20
3.2. Identifying relevant firms .....	20
3.3. Data and estimates.....	25
3.4. Concluding remarks.....	49
<b>Chapter 4. Data and firm values .....</b>	<b>50</b>
4.1. Introduction .....	50
4.2. Linking data holdings with firm valuations.....	50
4.3. Identifying “data-driven firms” .....	50
4.4. Findings .....	52
4.5. Concluding remarks.....	54
<b>Chapter 5. Perspectives on the value of cross-border data flows.....</b>	<b>56</b>
5.1. Introduction .....	56
5.2. Trade in products delivered through data flows .....	58
5.3. Concluding remarks.....	80
<b>References .....</b>	<b>83</b>
Annex 5.A. Industry to product classification concordance.....	87
<b>Notes .....</b>	<b>98</b>

## **Tables**

Table 2.1. Data storage and database management software product categories .....	12
Table 2.2. Available data for data storage and database software products, United States, 2012-2017.....	14
Table 2.3. Estimated use of data storage hardware, software, and services, United States, 2012-2017.....	15
Table 2.4. Estimated expenditure on data storage devices hardware and media, United States, 2012-2016.....	17
Table 3.1. ISIC Rev.4. Industries potentially related to database compilation and distribution.....	21
Table 3.2. CPC v2.1 products related to database compilation and distribution .....	23
Table 3.3. Definitions of “residual” activities .....	24
Table 3.4. ISIC Rev.4 to US NAICS 2017 conversion .....	26
Table 3.5. Data availability and detail in different sources .....	26
Table 3.6. Data on industries compiling and selling datasets, United States, 2012-2017. ....	28
Table 3.7. ISIC Rev.4 to Canada NAICS 2017 conversion.....	40
Table 3.8. Estimated supply of products related to compiling and selling databases. Canada, 2011-15.....	42
Table 5.1. ISIC rev. 4 Section J Information and Communication services.....	60
Table 5.2. UNCTAD “potentially ICT-enabled services” .....	72

## Figures

Figure 2.1. Investment in computer software and databases and ICT equipment, 2017 .....	10
Figure 2.2. Estimated use of data storage hardware, software, and services, United States, 2012-2017.....	16
Figure 2.3. Estimated expenditure on data storage devices, hardware and media, United States, 2012-2016 .....	18
Figure 3.1. 51114 Directory and mailing list publishers' revenue, by media type, 2012-2017.....	29
Figure 3.2. 51114 Directory and mailing list publishers' revenue, by product type, 2012-2017 .....	30
Figure 3.3. 51114 Directory and mailing list publishers' revenue, by detailed product type, 2012 .....	31
Figure 3.4. 51114 Directory and mailing list publishers' revenue from activities related to compiling and selling databases, 2012-2017 .....	32
Figure 3.5. 519130 Internet Publishing and Broadcasting and Web Search Portals' revenue, by product type, 2012-2017.....	33
Figure 3.6. 519130 Internet Publishing and Broadcasting and Web Search Portals' revenue, by detailed product type, 2012 .....	34
Figure 3.7. 519130 Internet Publishing and Broadcasting and Web Search Portals' estimated revenue from activities related to compiling and selling databases, 2012-2017 .....	34
Figure 3.8. 5415 Computer systems design and related services' revenue, by product type, 2012-2017 .....	35
Figure 3.9. 5415 Computer systems design and related services' revenue, by product type, 2012.....	36
Figure 3.10. 5415 Computer systems design and related services' estimated revenue from activities related to compiling and selling databases, 2012-2017 .....	37
Figure 3.11. 561450 Credit Bureaus' revenue, by detailed product type, 2012 .....	38
Figure 3.12. 561450 Credit Bureaus' estimated revenue from activities related to compiling and selling databases, 2012-2017.....	38
Figure 3.13. Estimated revenue from activities related to compiling and selling databases, by industry, United States, 2012-2017 .....	39
Figure 3.14. 5415 Computer systems design and related services industry revenue, from database design and development services, Canada, 2013-2017.....	41
Figure 3.15. Estimated supply of products relating to compiling and selling databases, by industry, Canada, 2012-2017.....	43
Figure 3.16. Directory and mailing list publishers' turnover, European countries, 2017 or latest .....	44
Figure 3.17. Computer programming activities' turnover, European countries, 2017 or latest .....	45
Figure 3.18. Computer programming activities' estimated turnover from compiling and selling databases, European countries, 2017 or latest.....	46
Figure 3.19. Activities of collection agencies and credit bureaus' turnover, European countries, 2017 or latest .	47
Figure 3.20. Estimated turnover from activities related to compiling and selling databases, by industry, European countries, 2017 or latest.....	48
Figure 3.21. Total estimated turnover from activities related to compiling and selling databases, European countries, 2017 or latest.....	48
Figure 4.1. Market capitalisation of "data-driven firms", 1985-2020 .....	52
Figure 4.2. Average market capitalisation of "data-driven firms", 1985-2020 .....	53
Figure 4.3. Index of average growth of market capitalisation of "data-driven firms", 1985-2020.....	54
Figure 5.1. The conceptual framework for digital trade .....	59
Figure 5.2. Exports by firms in ISIC section J – Information and communication, USD billions .....	61
Figure 5.3. Exports by firms in ISIC section J – Information and communication, export share.....	62
Figure 5.4. International trade in ISIC J Information and Communication services – bi-lateral flows .....	63
Figure 5.5. Exports of services produced by firms in the Information and Communication services sector (ISIC section J).....	66
Figure 5.6. Imports of services produced by firms in the Information and Communication services sector (ISIC section J).....	66
Figure 5.7. Exports of services produced by firms in the Information and Communication services sector (ISIC section J) – including estimations .....	67
Figure 5.8. Imports of services produced by firms in the Information and Communication services sector (ISIC section J) – including estimations .....	67
Figure 5.9. Exports of services produced by firms in the Information and Communication services sector (ISIC section J).....	69

Figure 5.10. Imports of services produced by firms in the Information and Communication services sector (ISIC section J) .....	70
Figure 5.11. Exports of digitally deliverable services .....	73
Figure 5.12. Imports of digitally deliverable services .....	74
Figure 5.13. Exports of <i>digitisable products</i> .....	76
Figure 5.14. Imports of <i>digitisable products</i> .....	77
Figure 5.15. International trade underpinned by data flows .....	79
Figure 5.16. “Linking” data on international data flows to international trade .....	81

## *Perspectives on the value of data and data flows*

### **Executive Summary**

Data are becoming increasingly central to business activities. Associated with this, databases are becoming an increasingly valuable form of asset; even relatively minimal data flows can create considerable economic value.

Despite this, attempts to conceptualise and measure the value of data remain relatively underdeveloped. This paper explores four different perspectives from which the value of data, databases, and data flows can be conceptualised and measured:

- The first perspective takes data from use tables (part of the supply-use framework) and business statistics databases (expenditures on data storage products – hardware, software, and services) as a potential avenue for gaining an alternative measurement perspective. It is estimated that firms in the United States spent over USD 36 billion on storing data in 2017 (comprising 0.25% of total intermediate consumption), and that overall expenditure on data storage is increasing over time. These estimates are possible because the United States publishes detailed expenditure by product line data, as well as supply-side product detail. Extending the approach to other countries will rely on additional details becoming available.
- A second perspective uses business statistics sources to look at the revenues generated by firms creating explicit value from data – those collecting, compiling, and selling databases. Activities related to compiling and selling databases are estimated to have been worth over USD 60 billion in the United States in 2017, around USD 1.4 billion in Canada, and around EUR 19-50 billion in the European Union in 2016. This wide range reflects differing estimation parameters arising from the relatively less detailed information on these activities being available for European countries.
- A third perspective considers whether data holdings and analysis may be contributing to firm valuations. “Data-driven firms” are identified from web-scraped information, matching these to stock market tickers, and web-scraping market capitalisation data for those firms. On average, the value of the identified data-driven firms has grown faster than other firms listed on the NASDAQ and NYSE over the period from 1986-2020 and the identified firms have a combined market capitalisation of USD 5 trillion in 2020. Further work is needed to extend the approach to firms listed on other stock markets. Other measures of firm performance could also be used including enterprise value, profitability, and productivity.
- Finally, the fourth perspective looks at potential links between trade flows and data flows, comparing different measures of the trade in “digitally deliverable products” – defined from several perspectives. Overall the evidence suggests that there are a number of global hubs for trade in these products. The United States appears to be the largest by quite some margin but others include Germany, India, Ireland, France, the Netherlands, and the United Kingdom.

Overall, it appears that businesses are spending increasing amounts on storing data and that firms are effectively compiling and monetising data-based products. Furthermore, firms in

a number of countries are playing a key role in the international trade in digitally deliverable products. Such data-driven business models appear likely to support firm performance in terms of market valuations.

This adds to evidence on the importance of data-based business models for businesses and national economies. It therefore reinforces the need for policies that support the development of digital infrastructure needed to power innovation in products and business models. It also points to the need to balance firms' abilities to access and analyse data alongside key issues around privacy and the rights individuals and firms have over data that relate to them (OECD, 2019<sup>[1]</sup>).

The approaches outlined can be improved and enhanced. Notably, an improvement of the product detail available in many countries' statistical sources would greatly help in extending the coverage of the perspectives set out in Chapter 1. and Chapter 2. to more countries (or reducing the need for estimation based on information from other countries). This detail does not necessarily need to be published but could be shared directly with the Secretariat. The United States provides a good example of the considerable detail published in supply-use tables and underlying business statistics.

Chapter 4. sets out a prototype looking at the valuations of "data-driven firms" using web-scraped data on market capitalisation. As well as the potential to enhance this analysis by looking at additional metrics such as enterprise value, profitability, and productivity, it would benefit from developing better ways to identify "data-driven firms" and relevant data sources – especially for countries beyond the United States. In particular, responses on business ICT usage surveys could provide one approach for identifying relevant firms.

Finally, Chapter 5. set out several ways in which products that can be delivered through international data flows ("digitally deliverable products") can be identified. Each of these has merits and limitations in terms of how well they focus in on the relevant products. This could be built upon by trying to move on from dimply identifying products that can be digitised for delivery to estimate the "data intensity" of different activities – providing more of a link to the volume of underlying data flows. Another enhancement could be to broaden the scope to identify economic activities that *rely on* international data flows (e.g. for scale, profitability, etc.).



## Chapter 1. Introduction

Data has been described as a vital *infrastructure* or even like *sunlight* (because it is everywhere and underpins everything) (The Economist, 2020<sup>[2]</sup>). Whatever your preferred analogy, it is evident that firms are managing to draw together large volumes of data, apply advanced analytics to spot patterns and gain insights, and monetise these in various ways. This is exemplified by a relatively small number of well-known “superstar” firms, but there are undoubtedly many more firms that are embracing the power of data to enhance or power their business models.

For firms selling products that are digitally-delivered, databases and data flows are an inextricable part of their business models. Take, for example, a streaming music service. It gathers data – in the form of music files and metadata - to create a database of content, and then sells access to that content to subscribing customers. That access creates, and relies on, a bi-directional flow of data between the customer and the streaming music company: searches for songs, instructions (e.g. play, pause, etc.), streaming of the music itself, “like” button clicks, etc.

Nevertheless, businesses that do not deliver their products by digital means are also increasingly realising the value of data. Companies which are not “data-enabled” are becoming increasingly “data enhanced” by using data to improve production processes, efficiency, or end products, or even by selling or licencing data they hold (Nguyen and Paczos, 2020<sup>[3]</sup>).

Data can clearly be of value – most starkly illustrated by the stock-market valuations of companies built almost entirely around collating and monetising databases (e.g. Web Search providers, online social networks). However, that value is highly context-dependent, being affected not just by the information content of the data itself but also by the context in which it was collected, and the context in which the data are to be used. The value, therefore, may be only loosely related to the volume of data held or transmitted.

Furthermore, the economic value of data for businesses is challenging to conceptualise and measure directly. This paper explores several perspectives from which the value of data, databases, and data flows might be conceptualised and measured. The following chapters address each of these in turn:

- The first uses data from supply-and-use tables and business statistics databases to look at expenditures on data storage products – hardware, software, and services – which underpin databases and other holdings of data.
- The second uses business statistics sources to look at the revenues generated by firms creating explicit value from data – those collecting, compiling, and selling database.
- The third perspective tries to identify “data-driven firms” and examine whether their market capitalisation out-performs other firms.
- The final perspective examines the economic value associated with international trade underpinned by data flows.



## Chapter 2. Data storage

### 2.1. Introduction

Databases are undoubtedly an increasingly important source of economic value for businesses and others. That said, databases are only briefly mentioned as an example of intangible assets in the International (business) Accounting Standard on intangible assets (IAS 38). Furthermore, under the standard, they would only be recorded as assets in business financial statements “if, and only if: i) it is probable that the future economic benefits that are attributable to the asset will flow to the entity; and ii) the cost of the asset can be measured reliably”. This is straightforward in cases where databases are purchased through market transactions. However, it is likely that the vast majority of databases are produced in-house. Unlike for software, Research and Development, internally generated brands, the standard does not provide any particular specification on how the costs of internally developed databases should be recorded and valued. As such it is unlikely that firms are routinely recognising databases as assets in their own accounts.

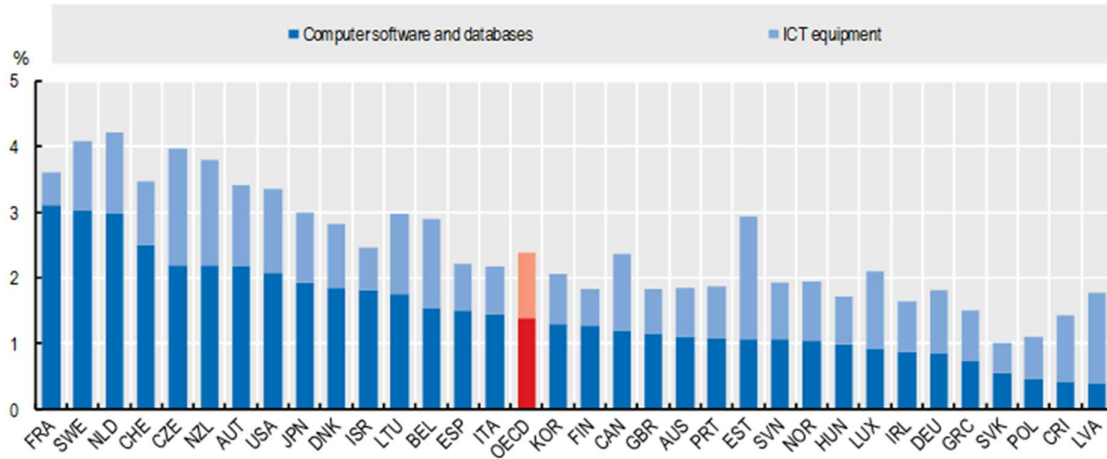
Nevertheless, databases are considered assets in National Accounts. The 2008 System of National Accounts (SNA) para. 10.112-10.114 (UN, 2010<sup>[4]</sup>), states that all databases holding data with a useful life of more than one year should be recorded as Intellectual Property Products assets providing they meet the general definition of an asset (i.e. are expected to provide benefits to their owners and over which ownership rights are exercised). However, it is not completely clear what this means in practice. Firms have all kinds of databases: for HR purposes, for clients, for inventories, etc. The OECD Handbook on Deriving Capital Measures of Intellectual Property Products (OECD, 2009<sup>[5]</sup>) provides the following recommendation (number 38): “a database should be recorded as a fixed asset if a typical datum is expected to be stored on the database, or archived on a secondary database, for more than a year”. In a world of rapidly decreasing data storage costs, as well as growing awareness of the value of large datasets, it is likely that very few databases do not meet this criterion.

As such, National Accounts figures should, in principle, include provide broad coverage of a wide range of databases. **Figure 2.1** presents the National Accounts figures covering databases.

However, there are some important limitations to these figures. The most striking is that, in practice, databases are usually incorporated in a single series along with software. Only a few countries make separate estimates for “software” and “databases” (Eurostat-OECD Task Force on Land and Other non-financial assets, 2020<sup>[6]</sup>) and these estimates are not included in the OECD or Eurostat databases<sup>1</sup>.

**Figure 2.1. Investment in computer software and databases and ICT equipment, 2017**

As a percentage of GDP



Note: Investment is based on gross fixed capital formation. For Germany, Korea and Spain, ICT equipment are estimates based on last available share. Data for Poland relate to 2015.

Source: Adapted from OECD (2019<sup>[7]</sup>); based on OECD, National Accounts Statistics; Eurostat, National Accounts Statistics and national sources, February 2019.

A second, foundational challenge relates to the way in which investment in databases, even if available separately, is estimated. The standard SNA practice in cases where value cannot be observed directly through a market transaction (e.g. a sale) is to use the “sum-of-costs” approach. This involves adding up the costs associated with producing the assets in question. The underlying assumption is that the resulting assets must be worth at least as much as is spent (by rational economic agents with full information) on producing them.

This naturally leads to the question “what costs should be included?”. According to 2008 SNA para. 10.113:

*“The creation of a database will generally have to be estimated by a sum-of-costs approach. The cost of the data base management system (DBMS) used should not be included in the costs but be treated as a computer software asset unless it is used under an operating lease. The cost of preparing data in the appropriate format is included in the cost of the database but not the cost of acquiring or producing the data. Other costs will include staff time estimated on the basis of the amount of time spent in developing the database, an estimate of the capital services of the assets used in developing the database and costs of items used as intermediate consumption.”*

As such, the database management system (DBMS) is excluded, on the basis of being software. Furthermore, although the SNA definition explicitly mentions ‘files of data’, para. 10.113 excludes the cost of “acquiring or producing the data”. This leads to the observation that intellectual property asset sub-item AN.11732 Databases consists of neither the database management system nor the content (data)<sup>2</sup>. The following elements are those included: i) the cost of preparing data in the appropriate format; ii) staff time spent on developing the database; iii) capital services of the assets used in developing the database; and iv) costs of items used as intermediate consumption.

The “Report on Intellectual Property Products” of the Joint Eurostat – OECD Task Force on Land and other non-financial assets (2020<sup>[6]</sup>) provides a more detailed explanation of

the challenges associated with this, including highlighting that it is likely to be very difficult to delineate those costs that should be included from those that should be excluded. Furthermore, “staff time”, the main component of the costs covered is often only roughly estimated based on occupational data and parameters which may not be strongly based in empirical reality. For the purposes of this paper, it suffices to note that, regardless of such measurement challenges, the relatively narrow boundary around the costs to be included is likely to lead to National Accounts figures providing a relatively restricted estimate of the overall value of databases.

Furthermore, because databases are treated as assets in National Accounts, only the expenditures used in initially producing or materially improving the database are included. Nevertheless, there can be significant on-going costs to maintaining a large database, even if no additional data is added to it. It is still necessary to keep the server on which the database is stored running or, increasingly often, to pay for its storage in the cloud<sup>3</sup>. This chapter will therefore seek an alternative perspective on firms’ expenditures related to databases by examining their spending on data storage equipment, media, services, and related software.

The overarching assumption of such an approach is that rational businesses will only spend money on storing data if the value they expect in return equals or exceeds the cost of storage, database management etc. Nevertheless, it should be borne in mind that, in a world of plummeting data storage costs and increasing understanding of the (potential) value of data, the decision to collect and store data is ever less-likely to be constrained by costs in comparison to other constraints, such as privacy legislation.

## 2.2. Classification of data storage expenditures

A number of standard statistical classifications of products are used internationally. The central UN standard is the Central Product Classification (CPC) (United Nations, 2015<sup>[8]</sup>), which also forms the basis for the Eurostat Classification of Products by Activity (CPA) (Eurostat, 2015<sup>[9]</sup>), and provides a key foundation for the North American Product Classification System (NAPCS) used by the US Census Bureau (US Census Bureau, n.d.<sup>[10]</sup>), and Statistics Canada (Statistics Canada, 2018<sup>[11]</sup>)<sup>4</sup>. Several other countries, such as Japan, implement bespoke product classifications.

**Table 2.1** begins with the CPC 2.1, seeking to identify the product classes which contain data storage hardware and media products, database management software, and data storage services (e.g. cloud storage) products. These are then mapped to the CPA 2.1 and 2017 NAICS (as well as the precursor “2012 product classification” used in the United States).

Most of the products identified in the CPC map straightforwardly to the CPA and NAICS, although the extent of breakdowns varies. However, the NAICS tends to adopt a different breakdown at the lower levels. In particular, the NAICS includes a specific product class for “servers”. By contrast, the recording of servers as a form of product is not mentioned in the CPC or CPA.

Having identified the products to be targeted, it is next necessary to seek data on these.

**Table 2.1. Data storage and database management software product categories**

CPC, CPA, NAPCS.

	Code	Product description	Detailed description	Other product classification	ISIC4
Storage hardware	4527	Storage units		CPA 26.20.21  US NAPCS 2011625000 / 2041900000 Computer storage devices (334112)  CAN NAPCS 3611122 Data storage devices (including flash media)	2620
	45271	Fixed media storage units	- hard disk drives and other storage units where the storage media is physically inseparable from its reading or writing device	CPA na? US NAPCS uses different breakdown. CAN NAPCS no breakdown.	2620
	45272	Removable media storage units	- storage units designed for removable media: • disk storage units • magnetic tape storage units • compact disk (CD) drives • DVD drives • flash card readers	CPA na? US NAPCS uses different breakdown CAN NAPCS no breakdown.	2620
				US NAPCS 2011500006 Manufacture of Servers (3341111105) CAN NAPCS 3611115 Servers	
Storage media	475	Disks, tapes, solid-state non-volatile storage devices and other media, not recorded		CPA 26.80 Magnetic and optical media  NAPCS 2046800000 Magnetic and optical recording media (334613)  CAN NAPCS 3622131 unrecorded optical and magnetic media, and media n.e.c. (includes "optical and magnetic disc packs and media, for data processing")	
	47530	Magnetic media, not recorded, except cards with a magnetic stripe	This subclass includes: - magnetic tapes, unrecorded: • for recording sound • for recording video • for other recordings - magnetic tape cartridges, unrecorded - cassettes, unrecorded: • for recording video • for recording sound • for other recordings - magnetic floppy disks, unrecorded	CPA 26.80.11 NAPCS 2046800006 Magnetic recording media, including parts (3346130802)	2680
	47540	Optical media, not recorded	This subclass includes: - unrecorded disks prepared for laser-beam recording: • compact disks (CDs) • DVDs • other optical media	CPA 26.80.12 NAPCS 2046800003 Optical discs (3346130801)	2680
	47550	Solid-state non-volatile storage devices	Solid-state, non-volatile data storage devices for recording data from an external source [flash memory cards or flash electronic storage cards]	CPA 26.20.22	2620

Storage services (e.g. cloud storage)	83159	Other hosting and IT infrastructure provisioning services	This subclass includes: <ul style="list-style-type: none"> <li>- co-location services.</li> <li>- <b>data storage services, i.e. managing or administrating the storage and back-up management of data such as remote back-up services, storage, or hierarchical storage management (migration)</b></li> <li>- data management services</li> <li>- video and audio streaming services</li> <li>- other IT hosting or infrastructure provisioning services</li> </ul>	CPA63.11.19 Other hosting and IT infrastructure provisioning services  US NAPCS 7014350000 / 77107010305 Data storage services (36140)  CAN NAPCS 7511142 Data storage infrastructure provisioning services	6311
Database management software	47813	Database management software, packaged	This subclass includes: <ul style="list-style-type: none"> <li>- published collections/suites of software programs that enable storage, modification and extraction of information from a database. There are many different types of DBMSs ranging from small systems that run on personal computers to large systems that run on mainframes</li> </ul>	CPA 58.29.13 NAPCS 7008975009 Database management software publishing (34883)  CAN NAPCS 4821111 System and programming software (except commercial use licences), on physical media (includes database management software)	5820

*Note:* codes in brackets relate to the US “2012 product codes” used in various sources.

*Source:* OECD based on (United Nations, 2015<sup>[8]</sup>), (Eurostat, 2015<sup>[9]</sup>), (US Census Bureau, n.d.<sup>[10]</sup>), (US Census Bureau, 2019<sup>[12]</sup>), (Statistics Canada, 2018<sup>[11]</sup>)

## 2.3. Data and estimates

This section examines the data available in various countries with the aim of deriving estimates of the extent of expenditures on or use of products that underpin databases (data storage and database management software). As the focus is on the use of products, rather than the production of those products, the use tables (from the supply-use framework) and underlying data on firm expenditures are a natural starting-point.

### 2.3.1. United States

The US Census Bureau publishes several data sources which complement the US Bureau of Economic Analysis annual supply-use tables, as well as the detailed supply-use tables published every five years (US Bureau of Economic Analysis, n.d.<sup>[13]</sup>). These include the annual Service Annual Survey (SAS) Tables (US Census Bureau, 2018<sup>[14]</sup>), and the Economic Census product line data (US Census Bureau, 2012<sup>[15]</sup>). The latter provide the greatest degree of detail, although the latest product line data currently available relate to 2012. The agedness of much of the detailed data available is a key limitation in this analysis as it is likely that the situation has evolved markedly since 2012. Data for 2017 are due to be published in 2020 and will provide a valuable additional source with which to enrich the analysis.

**Table 2.2** presents the available data, which cover all the targeted products. It is important, however, to note that the measurement approaches and concepts underpinning these figures can vary. The aim is to estimate the value of data storage and database management software products *used* or *purchased* in the United States. However, the Economic Census

Product line data come from the *supply-side*, relating to sales receipts for the relevant products. The same is true for some data from the SAS.

**Table 2.2. Available data for data storage and database software products, United States, 2012-2017**

USD millions, current prices

	Product	Measure	Source		2012	2013	2014	2015	2016	2017
Hardware and media	334 Computer and Electrical Products	Intermediate use + non-residential investment	Annual use tables	a	625 962	641 785	660 879	684 388	687 449	758 061
	334 Computer and Electrical Products	Product shipments value (supply-side)	2012 Economic Census by product line	b	294 542					
	334112 Data storage devices	Intermediate use + non-residential investment	2012 detailed use table	c	12 035					
	334112 Data storage devices	Product shipments value (supply-side)	2012 Economic Census by product line	d	8 619					
	334111105 Servers			e	1 250					
	334618 Magnetic and optical media			f	321					
	Share c/a	Intermediate use + non-residential investment	Supply-use tables		1.9%					
	Share d/b	Product shipments value (supply-side)	2012 Economic Census by product line		2.9%					
	Share e/b				0.4%					
	Share f/b				0.1%					
Software	511 Publishing, except Internet (includes software)	Intermediate use + non-residential investment	Annual use tables		234 044	246 947	264 847	277 110	298 038	319 241
	Database management software	Total receipts from products (supply-side)	Services Annual Survey	g	272 152	279 253	291 575	293 771	308 489	317 850
				h	11 293	12 737	13 364	13 838	13 913	14 294
	Share h/g				4.1%	4.6%	4.6%	4.7%	4.5%	4.5%
Services	514 Data processing, Internet publishing, and other information services	Intermediate use	Annual use tables		117 688	128 127	138 905	148 228	162 483	175 770
		Total receipts from products (supply-side)	2012 Economic Census by product line		99 204					
	Data processing and other purchased computer services	Estimated Selected Expenses for Employer Firms, select services industries (use-side)	Services Annual Survey		54 661	64 322	55 915	82 224	88 436	96 933
	Data processing, hosting, and related services	Total receipts from products (supply-side)	2012 Economic Census by product line	i	70 298					
	Data storage services			j	2 435					
	Share j/i				3.5%					

Source: OECD, based on (US Bureau of Economic Analysis, n.d.<sub>[13]</sub>), (US Census Bureau, 2018<sub>[14]</sub>), (US Census Bureau, 2012<sub>[15]</sub>).

That there are differences between the measures available from these sources is clearly illustrated by the disparities between several of the series. The total intermediate and non-residential investment use in 2012 of the parent product category “334 Computer and Electrical Products” was USD 626 billion, while the Product shipments value available from the 2012 Economic Census by product line was less than half that at USD 300 billion. However, these come from opposite sides of the supply-use framework as well as from different sources based on different concepts. One key difference is that use also takes account of imports of the relevant products, and it is likely that much of the storage hardware used is produced outside the US. Indeed OEC data show that the United States is the top importer of computer data storage units, while the People’s Republic of China (hereafter “China”) is the top exporter (OEC, n.d.<sup>[16]</sup>).

For that reason, it is preferable to use use-side information (such as row c) when making imputations. However, in several cases, only supply side information is available. This may be relatively less of an issue when it comes to software and cloud services products, in which the United States is comparatively specialised. Nevertheless, applying supply-side shares in order to impute product breakdowns relies on the strong assumption that the products produced domestically are representative of all those consumed, including imports. That is especially unlikely in hardware; therefore, servers and storage media will be omitted from the analysis - only the share of data storage devices in the use of computer and electronic products, available from the 2012 detailed use table, will be used to derive estimates for hardware.

With those caveats in hand, **Table 2.3** presents estimates for data storage hardware and media, database management software, and storage services respectively.

**Table 2.3. Estimated use of data storage hardware, software, and services, United States, 2012-2017**

USD millions, current prices								
	Product	Source	2012	2013	2014	2015	2016	2017
Hardware	334 Computer and electronic Products	Annual use tables	625 962	641 785	660 879	684 388	687 449	758 061
	2012 detailed use table share	2012 detailed use table	1.9%					
	Data storage devices		12 035	12 339	12 706	13 158	13 217	14 575
Software	511 Publishing, except Internet (includes software)	Annual use tables	234 044	246 947	264 847	277 110	298 038	319 241
	Average share in receipts, 2012-2017	Services Annual Survey	4.5%					
	Database management software		10 537	11 117	11 923	12 475	13 417	14 372
Services	514 Data processing, Internet publishing, and other information services	Annual use tables	117 688	128 127	138 905	148 228	162 483	175 770
	2012 receipts share adjusted for average excess growth in data processing and other purchased computer services*	Economic Census product line data and SAS	3.5%	3.7%	3.8%	4.1%	4.3%	4.5%
	Data storage services		4 076	4 677	5 345	6 012	6 946	7 920

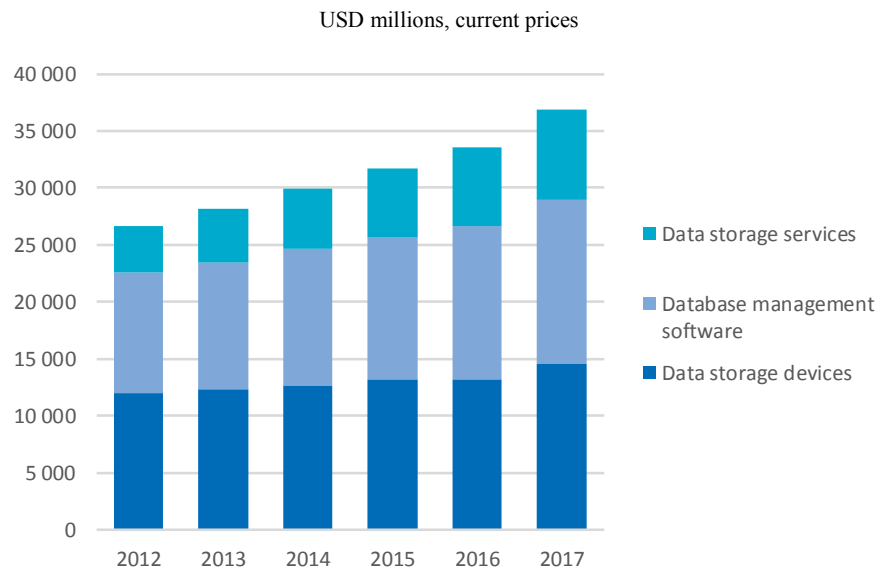
Note: \* adjusts for the average extent that annual growth of firms’ expenditure on “data processing and other computer services” (14% annual average growth) exceeds the annual growth in use figures for “data processing, Internet publishing, and other information services” (8% annual average growth).

Source: OECD based on (US Bureau of Economic Analysis, n.d.<sup>[13]</sup>), (US Census Bureau, 2018<sup>[14]</sup>), (US Census Bureau, 2012<sup>[15]</sup>).



The resulting estimates are presented in **Figure 2.2**, which shows that use of data storage products in the United States is estimated to have grown rapidly from around USD 26 billion in 2012 to over USD 36 billion in 2017 – comprising 0.25% of total intermediate consumption. Furthermore, as no adjustment is made for price changes; it is likely that prices per unit for these products have generally been decreasing, these estimates may understate the extent of growth over the period. The 2017 US Economic Census by product line data, due to become available in November 2020 will provide new insights into the extent of growth in these products.

**Figure 2.2. Estimated use of data storage hardware, software, and services, United States, 2012-2017**



Source: OECD based on (US Bureau of Economic Analysis, n.d.<sup>[13]</sup>), (US Census Bureau, 2018<sup>[14]</sup>), (US Census Bureau, 2012<sup>[15]</sup>).

Data storage devices and database management software are approximately equal components of the estimate. Meanwhile, data storage services appear relatively smaller. However, it should be borne in mind that this product class is likely to relate mainly to those services (including cloud services) specialised in data storage. Many online services include some aspect of online data storage. For example, Microsoft and Google online Office tools (document, spreadsheet, etc.) would likely be classified in “application service provisioning” but undoubtedly provide subscribed businesses with the ability to store data in the cloud.

The above offer an initial set of estimates for the United States. The approach could also be extended down to the industry level. However, the limitations of relying on detailed product data from 2012, and of using supply-side information to estimate product use are clear.

An alternative approach which can work at the whole-economy level, is to take total domestic output of the relevant products from the supply tables, subtract exports, and add imports. While this would still rely on detailed product breakdowns from other sources, the relevant calculations at least occur on the supply-side. However, this approach requires

sufficiently detailed import and export data allowing the relevant products to be identified. Investigations only found such data for trade in data storage units and unrecorded magnetic and optical storage media. **Table 2.4** presents the derivation of estimates based on the available data.

It should be noted that, the import and export flows include a large volume of re-exports. On average, only one-third of computer storage devices and 45% of magnetic and optical recording media exported are manufactured in the USA. This can cause the volume of exports to be greater than the value of domestic production.

Regrettably, figures for imports and exports of servers are not available, so an aggregate including all the hardware and media product classes identified cannot be derived.

**Table 2.4. Estimated expenditure on data storage devices hardware and media, United States, 2012-2016**

USD millions, current prices

Product	Measure	Source	2012	2013	2014	2015	2016
			Product shipments value (domestic production)				
Computer storage devices	Product shipments value	Annual Survey of Manufactures	8 619	9 072	8 138	7 441	6 631
Magnetic and optical recording media			321	233	452	446	345
Host computers, multiusers (mainframes, super computers, midrange servers, blade servers, UNIX servers, PC servers)			4 769	4 559	5 464	5 724	5 372
Servers		2012 Economic Census product line data	4 769				
2012 servers share			1 250				
2012 servers share			26%				
Servers (estimate based on 2012 share)				1 250	1 195	1 433	1 501
			Exports				
Computer storage devices	Total value of goods exported	USA Trade online	-6 457	-6 096	-6 003	-5 417	-5 840
Magnetic and optical recording media			-1 201	-1 299	-1 411	-1 385	-1 669
Servers				na	na	na	na
			Imports				
Computer storage devices	Customs Import Value (General imports)	USA Trade online	13 141	11 998	11 483	11 127	10 005
Magnetic and optical recording media			3 362	3 829	4 894	5 920	7 364
Servers				na	na	na	na
			Estimated expenditure on storage hardware and media				
Computer storage devices			15 304	14 974	13 619	13 152	10 795
Magnetic and optical recording media			2 482	2 763	3 935	4 980	6 039
Servers			na	na	na	na	na

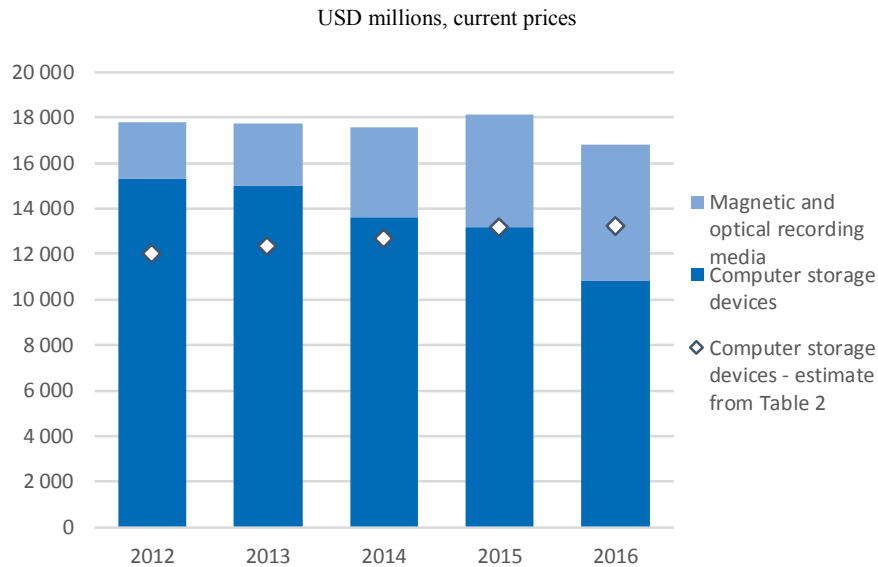
Note: na="not available". In some cases the volume of exports is greater than the value of domestic production. This arises because the import and export flows include a large volume of re-exports. On average, only one third of computer storage devices and 45% of magnetic and optical recording media exported are manufactured in the USA.

Source: OECD based on (US Census Bureau, 2018<sup>[17]</sup>), (US Census Bureau, 2012<sup>[15]</sup>), (US Census Bureau, n.d.<sup>[18]</sup>).

**Figure 2.3** presents these estimates, which total around USD 17-18 billion annually between 2012 and 2016. For comparison, the figure also presents the computer storage devices estimates from Table 2.2 / Figure 2.2. It is notable that both sets of estimates are of the same order of magnitude, each averaging around USD 13 billion over the period. However, the trends over time move in opposite directions, with the estimates based on the

use tables increasing over time while the estimates based on revenue, exports, and imports decline over time.

**Figure 2.3. Estimated expenditure on data storage devices, hardware and media, United States, 2012-2016**



Source: OECD based on (US Census Bureau, 2018<sup>[17]</sup>), (US Census Bureau, 2012<sup>[15]</sup>), (US Census Bureau, n.d.<sup>[18]</sup>).

### 2.3.2. Other countries

Use tables available from Statistics Canada and Eurostat present only higher level aggregates such as CPA 26 “computer, electronic, and optical products”, CPA 58 “Publishing services”, and CPA 62\_63 “computer programming, consultancy, and related services; information services”. Furthermore, it does not appear that detailed product-line data are published. This work was presented to the OECD Working Party on Measuring and Analysing the Digital Economy in November 2019 and delegates were encouraged to share any similar sources available for their countries. However, this did not yield any submissions with comparable detail. As such, extension of this approach across countries will rely on additional detail becoming available.

## 2.4. Concluding remarks

The above offers two perspectives on data storage expenditures in the United States. To the extent that storage capacity is, presumably, being used to store data, it also gives an implicit view on the annual appetite for data storage products. However, neither approach offers full coverage of the relevant product classes identified in Table 2.1. Furthermore, while “computer storage devices” is covered in both approaches, there is an important caveat to consider. Unlike servers, database management software, and intermediate consumption of cloud storage services, which are most likely to be directly consumed by the businesses concerned, computer storage units may be used directly for storage *or*

incorporated into products for sale (e.g. personal computers, smartphones, etc.). Further development is needed to try to identify the size of the latter component.

Conversely, it is also possible that a business might not store their data(bases) using specific data storage products but may, for example, simply buy a desktop or laptop computer with sufficient hard drive space (this is, perhaps, most likely in small businesses). Such products are not captured above, being recorded instead as purchases of computer hardware. Nevertheless, the databases in such cases may no doubt be a crucial asset enabling the business to operate (e.g. a database of customers and prospects).

Nevertheless, this analysis adds a new measurement perspective in an area that is somewhat underserved by existing statistics (e.g. the combined “software and databases” measures from National Accounts). Further work is needed to develop appropriate approaches for application to other countries.

## Chapter 3. Output of firms compiling and selling databases

### 3.1. Introduction

In the examples given in **Chapter 1** and in many other cases, “data” are a means-to-an end rather than the core product itself. In contrast, this section will focus on activities where data is “the product”: the production and sale of compilations of facts and information – i.e. databases.

Those activities may take several forms. Firstly, the compilation may directly involve gathering individual data items to produce a database or may instead (or as well) focus on procuring and re-using existing databases. Meanwhile, sales may take various forms including a straightforward sale of the database, subscription access, licencing etc. For convenience, these different forms will be referred to as selling or monetising the database.

Indeed, access to some of the largest databases so far curated – Internet search providers’ indexes of Internet addresses and content; Social Networks’ vast holdings of individuals information and content – is only “sold” implicitly. Users agree (knowingly or unknowingly) that, in return for access to the service, their searches and other information can be used to sell advertising services to companies. However, this is different from direct sales of the databases and is not the focus of this analysis.

The following section seeks to identify the industries in which firms engaging in compiling and selling databases are classified. Following this, the data available for various countries is considered and estimates of the scale of activities related to compiling and selling databases are derived. The final section draws out conclusions, highlights key challenges and areas of for further work.

### 3.2. Identifying relevant firms

There are several ways to try to identify firms engaged in compiling and selling databases. The first is to begin with the International Standard Industrial Classification of all Economic Activities (ISIC) Rev.4. (United Nations, 2008<sup>[19]</sup>) and identify relevant industry descriptions.

There are three ISIC industry descriptions which specifically mention databases. Furthermore, there are several other industries which may potentially also be relevant, as summarised in **Table 3.1**.

“5812 Publishing of directories and mailing lists” specifically states that this industry is characterised by publishing databases that are “protected in their form, but not in their content”. It is not completely clear where a firm publishing similar databases, but for which the content itself is subject to Intellectual Property Protection (IPP), would be classified.

“6201 Computer Programming activities” covers “designing the structure and content of, and/or writing the computer code necessary to create and implement [...] databases”. However, it is not completely clear whether the phrase “structure and content of” refers to the computer code, or the database itself. Only in the latter case would this activity appear to be sufficiently related to database compilation and production.

“6312 web portals” also mentions databases in the specific context of search engines and other portals and media sites. In the business models of Internet search engines, web portals (e.g. MSN, Yahoo!), content portals (e.g. YouTube), Social networks, and the like, these databases are not sold directly but used to target advertising at content consumers.

**Table 3.1. ISIC Rev.4. Industries potentially related to database compilation and distribution**

No.	Name	Description	Notes
5812	Publishing of directories and mailing lists	Publishing of lists of facts/information ( <b>databases</b> ) that are protected in their form, but not in their content. These lists can be published in printed or electronic form. This class includes: <ul style="list-style-type: none"> <li>publishing of mailing lists</li> <li>publishing of telephone books</li> <li>publishing of other directories and compilations, such as case law, pharmaceutical compendia, etc.</li> </ul>	Specifies that underlying content is not protected.
5819	Other publishing activities	<ul style="list-style-type: none"> <li>publishing (including on-line) of: catalogues; photos; engravings and postcards; greeting cards; forms; posters; reproduction of works of art; advertising material; other printed matter</li> <li><b>on-line publishing of statistics or other information</b></li> </ul> <i>Excludes: retail sale of software, publishing advertising newspapers, on-line software.</i>	Appears to relate to statistics rather than datasets.
6201	Computer programming activities	The writing, modifying, testing and supporting of software. <ul style="list-style-type: none"> <li>designing the structure and content of, and/or writing the computer code necessary to create and implement: systems software (including updates and patches); software applications (including updates and patches); <b>databases</b>; web pages</li> <li>customizing of software, i.e. modifying and configuring an existing application so that it is functional within the clients' information system environment.</li> </ul> <i>Excludes publishing packaged software.</i>	Focused primarily on (database) software design and structure.
6311	Data processing, hosting and related activities	<ul style="list-style-type: none"> <li>provision of infrastructure for hosting, data processing services and related activities</li> <li>specialized hosting activities such as: Web hosting; streaming services; application hosting</li> <li>application service provisioning</li> <li>general time-share provision of mainframe facilities to clients</li> <li>data processing activities: complete processing of data supplied by clients; generation of specialized reports from data supplied by clients</li> <li><b>provision of data entry services</b></li> </ul>	Focused on infrastructure, processing and analysis of data supplied by clients. Provision of data entry services (i.e. to populate/maintain customer's database).
6312	Web portals	<ul style="list-style-type: none"> <li>operation of web sites that use a search engine to generate and maintain extensive <b>databases</b> of Internet addresses and content in an easily searchable format</li> <li>operation of other websites that act as portals to the Internet, such as media sites providing periodically updated content</li> </ul>	These databases are not typically a product for sale.
8291	Activities of collection agencies and credit bureaus	<ul style="list-style-type: none"> <li>collection of payments for claims and remittance of payments collected to the clients, such as bill or debt collection services</li> <li><b>compiling of information</b>, such as credit and employment histories on individuals and credit histories on businesses and <b>providing the information</b> to financial institutions, retailers and others who have a need to evaluate the creditworthiness of these persons and businesses</li> </ul>	Clearly involves compiling databases and selling them/extracts from/access to them.

Source: Adapted from (United Nations, 2008<sub>[19]</sub>).

Several other industry descriptions also include activities which may be of relevance. “5819 Other publishing activities” includes “on-line publishing of statistics or other information, although this appears focussed more on statistics than datasets so will be excluded from further analysis. “6311 Data processing, hosting and related activities” mainly relates to infrastructure and provision of services where client-supplied data are processed, analysed, and reported on. Nevertheless, it does include “provision of data entry services” which could relate to populating a client’s database. However, as this is one specific activity among many listed and it is highly unlikely that a data source separating out data entry can be found, this will also be excluded from further analysis at this time.

Finally, within “8291 Activities of collection agencies and credit bureaus” credit bureaus are clearly involved in compiling and selling databases (or extracts from/access to them). The availability of separate data on this activity alone will be investigated.

A second approach starts with the products that firms produce. The Central Product Classification (CPC) v2.1 (United Nations, 2015<sup>[8]</sup>), describes a number of products that appear to be relevant, as summarised in **Table 3.2**. As well as industries “5812 Publishing of directories and mailing lists”, “6201 Computer programming activities”, and “6312 Web portals”, this approach suggests several other industries are worthy of consideration.

With respect to “5820 Software Publishing”, the products covered appear to be database management software rather than databases populated with data. For this reason, and for the reason that this product category includes all other forms of software publishing, this industry will not be considered further in the analysis. A further, practical reason, is that US Economic Census product line data do not appear to be available for this industry.

“7490 Other professional, scientific and technical activities n.e.c” includes the activity “compilation services of facts and information (i.e. databases), other than mailing lists”. A clear implication of this wording is that mailing lists are also databases and hence also relevant.

A challenge with these products is that they map to residual, “not elsewhere classified” (n.e.c.) industries as defined in **Table 3.3**. Furthermore, it is not necessarily clear which activity listed within these industries the products map across to. On this basis, these industries will not be further considered in this analysis.



**Table 3.2. CPC v2.1 products related to database compilation and distribution**

No.	Name	Description	Associated ISIC
47692	Text-based disks, tapes or other physical media	Physical media (CD-ROM, diskette, microfilm, microfiche etc.) containing non-audio recordings of text material, such as: <ul style="list-style-type: none"> <li>from periodicals, textbooks, reference books and other sources</li> <li>directories and mailing lists</li> <li><b>databases</b> and similar compilations</li> </ul>	5812 Publishing of directories and mailing lists
73312	Licensing services for the right to use databases	Licensing services for the right to reproduce, distribute or incorporate <b>databases</b> (i.e. compilations of facts/information) in other <b>databases</b> or applications. This applies to various levels of licensing rights: <ul style="list-style-type: none"> <li>rights to reproduce and distribute the <b>database</b></li> <li>rights to use <b>database</b> components for the creation of and inclusion in other databases and applications</li> </ul> <i>Excludes: packaged (non-customized) databases, cf. 478 [packaged software]; licensing services for the right to use database software.</i>	
83940	Original compilations of facts/information	Original compilations of facts or information (i.e. <b>databases</b> ) organised for retrieval and consultation, including mailing lists. These compilations are protectable in their presentation, but not their content. <i>Excludes: compilation services of facts and information, cf. 83990; mailing list compilation services, cf. 85952</i>	
84313	On-line directories and mailing lists	<ul style="list-style-type: none"> <li>on-line directories and mailing lists, including telephone books</li> <li>other on-line collections of facts/information (<b>databases</b>)</li> </ul> <i>Excludes: web search portal content, cf. 84394</i>	
47813	Database management software, packaged	Published collections/suites of software programs that enable storage, modification and extraction of information from a <b>database</b> . There are many different types of DBMSs ranging from small systems that run on personal computers to large systems that run on mainframes.	5820 Software Publishing
83141	IT design and development services for applications	Services of designing the structure and/or writing the computer code necessary to create and/or implement a software application, such as: <ul style="list-style-type: none"> <li>designing the structure and content of a web page and/or writing the computer code necessary to create and implement a web page</li> <li><b>designing the structure and content of a database and/or writing the computer code necessary to create and implement a database</b></li> <li>designing the structure and writing the computer code necessary to design and develop a custom software application</li> <li>customization and integration, adapting (modifying, configuring, etc.) and installing an existing application so that it is functional within the clients' information system environment</li> </ul> <i>Excludes service contracts where design and development is bundled with on-going hosting and management.</i>	6201 Computer programming activities
83990	Other professional, technical and business services, n.e.c.	<ul style="list-style-type: none"> <li>Drafting services (detailed layouts, drawings, plans and illustrations of buildings, structures, systems or components from engineering and architectural specifications, done by architectural draftsmen or engineering technicians)</li> <li>compilation services of facts and information (i.e. <b>databases</b>), other than mailing lists</li> </ul> <i>Excludes mailing list compilation services, cf. 85952</i>	7490 Other professional, scientific and technical activities n.e.c
84394	Web search portal content	Content provided on web search portals, i.e. extensive <b>databases</b> of Internet addresses and content in an easily searchable format	6312 Web portals
85952	Mailing list compilation services	Services consisting of compiling lists of names and addresses from telephone directories and other sources <i>Excludes compilation services of facts and information, cf. 83990</i>	8299 Other business support service .

Source: Adapted from (United Nations, 2015<sup>[8]</sup>).

**Table 3.3. Definitions of “residual” activities**

7490 Other professional, scientific and technical activities n.e.c	8299 Other business support service activities n.e.c.
<p>This class includes a great variety of service activities generally delivered to commercial clients. It includes those activities for which more advanced professional, scientific and technical skill levels are required, but does not include ongoing, routine business functions that are generally of short duration.</p> <p>This class includes:</p> <ul style="list-style-type: none"> <li>• translation and interpretation activities</li> <li>• business brokerage activities, i.e. arranging for the purchase and sale of small and medium-sized businesses, including professional practices, but not including real estate brokerage</li> <li>• patent brokerage activities (arranging for the purchase and sale of patents)</li> <li>• appraisal activities other than for real estate and insurance (for antiques, jewellery, etc.)</li> <li>• bill auditing and freight rate information</li> <li>• activities of quantity surveyors</li> <li>• weather forecasting activities</li> <li>• security consulting</li> <li>• agronomy consulting</li> <li>• environmental consulting</li> <li>• other technical consulting</li> <li>• activities of consultants other than architecture, engineering and management consultants</li> <li>• This class also includes:</li> <li>• activities carried on by agents and agencies on behalf of individuals usually involving the obtaining of engagements in motion picture, theatrical production or other entertainment or sports attractions and the placement of books, plays, artworks, photographs etc., with publishers, producers etc.</li> </ul>	<p>This class includes:</p> <ul style="list-style-type: none"> <li>• providing verbatim reporting and stenotype recording of live legal proceedings and transcribing subsequent recorded materials, such as: court reporting or stenotype recording services; public stenography services</li> <li>• real-time (i.e. simultaneous) closed captioning of live television performances of meetings, conferences</li> <li>• address bar coding services</li> <li>• bar code imprinting services</li> <li>• fundraising organization services on a contract or fee basis</li> <li>• mail pre-sorting services</li> <li>• repossession services</li> <li>• parking meter coin collection services</li> <li>• activities of independent auctioneers</li> <li>• administration of loyalty programmes</li> <li>• other support activities typically provided to businesses not elsewhere classified</li> </ul>

Source: Adapted from (United Nations, 2008<sup>[19]</sup>).

As such, the list of industries that will be focussed on for this exploration is:

- “5812 Publishing of directories and mailing lists”
- “6201 Computer programming activities”
- “8291 Activities of collection agencies and credit bureaus”

Even so, there are some clear challenges and issues with each of these:

- A wide range of similar activities not related to databases are included in “6201 Computer programming activities”. In order to focus on databases it is necessary to find a way to focus in on them. A similar issue arises for “8291 Activities of collection agencies and credit bureaus”, within which only credit bureau activities are relevant.
- It seems clear that, “5812 Publishing of directories and mailing lists” covers only databases where the content is not subject to IPP. It is not clear where firms publishing databases with contents protected by IPP would be classified.
- It should also be considered that, based upon the mapping of products to industries given in the CPA, it appears the actual act of compiling the directories and mailing lists might, in theory, be classified elsewhere (in industry 7490 or 8299). However,

in practice this would only be likely to be a problem if compilation and publication took place in separate institutional units. As this seems relatively unlikely, and publication is likely to be the main driver of value added, these units would usually be classified in industry 5812.

- Nevertheless, this alludes to a wider issue with choosing to base analysis on industries when it is products (in this case databases) that are of primary interest. Firms are classified to industries based on the activities that are their main source of value-added (though sometimes revenue or headcount are used as proxies). However, it is common for firms to produce secondary products which are not their main source of value-added and hence may not be reflected in their industry classification. There is considerable evidence that database compilation and indeed monetisation are often additional or ancillary to firms' main activities and are likely to be such secondary activity (Nguyen and Paczos, 2020<sup>[3]</sup>).
- An approach based on industry classification will, by definition, focus only upon those firms for which database compilation and monetisation is a primary activity and hence may significantly understate the overall extent of the phenomenon. However, an advantage is that there tends to be wider availability of data broken down across industries than for products in general.

With those challenges and caveats in mind, the following section examines the availability of data on these activities.

### 3.3. Data and estimates

This section examines the data available in various countries with the aim of deriving estimates of the scale of activities related to compiling and selling databases. The starting point is the United States, which as well as being home to a vibrant data-driven economy, also offers a variety of sources offering detailed industry and product information.

#### 3.3.1. United States

As a first step toward using the US data it is necessary to assess how the ISIC industries identified map across to the North American Industry Classification System (NAICS). This is set out in **Table 3.4**, which is based on the Census Bureau concordance table (US Census Bureau, 2017<sup>[20]</sup>). It can be seen that NAICS offers both challenges and opportunities in comparison to the ISIC structure. Industry “5812 Publishing of directories and mailing lists” maps across to NAICS “511140 Directory and Mailing List Publishers”, however elements relating to online directory, collection, and mailing list publishing are mapped to “519130 Internet Publishing and Broadcasting and Web Search Portals”. The latter is a very broad industry encompassing a wide range of activities so it will be necessary to find a way to focus in on relevant activities.

ISIC “6201 Computer programming activities” maps directly to NAICS “541511 Custom computer programming services”. Finally, NAICS splits the activities of credit bureaus from those of collection agencies, offering an opportunity to focus on the most relevant of these activities.

**Table 3.4. ISIC Rev.4 to US NAICS 2017 conversion**

ISIC code	ISIC title	NAICS code	NAICS title
5812	Publishing of directories and mailing lists	511140	Directory and Mailing List Publishers
		519130	Internet Publishing and Broadcasting and Web Search Portals (on-line directory, collection, and mailing list publishing)
6201	Computer programming activities	541511	Custom Computer Programming Services
8291	Activities of collection agencies and credit bureaus	561440	Collection Agencies
		561450	Credit Bureaus

Source: OECD, based on (US Census Bureau, 2017<sup>[20]</sup>)

1. Having identified the relevant industries, it is necessary to examine the availability of relevant data, as well as the level of detail or “closeness” to the specific NAICS sub-industries of interest available across different sources. **Table 3.5** shows that the Services Annual Survey (SAS) provides total operating figures for the two NAICS related to publishing of directories and mailing lists. A key benefit of this source is that it provides annual data. By contrast, the other sub-industries identified are only isolated in the Economic Census data, available for 2012 and 2017, or the detailed supply-use tables, available for 2012 only.

**Table 3.5. Data availability and detail in different sources**

NAICS industry	Annual SUT	Services Annual Survey	Detailed SUT (2012)	Economic Census
	Domestic supply of commodities by industries	Revenue	Domestic supply of commodities by industries	“Sales, value of shipments, or revenue”
511140 Directory and Mailing List Publishers	511 Publishing industries, except Internet (includes software)	511140 Directory and Mailing List Publishers	5111A0 Directory, mailing list, and other publishers	511140 Directory and Mailing List Publishers
519130 Internet Publishing and Broadcasting and Web Search Portals	518-519 Data processing, Internet publishing, and other information services	519130 Internet Publishing and Broadcasting and Web Search Portals	519130 Internet publishing and broadcasting and Web search portals	519130 Internet Publishing and Broadcasting and Web Search Portals
541511 Custom Computer Programming Services	5415 Computer systems design and related services	5415 Computer Systems Design and Related Services	541511 Custom computer programming services	541511 Custom computer programming services
561450 Credit Bureaus	561 Administrative and support services	561 Administrative and support services	5614 Business support services	561450 Credit Bureaus

Source: OECD, based on (US Bureau of Economic Analysis, n.d.<sup>[13]</sup>), (US Census Bureau, 2018<sup>[14]</sup>), (US Census Bureau, 2012<sup>[15]</sup>)

**Table 3.6** presents the most detailed data available from these sources, as well as figures for the parent industry aggregates found in the Annual Supply Tables, allowing disparities in figures from different sources to be seen and also for the share of the specific industry of interest in the parent industry aggregate to be calculated.

It should be noted that these sources do not measure precisely the same concepts. The Supply Tables provide estimates of the value of total production by these industries according to National Accounts concepts. These include various adjustments and reclassifications in comparison to underlying total operating revenue data from the SAS or Economic Census. In general, the supply figures are around 10% greater than the underlying revenue amounts.

Table 3.6 also shows that the sub-industries of interest can comprise vastly different shares of the parent industry aggregates. “511140 Directory and Mailing List Publishers” only makes up around 1.5-3.7% of “5111 Publishing Industries except Internet”. Similarly, “561450 Credit Bureaus” drive only around 1% of total activity in “561 Administrative and support services”.

By contrast, “541511 Custom computer programming services” makes up around 40% of “5415 Computer systems design and related services” and “519130 Internet Publishing and Broadcasting and Web Search Portals” is around half of “518-519 Data processing, Internet publishing, and other information services” despite the latter combining two high level sectors. However, this is likely to be related to the fact that, even at the most specific levels, these industries are likely to contain a considerable amount of activity that is not related to compiling or selling data and databases. In order to investigate further, we will now look at each of the selected industries in turn using the most detailed sources available for each.

**Table 3.6. Data on industries compiling and selling datasets, United States, 2012-2017.**  
USD Millions, current prices

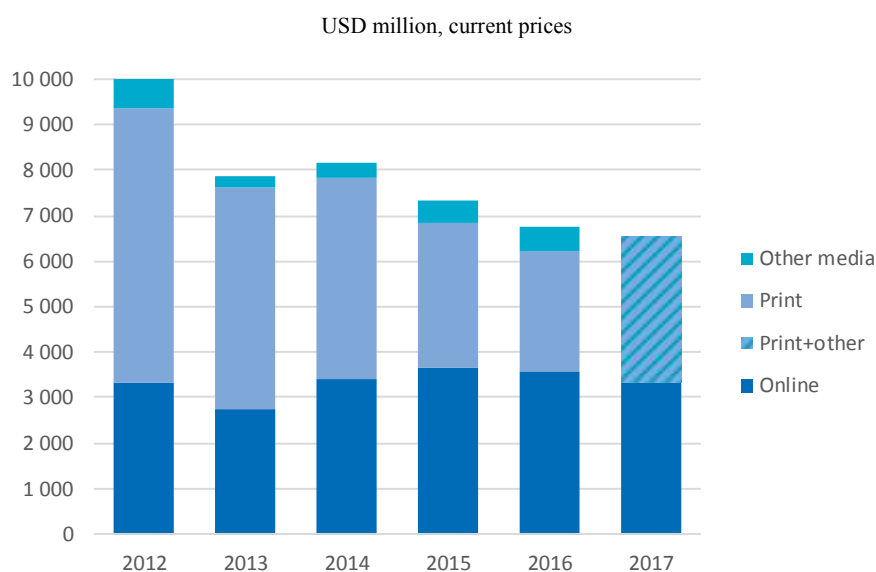
Code	NAICS description	Source	Measure		2012	2013	2014	2015	2016	2017
511	Publishing industries, except Internet (includes software)	Annual supply table	Total industry supply		295 201	307 621	319 773	320 870	333 622	363 500
		Services Annual Survey (SAS)	Total industry revenue	a	272 152	279 253	291 575	293 771	308 489	317 850
51114	Directory and Mailing List Publishers			b	10 118	8 265	8 602	8 214	7 777	7 533
	Share (b/a)				3.7%	3.0%	3.0%	2.8%	2.5%	2.4%
511	Publishing industries, except Internet	US Economic Census (EC)	"Sales, value of shipments, or revenue"	c	272 152					360 239
51114	Directory and Mailing List Publishers			d	10 118					5 370
	Share (d/c)				3.7%					1.5%
518-519	Data processing, Internet publishing, and other information services	Annual supply table	Total industry supply		220 654	240 220	260 193	289 503	321 784	357 627
		Services Annual Survey	Total industry revenue	e	202 021	219 379	236 332	262 697	294 711	333 915
519130	Internet Publishing and Broadcasting and Web Search Portals			f	87 946	96 951	109 414	125 868	148 039	170 781
	Share (f/e)				44%	44%	46%	48%	50%	51%
518-519	Data processing, Internet publishing, and other information services	US Economic Census	"Sales, value of shipments, or revenue"	g	202 021					346 279
519130	Internet Publishing and Broadcasting and Web Search Portals			h	114 074					164 106
	Share (h/g)				56%					47%
5415	Computer systems design and related services	Annual supply table	Total industry supply		350 951	355 201	373 810	377 411	403 985	435 095
		Detailed supply table	Total industry supply	i	350 951					
541511	Custom computer programming services			j	134 795					
	Share (j/i)				38%					
5415	Computer systems design and related services	US Economic Census	"Sales, value of shipments, or revenue"	k	326 031					445 236
541511	Custom computer programming services			l	121 232					185 289
	Share (l/k)				37%					42%
561	Administrative and support services	Annual supply table	Total industry supply		656 909	685 240	752 454	802 593	858 025	907 409
		Services Annual Survey	Total industry revenue		593 835	619 716	683 527	728 883	778 178	832 981
		Economic Census	Total industry receipts	m	593 837					916 730
561450	Credit bureaus			n	7 062					10 104
	Share (n/m)				1.2%					1.1%

Note: Total supply figures are in basic prices. NAICS "518-519" is labelled "514" in the annual supply tables.  
Source: OECD, based on (US Bureau of Economic Analysis, n.d.<sub>[13]</sub>), (US Census Bureau, 2018<sub>[14]</sub>) & (2012<sub>[15]</sub>)

### 51114 Directory and Mailing List Publishers

This industry houses firms focussed on “publishing of mailing lists, publishing of telephone directories, and publishing of other directories and compilations, such as case law, pharmaceutical compendia, etc.” (United Nations, 2008<sup>[19]</sup>). However, a significant (although declining) portion of revenues come from publications in printed form – around 40% in 2016, as shown in **Figure 3.1**. Despite having some similarities, such printed outputs are fundamentally different from, and likely to perform a very different function than the electronic databases that are the focus of this analysis. As a result, only revenues from online and other (non-print) products should be included in estimates of the scale of activities related to compiling and selling databases.

**Figure 3.1. 51114 Directory and mailing list publishers’ revenue, by media type, 2012-2017**



Note: “Print+other” estimated by residual in 2017. “Other media” estimated by residual in 2012.

Source: OECD, based on (US Census Bureau, 2018<sup>[14]</sup>)

It is also case that these revenues arise from selling several different products, as shown in **Figure 3.2**. Subscriptions and sales of directories, databases, and other collections of information, together with rental or sale of mailing lists (a specific form of database), accounted for only 46% of revenues in 2017.

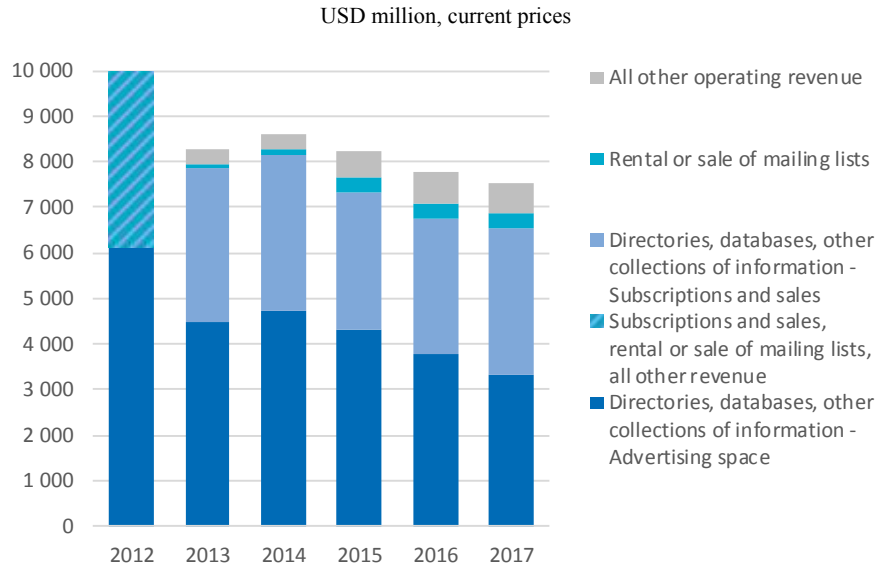
The main individual component comes from selling advertising space e.g. in directories. However, advertising revenue is derived from selling ancillary products (advertising services) rather than from producing and selling the database itself. As that is the focus here, advertising activities should be excluded and considered in a separate analysis alongside the similar activities of web portals or online social networks which leverage a database to offer users services but monetise this by showing them adverts.

At this point a challenge arises because the combined breakdown (i.e. revenue by media and product type is not available). The advertising component has been declining, and this is likely to be related to the simultaneous decline in print media revenues seen in Figure 3.1. It is therefore not reasonable to assume that the share of different media channels is



the same across the different product types and it is necessary to decide whether to give primacy to the media type or product type distinction when producing over all estimates of the scale of activities related to compiling and selling databases.

**Figure 3.2. 51114 Directory and mailing list publishers' revenue, by product type, 2012-2017**



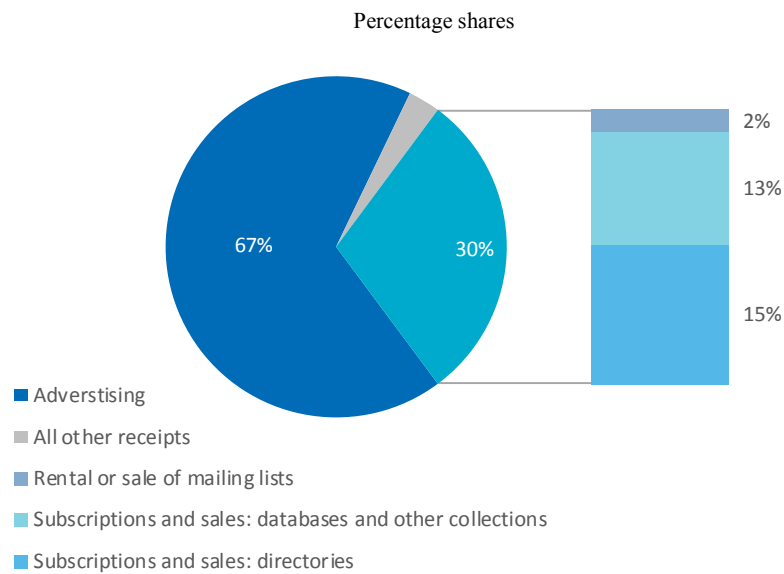
Note: “Subscriptions and sales, rental or sale of mailing lists, all other revenue” estimated by residual for 2012. “Rental or sale of mailing” lists estimated by residual for 2013 and 2014.

Source: OECD, based on (US Census Bureau, 2018[5])

The 2012 Economic Census by product line data offer some further insight into the specific products being produced (**Figure 3.3**). These support the view that a majority of revenues arise from advertising, although the share in the Economic Census data is higher (67%) than that in the SAS data (61%).

Subscriptions and sales of directories generate 15% of revenue, while subscriptions and sales of databases generate 13%. Rental or sale of mailing lists generate a further 2% of revenue. Together, these account for 30% of the total. A comparable figure is not available from the 2012 SAS, due to the suppression of relevant items in the tables. However, the share of “all other operating revenue” in the SAS figures ranges from around 4-9% in the years 2013-17 so the shares shown by the Economic Census data do not appear greatly out of line.

**Figure 3.3. 51114 Directory and mailing list publishers' revenue, by detailed product type, 2012**

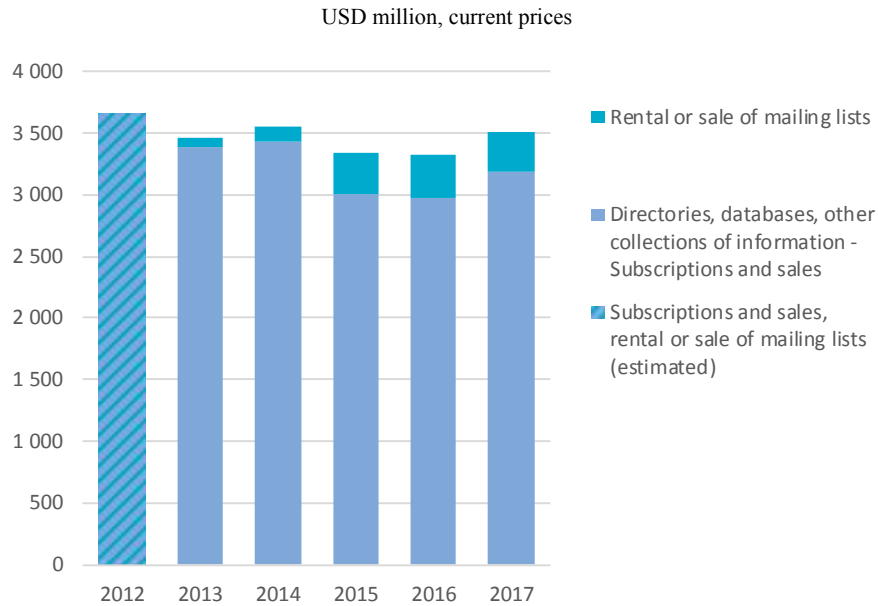


Source: OECD based on (US Census Bureau, 2012<sup>[15]</sup>)

As noted above, it is necessary to decide what to include or exclude when estimating the scale of activities related to compiling and selling databases. For the purposes of this analysis, the elements relating to Subscriptions to or sales of directories, databases, and other information collections and also rental or leasing of mailing lists will be retained, as shown in **Figure 3.4**. However, although excluding advertising revenue is likely to limit the inclusion of revenues arising from printed directories, it is likely that some print-generated monies will still be included as a breakdown by media and product type is unavailable.

It is therefore estimated that Directory and mailing list publishers' revenues related to compiling and selling data and databases amounted to around USD 3.5 billion annually between 2012 and 2017.

**Figure 3.4. 51114 Directory and mailing list publishers' revenue from activities related to compiling and selling databases, 2012-2017**



Note: 2012 figure estimated based on 2013 product share.

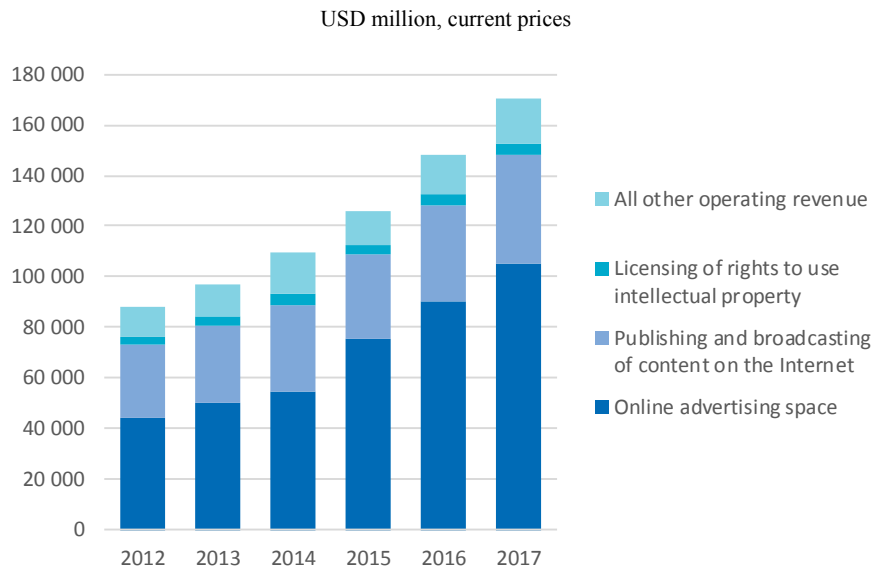
Source: OECD, based on (US Census Bureau, 2018[5])

### *519130 Internet Publishing and Broadcasting and Web Search Portals*

This industry covers a wide range of online publishing activities. It “comprises establishments primarily engaged in (1) publishing and/or broadcasting content on the Internet exclusively or (2) operating Web sites that use a search engine to generate and maintain extensive databases of Internet addresses and content in an easily searchable format (and known as Web search portals). The publishing and broadcasting establishments in this industry do not provide traditional (non-Internet) versions of the content that they publish or broadcast. They provide textual, audio, and/or video content of general or specific interest on the Internet exclusively. Establishments known as Web search portals often provide additional Internet services, such as email, connections to other Web sites, auctions, news, and other limited content, and serve as a home base for Internet users” (US. Office of Management and Budget, 2017<sup>[21]</sup>). It is therefore necessary to find a way to focus in on publishing of relevant products.

The SAS provides a breakdown of revenue by product, presented in **Figure 3.5**. Once again, advertising is the main driver of revenue, having risen from around a 50% share in 2012 to 62% in 2017. Following the logic above, these revenues should be excluded.

**Figure 3.5. 519130 Internet Publishing and Broadcasting and Web Search Portals' revenue, by product type, 2012-2017**

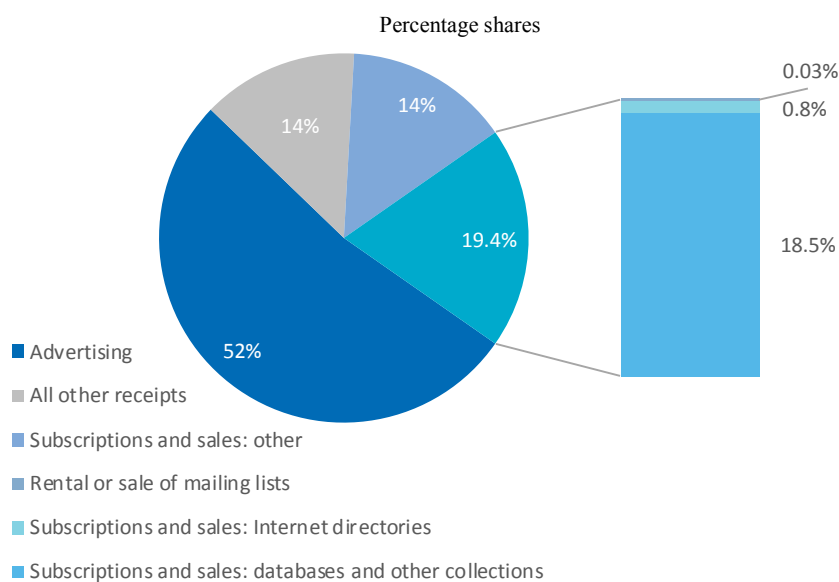


Source: OECD, based on (US Census Bureau, 2018[5])

However, it is still not clear how much of the other revenue streams presented in the SAS tables relate to compiling and selling databases. Nevertheless, the 2012 Economic Census product line data, presented in Figure 3.6, can provide additional insight. In these data, advertising generates 52% of revenues, about the same as in the 2012 SAS (50%). “Subscriptions and sales: other” covers a wide range of print and other non-Internet subscriptions and sales, Internet newspapers, periodicals, books, online gaming, and other content. This excludes, Internet directories, databases, and other collections which, together with rental or sales of mailing lists, comprise almost 20% of total revenue.

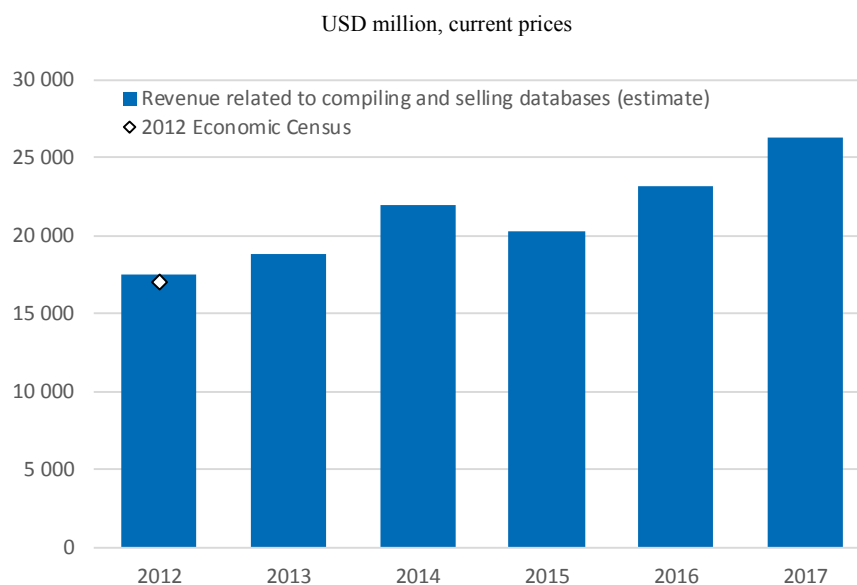
This is equivalent to 40% of the non-advertising revenue in the 2012 Economic Census data. Applying this share to the non-advertising revenue in the SAS data yields estimates of the revenue arising from compiling and selling databases between 2012 and 2017. The results are presented in **Figure 3.7**, which shows that businesses classified in the Internet Publishing and Broadcasting and Web Search Portals industry have seen revenues related to compiling and selling data and databases rise significantly since 2012, to reach over USD 25 billion in 2017.

**Figure 3.6. 519130 Internet Publishing and Broadcasting and Web Search Portals' revenue, by detailed product type, 2012**



Source: OECD based on (US Census Bureau, 2012<sup>[15]</sup>)

**Figure 3.7. 519130 Internet Publishing and Broadcasting and Web Search Portals' estimated revenue from activities related to compiling and selling databases, 2012-2017**



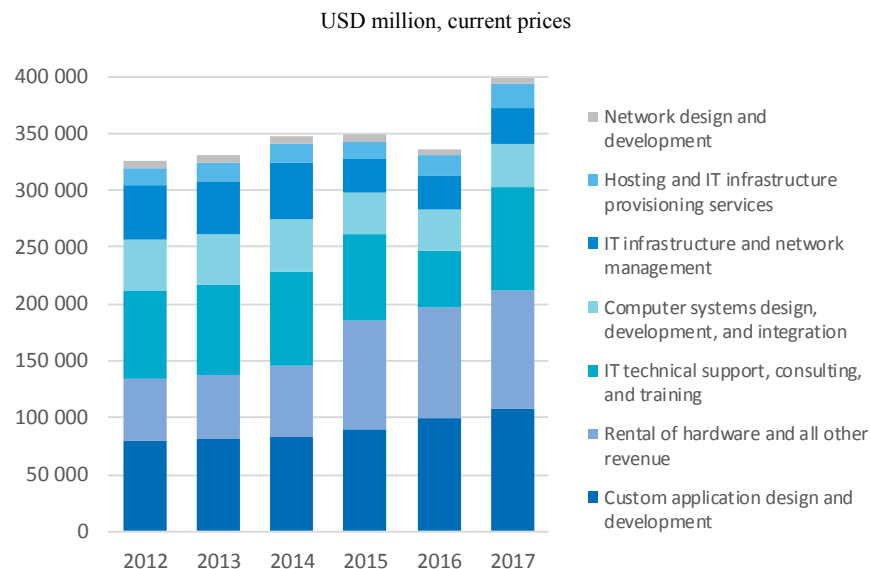
Source: OECD, based on (US Census Bureau, 2018<sup>[5]</sup>), (US Census Bureau, 2012<sup>[15]</sup>)

### 541511 Custom Computer Programming Services

This industry “comprises establishments primarily engaged in writing, modifying, testing, and supporting software to meet the needs of a particular customer” (US. Office of Management and Budget, 2017<sup>[21]</sup>). It is clear that a considerable share of this activity is likely to relate to software programming that does not relate to databases. It is therefore necessary to seek further information on the products involved.

The SAS provides a revenue breakdown for the parent Industry “5415 Computer systems design and related services”, presented in **Figure 3.8**. Several items such as network design and development, IT infrastructure and network management, and hardware rental appear unlikely to be relevant. However, it is not clear which of the other categories are likely to contain activities related to compiling and selling databases. Furthermore, it is likely that the relevant activities occur across the multiple sub-industries within 5415 rather than in 541511 alone.

**Figure 3.8. 5415 Computer systems design and related services’ revenue, by product type, 2012-2017**



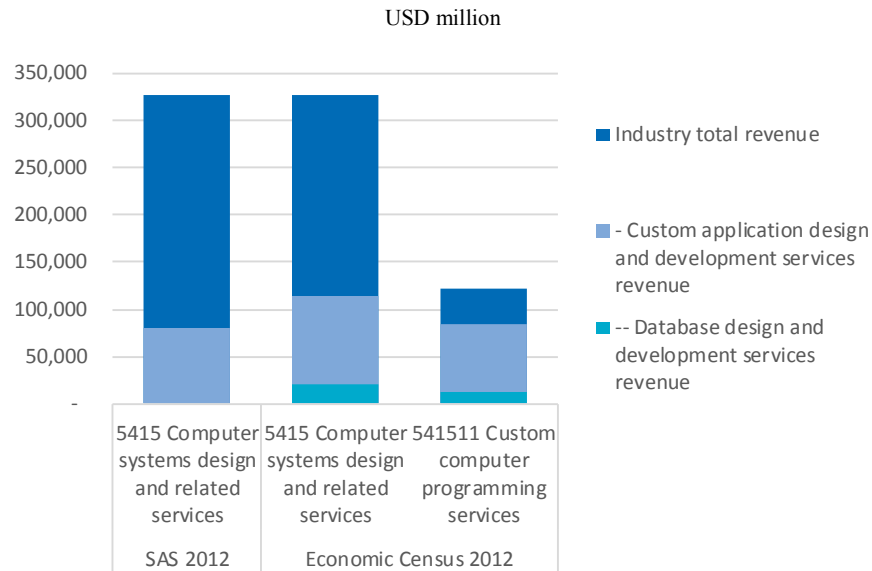
Source: OECD, based on (US Census Bureau, 2018<sup>[5]</sup>)

The 2012 Economic Census by product line data (US Census Bureau, 2012<sup>[15]</sup>) provide considerable product detail including a specific product line for “database design and development services”. This is within the “custom application design and development services” product available from the SAS.

**Figure 3.9** compares the data available from these sources. It shows that the total industry revenue for “5415 Computer systems design and related services” is the same in both sources, although the portion revenue from “custom application design and development services” is markedly different, being around one third smaller in the SAS. The specific “database design and development services” product is not available from the SAS. However, the Economic Census shows that it comprises 18% of revenue in industry 5415

and just 14% of revenue in the specific sub-industry “541511 Custom computer programming services”.

**Figure 3.9. 5415 Computer systems design and related services’ revenue, by product type, 2012**



Note: Database design and development services revenue sub-product not available in SAS 2012 table.

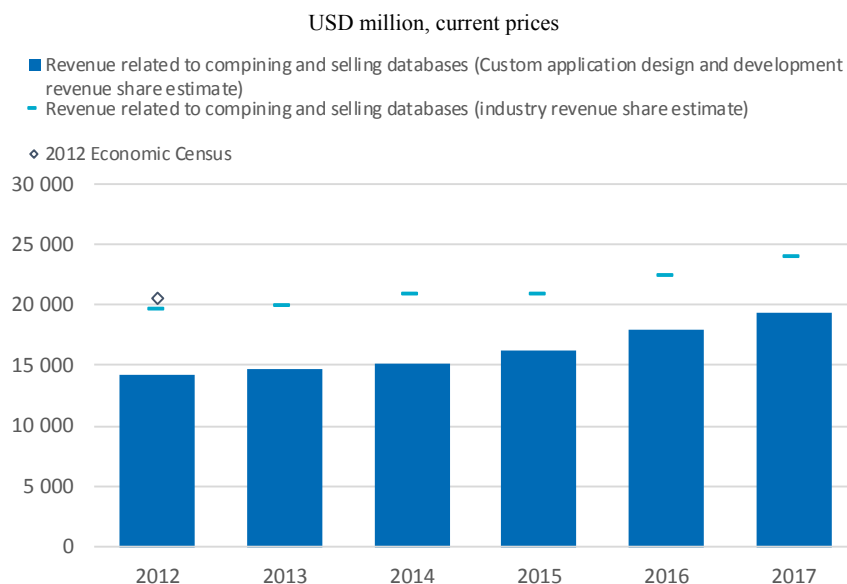
Source: OECD, based on (US Census Bureau, 2018<sup>[14]</sup>), (US Census Bureau, 2012<sup>[15]</sup>)

This confirms that database design and development takes place across multiple sub industries within 5415. Indeed, less than two-thirds occurs in firms classified within “541511 Computer programming services”. For this reason, and for reasons of wider data availability, it is therefore proposed to use the SAS data for “5415 Computer systems design and related services” adjusted using the 18% share of database design and development services in revenue, as shown by the Economic Census product line data, to derive an estimate of estimated revenue from activities related to compiling and selling databases, as presented in **Figure 3.10**.

However, due to the discrepancy between the two sources in custom application design and development services revenue, this approach results in the estimate being significantly below the Economic Census observation for 2012. Figure 3.10 therefore includes a series based on the share of database design and development services revenue in total industry revenue (6%). This results in estimates around a quarter greater.

Taking the latter estimates as they are closest to the only observation available (the Economic Census figure from 2012), this shows that revenues related to compiling and selling databases have increased from around USD 20 billion in 2012, to USD 24 billion in 2017.

**Figure 3.10. 5415 Computer systems design and related services' estimated revenue from activities related to compiling and selling databases, 2012-2017**



Note: Estimates based on the share of “database design and development services” revenue in industry total revenue or revenue from “custom systems design and related services.”

Source: OECD, based on (US Census Bureau, 2018[5]), (US Census Bureau, 2012[6])

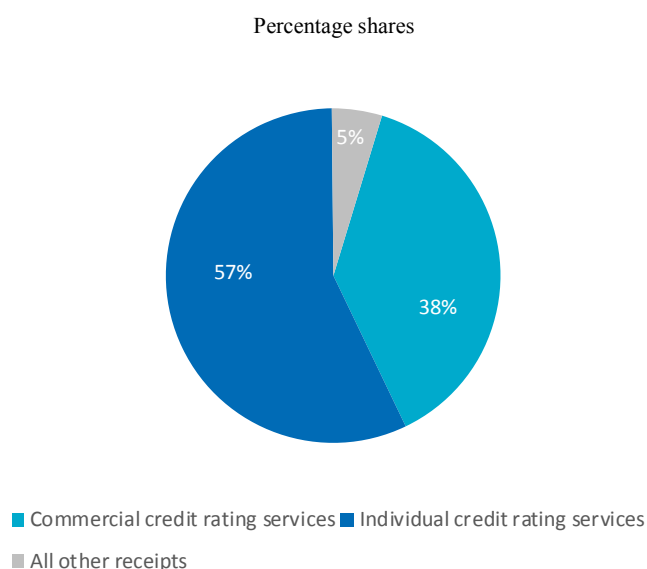
### 561450 Credit Bureaus

“This industry comprises establishments primarily engaged in compiling information, such as credit and employment histories, and providing the information to financial institutions, retailers, and others who have a need to evaluate the creditworthiness of individuals and businesses” (US. Office of Management and Budget, 2017<sup>[21]</sup>). As such, this industry appears narrowly focussed on activities related to developing and selling data and databases. However, it should be noted that the precise nature of the access to the relevant databases may vary from selling of copies or live access (e.g. via an API) allowing banks etc. to make their own credit checks according to their own algorithms, to the selling of ancillary services where the credit bureau itself performs a check and produces a report. Similar to the case where a database is monetised via ancillary advertising services, the latter might be considered for exclusion.

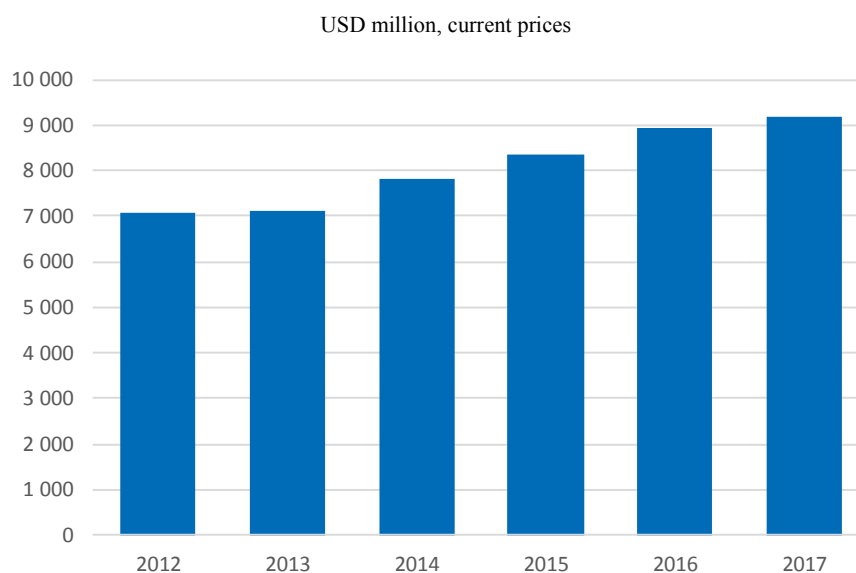
The only sources found for data at the level of the industry are the 2012 and 2017 Economic Census data, as presented in Table 3.6. The 2012 Economic Census product line data provides further detail on the products produced. **Figure 3.11** shows that in 2012, 95% of revenue for the industry related to credit rating services. Items which appear to be services derived from the database, such as “investigation and personal background checks” are within “all other receipts” and make up a very small fraction of the overall activity.

It is therefore proposed to estimate the revenue from activities related to compiling and selling databases by taking applying average of the 2012 (1.2%) and 2017 (1.1%) shares presented in Table 3.6 (i.e. 1.15%) and applying it to the parent industry figures available from the SAS. The results are presented in **Figure 3.12** and show that industry revenues are estimated to have risen from USD 7 billion in 2012 to over USD 9 billion in 2017.



**Figure 3.11. 561450 Credit Bureaus' revenue, by detailed product type, 2012**

Note: All other receipts includes revenues from “collection services and debt recovery”, “investigation services including personal background checks”, “convention and trade show organisation”, and “sale of merchandise”.  
Source: OECD based on (US Census Bureau, 2012<sup>[15]</sup>)

**Figure 3.12. 561450 Credit Bureaus' estimated revenue from activities related to compiling and selling databases, 2012-2017**

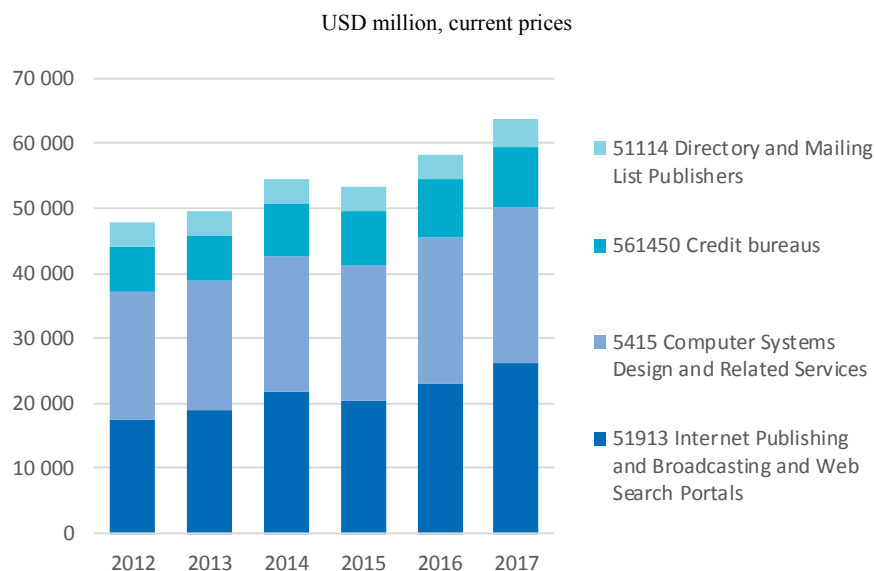
Source: OECD, based on (US Census Bureau, 2018<sup>[5]</sup>), (US Census Bureau, 2017<sup>[22]</sup>), (US Census Bureau, 2012<sup>[6]</sup>).

### *Estimating total revenue from compiling and selling databases*

The foregoing sections developed individual estimates for each industry's revenue from compiling and selling data and databases. Combining these in **Figure 3.13** gives an overall estimate of such revenue for the United States.

It is estimated that total revenue from compiling and selling data and databases reached over USD 60 billion in 2017, having increased by around one third since 2012. The contribution from all industries grew in this time, with the exception of “51114 Directory and Mailing List Publishers” which remained fairly flat over the period.

**Figure 3.13. Estimated revenue from activities related to compiling and selling databases, by industry, United States, 2012-2017**



Source: OECD based on (US Census Bureau, 2018<sup>[14]</sup>), (US Census Bureau, 2017<sup>[22]</sup>), (US Census Bureau, 2012<sup>[15]</sup>)

It is notable that “51913 Internet Publishing and Broadcasting and Web Search Portals” is the largest single component of the total. This industry was only included because it includes the on-line component of directory, database, and mailing list publishing. Analysis of the 2012 Economic Census product line data found that almost 20% of total revenues in this industry were subscriptions and sales related to those products.

The US Services Annual Survey and Economic Census data offer a wealth of detail at industry and product level. This allows fairly specific estimates of the revenue from compiling and selling data and databases to be derived for the United States. The following sections examine data availability in other countries.

### **3.3.2. Canada**

Canada also uses the NAICS. However, there are some differences between the US and Canadian NAICS classifications which mean that, while they are compatible to a degree, there can be differences at the sub-industry level. There are also differences in the

concordance with ISIC rev.4 industries, as set out in the Statistics Canada NAICS-ISIC concordance table (Statistics Canada, 2019<sup>[23]</sup>).

The concordance table shows that ISIC 5182 “Publishing of directories and mailing lists” maps across not only to “511140 Directory and Mailing List Publishers”, as it does in the US, but the part related to radio guide and schedule publishing maps to “511120 Periodical publishers”. As the latter is likely to have limited relevance to the database compilation and sales, and also due to the large range of other periodicals included in the category, this will be ignored in the analysis.

This highlights an important point of difference in comparison to the US NAICS: that “519130 Internet Broadcasting and Web Search Portals” does not map across to the relevant ISIC industry (5182, instead mapping to the broadcasting industries and Web portals industry)<sup>5</sup>. The reason for this appears to be that “511140 Directory and Mailing List Publishers” explicitly includes online publishing whereas it is specifically excluded in the NAICS. The analysis will therefore focus only on this NAICS industry.

Ideally, sales of the “519130 Internet Publishing and Broadcasting and Web Search Portals” would be analysed in the same way as those in the US, to identify sales related to directories, databases, and mailing lists. However, in the published series these are hidden within the product group “other published works” which does not appear to be available broken down by industry.

One key difference in comparison to the US NAICS is that the Canadian NAICS industry 519130 relates only to “Internet broadcasting and web search portals” whereas the US version also includes Internet Publishing. As such, it does not map across to the relevant ISIC industry (5182), instead mapping to the broadcasting industries and Web portals.

Another divergence between the US and Canadian NAICS is that industry “541511 Custom Computer Programming Services” does not exist in the latter. ISIC 6201 maps instead to “541514 Computer systems design and related services (except video game design and development)” and “541515 Video game design and development services”. However, in the US estimates presented above, the parent category “5415 Computer Systems Design and Related Services” was used instead due to evidence from the 2012 Economic Census product line data showing that “database design and development services” were produced across the sub-industries therein. As the same is likely to be true in Canada, the same approach will be taken here. This is summarised in **Table 3.7**, which also shows that the mapping of credit bureaus is the same as in the US NAICS.

**Table 3.7. ISIC Rev.4 to Canada NAICS 2017 conversion**

ISIC code	ISIC title	NAICS code	NAICS title
5812	Publishing of directories and mailing lists	511140	Directory and Mailing List Publishers
		511120	Periodical publishers (radio guide and schedule publishers or publishing)
6201	Computer programming activities	5415	Computer Systems Design and Related Services
8291	Activities of collection agencies and credit bureaus	561440	Collection Agencies
		561450	Credit Bureaus

Source: OECD, based on (Statistics Canada, 2019<sup>[23]</sup>)

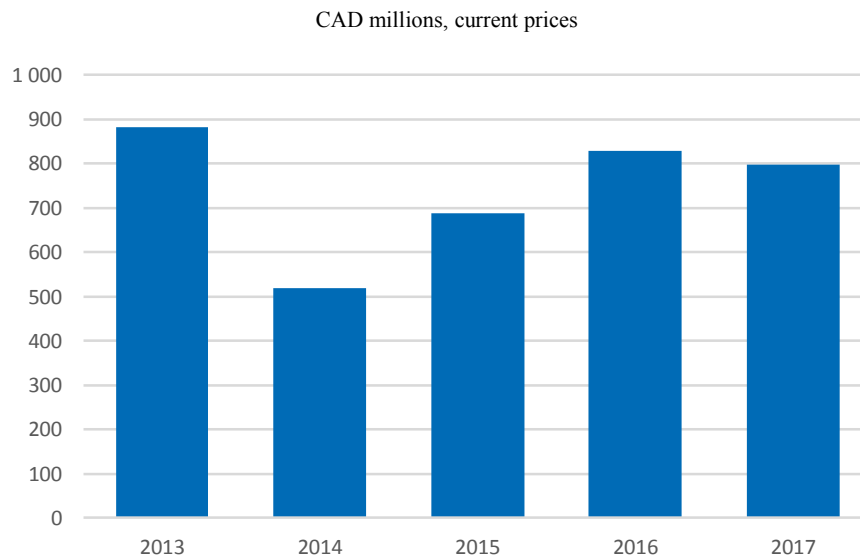
The following sections examine the availability of data for these categories.

### 5415 Computer systems design and related services

Statistics Canada publishes both operating revenue (Statistics Canada, 2019<sup>[24]</sup>), and sales shares by product (Statistics Canada, 2019<sup>[25]</sup>) for the “Computer systems design and related services” industry. Available for 2013-2017, the share of sales revenue from “database design and development” varies between 1.2 and 2.2% annually. For comparison, the equivalent figure for USA in 2012 was considerably greater, at 6.3% (Figure 3.9).

Combining revenue and shares gives the annual revenue from database design and development as shown in **Figure 3.14**, which shows that these revenues were around 0.8 billion CAD annually between 2013 and 2017.

**Figure 3.14. 5415 Computer systems design and related services industry revenue, from database design and development services, Canada, 2013-2017**



Source: OECD based on (Statistics Canada, 2019<sup>[24]</sup>), (Statistics Canada, 2019<sup>[25]</sup>).

### Other industries

Such revenue figures and/or breakdowns do not appear to be available for the other industries of interest here. The only source for detailed data found is the annual Supply Tables (Statistics Canada, 2019<sup>[26]</sup>). **Table 3.8** presents the data for the relevant NAICS industries, or the parent industry grouping when more specific series are unavailable. It also presents the estimated portion generated by compiling and selling databases. Shares from the US Economic Census are used when Canadian data are unavailable.

It is important to note that “supply” is a National Accounts concept and differs from revenue or turnover as presented in the US and European Countries sections. The latter is a measure of money received from sales and other sources in the year. Meanwhile, supply excludes any prepayments but includes the build-up of inventories and work in progress as well as own-account production (e.g. of capital goods). As a point of comparison, the total product of industry “5415 Computer systems design and related services” in 2017, as indicated by the supply tables was CAD 45.4 billion. Meanwhile, the total operating

revenue from the business statistics was nearly one-fifth greater at CAD 53.1 billion. Correspondingly, when the “database design and development” shares from the survey breakdown data (Statistics Canada, 2019<sup>[25]</sup>) are applied to the supply table data, the resulting estimates (shown in Table 3.8) are lower than the revenue-based estimates.

**Table 3.8. Estimated supply of products related to compiling and selling databases. Canada, 2011-15**

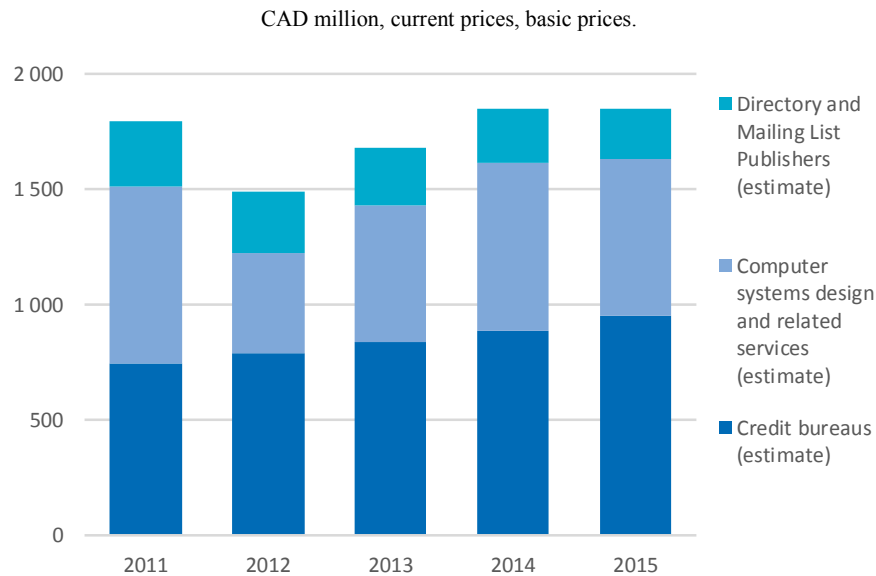
CAD million, current prices, basic prices

	2011	2012	2013	2014	2015	Source / explanation
Periodical, book and directory publishers BS5111A0 [a]	5 597	5 386	5 021	4 774	4 365	Supply table.
US 2012 Economic Census receipts share [b/a]	5%	5%	5%	5%	5%	Based on share in US 2012 Economic Census product line data.
- of which, directories, databases, and mailing list subscriptions (estimate) [b]	280	269	251	239	218	
Computer systems design and related services BS5415 [c]	35 010	36 160	39 403	42 889	45 415	Supply table
CAN sales shares [d/c]	2.2%	1.2%	1.5%	1.7%	1.5%	Based on Canada Software development and computer services, breakdown of sales
- of which, "database design and development services" (estimate) [d]	770	434	591	729	681	
Business support services BS561400 [e]	6 205	6 568	6 990	7 375	7 929	Supply table
US average 2012 and 2017 Economic Census share [f/e]	12%	12%	12%	12%	12%	Based on average share in US economic Census 2012 and 2017.
- of which, credit bureaus (estimated) [f]	745	788	839	885	952	

Source: OECD based on (Statistics Canada, 2019<sup>[26]</sup>), (Statistics Canada, 2019<sup>[25]</sup>), (US Census Bureau, 2017<sup>[22]</sup>), (US Census Bureau, 2012<sup>[15]</sup>).

**Figure 3.15** compiles these components into an estimate of the total supply of products related to compiling and selling databases. These contrast markedly with the US estimates in Figure 3.13, with credit bureaus being the largest individual component in Canada. The most striking difference, though, is the lack of a contribution from the Internet [publishing and] broadcasting and Web Portals industry. If data on sales by product were available for this industry the relevance of making an estimation for that industry could be assessed.

**Figure 3.15. Estimated supply of products relating to compiling and selling databases, by industry, Canada, 2012-2017**



Source: OECD based on (Statistics Canada, 2019<sup>[26]</sup>), (Statistics Canada, 2019<sup>[25]</sup>), (US Census Bureau, 2017<sup>[22]</sup>), (US Census Bureau, 2012<sup>[15]</sup>).

Although revenue by product is available for one of the industries in this analysis, it is necessary to resort to supply tables as they provide data for the other industries. This entails a move from the revenue or turnover concepts used in business statistics sources to output and supply – concepts from the National Accounts framework. Furthermore, where relevant details are unavailable, it is necessary to use other sources – namely the breakdowns available in US data. This, of course, makes the assumption that the industry and output structure is sufficiently similar across the two countries. If similar detail became available for Canada itself more robust results could be achieved.

### 3.3.3. European Countries

European Countries implement the “Statistical classification of economic activities in the European Community” (NACE Rev.2) (Eurostat, 2008<sup>[27]</sup>), which is aligned with ISIC Rev.4 with regards to the numbering and definition of the industries used in this analysis.

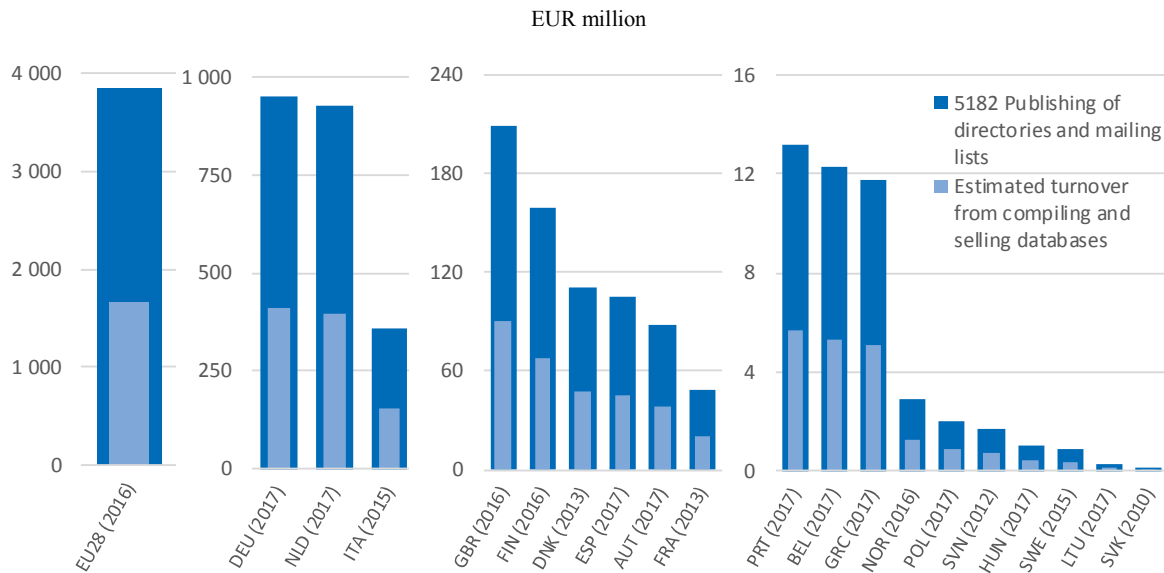
The “Eurostat Annual Detailed Enterprise Statistics for Services” (Eurostat, 2019<sup>[28]</sup>) provide various measures, including production value and turnover – which is similar to the revenue measures available for the US. These are broken down across over 200 industries and sub-industries. Specific series are available for each of the industries identified for this analysis. Nevertheless, data availability varies between countries. The following sections examine the available data.

#### 5812 Publishing of directories and mailing lists

Data are available for 19 countries, as well as for the EU28 as a whole (**Figure 3.16**). However, the detailed enterprise statistics for services do not include a product dimension. It is likely that, just as in the US, a considerable portion of turnover relates to advertising

(see Figure 3.2). In order to adjust for this the average US share of subscriptions and sales of directories, databases and other collections of information and from rental or sale of mailing lists in industry total revenue over the period 2015-17 – 43% - is used to derive an estimate of the turnover related to compiling and selling databases. This is estimated to be worth 1.7 billion EUR across the EU28, with Germany, the Netherlands, and Italy as the biggest contributors.

**Figure 3.16. Directory and mailing list publishers’ turnover, European countries, 2017 or latest**



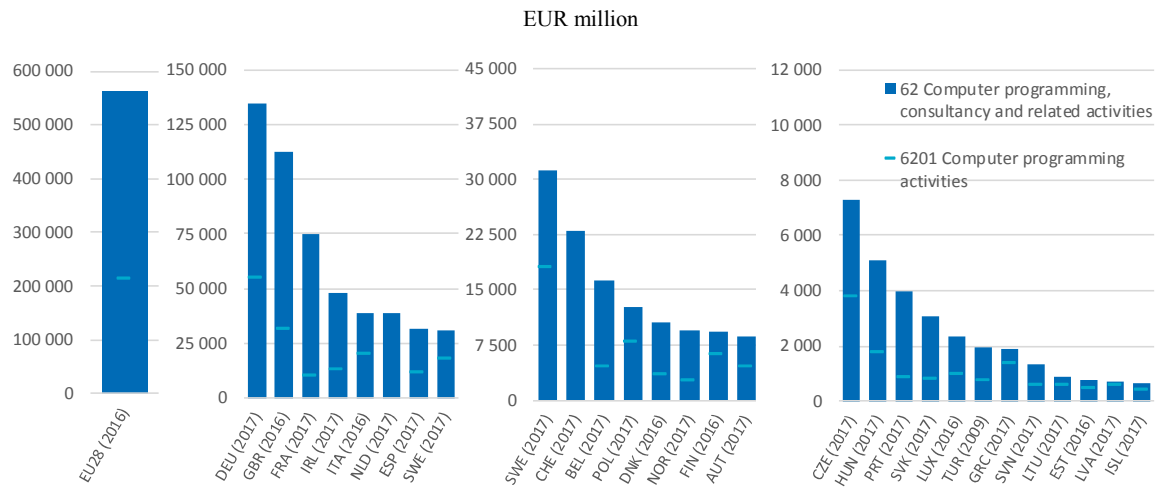
Note: Turnover from compiling and selling databases based on the 2015-2017 average US share (43%) of 51114 Directory and mailing list publishers’ revenue generated from Subscriptions and sales of directories, databases and other collections of information and from rental or sale of mailing lists according to the Services Annual Survey (see , ).

Source: OECD based on (Eurostat, 2019<sup>[28]</sup>), (US Census Bureau, 2018<sup>[14]</sup>)

## 62 Computer programming, consultancy, and related activities

The US analysis found that “database design and development services” are produced not only within (the NAICS equivalent of) “6201 Computer programming activities” but in multiple sub-industries within the parent industry. The same is likely to be true in European Countries and so this analysis will focus on “62 Computer programming, consultancy, and related activities”.

**Figure 3.17** presents the available data for both the specific industry 6201 and parent industry 62. The latter provides coverage of an additional two countries – 28 in total – and generates turnover of over EUR 560 billion across the EU28, with Germany, the United Kingdom, and France as the greatest contributors.

**Figure 3.17. Computer programming activities' turnover, European countries, 2017 or latest**

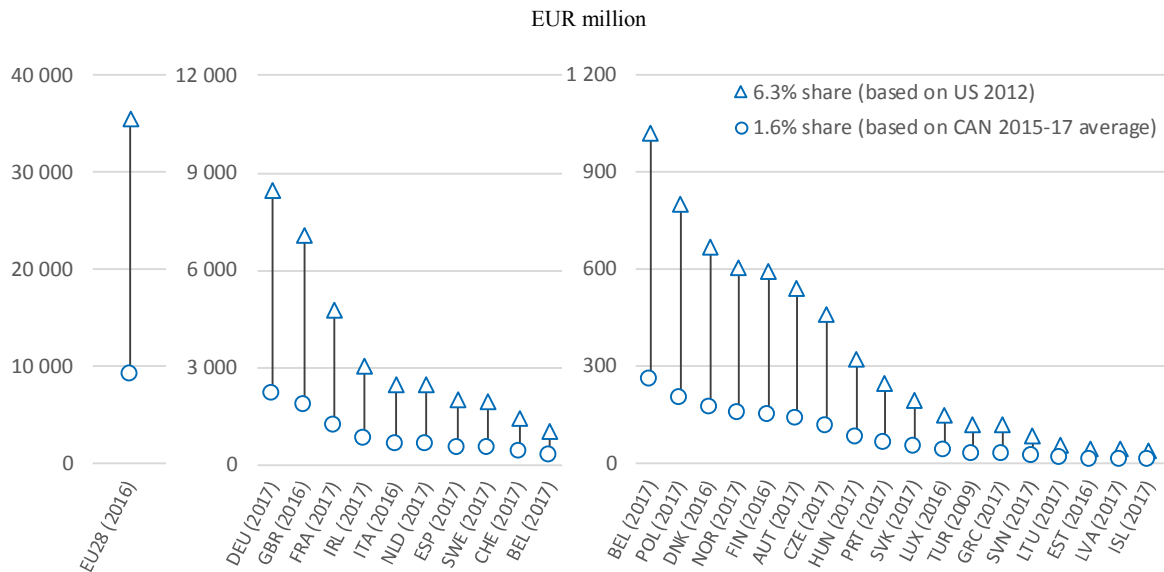
Source: OECD based on (Eurostat, 2019<sup>[28]</sup>).

The foregoing analysis has shown that these are broad industries in which the vast majority of activities are not related to compiling or selling databases. With no product detail available, the shares from the US or Canadian data can be used as a proxy. However, there is a large difference between the 2015-17 average share of database design and development services in Canada (1.6%) and the share found in the 2012 US Economic Census product line data (6.3%). **Figure 3.18** presents series based on each of these, resulting in an EU28 estimate of over EUR 35 billion based on the US share but only EUR 9 billion based on the Canadian share.

Arguments can be made in favour of using either one of these over the other. An alternative might be to take the average of them both. For the purposes of this analysis, estimates based on both will be presented.



**Figure 3.18. Computer programming activities' estimated turnover from compiling and selling databases, European countries, 2017 or latest**



Note: Turnover from compiling and selling databases based on the 2012 US (6.3%) and Canada 2015-17 average (1.6%) shares of revenue from database design and development services in 5415 Computer systems design and related services revenue according to the 2012 Economic census product line data (see ).

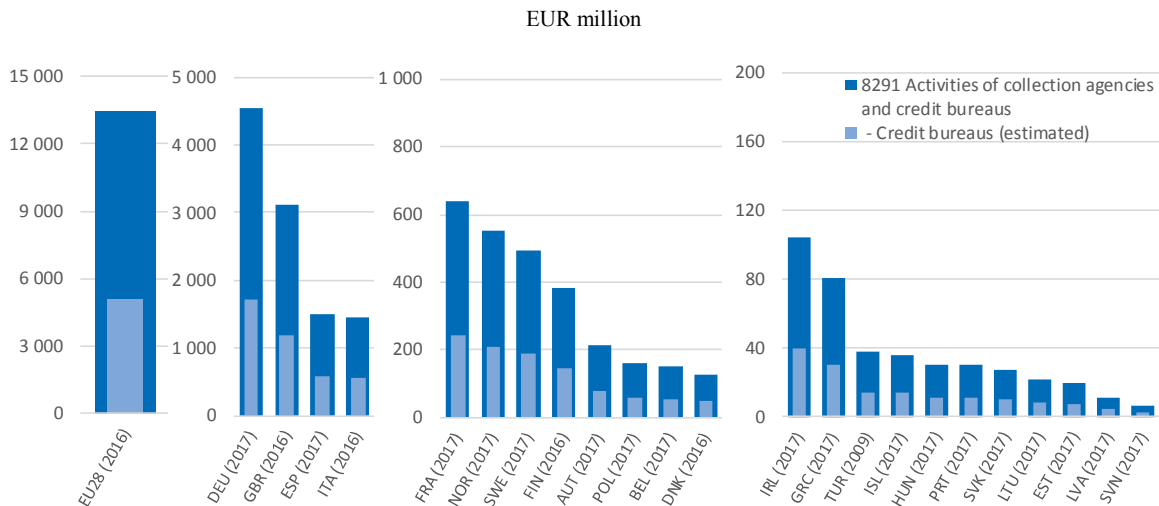
Source: OECD based on (Eurostat, 2019<sup>[28]</sup>), (Statistics Canada, 2019<sup>[25]</sup>), (US Census Bureau, 2012<sup>[15]</sup>).

### 8291 Activities of collection agencies and credit bureaus

NACE rev.2, just like ISIC rev.4, combines the activities of credit bureaus with those of collection agencies. However, it is only the credit bureaus whose core business relates to compiling databases (of a wide variety of information on people and business' creditworthiness) and selling copies of, or access to, those databases so that financial institutions can make lending decisions. The Canadian data further aggregate both these activities with other business support services but the US Economic Census data – available for both 2012 and 2017 – allow them to be split. On average across the two years, credit bureaus accounted for 37% of the combined revenue (36% in 2012 and 38% in 2017).

Applying this share to the turnover figures in **Figure 3.19**, yields the estimate that credit bureaus generated turnover of just over EUR 5 billion in the EU28 in 2016, with Germany, the United Kingdom, Spain, and Italy as the main contributors.

**Figure 3.19. Activities of collection agencies and credit bureaus' turnover, European countries, 2017 or latest**



Note: Turnover from compiling and selling databases based on the 2017 US share (38%) of 561450 Credit Bureaus' revenue in the sum of revenue for 561440 Collection Agencies *plus* 561450 Credit Bureaus according to the 2017 US Economic Census basic tables.

Source: OECD based on (Eurostat, 2019<sup>[28]</sup>), (US Census Bureau, 2017<sup>[22]</sup>).

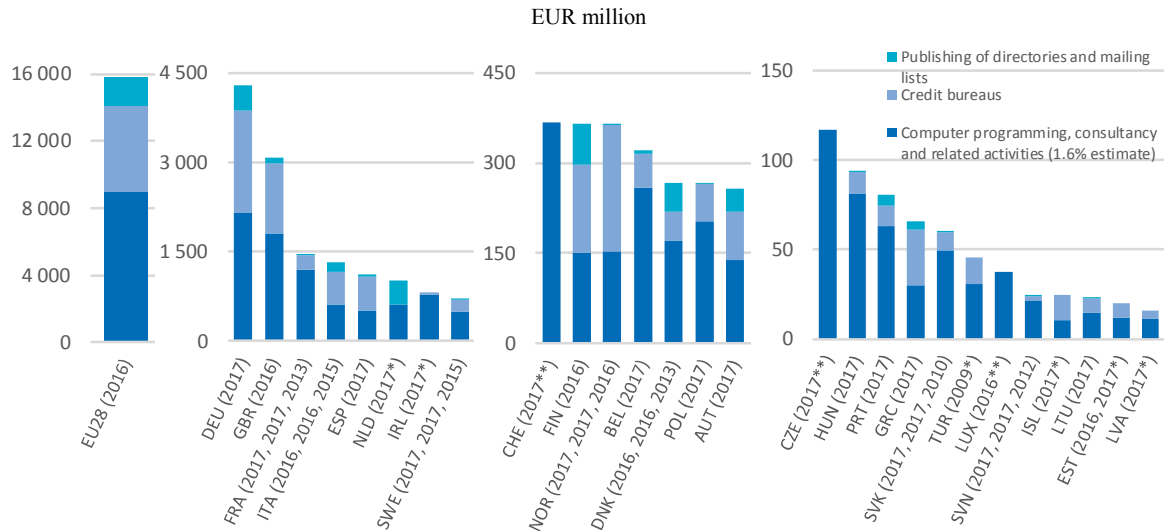
#### *Estimated total turnover from compiling and selling databases in Europe*

**Figure 3.20** brings together the components estimated above to estimate that total turnover from activities related to compiling and selling databases amounts to just below EUR 16 billion across EU28 countries. Germany and the United Kingdom are by far the two greatest contributors, while in Luxembourg, Lithuania, Estonia, and Latvia, turnover amounts to less than EUR 50 million.

The breakdown of constituent components bears more similarity to the United States estimates than those for Canada. As in the case of Canada, a key difference in comparison to the US is that, in the US case a large contribution comes from the “Internet publishing and broadcasting and web search portals” industry. In the absence of a product breakdown it is challenging to ensure full comparability between estimates for the US and other countries.

The estimates presented in Figure 3.20 rely on the 1.6% estimate of the share of database design and development services which comes from the Canadian Survey data. However, as noted in Figure 3.18, it is also possible to use the higher series based on the US share. **Figure 3.21** presents total estimated turnover from activities related to compiling and selling databases based on each of these parameters, finding that the total increases from around EUR 16 billion to over EUR 40 billion across the EU28. Furthermore, the country ordering is altered somewhat.

**Figure 3.20. Estimated turnover from activities related to compiling and selling databases, by industry, European countries, 2017 or latest**

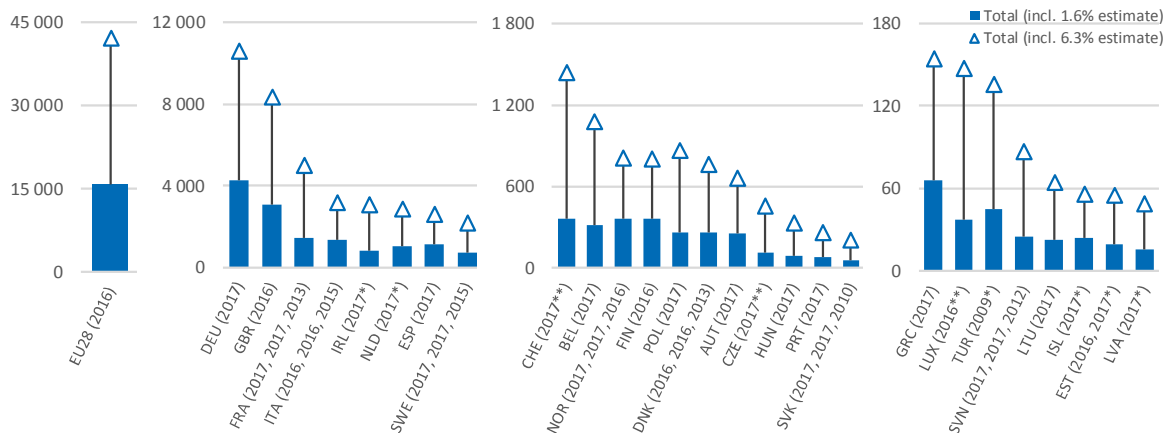


Note: \* indicates one component missing, \*\* indicates two components missing. In both cases underestimation will occur. In some cases, individual components come from different years.

Source: OECD based on (Eurostat, 2019<sub>[28]</sub>), (US Census Bureau, 2018<sub>[14]</sub>), (US Census Bureau, 2017<sub>[22]</sub>), (US Census Bureau, 2012<sub>[15]</sub>).

**Figure 3.21. Total estimated turnover from activities related to compiling and selling databases, European countries, 2017 or latest**

Including upper (6.3%) and lower (1.6%) contributions from Computer programming, consultancy, and related activities; EUR million



Note: \* indicates one component missing, \*\* indicates two components missing. In both cases underestimation will occur. In some cases, individual components come from different years.

Source: OECD based on (Eurostat, 2019<sub>[28]</sub>), (US Census Bureau, 2018<sub>[14]</sub>), (US Census Bureau, 2017<sub>[22]</sub>), (US Census Bureau, 2012<sub>[15]</sub>).

The Eurostat Annual detailed enterprise statistics for services (Eurostat, 2019<sub>[28]</sub>) offer a wealth of data with good industry detail available for many countries. However, there is a

need to break out credit bureaus from the industry presented. Furthermore, the US analysis showed that it is necessary to go beyond the industry breakdown to look at specific products in order to focus in on activities related to compiling and selling databases, as these can often make up a relatively small portion of the total industry activity.

#### **3.3.4. Other countries**

This work was presented to the OECD Working Party on Measuring and Analysing the Digital Economy in November 2019 and delegates were encouraged to share any similar sources available for their countries. However, this did not yield any submissions with comparable detail. As such, extension of this approach across countries will rely on additional detail becoming available.

### **3.4. Concluding remarks**

This analysis has shown that it is possible to use business survey data to measure the output activities related to compiling and selling databases. However, the analysis of detailed data broken down by products has shown that focussing on industry classification only is likely to prove inexact in identifying these activities. Furthermore, it is likely that other industries engage in compiling and monetising databases as secondary activity. The US Economic Census by product line data – currently available for 2012 but with the 2017 vintage due out in November 2020 – would allow this wider phenomena to be analysed in future research<sup>6</sup>. However, there does not seem to be good availability of detailed product data in other countries, and an industry-based approach may remain necessary.

One key challenge relates to “Internet publishing and broadcasting and web search portals” which make a large contribution to the US estimates but are only included as one part of ISIC industry 5812 – that relating to on-line directory, collection, and mailing list publishing – is mapped across to it in the US Census Bureau ISIC to NAICS concordance table (US Census Bureau, 2017<sub>[20]</sub>). By contrast, the Canadian version of NAICS is slightly different and so does not map in the same way.

Although product line data were used to focus in on revenues arising from subscriptions and sales of directories, databases, and other collections of information, together with rental or sale of mailing lists, it seems likely that much of these revenues arise from different industries that were not included on the basis of the description of their activities (presented in Table 3.1) or the mapping of products to industries included in the CPC (Table 3.2). As such, the estimate seems likely to reflect such products being sold as secondary outputs. In an ideal world, detailed product data would be used to examine their production in web portal industries with the aim of identifying the best way of making the US and other countries’ estimates as comparable as possible. However, the scarcity of such detailed data is a significant challenge.

## Chapter 4. Data and firm values

### 4.1. Introduction

Chapter 2. looked at the value of business data holdings as implied by expenditure on storing and managing data. Chapter 3. went on to examine the economic value generated by firms directly selling data and data-based products. This chapter adds another perspective by considering the market valuation of “data driven firms”, defined as firms for which their core business involves collecting, storing, hosting, and/or analysing large quantities of data. The results provided below should be regarded as an illustration of this approach and call for further work to better identify the characteristics of “data driven” firms and expand the coverage of the analysis to firms outside the United States.

### 4.2. Linking data holdings with firm valuations

While databases are sometimes sold and purchased on the open market, it is most common for databases – including customer databases – to be developed and used internally within a company. While such databases are not often sold, their value can be a significant contributor to the price reached in company acquisitions.

Aside from observing firm value during an acquisition, there are various ways of measuring a firm’s value but the most common is “market capitalisation”: the number of shares outstanding *multiplied by* the share price . An individual firm’s market capitalisation is determined by a wide variety of factors – not only fundamentals in terms of reported revenues and investment, as well as expected future revenues – but, ultimately, reflects what investors *believe* a company to be worth.

Factors such as its business model, investment decisions, and the skills and capabilities the business has available are likely to influence earnings (or at least expectations of earnings). For (non-manufacturing) enterprises there is a positive association between intangible assets and market value – with data and databases increasingly being considered as part of that intangible capital (Reed, 2007<sup>[29]</sup>). In this way, firm characteristics such as adopting business models relying on data (e.g. online platforms), emphasising investment in building vast databases, and/or making intensive use of data analytics (e.g. to monetise those databases) are likely to feed through to the firm’s market valuation.

### 4.3. Identifying “data-driven firms”

There is no source that directly identifies “data-driven” firms. While it is straightforward to conceive criteria that could be used to identify these companies - for example the book value of databases relative to the total book value of the share of data scientists in total employees or payroll – it is not routine to separately itemise the information needed to calculate them.

In the most high-profile cases it, is quite clear that a firm’s value is substantially derived from its database(s) and ability to analyse and monetise the data therein. In particular, there are numerous major companies for which their business model entails compiling databases (e.g. of information, libraries of content, etc.) and either monetising these directly through subscriptions (Netflix) or indirectly through targeted online advertising (Google, Facebook, etc.).

For example, Google’s central business activity is online search, the core of which involves compiling and continuously updating a vast index of online content (i.e. a database), allowing users to search that database (and at the same time creating a database of information on those users’ interests, characteristics, etc.), and monetising that activity through advertising space shown alongside search results and in Google’s many other products. At the same time, Google also hosts libraries (databases) of other content which it uses to attract user time and monetises through advertising – e.g. the database of video content used to power YouTube. Advertising generates a majority of Google revenue (Alphabet, 2020<sup>[30]</sup>) and this revenue relies on having databases of information and content that attracts users.

Companies such as Amazon are clearly also built not only around vast databases of product listings but also the use of information on customers including search and purchase history, user journeys through the site (“people also searched for...”), and the like to both help users find products they are actively searching for but also to show them products they might be likely to buy.

In addition, companies such as Uber not only rely on a database of drivers and customers as a key foundation for its ride sharing and delivery services, but also on real-time analysis of location and traffic data to optimise customer pick-ups, reduce time spent in congestion, and determine pricing.

That said, it must clearly be acknowledged that the value of these firms is not entirely related to their data holdings. The firms’ “book value” will reflect the other assets they possess including physical assets (such as buildings and machines) and intangible assets (such as patents, brands), as well as financial assets including cash and stakes in other companies. As set out in Chapter 2., databases are not often recorded as assets in business accounts. In that way, any positive difference between the firms book value and its market value might be considered as attributable (at least in-part) to its data holdings (The Economist, 2020<sup>[31]</sup>). Furthermore, it is, of course, the case that companies such as Amazon and Uber, while relying on data-driven business models, deliver physical goods and services (rather than digitally-deliverable products).

Nevertheless, there are many more companies making extensive use of data, including their own customer databases, as a key foundation for value generation. Beyond the realm of “superstar” firms, though, there is no register of such data-driven firms. To try to identify them, this chapter draws upon lists indicating activity/proficiency in data and data analytics such as:

- “The cloud 100 list”, which covers firms in categories including “data platform”, “data analytics”, and “data management” (Konrad et al., 2020<sup>[32]</sup>)
- “The top 100 Big Data companies” (The Manifest, 2020<sup>[33]</sup>)
- “Top 30 big data companies” (Patrizio, 2019<sup>[34]</sup>)
- “34 Big data companies helping us make sense of the world” (Schroer, 2019<sup>[35]</sup>), and
- “Top Big Data analytics companies”. (Goodfirms, 2020<sup>[36]</sup>)

From those, a list of almost 3 000 target firms globally was compiled using web-scraping to capture firm names and locations (when available). However, identifying those which are listed firms (i.e. with shares traded on a stock market) and matching them to their stock market tickers is challenging to automate for several reasons and so a full matching exercise

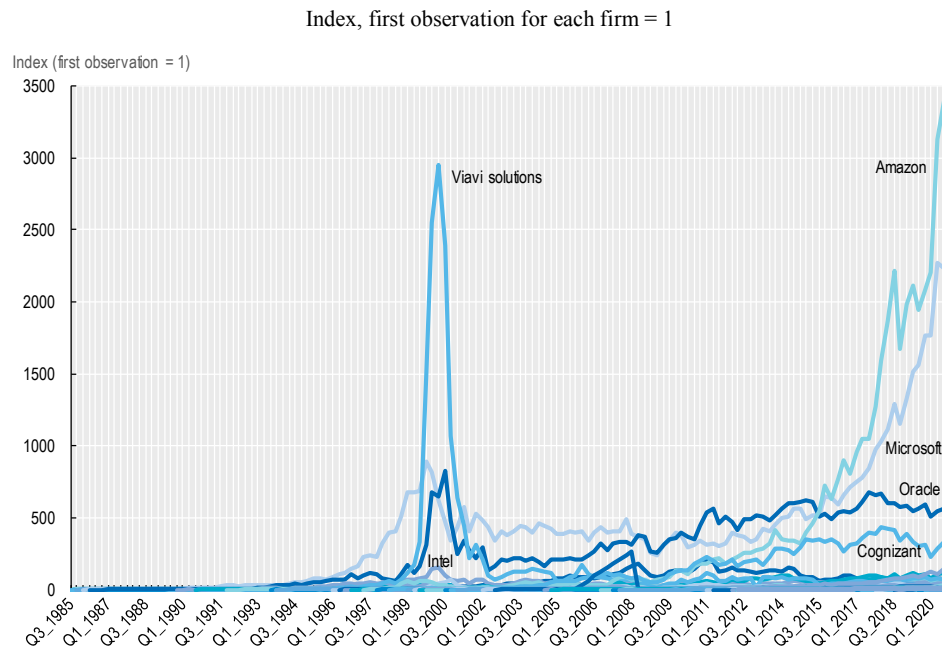
is left for further work. Nevertheless, in order to illustrate the market valuation approach we focussed on firms identified as being located in the USA (around 1100 firms) and looked up the stock market tickers for these through a combination of automated and manual matching. The vast majority of these searches found that the company was private, while some were entirely absent from the Yahoo! Finance database being searched. The result was 44 firms matched to listings on the NASDAQ and 32 firms matched to listings on the NYSE.

With the stock market tickers in-hand, web scraping was used to put together a dataset of market capitalisation for 64 firms going from Q3 (quarter 3) 2020 all the way back to Q3 1985<sup>7</sup>.

#### 4.4. Findings

The combined market capitalisation of the firms identified was over USD 5 trillion in Q3 2020. In Q3 1985, the combined market capitalisation of the six companies that existed at that time (Apache Corp., DXC Technology Company, Intel, IBM, Spectrum, and Unisys) was USD 88 billion, or 200 billion in 2020 dollars. The vast increase in the intervening period is driven by many additional firms listing – with some of these exhibiting strong growth in market capitalisation, as shown in **Figure 4.1**.

**Figure 4.1. Market capitalisation of “data-driven firms”, 1985-2020**



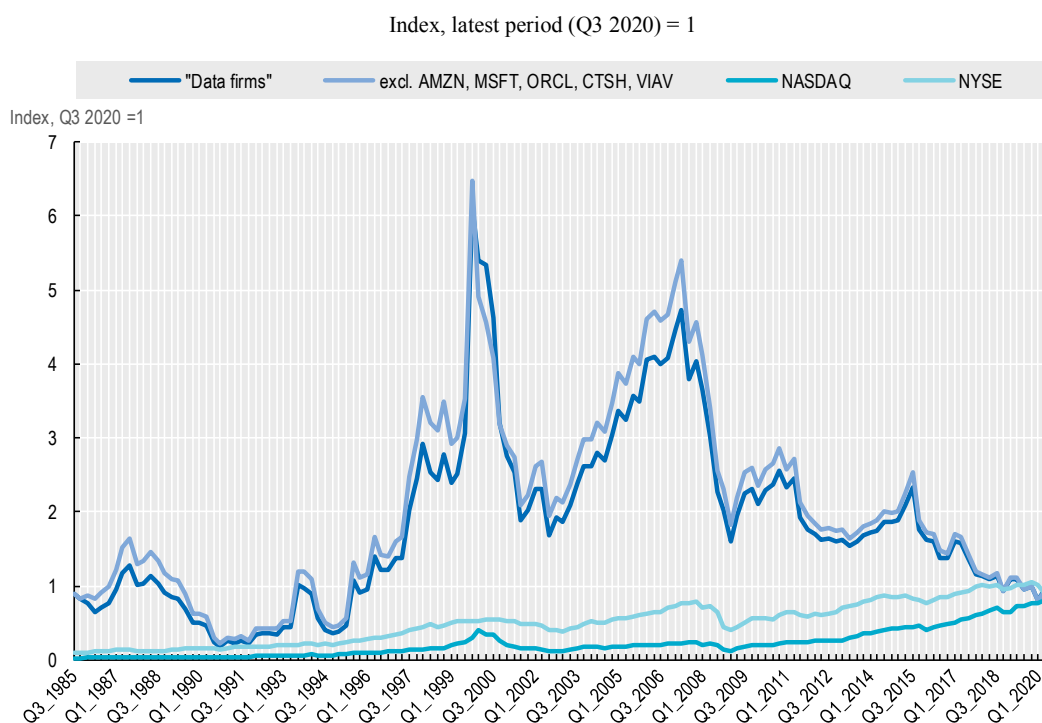
Source: OECD, based on data from Yahoo! Finance, October 2020.

The mean period-on-period change across all of these companies is 119%. However, by comparison the median change is 18%. Though this is still a rapid rate of growth, the disparity between the two measures illustrates the extent of skewness caused by a few extremely strongly growing firms.

Indeed, the market capitalisation growth indices of a few companies stand out markedly in Figure 4.1. Amazon and Microsoft have experienced meteoric increases in market capitalisation since around 2015, while Oracle and Cognisant have experienced sustained periods of market capitalisation growth that sets them apart from other firms in the sample. Several of these firms saw marked upticks in value around the time of the Dotcom bubble in the late 1990s, though this is dwarfed by the spike in market capitalisation of Viavi Solutions, formally JDS Uniphase, a network test, measurement, and assurance technology company based in California.

Taking the average of the company indices in each period, and rebasing this to express the series relative to Q3 2020, shows that the identified “data-driven firms” exceeded the growth rate of market capitalisation of both the overall NASDAQ composite and NYSE index throughout almost the whole period – and generally did so by a considerable margin (Figure 4.2).

**Figure 4.2. Average market capitalisation of “data-driven firms”, 1985-2020**



Note: “Data firms” based on the average index across all firms existing in each period.

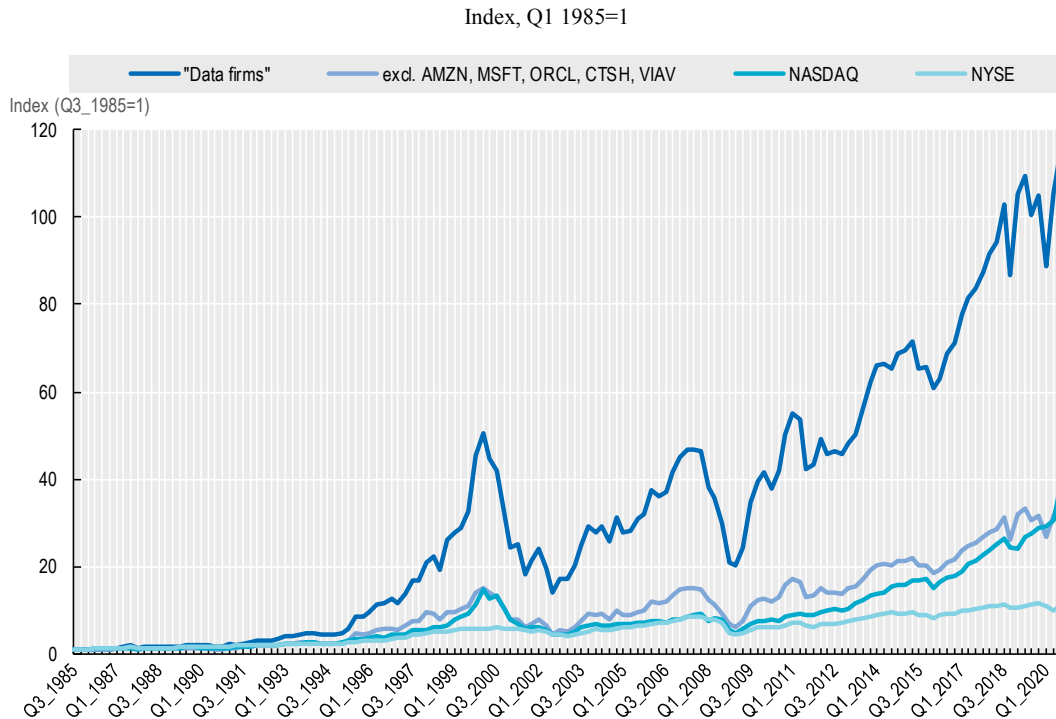
Source: OECD, based on data from Yahoo! Finance, fred.stlouisfed.org, and OECD Main Economic Indicators database, October 2020.

However, once the firms with unusually strong growth, as identified in Figure 3.1, are omitted the performance gap narrows significantly. Nevertheless, this sub-group does appear to outperform the overall NASDAQ and NYSE indices in terms of value growth in most of the period shown (Figure 4.3). These firms have increasingly outperformed the overall market indices since the early 2000s – a time period which roughly accords with the rise of the technologies needed to make large volumes of data manageable (broadband and high speed networks, cloud services, falling processing and data storage costs),



associated techniques and skills (Big Data analysis, “Data Scientist” as the “most sexy job of the 21<sup>st</sup> Century” (Davenport and Patil, 2012<sup>[37]</sup>), and a “social data revolution” that has seen individuals (knowingly or otherwise) take to mass participation in data and content generation (Wiegand, 2009<sup>[38]</sup>) (van Rijmenam, 2013<sup>[39]</sup>). That said, the difference between the overall NASDAQ index and the selected firms has diminished in recent periods.

**Figure 4.3. Index of average growth of market capitalisation of “data-driven firms”, 1985-2020**



Note: These figures are derived taking the natural logarithm transformation of the firm market capitalisation series and then taking period-period differences. Ln differences approximate the period-on-period percentage change while also allowing the sum of the period-on-period changes to equal the total change between the first and latest observation for each firm. The series are calculated by averaging the Ln differences across all “data-driven firms” and then cumulating these changes. Finally, the average Ln difference series are transformed back into normal index space by inversion (i.e. taking the exponential of the Ln difference values). The result is best understood as an “index of growth”.

“Data firms” based on the average index across all firms existing in each period.

Source: OECD, based on data from Yahoo! Finance, fred.stlouisfed.org, and OECD Main Economic Indicators database, October 2020.

#### 4.5. Concluding remarks

“Data-driven firms” appear to be worth over USD 5 trillion in the United States alone in 2020. On average, the firms do appear to build market value - as reflected by their market capitalisation - than other firms. That capital can support them in further investing in developing and expanding databases, and the complementary capabilities to analyse and monetise them.

However, due to the way the selected firms have been identified, it is not completely clear to what extent they hold and make use of data in their own activities – especially beyond the most high-profile “household names”. It is likely that at least some of the businesses identified are mainly involved in providing data analysis and consultancy services for other firms rather than building datasets of their own. Detailed research into each individual firm would be needed to gain greater understanding and finesse the company selection. It is possible that government agencies, industry bodies, and business intelligence firms might have the knowledge needed to improve the target list. For example, it would be possible to identify firms performing data analytics based on national ICT usage surveys. This would build on a previous collaboration between countries and the OECD, which looked at the value added, employment, and productivity of “big data industries” – though this new approach would require working directly with countries to share company names in order to scrape the relevant market valuation data. Such additional approaches for identifying the target firms could be particularly helpful in expanding the analysis to other countries as the company lists available seem to be somewhat US-centric.

Furthermore, other measures of firm performance could be analysed. As well as “enterprise value” measures, which complement market capitalisation measures by taking business assets and liabilities into account, measures of profitability or productivity could also be used to improve the analysis.

Nevertheless, the analysis suggests that data may be playing (potentially significant) role in firm valuations by the market and provides an additional perspective on the economic value linked to data and data flows. Future work is needed to extend and refine the approach to gain further insights.

## Chapter 5. Perspectives on the value of cross-border data flows

### 5.1. Introduction

Another important characteristic of data, besides the ability to be stored (i.e. held as a stock), is that it can be shared, transferred and “flow” between entities and national jurisdictions - both “physically” or in a purely legal sense while remaining within a single data centre (OECD, 2019<sup>[7]</sup>). As noted by Nguyen and Paczos (2020<sup>[3]</sup>), this is accelerated by the fact that the different steps of collection, storage and subsequent use of data can take place across several countries, where firms find it most convenient or profitable for their business.

As set-out in preceding chapters and in Nguyen and Paczos, the value of data is highly dependent on the information contained therein, the features of the data, and the context of its collection and use. This poses a considerable range of measurement challenges.

Although there are theoretical and practical measurement challenges, the volume and value associated with international data flows is of increasing measurement and policy interest (Casalini and Lopez Gonzalez, 2019<sup>[40]</sup>), (Lopez Gonzalez, 2019<sup>[41]</sup>). This chapter seeks perspectives on this.

At least three such perspectives related to international data flows might be considered:

1. Measures of the *volume* of international data/traffic flows related to trade
2. Measures of the *value* of trade in products delivered through data flows (trade in digitally deliverable products)
3. The value of economic activities *depending on* international data flows.

Clearly, these vary in scope, and the first could be viewed as underpinning the others.

The third of these options would require a way of identifying those economic activities (industries) and products that are not only data-intensive (i.e. make significant use of data inputs) but, since the focus here is on international data flows, more specifically are heavily reliant on flows of data from abroad for their business model and profitability. Furthermore, it might also further consider all e-commerce activities involving customers abroad – though there would be a need to work out how to attribute value to the data flows involved (e.g. in communicating the products for sale, establishing the order, where items should be delivered, etc.) in cases such as when the underlying product being traded is a good (that has intrinsic value itself irrespective of the data flows related to it). While such assessments could be valuable, they are beyond the scope of this work. Nevertheless, these could be avenues for further developing and building upon this work in future.

Meanwhile, the first option appears relatively straightforward. In principle, “international data flows” appear straightforward to define and readily measureable. Data flows are routinely measured in terms of the volume of data transiting a network in a given period of time – and typically communicated in kilobits/megabits/gigabits per second etc. (i.e. kbps/mbps/gbps). It is logical that cumulating these quantities over a longer time period should lead to a measure of the total traffic transiting the relevant network e.g. the Internet. Furthermore, it also appears conceptually straightforward to define “international data flows” as those where data move from a machine or storage device located in one country to a device in another country.

Despite this, no directly observed measures of total Internet traffic flows are available. There is a good reason for this: “the Internet” is not a singular network, it is a complex “network of networks” with traffic being routed in myriad ways that do not provide a single point through which to measure overall traffic. Furthermore, the architecture of the Internet does not lend itself to tracking data flows across geographical space in a meaningful way. As such, while defining and measuring “international data flows” appears conceptually straightforward, it quickly emerges to be very difficult in practice.

Nevertheless, there are some estimates available. For example, Cisco (2018<sub>[42]</sub>) estimates that global Internet traffic was around 160 Exabyte (i.e. 160 billion Gigabyte) per month in 2019, which would equate to around 1 900 Exabyte over the year. However, this is not a direct measurement of observed data traffic through the Internet; it is based on a combination of “independent analyst forecasts and real-world network usage data”, with assessments of factors such as connection numbers, adoption of a particular “application” type (e.g. “Internet video”), usage, and bitrates being used to estimate Internet traffic volumes.

Even if such an estimate can be taken as representative of total Internet traffic, the question at hand for this analysis is not simply “what is the volume of Internet traffic?” but “how much of that traffic relates to international trade?” The complexities of the real world are strong hurdles to answering these questions. A considerable volume of data flows take place within data centres (OECD, 2019<sub>[7]</sub>), meanwhile, companies offering services that generate considerable volumes of data flow – such as streaming video providers – take steps to reduce or mitigate the impact of these flows on the Internet as a whole. For example, Netflix uses AWS (Amazon Web Services) to host its content library in AWS datacentres across at least 17 countries worldwide (AWS, 2020<sub>[43]</sub>) allowing it to put content closer to users in different countries. This reduces the physical distance a video stream needs to travel to reach a viewer, improving latency and reducing how much of the Internet a stream needs to traverse from server to viewer (Netflix, 2016<sub>[44]</sub>). Netflix also reports works with Internet Exchange Points (IXPs) and individual Internet service providers (ISPs) to install content servers *within* ISPs’ own networks. Because of this, Netflix explains that video stream can be served to the viewer entirely inside the ISP’s network, limiting the need for transiting multiple networks (Netflix, 2019<sub>[45]</sub>).

As well as the challenges of identifying “international data flows”, even if the volume of data flows could be reliably monitored this tells us nothing of their content. Whether a data flow comprises streaming video, product blueprints, e-commerce orders, or online gameplay (for example) will have implications for the economic value associated with the data flow.

Furthermore, it does not appear likely that economic value is directly related to data volumes. Streaming video is estimated to account for the large majority of Internet traffic – 75% of traffic in 2017 (Cisco, 2018<sub>[42]</sub>), and undoubtedly generates vast sums of economic value (through subscription payments, advertising revenue, etc.). However, the relatively small data flows associated with most e-commerce transactions (i.e. presenting products for sale, transmission of information during the order process) are likely to underpin a more significant portion of economic activity. For example, US Census Bureau data show that e-commerce sales were worth over USD 0.5 trillion in 2018, around 10% of total retail sales (US Census Bureau, 2020<sub>[46]</sub>) – and appear to have risen sharply in 2020 (US Census Bureau, 2020<sub>[47]</sub>).

Indeed, Netflix is *estimated* to account for around 13% of downstream Internet traffic in 2019 (Sandvine, 2019<sub>[48]</sub>)<sup>8</sup> - and through this generated USD 20.2 billion in revenue

(Netflix, 2020<sup>[49]</sup>). Nevertheless, in the aforementioned case where Netflix has installed content servers within a subscriber's ISP's network, there might be an international trade flow reflecting a non-US subscriber's payment to Netflix for an export of streaming services<sup>9</sup> but the delivery of that service will result in negligible Internet traffic beyond the ISP's network as the data flow delivering the service to the subscriber occurs mainly within that.

Thus, even in the case of Netflix, a company with homogeneous product<sup>10</sup>, and responsible for significant Internet traffic flows, it is challenging to relate the estimated data flow to flows of economic value. For other companies, with more heterogeneous revenue streams – for example, spanning data-based products, operating systems and applications, and hardware (as in the cases of Alphabet and Facebook) – inferring how much of their revenue is related to data flows is even more challenging. Furthermore, there is even less availability of information and estimates of the data flows relating to the great many smaller firms that are producing digitally deliverable products.

Overall, an approach looking at the volume of international data flows does not appear to be fertile ground for development at this time for three key reasons:

1. A lack of reliable and verifiable measures (or proxy for measures) of total Internet traffic
2. An inability to distinguish those Internet traffic flows that cross international borders or move data between legal jurisdictions.
3. The challenge of relating data flows to economic value generated.

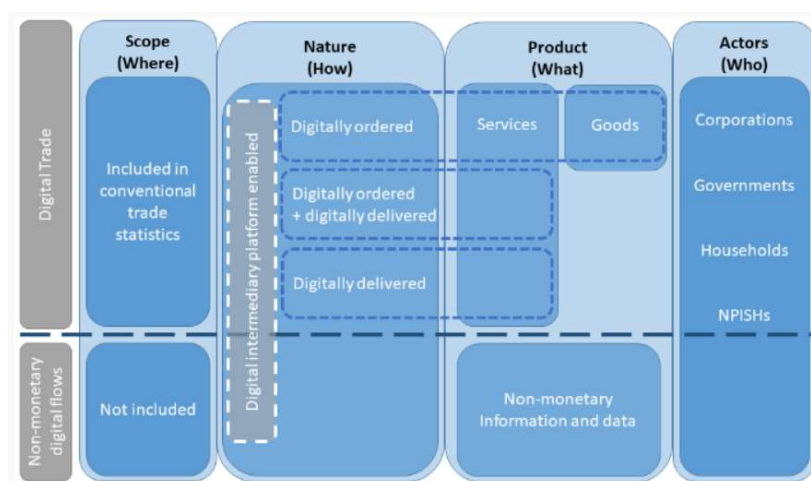
The next section looks in detail at the second approach – based around identifying trade flows in digitally deliverable products.

## 5.2. Trade in products delivered through data flows

As the digital transformation continues, an ever-increasing range of products is being embodied in digital form and delivered through Internet data flows. This ranges from relatively “new” services born of digital technology such as the streaming media and cloud computing discussed above, through communications services (e.g. voice services carried through the Internet), knowledge services such as Research and Development, and design services (with outputs such as designs, blueprints, reports, and databases being delivered in digital form), financial services (e.g. online banking and investment services), to services that have long been associated with physical provision such as education services (e.g. online lessons) and even health services (e.g. telemedicine). These products are often referred to as “digitally deliverable” (OECD-WTO-IMF, 2019<sup>[50]</sup>), meaning that they can, and are increasingly, being delivered in the form of data flows.

Countries routinely publish information on services imports and exports; in principle these might be used to gain a perspective on international data flows through the lens of trade in products delivered through data flows. However, trade statistics do not currently delineate such “digitally delivered” services. Nevertheless, efforts are being made to implement such distinctions in line with the framework set out in the OECD-WTO-IMF Handbook on Measuring Digital Trade (2019<sup>[50]</sup>) and reproduced in **Figure 5.1**.

Nevertheless, achieving implementation across a “critical mass” of countries is likely to take several years. In the meantime, an alternative approach is to identify which products are *likely to be delivered in digital form* (i.e. through data flows).

**Figure 5.1. The conceptual framework for digital trade**

Note: Digital Intermediation Platforms (DIPs) are also an important component of Actors. Their current explicit inclusion in the nature of transactions (which may change depending on how measurement efforts evolve) reflects the scope for measuring modes of digital delivery and/or ordering through targeted surveys of DIPs. For a more detailed description of non-monetary information and data, and more generally measurement challenges related to information and data (paid, which are included in the scope of digital trade, or non-monetary, which, currently, are not) see below. Deliberations continue on the precise terminology concerning non-monetary flows. Future versions of this Handbook may introduce different terminology.

Source (OECD-WTO-IMF, 2019<sub>[50]</sub>).

There is no generally accepted standard list of “digitally deliverable services” i.e. services which are typically delivered through flows of data. Nevertheless, attempts have been made to identify such items within the classifications used in several sources of trade data. The following sections try to identify “rough proxies” of cross-border data flows.

### **5.2.1. Information and communication services firms’ international trade**

In data broken down by the International Standard Industry Classification (ISIC rev. 4), trade in the products of firms in industry “Section J: Information and Communication” might provide one perspective on cross-border data flows. The full classification structure is presented in **Table 5.1**.

**Table 5.1. ISIC rev. 4 Section J Information and Communication services**

Firms are classified to industries based on the business activity that provides their main source of value added (sometimes proxied by revenue or headcount)

ISIC class	Description	Notes (based on definition):
58	Publishing activities	
-5811	Book publishing	Includes e-books, audio books.
-5812	Publishing of directories and mailing lists	Includes those published in electronic form, including in the Internet
-5813	Publishing of newspapers, journals, and periodicals	
-5819	Other publishing activities	Including on-line
-5820	Software publishing	On-line provision of software classified in 6311.
59	Motion picture, video and television programme production, sound recording and music publishing activities.	
-5911	Motion picture, video and television programme production activities	Activities of making films, programmes, commercials.
-5912	Motion picture, video and television programme post-production activities	Activities such as editing, subtitling, graphics and special effects, developing film, reproduction for distribution.
-5913	Motion picture, video and television programme distribution activities	Definition specified distribution in various physical forms "film, video tapes, DVDs, and similar" to broadcasters and exhibitors. Does not appear to cover streaming distribution (6311).
-5914	Motion picture projection activities	Cinemas and similar.
-5920	Sound recording and music publishing activities	Does not appear to include publishing via streaming distribution (6311)
60	Programming and broadcasting activities	
-6010	Radio broadcasting	Includes Internet radio
-6020	Television programming and broadcasting activities	Does not mention Internet broadcast/streaming
61	Telecommunications	
-6110	Wired telecommunications activities	Includes provision of Internet access by the operator of the wired/wireless/satellite infrastructure
-6120	Wireless telecommunications activities	
-6130	Satellite telecommunications activities	
-6190	Other telecommunications activities	Specifically includes VOIP, 3 <sup>rd</sup> party ISPs (which do not own the telecoms infrastructure)
62	Computer programming, consultancy, and related activities	
-6201	Computer programming activities	The writing, modifying, testing, and supporting of software (excl. packaged software publishing)
-6202	Computer consultancy and computer facilities management activities	Planning and design of computer systems; on-site management, operation, and support
-6209	Other information technology and computer service activities	Includes software installation services
63	Information service activities	
-6311	Data processing, hosting, and related activities	Includes web hosting, streaming services, and cloud services.
-6312	Web portals	Includes search engines, and portals such as "media sites providing periodically updated content"
-639	Other information service activities	
-6391	News agency activities	
-6399	Other information service activities nec.	

Source: OECD based on

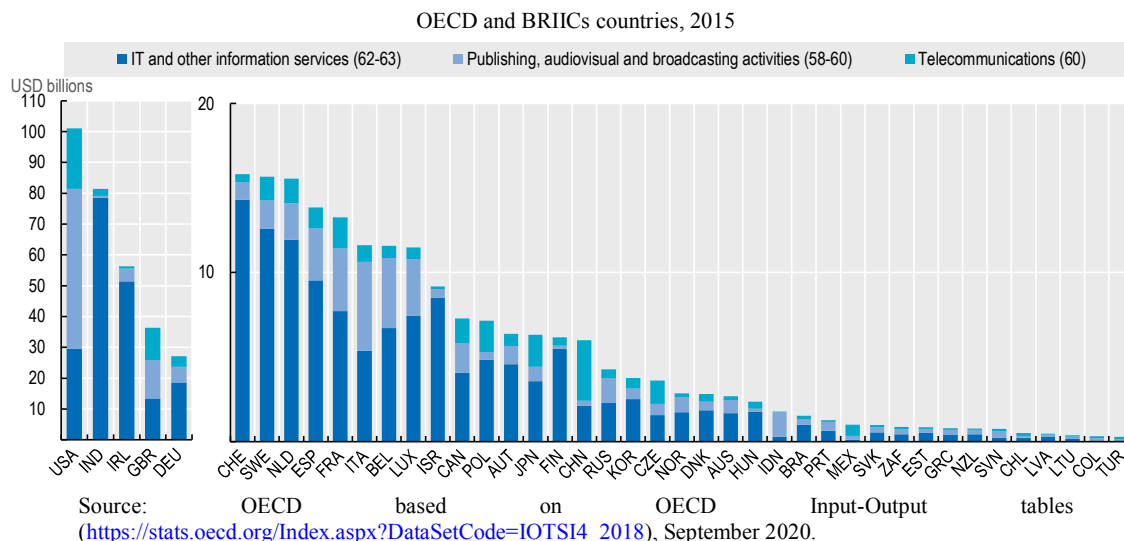
[https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC\\_Rev\\_4\\_publication\\_English.pdf](https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_publication_English.pdf)

Several classes within this comprise firms mainly producing products that can only be delivered through data flows (web portal services, software publishing – even software traded on physical media would still represent a cross-border data flow). Meanwhile, many of the other industries listed deal in products that may be delivered in physical form or through data flows and this is specified in the class definitions. For example, the “book publishing” class explicitly includes both e-books and audiobooks as well as books in print form.

This means that the economic value of products these industries deliver through data flows is included in trade figures, but it also means that they are not easily distinguished from those relating to physical (non-data) flows. As such, it is not possible to use the industry data to isolate the production and trade in digitally deliverable products and so this will provide only one *rough approximation* of the value associated with some data flows.

**Figure 5.2** presents figures from the OECD Input-output tables, which shows that firms in ICIC section J made exports worth almost 400 billion USD across the OECD in 2015. In BRIICS<sup>11</sup> countries, the Information and communication sector made almost a further 100 billion of exports. The majority of these, around two-thirds overall, were generated by firms classified in the “IT and information services” sector (ISIC rev.4 divisions 62-63). Meanwhile, Publishing, audiovisual, and broadcasting activities (62-63) accounts for around 20% of these exports, and Telecommunications for the remaining 15% - although the relative shares vary considerably between countries.

**Figure 5.2. Exports by firms in ISIC section J – Information and communication, USD billions**

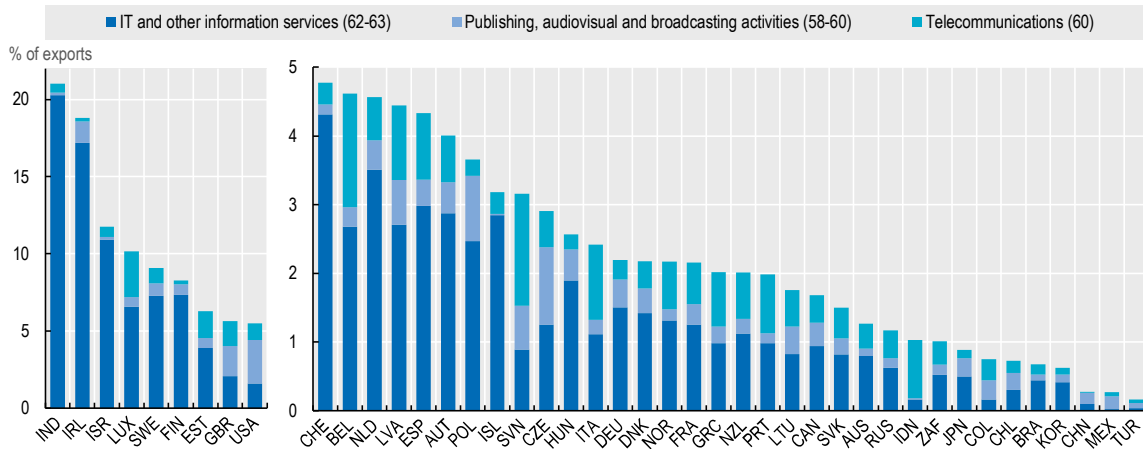


India and Ireland stand out with exports by firms classified in industry section J (Information and communication services) accounting for around 20% of total export revenues (**Figure 5.3**). They also account for a considerable share of total exports (5% or more) in many countries including Israel, Luxembourg, the United Kingdom, and the United States. They make up the lowest share – less than 0.3% – in China, Mexico, and Turkey.



**Figure 5.3. Exports by firms in ISIC section J – Information and communication, export share**

OECD and BRIICs countries, 2015



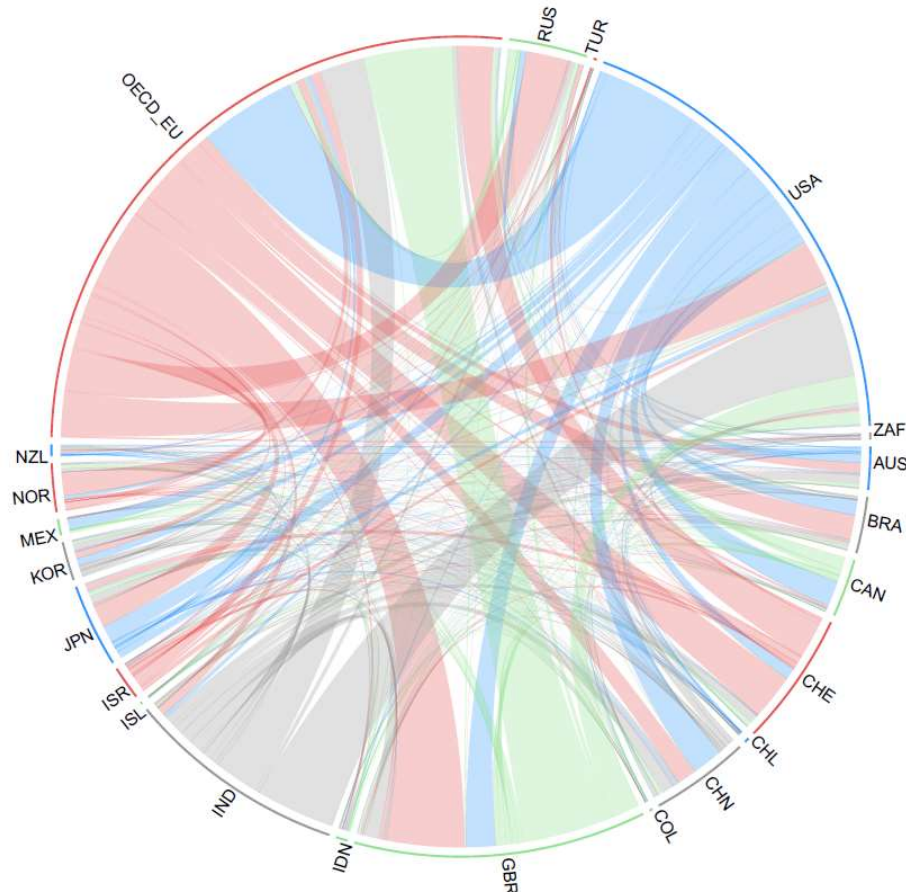
Source: OECD based on OECD Input-Output tables ([https://stats.oecd.org/Index.aspx?DataSetCode=IOTS14\\_2018](https://stats.oecd.org/Index.aspx?DataSetCode=IOTS14_2018)), September 2020.

While these figures do not indicate the absolute volume of data flows exiting each country, they might provide one perspective on the differing relative scale of such flows from each country – with countries such as the United States, India, Ireland, United Kingdom, and Germany appearing likely to be party to relatively large data flows, and countries such as South Africa, Chile, Lithuania, Colombia, and Turkey appearing less internationally connected in terms of data flows.

A defining feature of any flow is that it has a discrete origin and end-point. The OECD inter-country input-output tables (ICIO, <https://oe.cd/icio>) can be used to extend this approach to look at bi-lateral international trade flows of output produced by firms in ISIC section J, as shown in **Figure 5.4**. The largest flows of products produced by firms in the Information and communication industry comprise exports from the USA to EU countries, Exports from India to the United States, and trade in both directions between EU countries and the United Kingdom. As some of the products underlying these trade flows are “digitally deliverable”, it is likely that the flows plotted bear some relation to data flows.

**Figure 5.4. International trade in ISIC J Information and Communication services – bilateral flows**

OECD and BRIICS countries, 2015



How to read this chart: the outer ring colour for each country is applied to that country's export flows. For example, the outer ring colour for the USA is blue, and the large blue mass of flows emanating from the USA segment represents its exports. The flows of various colours going to the USA segment represent imports. Note: "OECD\_EU" refers to OECD countries that are also EU member states.

Source: OECD based on Inter-Country Input-Output database, 2015.

However, by focussing on an industry perspective, the analysis is further removed from the underlying phenomenon of interest than is necessary. *Being delivered through data flows is an attribute of products rather than industries* and while firms are classified to industries based on the products that are their main source of value added/revenue, it is crucial to note that they can (and often do) produce other products. This is especially true of many of the companies likely to be involved in generating considerable volumes of data flow.

Take the US company Alphabet for instance. Its primary source of revenue is from selling online advertising space, including on its Google search engine, YouTube video service, gmail email service, and more. However, Alphabet also has divisions producing and licensing the Android mobile operating system, designing hardware such as phones and smart devices, selling telecommunications services, and even working on autonomous car technologies. Individually, these activities would belong in different ISIC industries.

Furthermore, Alphabet itself is a holding company so its direct activities should be recorded in ISIC rev.4 section K (financial services). The extent to which the activities of its constituent divisions (Google, YouTube, Android, Waymo, Nest, etc.) are split out and classified to separate industries in practice depends on whether those producing the statistics are able to get the necessary component information – which is not provided separately in Alphabet’s financial reports. Furthermore, Alphabet is an extremely large and high profile case and so likely to command attention and investigation from National Statistics agencies’ “large case units”, but there are many smaller companies with simpler structures that produce a range of products, including those associated with other industries.

The outcome is that, by focussing on trade broken down by industry, the figures will include secondary production of products that are not Information and Communication services and at the same time will exclude Information and Communication services produced as secondary products by firms classified in other industries.

A focus on products would avoid some of the distortion potentially introduced with industry data. Perhaps more importantly, most trade data sources report on the basis of products rather than the industry the producer is classified in. As such, a product-based approach opens up the prospect of using richer data sources.

### ***5.2.2. Trade in products mainly produced by Information and Communication Service firms***

The UN Central Product Classification (CPC 2.1) is a sister classification to the ISIC and the statistical product classification most closely related to it since “each subclass of the CPC consists of goods or services that are predominantly produced in one class of ISIC” (United Nations, 1991<sup>[51]</sup>). Despite this, the concordance of ISIC section J Information and Communication services, with CPC 2.1 is not straightforward. The 23 individual industries in ISIC section J concord with over 100 unique CPC product classes, as shown in **Annex Table 5.A.1**. As should be expected given the above, many of these products themselves concord with several industries (i.e. are produced across multiple industries) - the problem with focussing on an industry perspective. Furthermore, the concurring products are scattered across various sections of the CPC structure. This matters because data are not typically available at the level of individual products – instead being aggregated into product groups in order to manage disclosiveness and robustness of the data. This is likely to make populating the full mapping challenging.

Additionally, while the CPC is the product classification most complementary to ISIC, it is not routinely used in international trade data. The main classification system for trade in services is the Extended Balance of Payments Services classification (EBOPS) (OECD, 2010<sup>[52]</sup>). Unlike for the CPC, there does not appear to be an official concordance table available between ISIC and EBOPS. Nevertheless, as EBOPS is a product classification it can be linked to the CPC (UNSD, 2010<sup>[53]</sup>). As such, it is possible to take the CPC products identified as being outputs of the Information and Communication sector (ISIC section J) and map that through to EBOPS 2010 products, as also shown in Annex Table 5.A.1. With this approach, most of the outputs produced by firms in ISIC J can be linked to EBOPS products. Furthermore, almost all of the few products that do not map through are clearly goods rather than services (as shown by CPC descriptions stating things such as “in print” or “on physical media”), and hence are unlikely to be generating cross-border flows of digital data.

However, it should be noted that in practice it is necessary to include only those EBOPS classes where CPC products mapped through from firms in ISIC J account for all or almost

all of the CPC products concurring with those EBOPS classes, while EBOPS product classes where the relevant CPC products are only one or two among many concurring CPC classes are excluded to avoid inflating the estimate. As a practical example, EBOPS class SK11 “Audiovisual services” is included in the following estimates despite, one of the four concurring CPC codes (73220 rental of videotapes and disks) not mapping through from firms in ISIC J. By contrast, EBOPS class SK12 “artistic related services” is not included because only one concurring CPC product out of three (84311 On-line books) was mapped through from firms in ISIC J (CPC 962 “performing arts and other live entertainment and promotion thereof”, and CPC 963 “Services of performing and other artists” were not and furthermore seem unlikely to be strongly related to international data flows).

With this in hand, the OECD EBOPS database can be used to examine exports and imports of these products. **Figure 5.5** shows the available exports data, while **Figure 5.6** presents the available data on imports of the identified products. In both cases, the United States and Germany are engaged in by far the greatest trade flows. However, these data are not available for several other economies that featured strongly in Information and Communication sector trade as shown in Figure 5.2 – such as India and the United Kingdom - and so the country ordering can’t be meaningfully compared.

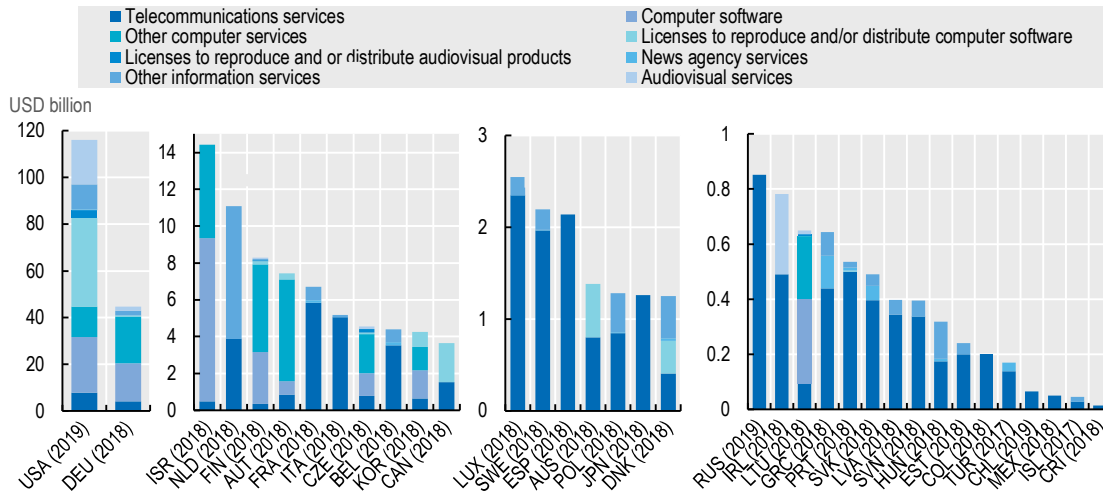
Furthermore, most countries are missing at least some components: only Finland, Lithuania, and the United States have data for all categories available. This limits the ability to make meaningful comparisons amongst countries. To address this, **Figure 5.7** and **Figure 5.8** takes “Telecommunications, computer, and information services”, which combines several of the more detailed service classes and is available for more countries (all of the countries shown in the figures). In addition, where a component is missing for a country but the parent product class is available, the portion relating to the product of interest is estimated based on the average share of that product in the parent class across those countries for which both are available.

The estimation process materially impacts the figures for many countries and the ordering. For example, exports of around USD 11 billion of Telecommunications services and Information services from the Netherlands are supplemented by around USD 16 billion of Computer Services included in the parent aggregate but not split out in the database, almost USD 19 billion of estimated licensing services, and USD 1.6 billion estimated Audiovisual services. This leads to a total of USD 47 billion of exports and moves the Netherlands from 4<sup>th</sup> to 3<sup>rd</sup> in terms of export volume. Similarly, the figure for Ireland goes from USD 0.8 billion of Telecoms and Audiovisual services to over USD 100 billion once the parent aggregate, which covers computer and information services missing from the more detailed breakdown, is included. This takes both countries above Germany, which moves down to 4<sup>th</sup> in terms of export volume.

Similarly, making imputations and taking the composite of Telecommunications, computer, and information services in place of the available sub-components substantially increases the import figures.

**Figure 5.5. Exports of services produced by firms in the Information and Communication services sector (ISIC section J)**

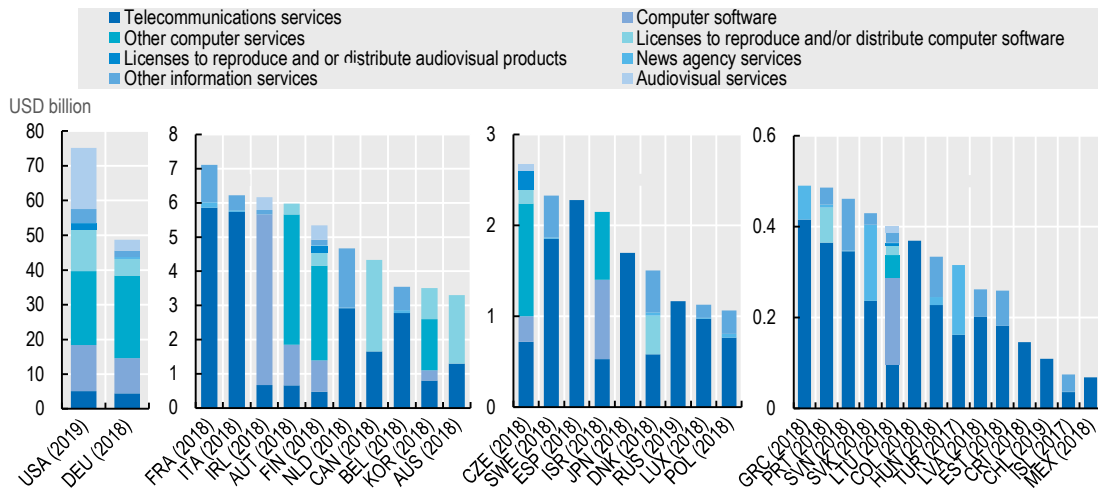
USD billions, 2019 or latest



Note: Components not shown are missing. Also includes production by firms not classified in ISIC section J.  
Source: OECD, based on OECD EBOPS database, October 2020.

**Figure 5.6. Imports of services produced by firms in the Information and Communication services sector (ISIC section J)**

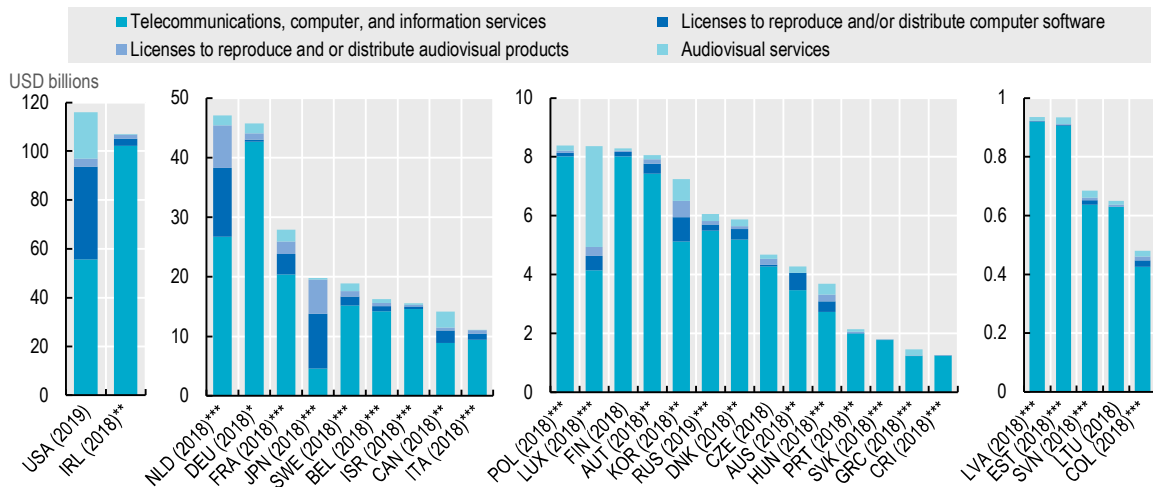
USD billions, 2019 or latest



Note: Components not shown are missing. Also includes production by firms not classified in ISIC section J.  
Source: OECD, based on OECD EBOPS database, October 2020.

**Figure 5.7. Exports of services produced by firms in the Information and Communication services sector (ISIC section J) – including estimations**

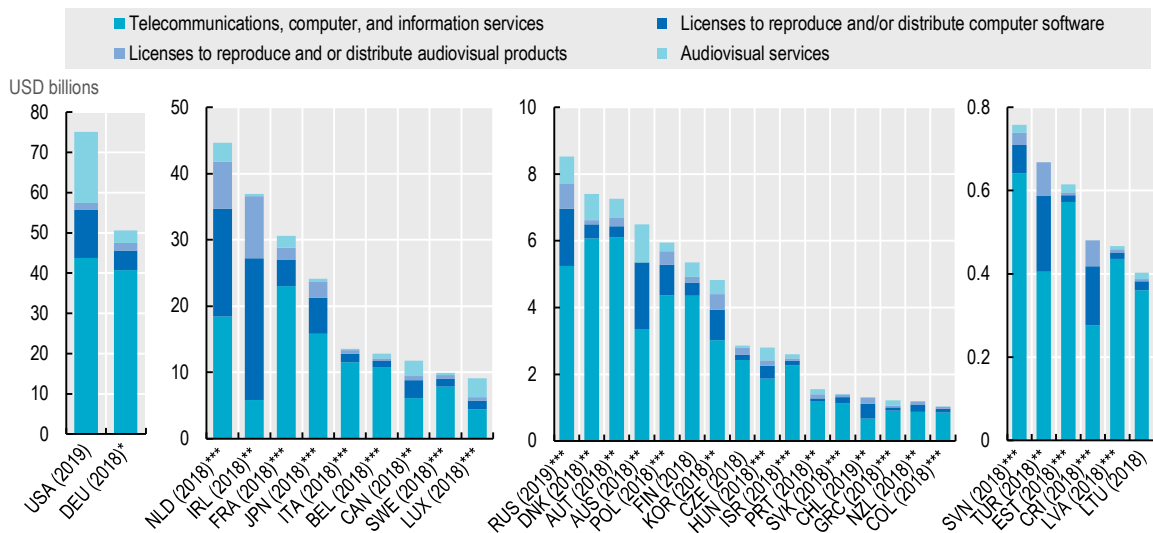
USD billions, 2019 or latest



Note: Asterisks (\*) indicate the number of components estimated using the average across countries with available data. Telecommunications, computer, and information services is available for all countries but for Lithuania excludes information services. Figures include production by firms not classified in ISIC section J. Source: OECD, based on OECD EBOPS database, October 2020

**Figure 5.8. Imports of services produced by firms in the Information and Communication services sector (ISIC section J) – including estimations**

USD billions, 2019 or latest



Note: Asterisks (\*) indicate the number of components estimated using the average across countries with available data. Telecommunications, computer, and information services is available for all countries. Figures include production by firms not classified in ISIC section J. Source: OECD, based on OECD EBOPS database, October 2020

Following the approach established above, while these trade flows do not directly indicate the scale of data inflows and outflows across jurisdictional borders they might nevertheless provide an indication of the relative scale of these flows with respect to different countries. Furthermore, by moving from purely industry-based flows to focussing on the subset of ISIC J exports and imports comprising products that are not explicitly delivered in physical form, these estimates will more closely represent the value generated through underlying international data flows.

The OECD EBOPS database also benefits from having trading partner information for many of the countries shown above. This allows the bi-lateral export and import flows to be visualised in **Figure 5.9** and **Figure 5.10**, respectively. While in principle the exports reported by country X as going to Y should equal the imports reported by country Y as coming from country X, there are various reasons why this is not the case including time of recording, valuation, product classifications, and others. Some databases use statistical methods and adjustments to “balance” import and export flows (this is the case for the ICIO database used for **Figure 5.4**, for example). However the OECD EBOPS database is unbalanced and so separate chord diagrams are presented for both reported exports by customer country and reported imports by seller country – with differences in the resulting picture found.

When focussing on exports, with recipients reported by the exporting country (Figure 5.9), the most striking feature is a large flow from Ireland to countries in the Rest of the World (ROW) grouping. However, it is important to note that, for some of the relevant products Ireland does not individually report exports to several of the OECD+BRIICS countries itemised separately in the diagram. As such these amounts also end up in the “ROW” figure, which is best interpreted as “ROW/unspecified”. As an illustrative example, Ireland reported USD 102 billion exports of Telecommunication, computer, and information services worldwide. Of these, an aggregate of USD 44 billion was reported as going to EU countries but only USD 33 billion was itemised as going to separate EU countries. The residual, USD 11 billion will be subsumed into the ROW figure, along with any amounts going to other OECD countries not itemised separately including AUS, CHL, COL, JPN, NZL and BRIICS countries other than IDN (which is itemised individually).

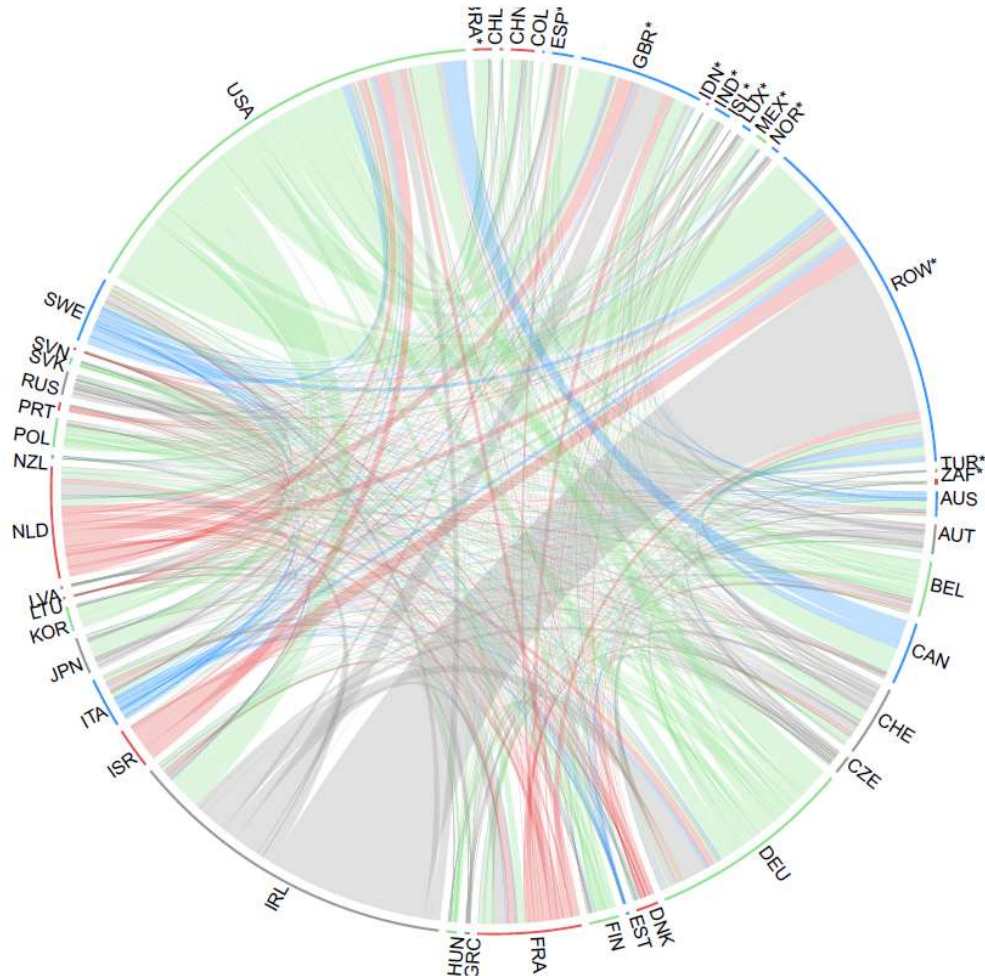
Flows to the ROW, then, should be interpreted carefully. However, flows shown as going from one specific country to another, including the notable flow between Ireland and the United Kingdom, and the similarly sized flow to Germany, are all as itemised by the reporting (exporting) country. Relatively large export flows can also be observed from the United States to the United Kingdom, Canada, and Germany. The latter two, report similarly sized export flows going *to* the US (GBR does not report exports by recipient and hence is included only as a customer country). There is also a sizeable export flow from the United States to Ireland – several times greater than the flow going the other way (i.e. IRL to USA). The general picture otherwise is of many relatively small flows occurring amongst the countries shown.

The same is also largely the case when looking at imports, with the origin countries as reported by the importing country. However, an import reporting basis does provide a somewhat different picture – with India appearing as an especially important supplier of these services to the United States alongside Ireland, the United Kingdom, Canada, as well as Israel.



**Figure 5.9. Exports of services produced by firms in the Information and Communication services sector (ISIC section J)**

Purchasing countries as reported by exporting country, 2018



How to read this chart: The outer ring colour for each country is applied to that country's reported *export flows*. For example, the outer ring colour for the USA is green, and the large green mass of flows emanating from the USA segment represents exports reported by the US as going to other countries. Meanwhile, the flows of various colours going to the USA segment represent exports reported by other countries reported by them as going to the US.

Note: Also includes production of these products by firms not classified in ISIC section J.

Countries marked with an asterisk (\*) are only available as a recipient reported by the exporting country, their own exports are unavailable.

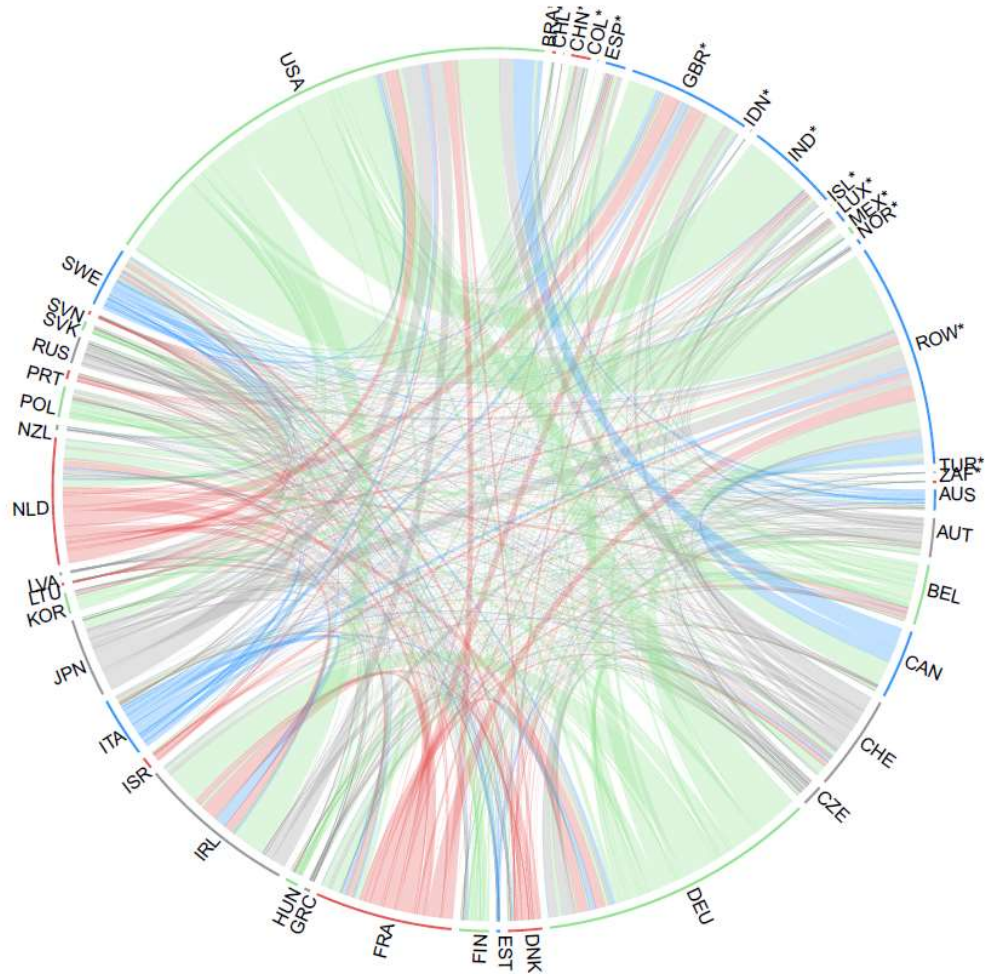
In principle "ROW" refers to Rest of the World i.e. recipient countries not shown separately. **However**, some reporting countries do not provide data for all OECD+BRICS recipient countries individually. Exports to countries not itemised individually will also be included in the ROW figure. ROW should therefore be interpreted as "ROW/unspecified".

Source: OECD, based on OECD EBOPS database, October 2020.



**Figure 5.10. Imports of services produced by firms in the Information and Communication services sector (ISIC section J)**

Selling countries as reported by importing country, 2018



How to read this chart: The outer ring colour for each country is applied to that country's reported import flows. For example, the outer ring colour for the USA is green, and the large green mass of flows terminating at the USA segment represents imports reported by the US as coming from various countries. Meanwhile, the flows of various colours going from the USA segment to other countries represent imports reported by those partner countries as coming from the US.

Note: Also includes production of these products by firms not classified in ISIC section J.

Countries marked with an asterisk (\*) are only available as an origin reported by the importing country, their own imports are unavailable.

In principle "ROW" refers to Rest of the World i.e. supplier countries not shown separately. **However**, some reporting countries do not provide data for all OECD+BRICS supplier countries individually. Imports from countries not itemised individually will also be included in the ROW figure. ROW should therefore be interpreted as "ROW/unspecified".

Source: OECD, based on OECD EBOPS database, October 2020.

Overall, though, the main conclusion from these figures appears to be that the United States is the global leader among several countries that are global “hubs” for products mainly produced by firms classified in ISIC section J (Information and Communication services) and which appear likely to be delivered through data flows. Other such hubs include France, Germany, Ireland, India, and the Netherlands.

This industry-derived perspective represents one way to identify services product trade flows which are likely to be underpinned by, and hence to some degree correlated with, data flows. The following section takes an alternative approach based on looking at product definitions and identifying which products can be digitised and so *could* be delivered through data flows and/or are likely to be being delivered through flows of data.

### 5.2.3. Trade in “digitally deliverable services”

The digital transformation greatly affects trade. In particular, the Internet has made it easier to buy, sell and deliver certain services, such as telecommunications and audiovisual content, across borders. It has also enabled cross-border electronic delivery of financial, business and knowledge services, such as sales and marketing, management, administration, and back office services, engineering, R&D, and education (UNCTAD, 2015<sup>[54]</sup>). Furthermore, the Internet has also led to new categories of digitally-delivered services, such as web search.

The main approach here is one of looking at the detail of product definitions and descriptions, and using these to identify product categories that are likely to be delivered through flows of digital data. The flows identified products can then be measured using trade datasets.

For example, UNCTAD (2015<sup>[54]</sup>) identifies a range services they refer to as “potentially ICT-enabled services” that can be “delivered remotely” (i.e. through flows of data). As set out in OECD-WTO-IMF (2019<sup>[50]</sup>), these are analogous to “digitally deliverable services” – i.e. services that can be, and often are, delivered through flows of digital data. The identified products are set out in **Table 5.2** and go beyond those identified from the industry perspective above by also including a wider range of Information services, Insurance and financial services, management, administrative, and back-office services, a wider range of licensing services, Engineering, related technical services and R&D, and Education and training services.

The OECD EBOPS database is used to present export flows of these products in **Figure 5.11**, and imports in **Figure 5.12**. Although the total value of the services covered is much greater than in section 5.2.2 due to the broader scope of products covered, the overall patterns that emerge are not dissimilar with the United States, Germany, Ireland, the Netherlands, and France appearing to be hubs for imports and exports of digitally deliverable services while a considerable number of smaller flows occur with and between the other countries shown.

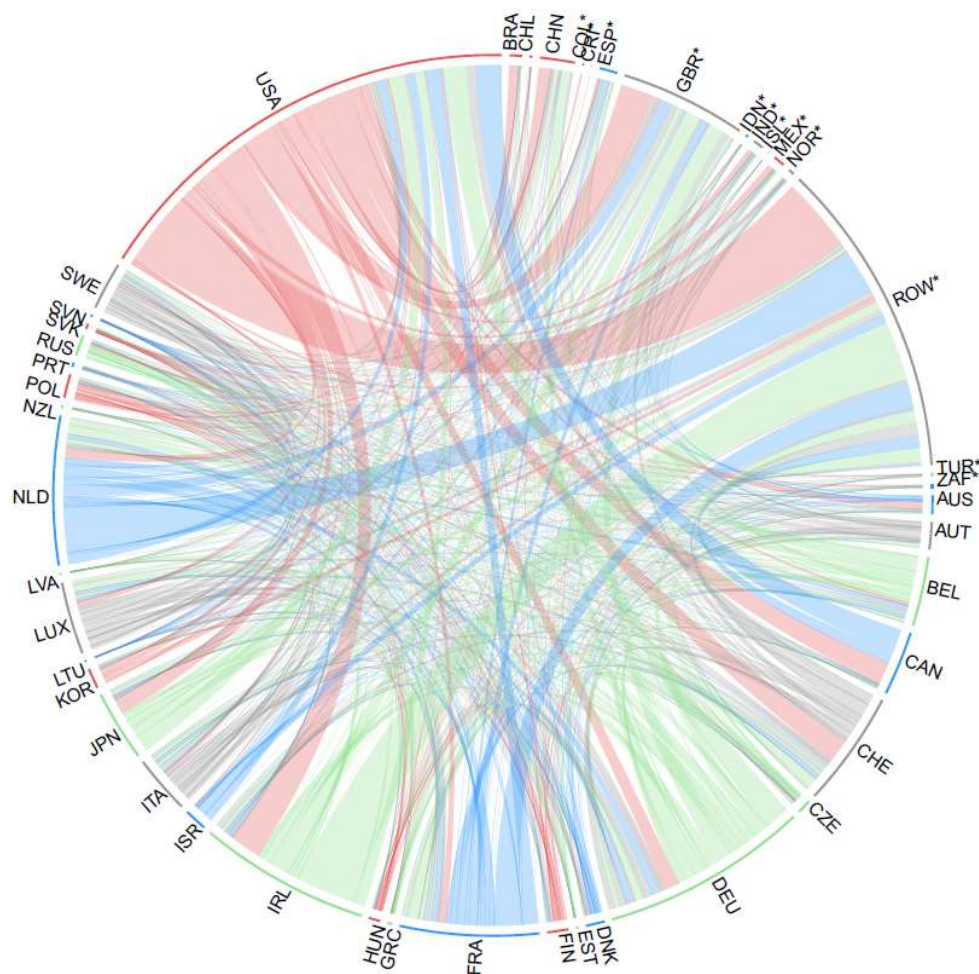
**Table 5.2. UNCTAD “potentially ICT-enabled services”**

Potentially ICT-enabled service categories	EBOPS 2010 class
Telecommunications (ICT services)	9.1 Telecommunications services (SI1)
Computer services (ICT services)	8.3 Licensing services for the right to use computer software (SH3) 9.2.1 Computer services – computer software (SI21) 9.2.2 Computer services – other computer services (SI22)
Sales and marketing services, not including trade and leasing services	10.2.2 Advertising, market research, and public opinion polling (SJ22)
Information services (note UNCTAD defines this more widely than the EBOPS category of the same name)	11.1.1 Audiovisual services (SK11) 11.1.2 Artistic relates services (SK12) 11.2.1 Health services (SK21) 11.2.3 Heritage and recreational services (SK23) 9.3.1 Information services – news agency services (SI31) 9.3.2 Information services – other information services (SI32)
Insurance and financial services	6.2 Reinsurance (SF2) 6.3 Auxiliary insurance services (SF3) 7.1 Financial services (SG) 7.2 Financial intermediation services indirectly measured (FISM) 6.1.1 Direct insurance – Life insurance (SF11) 6.1.2 Direct insurance – Freight insurance (SF12) 6.1.3 Direct insurance – Other direct insurance (SF13) 6.4.1 Pension services (SF41) 6.4.2 Standardised guarantee services (SF42) [= all of parent classes SF and SG]
Management, administration, and back office services	10.2.1.1 Legal services (SJ211) 10.2.1.2 Accounting, auditing, bookkeeping, and tax consulting services (SJ212) 10.2.1.3 Business and management consulting and public relations services (SJ213) 10.3.5 Other business services n.i.e (SJ35)
Licensing services	8.1 Franchises and trademarks licensing fees (SH1) 8.2 Licenses for the use of outcomes of research and development (SH2) 8.4.1 Licenses to reproduce and/or distribute audiovisual products (SH41) 8.4.2 Licenses to reproduce and/or distribute other products (SH42)
Engineering, related technical services and R&D	10.1.1.1 Provision of customised and non-customised R&D services (SJ111) 10.1.1.2.1 Patents (SJ1121) 10.1.1.2.2 Copyrights arising from research and development (SJ1122) 10.1.1.2.3 Industrial processes and design (SJ1123) 10.1.1.2.4 Other sales of proprietary rights arising from R&D (SJ1124) 10.1.2 Other research and development services (SJ12) [= all of SJ1] 10.3.1.1 Architectural services (SJ311) 10.3.1.2 Engineering services (SJ312) 10.3.1.3 Scientific and other technical services (SJ313) [=all of SJ31] 10.3.5 Other business service n.i.e (SJ35)
Education and training services	11.2.2 Education services (SK22)

Source: OECD based on UNCTAD (2015), “International trade in ICT services and ICT-enabled services”

**Figure 5.11. Exports of digitally deliverable services**

Purchasing countries as reported by exporting country, 2018



How to read this chart: The outer ring colour for each country is applied to that country's reported export flows. For example, the outer ring colour for the USA is red, and the large red mass of flows emanating from the USA segment represents exports reported by the US as going to other countries. Meanwhile, the flows of various colours going to the USA segment represent exports reported by other countries reported by them as going to the US.

Note: Countries marked with an asterisk (\*) are only available as a recipient reported by the exporting country, their own exports are unavailable.

In principle "ROW" refers to Rest of the World i.e. recipient countries not shown separately. **However**, some reporting countries do not provide data for all OECD+BRIICS recipient countries individually. Exports to countries not itemised individually will also be included in the ROW figure. ROW should therefore be interpreted as "ROW/unspecified".

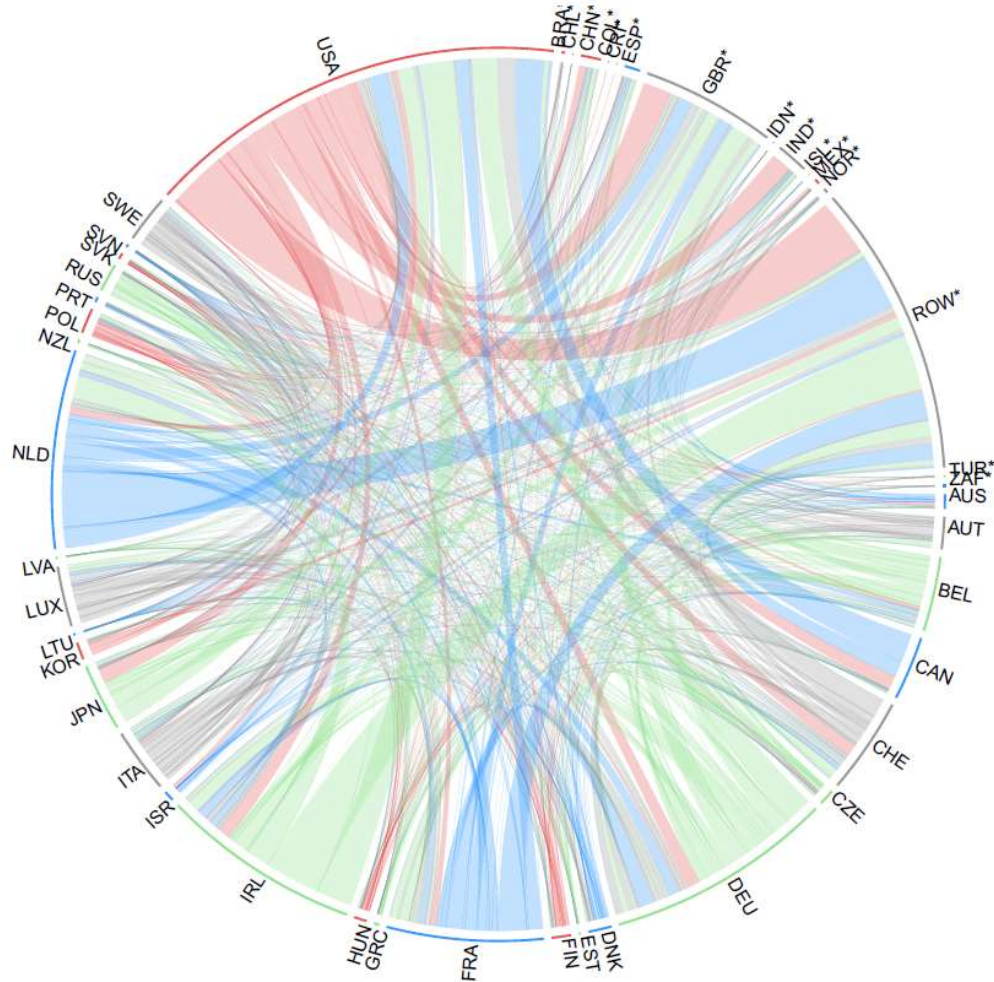
"Digitally deliverable services" refers to "potentially ICT-enabled services" per UNCTAD (2015).

Source: OECD, based on OECD EBOPS database, October 2020.



**Figure 5.12. Imports of digitally deliverable services**

Selling countries as reported by importing country, 2018



How to read this chart: The outer ring colour for each country is applied to that country's reported import flows. For example, the outer ring colour for the USA is red, and the large blue mass of flows terminating at the USA segment represents imports reported by the US as coming from various countries. Meanwhile, the flows of various colours going from the USA segment to other countries represent imports reported by those partner countries as coming from the US.

Note: Countries marked with an asterisk (\*) are only available as an origin reported by the importing country, their own imports are unavailable.

In principle "ROW" refers to Rest of the World i.e. supplier countries not shown separately. **However**, some reporting countries do not provide data for all OECD+BRIICS supplier countries individually. Imports from countries not itemised individually will also be included in the ROW figure. ROW should therefore be interpreted as "ROW/unspecified".

Source: OECD, based on OECD EBOPS database, October 2020.

These international trade flows of “digitally deliverable products”, also referred to as “potentially ICT-enabled services” are likely to be underpinned so a considerable degree by flows of data – and thus data flows are a key enabler of the economic value generated. However, there are alternative classifications and sources available. The following section looks at international trade flows of “digitisable products”. While these sound similar to the products in the foregoing analysis, importantly this selection is based on the Harmonised Commodity Description and Coding System (HS) which opens up the opportunity to compile estimates based on an alternative dataset.

#### 5.2.4. Trade in “digitisable products”

A recent UNCTAD research paper (Banga, 2019<sup>[55]</sup>) identifies a list of “Digitisable products” – those products which are “traded both in the physical form as well as online i.e. downloaded from the Internet e.g. music, e-books, software etc”. Banga goes on to explain that the “electronic transmissions” associated with these digital transmissions are increasingly important. “Digital technologies like 3D printing, robotics, and Big Data analytics depend heavily on electronic transmissions and can therefore lead to exponential growth in this trade”.

A list of 49 *digitisable products* were identified in the HS (combined HS codes). The list includes products previously identified in Telscher (2000<sup>[56]</sup>) and WTO (2018<sup>[57]</sup>). These fall under the categories of Photographic films, Cinematographic films, Printed matter, Music, Media, Software and Video games. A few new products which are increasingly being digitalized due to advancing technological developments were also added. These includes ‘other games’ like Playing cards; and Smart Cards. The full list is presented in **Annex Table 5.A.2.**

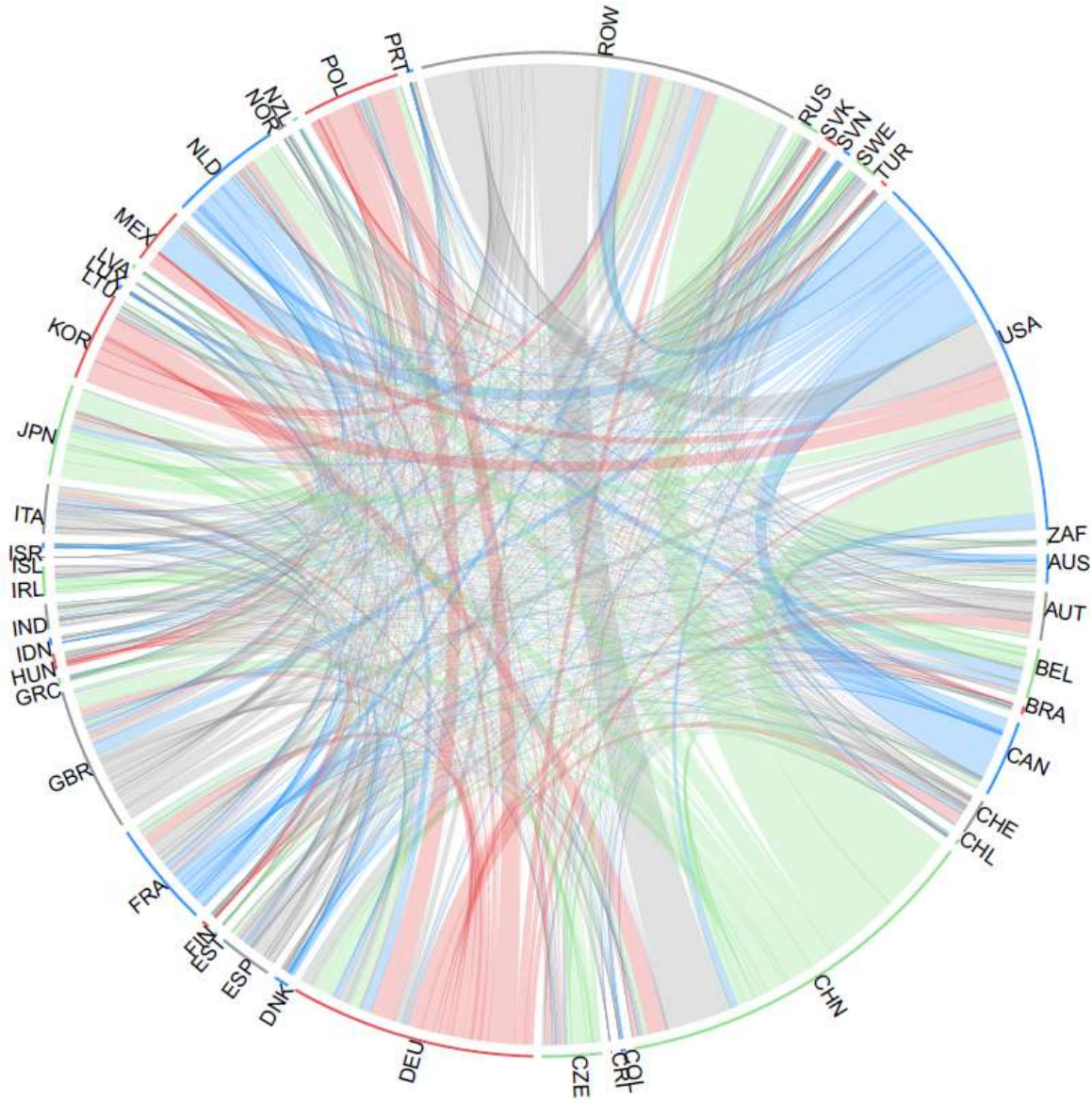
The World Customs Organisation Harmonised Commodity Description and Coding System (“HS”) is used by “more than 200 economies and Customs or Economic Unions [...] as a basis for their national Customs tariffs” (World Customs Organisation, n.d.<sup>[58]</sup>). There is a strong relationship between the CPC and the HS for transportable goods (where the CPC is almost always a grouping and rearrangement of complete categories of the HS) (United Nations, 1991<sup>[51]</sup>). However, the HS does not cover services products. As such, the list can be interpreted as a list of goods that can be digitised and delivered in digital form.

Data from the UN COMTRADE database (<https://comtrade.un.org>) are used to plot flows of these products in (exports) and Figure 5.14 (imports). A striking difference is the availability of data for China; by contrast China is only available as a declared partner in the EBOPS dataset used in other figures. **Figure 5.13** shows that China is a major exporter of digitisable products, and both the export figures reported by China and Imports reported by the United States point to the US as a key customer for Chinese production of these products.

Similarly, this data source also covers the United Kingdom fully instead of only as a trading partner reported by others. Again, this measure supports the view that the UK is one of the main hubs for international trade in products related to data flows.

**Figure 5.13. Exports of digitisable products**

Purchasing countries as reported by exporting country, 2019



How to read this chart: The outer ring colour for each country is applied to that country's reported *export flows*. For example, the outer ring colour for the USA is blue, and the large blue mass of flows emanating from the USA segment represents exports reported by the US as going to other countries. Meanwhile, the flows of various colours going to the USA segment represent exports reported by other countries reported by them as going to the US.

Note: "ROW" refers to Rest of the World.

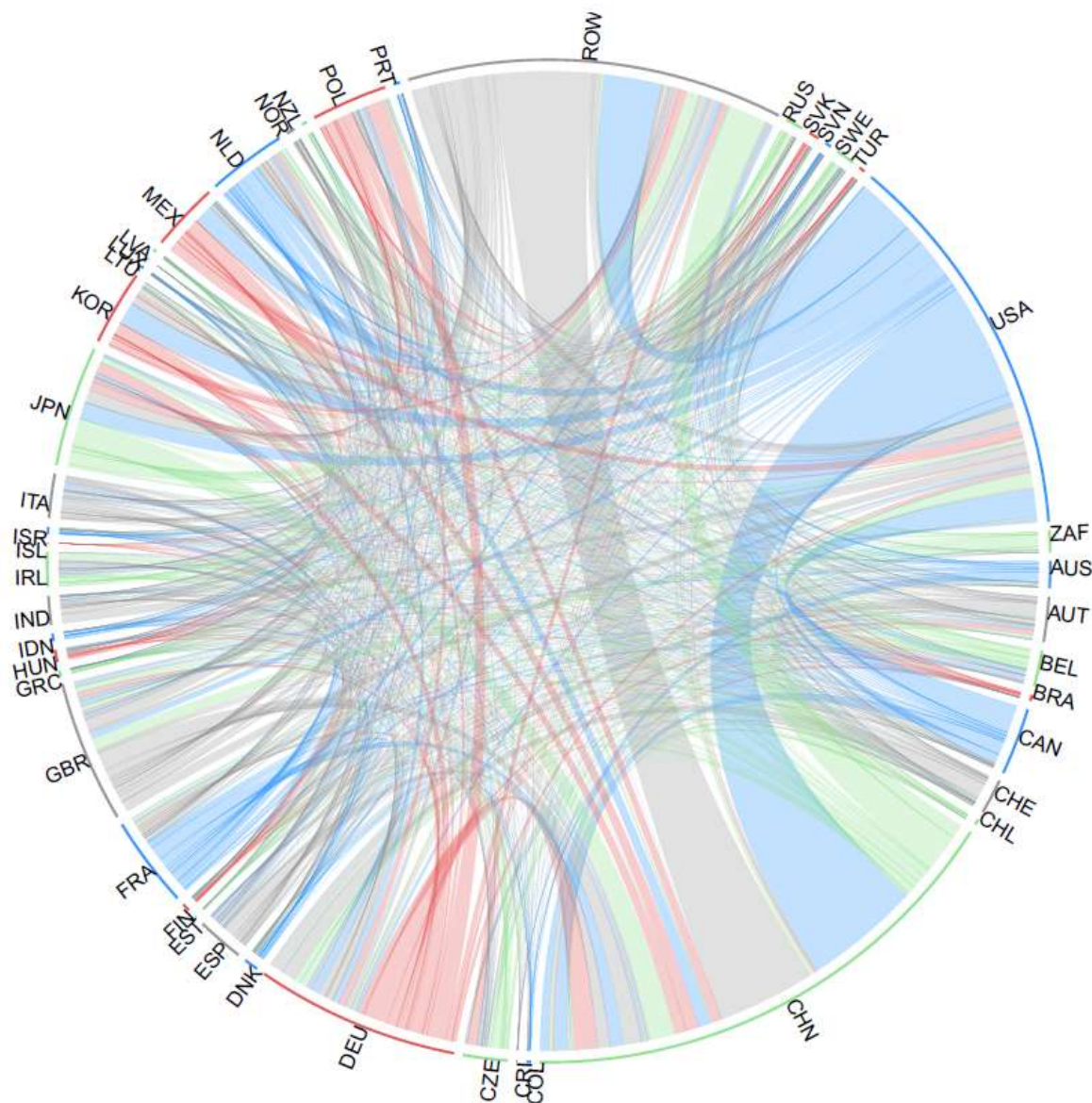
"Digitisable services" are identified in accordance with Banga (2019).

Source: OECD based on UN COMTRADE database, October 2020.



**Figure 5.14. Imports of digitisable products**

Selling countries as reported by importing country, 2015



How to read this chart: The outer ring colour for each country is applied to that country's reported import flows. For example, the outer ring colour for the USA is blue, and the large blue mass of flows terminating at the USA segment represents imports reported by the US as coming from various countries. Meanwhile, the flows of various colours going from the USA segment to other countries represent imports reported by those partner countries as coming from the US.

Note: Countries marked with an asterisk (\*) are only available as an origin reported by the importing country, their own imports are unavailable.

In principle "ROW" refers to Rest of the World.

"Digitisable services" are identified in accordance with Banga (2019).

Source: OECD based on UN COMTRADE database, October 2020.



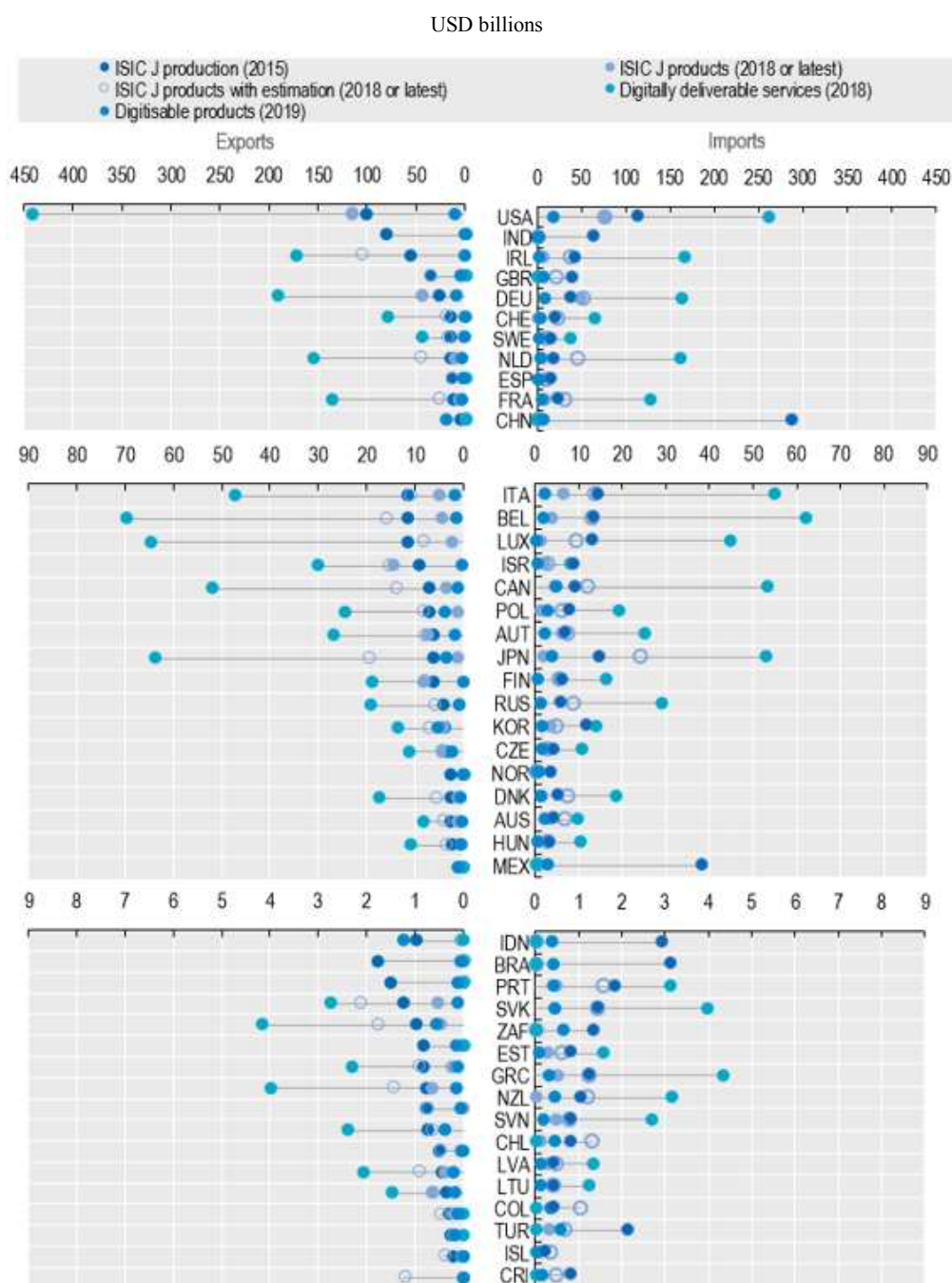
### **5.2.5. Overall findings**

Section 5.2 has examined several ways of identifying products that are likely to be digitised and therefore to be delivered in the form of data flows. The international trade flows of these products can therefore be considered as offering a perspective on the economic value generated through or associated with the underlying international data flows.

**Figure 5.15** brings together the various export and import figures for the product sets identified. There is wide variation between the different product scopes with flows “Digitally deliverable services” in particular being several times greater than exports by firms in ISIC industry section J (Information and Communication) or the trade in products associated with firms in ISIC section J - which also includes any production of those products by firms classified in other industries.

There is also wide variation in the magnitude of these trade flows across OECD and BRIICS countries. In all cases the United States appears to be a major hub for both exports and imports of the identified products. These measures also suggest that countries including India, Ireland, the United Kingdom, Germany, the Netherlands, Switzerland, and France are also especially active in trade underpinned by data flows and, as such, data flows appear likely to generate significant economic value for these economies.

Meanwhile, all OECD and BRIICS countries appear to be party to at least some data-related trade with many of the associated flows occurring with the key countries identified above.

**Figure 5.15. International trade underpinned by data flows**

Note: “ISIC J production” refers to trade in products produced by firms classified in ISIC section J. “ISIC J products” refers to trade in the products mainly associated with firms classified in ISIC section J but including production by firms classified in other sectors. “Digitally deliverable services” refers to “potentially ICT-enabled products” per UNCTAD (2015). “Digitisable products” are identified within the WTO HS commodity classification per Banga (2019).

Source: OECD based on OECD EBOPS database, OECD Input-Output database, and UN COMTRADE database, October 2020.

### 5.3. Concluding remarks

This chapter has sought perspectives on the volume and value associated with international data flows. In particular, the chapter focussed on products delivered through data flows - as opposed to considering all transactions facilitated through data flows (all e-commerce transactions rely on some exchange of digital information to establish the order, where it should be delivered, etc. but online sales of physical goods are not within the scope of this analysis as the data flow is not an integral part of the product being delivered).

At the firm-level, there are industry estimates, the quality of which is difficult to assess) available on the volume of data flows (in terms of bits and bytes) generated by some major technology firms. However, finding a meaningful way of relating these data flows to revenue and economic value generation is challenging even for a firm that is relatively homogeneous in terms of the product it produces such as Netflix. However, the information required in mandatory filings do not provide a strong basis for breaking down the Internet traffic generated by Netflix across countries.

For more complex companies such as Alphabet (Google), Facebook, and Microsoft, the challenge is even greater as they comprise multiple business units producing varied products that differ in the extent to which they are likely to be driving the company's Internet traffic and the revenue/economic value they generate for the company. Furthermore, these companies do not generally appear to go beyond mandatory filing requirements by reporting the revenues of these business units individually. These high profile companies are, due to their size, likely to attract more attention and investigation and so it is likely that even less information is available for the many smaller firms also delivering products through international data flows.

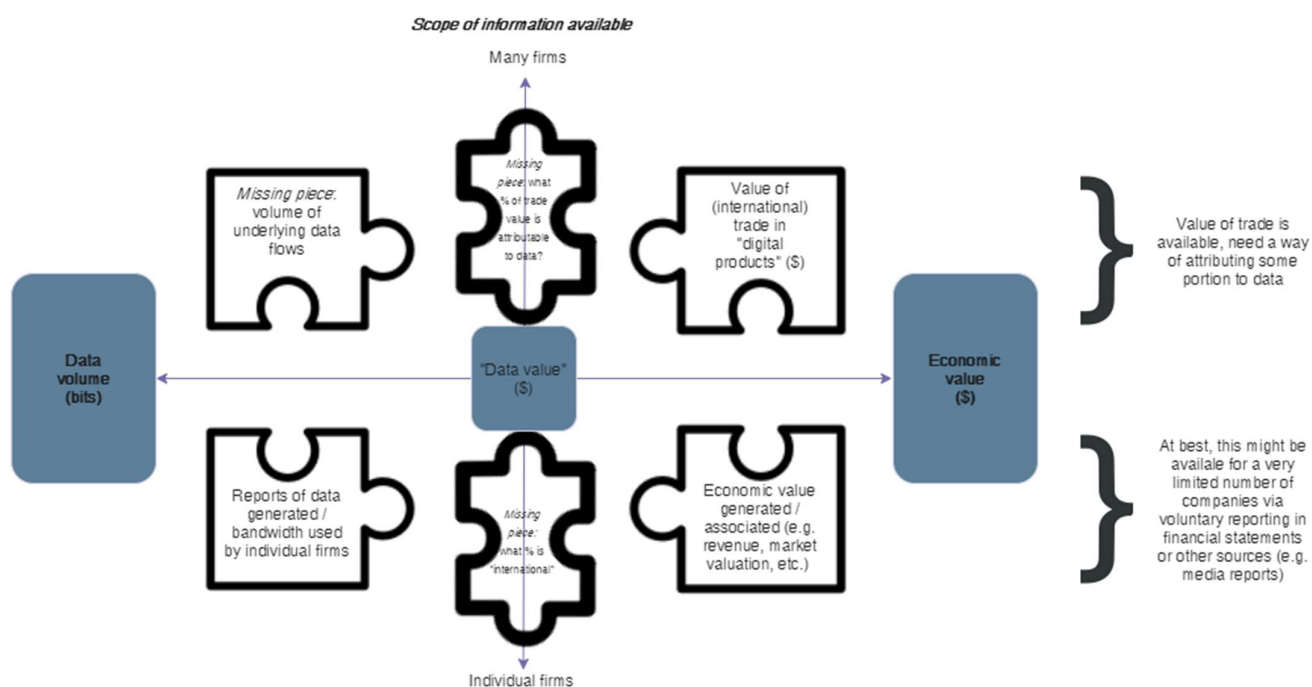
Overall, while there is some information on data flow volumes and economic values associated to some degree with them there is a “missing piece” in terms of the detailed information needed to relate one to the other and furthermore distinguish international (i.e. trade-related” elements (**Figure 5.16**), as well as additional issues with trying to scale to cover others than the most major companies. As such, this bottom-up approach does not appear ripe for further work at this time.

As an alternative, the chapter also took a “top-down approach” starting with measures of the economic value of trade in products that are likely to be delivered through data flows. While there is trade data available this approach relies on identifying which products should be counted and different inclusion criteria generate widely varying estimates. Nevertheless, it appears that the United States is a major hub for international trade in products delivered through data flows, while countries such as France, Germany, Ireland, the Netherlands, and the United Kingdom also appear likely to feature heavily in trade underpinned by data.

However, as Figure 5.16 shows we are also missing the pieces needed to make a link from the economic value of these trade flows to actual data flows. Indeed, the volume of data flows of Internet traffic related to trade in any way is unknown and difficult even to conceptualise. There are many reasons for this including simply the way the Internet is structured and works. The Internet is structured to minimise the distance that data need to travel to reach the end recipient. This is particularly the case for services that generate a lot of traffic. While a Netflix subscription in France may create a trade flow of audiovisual services from the United States to France (captured in trade statistics), it often will not, in fact, create any international Internet traffic flow as Netflix places content servers within regions and countries worldwide in order to optimise their service performance and avoid

placing too much burden on Internet infrastructure providers by minimising the physical distance that the content has to travel from Netflix's servers to end users. Furthermore, Netflix and others go as far as installing content servers *within* Internet Service Providers' networks. This means that in some cases the international trade results in no Internet data flow at all (the content being served entirely inside the ISPs network). Factors such as this mean that even when a product is entirely delivered in the form of data there need not be a correlation between the economic trade value of the product and the volume or even existence of Internet/international data flows.

Figure 5.16. "Linking" data on international data flows to international trade



Source: OECD.

There are, then, considerable conceptual and practical challenges. However, one way to build upon this analysis could be to try to identify the average "data flow intensity" of the digitally deliverable products measured above. This would be similar, in principle, to the way that Internet and mobile service providers use estimates averages to articulate data allowances in terms of what activities they allow "e.g. statements such as "1Gb of data will allow you to access 600 web pages, read 2 000 rich emails, stream 2 hours of SD video or download 25 apps". By identifying the main vehicles through which the product is digitally delivered (e.g. streaming video/audio, PDF documents, etc.) one could try to work out roughly how much data flow each unit of the product creates on average. However, this approach would rely on assumptions and require detailed information about what types of data flow different products are likely to entail.

Alternatively, one could accept that, due to the highly context-dependent nature of both the volume of data flows and associated economic value, it is not possible to meaningfully link data volumes and values in a "one-size-fits-all" way and instead to abstract from the problem by looking at the value of international trade in products that are digitisable and

ever more likely to be being delivered through data flows. Although the volume of the data flow remains unknown, this approach provides one perspective on the economic value arising from data flows.

## References

- Alphabet (2020), *FORM 10-Q, Q2*, [30]  
[https://abc.xyz/investor/static/pdf/20200731\\_alphabet\\_10Q.pdf?cache=f16f989](https://abc.xyz/investor/static/pdf/20200731_alphabet_10Q.pdf?cache=f16f989).
- AWS (2020), *AWS Global Infrastructure*, <https://aws.amazon.com/about-aws/global-infrastructure/>. [43]
- Banga, R. (2019), *Growing Trade in Electronic Transmissions: Implications for the South*, [55]  
[https://unctad.org/system/files/official-document/ser-rp-2019d1\\_en.pdf](https://unctad.org/system/files/official-document/ser-rp-2019d1_en.pdf).
- Casalini, F. and J. Lopez Gonzalez (2019), “Trade and Cross-Border Data Flows”, *OECD Trade Policy Papers*, [40]  
<https://doi.org/10.1787/b2023a47-en>.
- Cisco (2018), *Cisco Visual Networking Index (VNI) Complete Forecast Update 2017-2022*, [42]  
[https://www.cisco.com/c/dam/m/en\\_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1211\\_BUSINESS\\_SERVICES\\_CKN\\_PDF.pdf](https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1211_BUSINESS_SERVICES_CKN_PDF.pdf).
- Davenport, T. and D. Patil (2012), “Data Scientist: The Sexiest Job of the 21st Century”, *Harvard Business Review*, [37]  
<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>.
- Eurostat (2019), *Annual detailed enterprise statistics for services*, [28]  
[https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=sbs\\_na\\_1a\\_se\\_r2&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=sbs_na_1a_se_r2&lang=en) (accessed on October 2019).
- Eurostat (2015), *CPA Ver. 2.1 Structure and explanatory notes*, [9]  
<http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/classifications/other-classifications-used-within-national-statistics/other-national-and-international-classifications/cpa-ver--2-1-structure-and-explanatory-n>.
- Eurostat (2008), *NACE Rev. 2 Statistical classification of economic activities in the European Community*, [27]  
 Office for Official Publications of the European Communities,  
<https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF>.
- Eurostat-OECD Task Force on Land and Other non-financial assets (2020), *Report on Intellectual Property Products*, [6]  
<https://ec.europa.eu/eurostat/documents/24987/725066/Eurostat-OECD+Report+on+Intellectual+Property+Products.pdf>.
- Gebel, M. (2020), *Yes, Netflix still mails DVDs*, [60]  
<https://www.businessinsider.com/us/does-netflix-still-mail-dvds#:~:text=Yes%2C%20Netflix%20still%20mails%20DVDs,movies%20with%20no%20late%20fee&text=Yes%2C%20Netflix%20will%20still%20mail,for%20a%20DVD%20subscription%20plan>.
- Goodfirms (2020), *Top Big Data Analytics Companies*, [36]  
<https://www.goodfirms.co/big-data-analytics> (accessed on October 2020).
- Konrad, A. et al. (2020), *The Cloud 100*, <https://www.forbes.com/cloud100/#326b6dcb5f94> (accessed on October 2020). [32]
- Lopez Gonzalez, J. (2019), *DON'T PANIC! The hitchhiker's guide to cross-border data flows*, [41]  
<https://www.oecd.org/trade/hitchhikers-guide-cross-border-data-flows/>.

- Netflix (2020), *Form 10-k*, [https://s22.q4cdn.com/959853165/files/doc\\_financials/2019/ar/2019-10-K.pdf](https://s22.q4cdn.com/959853165/files/doc_financials/2019/ar/2019-10-K.pdf). [49]
- Netflix (2019), *Netflix Open Connect Overview*, <https://openconnect.netflix.com/Open-Connect-Overview.pdf>. [45]
- Netflix (2016), *Completing the Netflix Cloud Migration*, <https://about.netflix.com/en/news/completing-the-netflix-cloud-migration>. [44]
- Nguyen, D. and M. Paczos (2020), *Measuring the Economic Value of Data and Data Flows*, <https://doi.org/10.1787/6345995e-en>. [3]
- OECD (n.d.), *Computer data storage units*, <https://oec.world/en/profile/hs92/847193/> (accessed on October 2019). [16]
- OECD (2019), *Going Digital: Shaping Policies, Improving Lives*, <https://doi.org/10.1787/9789264312012-en>. [1]
- OECD (2019), *Measuring the Digital Transformation: A Roadmap for the Future*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264311992-en>. [7]
- OECD (2010), *Extended Balance of Payments Services Classification (EBOPS 2010)*, <https://www.oecd.org/sdd/its/EBOPS-2010.pdf>. [52]
- OECD (2009), *Handbook on Deriving Capital Measures of Intellectual Property Products*, OECD Publishing, Paris, <https://dx.doi.org/10.1787/9789264079205-en>. [5]
- OECD-WTO-IMF (2019), *Handbook on Measuring Digital Trade, Version 1*, <https://www.oecd.org/sdd/its/handbook-on-measuring-digital-trade.htm>. [50]
- Patrizio, A. (2019), *Top 30 Big Data Companies*, <https://www.datamation.com/big-data/big-data-companies.html> (accessed on October 2020). [34]
- Reed, D. (2007), “Database valuation: Putting a price on your prime asset”, *Journal of Database Marketing & Customer Strategy Management* volume, pp. 104-109, <https://doi.org/10.1057/palgrave.dbm.3250048>. [29]
- Sandvine (2019), *Global Internet Phenomena Report*, [https://www.sandvine.com/hubfs/Sandvine\\_Redesign\\_2019/Downloads/Internet%20Phenomena/Internet%20Phenomena%20Report%20Q32019%2020190910.pdf](https://www.sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/Internet%20Phenomena/Internet%20Phenomena%20Report%20Q32019%2020190910.pdf). [48]
- Schroer, A. (2019), *34 Big Data companies helping us make sense of the world*, <https://builtin.com/big-data/big-data-companies-roundup> (accessed on October 2020). [35]
- Sisson, P. (2020), *The perfect virtual video store isn't Netflix. It's DVD.com.*, <https://www.vox.com/culture/2020/4/23/21230324/netflix-dvd-rental-classic-movies>. [59]
- Statistics Canada (2019), *Concordance between the North American Industry Classification System (NAICS) Canada 2017 Version 3.0 and the International Standard Industrial Classification (ISIC), Rev. 4*, <https://www.statcan.gc.ca/eng/subjects/standard/naics/2017/v3/concordance-rev4> (accessed on October 2019). [23]
- Statistics Canada (2019), *Table 36-10-0478-01 Supply and use tables, detail level, provincial and territorial (x 1,000)*, <https://doi.org/10.25318/3610047801-eng> (accessed on October 2019). [26]

- Statistics Canada (2019), *Table 21-10-0210-01 Software development and computer services, breakdown of sales*, <https://doi.org/10.25318/2110021001-eng> (accessed on October 2019). [25]
- Statistics Canada (2019), *Table 22-10-0087-01 Software development and computer services, summary statistics (x 1,000,000)*, <https://doi.org/10.25318/2210008701> (accessed on October 2019). [24]
- Statistics Canada (2018), *North American Product Classification System (NAPCS) Canada 2017 Version 2.0*, <https://www.statcan.gc.ca/eng/subjects/standard/napcs/2017v2/index>. [11]
- Telscher, S. (2000), *Tariffs, Taxes, and Electronic Commerce: Revenue Implications for Developing Countries*, [http://UNCTAD\(2000\)Tariffs.TaxesandElectronicCommerce:RevenueImplicationsforDevelopingCountries\(Geneva:UnitedNationsConferenceonTradeandDevelopment\)](http://UNCTAD(2000)Tariffs.TaxesandElectronicCommerce:RevenueImplicationsforDevelopingCountries(Geneva:UnitedNationsConferenceonTradeandDevelopment)). [56]
- The Economist (2020), *Are data more like oil or sunlight?*, <https://www.economist.com/special-report/2020/02/20/are-data-more-like-oil-or-sunlight>. [2]
- The Economist (2020), *Who owns the web's data?*, <https://www.economist.com/business/2020/10/22/who-owns-the-webs-data>. [31]
- The Manifest (2020), *Top 100 Big Data Companies*, <https://themanifest.com/big-data/companies> (accessed on October 2020). [33]
- UD Census Bureau (2020), *E-Stats 2018: Measuring the Electronic Economy*, <https://www.census.gov/library/publications/2020/econ/2018-e-stats.html>. [46]
- UN (2010), *System of National Accounts 2008*, United Nations, New York, <https://dx.doi.org/10.18356/4fa11624-en>. [4]
- UNCTAD (2015), *International TRade in ICT Services and ICT-Enabled Services*, UNCTAD Technical Notes on ICT for Development, [https://unctad.org/en/PublicationsLibrary/tn\\_unctad\\_ict4d03\\_en.pdf](https://unctad.org/en/PublicationsLibrary/tn_unctad_ict4d03_en.pdf). [54]
- United Nations (2015), *Central Product Classification (CPC) Version 2.1*, <https://unstats.un.org/unsd/classifications/unsdclassifications/cpcv21.pdf>. [8]
- United Nations (2008), “ISIC Rev.4.”, in *International Standard Industrial Classification of All Economic Activities.*, United Nations, [https://unstats.un.org/unsd/publication/seriesm/seriesm\\_4rev4e.pdf](https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf). [19]
- United Nations (1991), *Provisional Central Product Classification*, <https://www.tralac.org/files/2013/12/UN-CPC-Provisional-central-product-classification.pdf>. [51]
- UNSD (2010), *Correspondence between the EBOPS 2010 and the Central Product*, [https://unstats.un.org/unsd/tradeserv/tfsits/msits2010/ebops2cpc\\_detailed.htm#ebops3](https://unstats.un.org/unsd/tradeserv/tfsits/msits2010/ebops2cpc_detailed.htm#ebops3). [53]
- US Bureau of Economic Analysis (n.d.), *Input-Output Accounts Data*, <https://www.bea.gov/industry/input-output-accounts-data> (accessed on October 2019). [13]
- US Census Bureau (2020), *Quarterly retail e-commerce sales 2nd quarter 2020*, [https://www.census.gov/retail/mrts/www/data/pdf/ec\\_current.pdf](https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf). [47]
- US Census Bureau (2019), *2017 NAPCS-Based Collection Code to 2012 Product Code*, <https://www.census.gov/programs-surveys/economic-census/guidance/understanding-napcs.html>. [12]



- US Census Bureau (2018), *Annual Survey of Manufactures (ASM) Tables*, [17]  
<https://www.census.gov/data/tables/2016/econ/asm/2016-asm.html> (accessed on October 2019).
- US Census Bureau (2018), *Service Annual Survey Tables*, [14]  
<https://www.census.gov/programs-surveys/sas/data/tables.html>.
- US Census Bureau (2017), *Economic Census Tables 2017*, [22]  
<https://www.census.gov/programs-surveys/economic-census/data/tables.2017.html> (accessed on October 2019).
- US Census Bureau (2017), *ISIC Rev.4 to 2017 NAICS*, [20]  
<https://www.census.gov/eos/www/naics/concordances/concordances.html>.
- US Census Bureau (2012), *Economic Census Tables 2012*, [15]  
<https://www.census.gov/programs-surveys/economic-census/data/tables.2012.html> (accessed on October 2019).
- US Census Bureau (n.d.), *North American Product Classification System*, [10]  
<https://www.census.gov/eos/www/napcs/more.html>.
- US Census Bureau (n.d.), *USA Trade Online*, [18]  
<https://usatrade.census.gov/> (accessed on October 2019).
- US. Office of Management and Budget (2017), *North American Industry Classification System*, Executive office of the President, [21]  
[https://www.census.gov/eos/www/naics/2017NAICS/2017\\_NAICS\\_Manual.pdf](https://www.census.gov/eos/www/naics/2017NAICS/2017_NAICS_Manual.pdf).
- van Rijmenam, M. (2013), *A Short History Of Big Data*, [39]  
<https://datafloq.com/read/big-data-history/239#:~:text=90%25%20of%20the%20available%20data,has%20been%20around%20much%20longer.>
- Wiegend, A. (2009), “The Social Data Revolution(s)”, *Harvard Business Review*, [38]  
<https://hbr.org/2009/05/the-social-data-revolution.html>.
- World Customs Organisation (n.d.), *HS Convention*, [58]  
[http://www.wcoomd.org/en/topics/nomenclature/instrument-and-tools/hs\\_convention.aspx](http://www.wcoomd.org/en/topics/nomenclature/instrument-and-tools/hs_convention.aspx).
- WTO (2018), *How Cyberspace “intrudes” on the physical space?*, [57]  
[https://docs.wto.org/dol2fe/Pages/FE\\_Search/FE\\_S\\_S009-DP.aspx?language=E&CatalogueIdList=243191,243201,237305,237306,234444,230236,230146,230135,230094,229772&CurrentCatalogueIdIndex=1&FullTextHash=&HasEnglishRecord=True&HasFrenchRecord=True&HasSpanishRec.](https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S009-DP.aspx?language=E&CatalogueIdList=243191,243201,237305,237306,234444,230236,230146,230135,230094,229772&CurrentCatalogueIdIndex=1&FullTextHash=&HasEnglishRecord=True&HasFrenchRecord=True&HasSpanishRec.)

## Annex 5.A. Industry to product classification concordance

Annex Table 5.A.1. ISIC rev.4 to CPC 2.1 concordance

ISIC 4		CPC 2.1		EBOPS 2010	
5811	Book publishing	32210	Educational textbooks, in print		
		32220	General reference books, in print		
		32291	Professional, technical and scholarly books, in print		
		32292	Children's books, in print		
		32299	Other books n.e.c., in print		
		32511	Maps and hydrographic or similar charts (including wall maps, topographical plans and maps for globes), printed, other than in book-form		
		47691	Audio books on disk, tape or other physical media		
		47692	Text-based disks, tapes or other physical media		
		73320	Licensing services for the right to use entertainment, literary or artistic originals	8.4.1/SH41	Licenses to reproduce and/or distribute audiovisual products Licences to reproduce and/or distribute audio-visual products covers fees and charges for the authorized reproduction and/or distribution, through licensing agreements, of produced audio-visual originals or prototypes (for example, cinematographic works and sound recordings). Also included are rights relating to the reproduction and/or distribution of recordings of live performances and radio, television, cable and satellite broadcast. Retransmission rights for sports events are also covered. Further information on audio-visual products is provided in the complementary grouping audio-visual transactions. (MSITS 2010, para. 3.220)
		83639	Sale of other advertising space or time (except on commission)	10.2.2/SJ22	Advertising; market research; and public opinion polling Advertising, market research, and public opinion polling services transacted between residents and non-residents, includes the design, creation and marketing of advertisements by advertising agencies; media placement, including the purchase and sale of advertising space; exhibition services provided by trade fairs; the promotion of products abroad; market research; telemarketing; and public opinion polling on various issues. (MSITS 2010, para. 3.242)

			<b>84311</b>	On-line books	<b>11.1.2/SK12</b>	Artistic related services Artistic related services includes the services provided by performing artists (actors, musicians, dancers, etc.), authors, composers and sculptors. It also includes services provided by independent models as well as set, costume and lighting designers. Transactions are included if the service providers are not employees of the entity making payments (otherwise, they represent compensation of employees). Also included are presentation and promotion services for performing arts and other live entertainment events. However, the recording of such events is included in audio-visual services. (MSITS 2010, para. 3.260)
			<b>89110</b>	Publishing, on a fee or contract basis	<b>10.3.5/SJ35</b>	Other business services n.i.e. includes distribution services for water, steam, gas and petroleum products as well as distribution services for electricity, where these are identifiable separately from transmission services (transmission of these products is recorded under transport); air conditioning supply; placement of personnel (the services provided by these personnel are recorded under the relevant services item); security and investigative services; translation and interpretation; photographic services; building cleaning; real estate services to businesses; and any other business services that cannot be classified to any of the business services listed above. (MSITS 2010, para. 3.252)
5812	Publishing directories and mailing lists		32230	Directories, in print		
			47692	Text-based disks, tapes or other physical media		

5813	Publishing newspapers, journals, periodicals	of and	73312	Licensing services for the right to use databases	8.3/SH3	Licenses to reproduce and/or distribute computer software Licences to reproduce and/or distribute computer software, which covers the charges for the authorized reproduction and/or distribution (through licensing agreements) of produced software originals. Distribution, in this sense, is not defined as wholesale and retail sale. The sale of a copy of a computer software package that is purchased by an individual or enterprise for individual use is not covered by a licence to distribute. The sale of the software for individual or personal use is recorded in computer services. Software sold on physical media with the right to perpetual use is included under goods. (MSITS 2010, para. 3.220)
			83631	Sale of advertising space in print media (except on commission)	10.2.2/SJ22	Advertising; market research; and public opinion polling (Definition above)
			83639	Sale of other advertising space or time (except on commission)		
			83940	Original compilations of facts/information	9.3.2/SI32	Information services - Other information services Other information services includes database services, such as database conception, data storage, and the dissemination of data and databases (including directories and mailing lists), both online and through magnetic, optical or printed media and web search portals (encompassing search engine services that find Internet addresses for clients who input keyword queries). Also included are: direct non-bulk subscriptions to newspapers and periodicals, whether by mail, electronic transmission or other means; other online content provision services; and library and archive services. (Bulk newspapers and periodicals are included under general merchandise.) (MSITS 2010, para. 3.232)
			84313	On-line directories and mailing lists		
			89110	Publishing, on a fee or contract basis	10.3.5/SJ35	Other business services n.i.e. (Description above)
			32300	Newspapers and periodicals, daily, in print		
			32410	General interest newspapers and periodicals, other than daily, in print		
			32420	Business, professional or academic newspapers and periodicals, other than daily, in print		
			32490	Other newspapers and periodicals, other than daily, in print		
			47692	Text-based disks, tapes or other physical media		

5819		<b>73320</b>	Licensing services for the right to use entertainment, literary or artistic originals	<b>8.4.1/SH41</b>	<b>Licenses to reproduce and/or distribute audiovisual products (Definition above)</b>
		<b>83631</b>	Sale of advertising space in print media (except on commission)		
		<b>83633</b>	Sale of Internet advertising space (except on commission)		
		<b>84312</b>	On-line newspapers and periodicals	<b>9.3.2/SI32</b>	Information services - Other information services (Definition above)
		<b>89110</b>	Publishing, on a fee or contract basis	<b>10.3.5/SJ35</b>	Other business services n.i.e. (Description above)
	Other publishing activities	32530	Printed or illustrated postcards; printed cards bearing personal greetings or messages, with or without envelopes or trimmings		
		32540	Printed pictures, designs and photographs		
		32620	Trade advertising material, commercial catalogues and the like		
		32630	Transfers (decalcomanias) and printed calendars		
		<b>83633</b>	Sale of Internet advertising space (except on commission)	<b>10.2.2/SJ22</b>	Advertising; market research; and public opinion polling (Definition above)
		<b>84393</b>	On-line adult content	<b>9.3.2/SI32</b>	Information services - Other information services (Definition above)
		<b>84399</b>	Other on-line content n.e.c.		
		<b>89110</b>	Publishing, on a fee or contract basis	<b>10.3.5/SJ35</b>	Other business services n.i.e. (Description above)
5820	Software publishing	38582	Software cartridges for video game consoles		
		47811	Operating systems, packaged		
		47812	Network software, packaged		
		47813	Database management software, packaged		
		47814	Development tools and programming languages software, packaged		
		47821	General business productivity and home use applications, packaged		
		47822	Computer game software, packaged		
		47829	Other application software, packaged		
		<b>73311</b>	Licensing services for the right to use computer software	<b>8.3/SH3</b>	Licenses to reproduce and/or distribute computer software
		<b>83143</b>	Software originals	<b>9.2.1/SI21</b>	
		<b>84341</b>	System software downloads		
		<b>84342</b>	Application software downloads		
		<b>84391</b>	On-line games		

		<b>84392</b>	On-line software		<b>Computer services - Computer software</b> Computer software includes: - Sales of customized software (however delivered) and related licences to use - Development, production, supply and documentation of customized software, including operating systems, made to order for specific users - Non-customized (mass-produced) software downloaded or otherwise electronically delivered, whether with a periodic licence fee or a single payment - Licences to use non-customized (mass-produced) software provided on a storage device such as a disk or CD-ROM with a periodic licence fee - Sales and purchases of originals and ownership rights for software systems and applications (MSITS 2010, para. 3.225)
		<b>89110</b>	Publishing, on a fee or contract basis	<b>10.3.5/SJ35</b>	Other business services n.i.e. (Description above)
5911	Motion picture, video and television programme production activities	38950	Motion picture film, exposed and developed, whether or not incorporating sound track or consisting only of sound track		
		47620	Films and other video content on disks, tape or other physical media		
		<b>73320</b>	Licensing services for the right to use entertainment, literary or artistic originals	<b>8.4.1/SH41</b>	<b>Audiovisual services</b> Audio-visual services relates to the production of motion pictures (on film, videotape, or disk or transmitted electronically), radio and television programmes (live or on tape) and musical recordings. Performing arts and other live entertainment event presentation and promotion services (namely, live performances such as concerts and plays) are excluded from this item and included in artistic related services. However, the recording of live performances is included in audio-visual services and for these recordings, the same treatment as for other audio-visual products applies. Included in audio-visual services are amounts receivable or payable for rentals of audio-visual and related products and charges for access to encrypted television channels (such as those offering cable and satellite services). (MSITS 2010, para. 3.256)
		<b>84331</b>	Films and other video downloads	<b>11.1.1/SK11</b>	
		<b>84332</b>	Streamed video content		
		<b>96121</b>	Motion picture, videotape and television programme production services		
		<b>96123</b>	Motion picture, videotape, television and radio programme originals		
5912	Motion picture, video and television programme post-production activities	<b>73320</b>	Licensing services for the right to use entertainment, literary or artistic originals	<b>8.4.1/SH41</b>	<b>Licenses to reproduce and/or distribute audiovisual products (Definition above)</b>
		<b>96131</b>	Audiovisual editing services	<b>11.1.1/SK11</b>	Audiovisual services (Definition above)
		<b>96132</b>	Transfers and duplication of masters services		

		<b>96133</b>	Colour correction and digital restoration services		
		<b>96134</b>	Visual effects services		
		<b>96135</b>	Animation services		
		<b>96136</b>	Captioning, titling and subtitling services		
		<b>96139</b>	Other post-production services		
5913	Motion picture, video and television programme distribution activities	<b>73320</b>	Licensing services for the right to use entertainment, literary or artistic originals	<b>8.4.1/SH41</b>	<b>Licenses to reproduce and/or distribute audiovisual products (Definition above)</b>
		<b>96140</b>	Motion picture, videotape and television programme distribution services	<b>11.1.1/SK11</b>	Audiovisual services (Definition above)
5914	Motion picture projection activities	96150	Motion picture projection services		
5920	Sound recording and music publishing activities	32520	Music, printed or in manuscript		
		47610	Musical audio disks, tapes or other physical media		
		47699	Other non-musical audio disks and tapes		
		<b>73320</b>	Licensing services for the right to use entertainment, literary or artistic original	<b>8.4.1/SH41</b>	<b>Licenses to reproduce and/or distribute audiovisual products (Definition above)</b>
		<b>84321</b>	Musical audio downloads	<b>11.1.1/SK11</b>	Audiovisual services (Definition above)
		<b>84322</b>	Streamed audio content		
		<b>89110</b>	Publishing, on a fee or contract basis	<b>10.3.5/SJ35</b>	Other business services n.i.e. (Description above)
		<b>96111</b>	Sound recording services, except live recording services	<b>11.1.1/SK11</b>	Audiovisual services (Definition above)
		<b>96112</b>	Live recording services		
		<b>96113</b>	Sound recording originals		
		<b>96122</b>	Radio programme production services		
		<b>96123</b>	Motion picture, videotape, television and radio programme originals		
		<b>96137</b>	Sound editing and design services		
6010	Radio broadcasting	<b>83632</b>	Sale of TV/radio advertising time (except on commission)		
		<b>84611</b>	Radio broadcast originals		
		<b>84621</b>	Radio channel programmes		
		<b>84631</b>	Broadcasting services		
		<b>96122</b>	Radio programme production services		
6020	Television programming and broadcasting services	<b>83632</b>	Sale of TV/radio advertising time (except on commission)	<b>10.2.2/SJ22</b>	Advertising; market research; and public opinion polling
		<b>84612</b>	Television broadcast originals	<b>9.1/SI1</b>	
		<b>84622</b>	Television channel programmes		

		<b>84631</b>	Broadcasting services		Telecommunications services Telecommunications services covers the broadcast or transmission of sound, images, data, or other information by telephone, telex, telegram, radio and television cable transmission, radio and television satellite, electronic mail, facsimile, etc., and includes business network services, teleconferencing and support services. It does not include the value of the information transported. Also included are mobile telecommunications services, Internet backbone services and online access services, including the provision of access to the Internet. Excluded are installation services for telephone network equipment (included in construction), and database services (included in information services). (MSITS 2010, para. 3.223)
		<b>96121</b>	Motion picture, videotape and television programme production services	<b>11.1.1/SK11</b>	Audiovisual services (Definition above)
6110	Wired telecommunications activities	<b>84110</b>	Carrier services	<b>9.1/SI1</b>	Telecommunications services (Definition above)
		<b>84120</b>	Fixed telephony services		
		<b>84140</b>	Private network services		
		<b>84150</b>	Data transmission services		
		<b>84190</b>	Other telecommunications services		
		<b>84210</b>	Internet backbone services		
		<b>84221</b>	Narrowband Internet access services		
		<b>84222</b>	Broadband Internet access services		
		<b>84290</b>	Other Internet telecommunications services		
		<b>84632</b>	Home programme distribution services, basic programming package		
		<b>84633</b>	Home programme distribution services, discretionary programming package		
		<b>84634</b>	Home programme distribution services, pay-per-view		
6120	Wireless telecommunications activities	<b>84110</b>	Carrier service		
		<b>84131</b>	Mobile voice services		
		<b>84132</b>	Mobile text services		
		<b>84133</b>	Mobile data services, except text services		
		<b>84140</b>	Private network services		
		<b>84150</b>	Data transmission services		
		<b>84190</b>	Other telecommunications services		
		<b>84221</b>	Narrowband Internet access services		
		<b>84222</b>	Broadband Internet access services		
		<b>84290</b>	Other Internet telecommunications services		
		<b>84632</b>	Home programme distribution services, basic programming package		
		<b>84633</b>	Home programme distribution services, discretionary programming package		
		<b>84634</b>	Home programme distribution services, pay-per-view		



6130	Satellite telecommunications activities	84140	Private network services		
		84150	Data transmission services		
		84190	Other telecommunications services		
		84221	Narrowband Internet access services		
		84222	Broadband Internet access services		
		84290	Other Internet telecommunications services		
		84632	Home programme distribution services, basic programming package		
		84633	Home programme distribution services, discretionary programming package		
6190	Other telecommunications activities	84140	Private network services		
		84150	Data transmission services		
		84190	Other telecommunications services		
		84221	Narrowband Internet access services		
		84222	Broadband Internet access services		
		84290	Other Internet telecommunications services		
6201	Computer programming activities	83141	IT design and development services for applications	9.2.2/SI22	<p>Computer services - Other computer services</p> <p>Other computer services includes:</p> <ul style="list-style-type: none"> <li>- Hardware and software consultancy and implementation services, including the management of subcontracted computer services</li> <li>- Hardware and software installation, including installation of mainframes and central computing units</li> <li>- Maintenance and repairs of computers and peripheral equipment</li> <li>- Data recovery services, and provision of advice and assistance on matters related to the management of computer resources</li> <li>- Analysis, design and programming of systems ready to use (including web page development and design) and technical consultancy related to software</li> <li>- Systems maintenance and other support services, such as training provided as part of consultancy</li> <li>- Data-processing and hosting services, such as data entry, tabulation and processing on a timesharing basis</li> <li>- Web page hosting services (that is, provision of server space on the Internet for hosting of clients' web pages)</li> <li>- Provision of applications, hosting clients' applications, and computer facilities management (MSITS 2010, para. 3.230)</li> </ul>
6202	Computer consultancy and computer facilities management activities	83131	IT consulting services		
		83132	IT support services		
		83142	IT design and development services for networks and systems		
		83161	Network management services		
		83162	Computer systems management services		

6209	Other information technology and computer service activities	87332	Installation services of personal computers and peripheral equipment	10.3.5/SJ35	Other business services n.i.e. (Definition above)
6311	Data processing, hosting and related activities	83151	Website hosting services	9.2.2/SI22	Computer services - Other computer services (Definition above)
		83152	Application service provisioning		
		83159	Other hosting and IT infrastructure provisioning services		
		83633	Sale of Internet advertising space (except on commission)	10.2.2/SJ22	Advertising; market research; and public opinion polling (Definition above)
6312	Web portals	83633	Sale of Internet advertising space (except on commission)		
		84394	Web search portal content	9.3.2/SI32	Information services - Other information services (Definition above)
6391	News agency activities	84410	News agency services to newspapers and periodicals	9.3.1/SI31	Information services - News agency services News agency services includes the provision of news, photographs and feature articles to the media (MSITS 2010, para. 3.232)
		84420	News agency services to audiovisual media		
6399	Other information service activities n.e.c.	85991	Other information services	10.3.5/SJ35	Other business services n.i.e. (Definition above)

Source: OECD based on [https://unstats.un.org/unsd/classifications/Econ/tables/CPC/CPCv21\\_ISIC4/cpc21-isic4.txt](https://unstats.un.org/unsd/classifications/Econ/tables/CPC/CPCv21_ISIC4/cpc21-isic4.txt),  
[https://unstats.un.org/unsd/publication/seriesm/seriesm\\_4rev4e.pdf](https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf),  
[https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/CPCv2.1\\_complete\(PDF\)\\_English.pdf](https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/CPCv2.1_complete(PDF)_English.pdf)

Annex Table 5.A.2. List of digitisable products

No.	HS code	HS description
<b>Photographic and Cinematographic Films</b>		
1	370510	For offset reproduction
2	370520	Microfilms
3	370590	Other
4	370610	Of a width of 35 mm or more
5	370690	Other Printed Matter
<b>Printed Matter</b>		
6	482110	Printed
7	490110	In single sheets, whether or not folded
8	490191	Dictionaries and encyclopaedias, and serial instalments thereof
9	490199	Other
10	490210	Appearing at least four times a week
11	490290	Other
12	490300	Children's picture, drawing or colouring books
13	490400	Music, printed or in manuscript, whether or not bound or illustrated
14	490510	Globes
15	490591	In book form
16	490599	Other
17	490600	Plans and drawings for architectural, engineering, industrial, commercial, topographical or similar purposes, being originals drawn by hand; handwritten texts; photographic reproductions on sensitised paper and carbon copies of the foregoing
18	490700	Unused postage, revenue or similar stamps of current or new issue in the country in which they have, or will have, a recognised face value; stamp-impressed paper; banknotes; cheque forms; stock, share or bond certificates and similar documents of title
19	490810	Transfers (decalcomanias), vitrifiable
20	490890	Other
21	490900	Printed or illustrated postcards; printed cards bearing personal greetings, messages or announcements, whether or not illustrated, with or without envelopes or trimmings
22	491000	Calendars of any kind, printed, including calendar blocks
23	491110	Trade advertising material, commercial catalogues and the like
24	491191	Pictures, designs and photographs
25	491199	Other
<b>Sound &amp; Media</b>		
26	852349	Other
27	852380	Discs, tapes, solid-state non-volatile storage devices, "smart cards" & other media for the recording of sound/of other phenomena, whether/not recorded, incl. matrices & masters for the production of discs, but excl. products of Ch.37., other n.e.s.
28	852410	Gramophone
29	852421	Records, tapes and other recorded media for sound or other similarly recorded phenomena, Of a width not exceeding 4 mm
30	852422	Records, tapes and other recorded media for sound or other similarly recorded phenomena, Of a width exceeding 4 mm but not exceeding 6.5 mm
31	852432	For reproducing sound only
32	852439	Other
33	852451	Of a width not exceeding 4 mm
34	852452	Of a width exceeding 4 mm but not exceeding 6.5 mm
35	852453	Of a width exceeding 6.5 mm
36	852460	Cards incorporating a magnetic stripe

37	852499	Other
<b>Software</b>		
38	852431	Data Processing Software on CD Roms for Reproducing Phenomena Other Than Sound Or Image
39	852440	Computer Software, Magnetic tapes for reproducing phenomena other than sound or image
40	852351	Flash memory cards or flash electronic storage cards, Semi-conductor media, solid-state non-volatile storage devices, for the recording of sound/of other phenomena, but excl. products of Ch. 37.
41	852352	'Smart cards'
42	852359	Other semi-conductor media, for the recording of sound/of other phenomena, but excl. products of Ch. 37., other than "Smart Cards" & Solid state non-volatile storage devices, proximity cards and tags
43	852491	For reproducing phenomena other than sound or image
44	854212	Cards incorporating an electronic integrated circuit ("smart" cards)
<b>Video Games</b>		
45	950450	Video game consoles and machines, other than those of subheading 950430
46	950430	Games; operated by coins, banknotes, bank cards, tokens or by other means of payment, other than billiard articles and accessories, and automatic bowling alley equipment
47	950440	Games; playing cards
48	950490	Games; articles for funfair, table or parlour games, including pintables, special tables for casino games, automatic bowling alley equipment, n.e.c. in heading 9504
49	950410	Video games of a kind used with a television receiver

Source: Banga (2019), Growing Trade in Electronic Transmissions: Implications for the South, UNCTAD Research Paper, [https://unctad.org/system/files/official-document/ser-rp-2019d1\\_en.pdf](https://unctad.org/system/files/official-document/ser-rp-2019d1_en.pdf)

## Notes

<sup>1</sup> The United States is not among these countries.

<sup>2</sup> However, para. 10.114 stipulates that databases for sale should be valued at their market price, which includes the value of the information content.

<sup>3</sup> Under the standard National Accounts treatment of ordinary repair and maintenance, the costs of repairing and maintaining an asset, (assuming it doesn't extend the assets life, which somewhat of a grey area) is treated as a current expense regardless of whether the asset being maintained is a basic fixed asset (e.g. equipment) or an Intellectual Property Product. Because of this, the intermediate inputs and capital services used in *developing* the database are included in the sum of costs but those for maintaining the database would not be included. For comparison, cost of intermediate and capital service inputs used in maintaining motor vehicles are also not capitalised.

<sup>4</sup> The NAPCS is also used by INEGI Mexico, although a Mexico-specific version does not appear to have been published ([http://naics-scian.inegi.org.mx/SICPAN/default\\_i.aspx](http://naics-scian.inegi.org.mx/SICPAN/default_i.aspx))

<sup>5</sup> Indeed, in the Canada NAICS 519130 is entitled "Internet Broadcasting and Web Search Portals" whereas in the US NAICS it is entitled "Internet Publishing and Broadcasting and Web Search Portals"

<sup>6</sup> Another area for potential further research relates to industries which compile databases but, rather than selling them or access to them, uses the database as a basis for selling derived services such as advertising. The US data suggest that such business models are likely to represent a large portion of the overall commercial value of data and databases in industries such as Web Portals.

<sup>7</sup> Despite being matched to stock market tickers, the database did not offer the needed data for some firms. Data were retrieved for 69 firms. However, four of these were listed in Q3 2020 and so are excluded from this analysis based on changes, while one firm was dropped due to long periods of missing data that could not reasonably be bridged.

<sup>8</sup> This is based on Sandvine's analysis of traffic through their installed networking infrastructure equipment and may not accurately represent wider Internet traffic.

<sup>9</sup> In practice the payment will often go to a regional Netflix subsidiary rather than directly to Netflix, Inc.

<sup>10</sup> Although Netflix does still make some revenue (around USD 60m in 2019) from offering DVD rentals by post in some markets, including the USA (Sisson, 2020<sup>[59]</sup>) (Gebel, 2020<sup>[60]</sup>)

<sup>11</sup> BRIICS countries: Brazil, Russian Federation, India, Indonesia, People's Republic of China, South Africa.