

Sargent, Timothy C.; Denniston, Laura

Working Paper

Valuing data: Where are we, and where do we go next?

CIGI Papers, No. 280

Provided in Cooperation with:

Centre for International Governance Innovation (CIGI), Waterloo, Ontario

Suggested Citation: Sargent, Timothy C.; Denniston, Laura (2023) : Valuing data: Where are we, and where do we go next?, CIGI Papers, No. 280, Centre for International Governance Innovation (CIGI), Waterloo, ON, Canada

This Version is available at:

<https://hdl.handle.net/10419/299977>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://www.creativecommons.org/licenses/by-nc-nd/3.0/>

CIGI Papers No. 280 – September 2023

Valuing Data

Where Are We, and Where Do We Go Next?

Tim Sargent and Laura Denniston



CIGI Papers No. 280 – September 2023

Valuing Data

Where Are We, and Where Do We Go Next?

Tim Sargent and Laura Denniston

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

Credits

Managing Director of Digital Economy **Robert Fay**

Director, Program Management **Dianna English**

Project Manager **Jenny Thiel**

Publications Editor **Susan Bubak**

Senior Publications Editor **Jennifer Goyder**

Graphic Designer **Abhilasha Dewan**

Copyright © 2023 by the Centre for International Governance Innovation

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Centre for International Governance Innovation or its Board of Directors.

For publications enquiries, please contact publications@cigionline.org.



This work is licensed under a Creative Commons Attribution — Non-commercial — No Derivatives License. To view this license, visit (www.creativecommons.org/licenses/by-nc-nd/3.0/). For re-use or distribution, please include this copyright notice.

Centre for International Governance Innovation and CIGI are registered trademarks.

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Table of Contents

vi	About the Authors
vi	Acronyms and Abbreviations
1	Executive Summary
1	Introduction
2	Current Treatment of Data in the National Accounts
3	Current Treatment of Data in the SNA
4	Experimental Estimates of the Value of Data by Statistical Agencies
8	Next Steps for National Accounts
9	Alternate Methods for Data Valuation
12	Conclusions and Directions for Future Research
14	Works Cited

About the Authors

Tim Sargent is a CIGI distinguished fellow with 28 years of experience with the Government of Canada. He has held senior roles at Global Affairs Canada, the Privy Council Office and the Department of Finance, giving him policy-making experience at the highest level, in particular in the areas of trade policy, international finance and macroeconomics.

Tim earned a Ph.D. in economics at the University of British Columbia, an M.A. at the University of Western Ontario and a B.A. in economics at the University of Manchester.

Laura Denniston is an honours economics student at the University of Ottawa. She was a coop student at the Centre for the Study of Living Standards in fall 2022 when this paper was drafted.

Acronyms and Abbreviations

AI	artificial intelligence
BEA	Bureau of Economic Analysis
CSLS	Centre for the Study of Living Standards
GFCF	gross fixed capital formation
IP	intellectual property
NAICS	North American Industry Classification System
PIM	perpetual inventory method
R&D	research and development
SNA	System of National Accounts
TFP	total factor productivity

Executive Summary

There is widespread agreement on the importance of data in advanced economies, but no consensus on exactly how to value the magnitude of this contribution. The authors of this paper look at recent attempts to value both the stock and flow of data in the national accounts and attempt to assess the way forward. They begin with separating out three different categories of data-related assets — data itself, databases and data science — and then outline some of the key national accounting concepts related to incorporating them as assets in the national accounts framework. They then examine in detail three recent studies by statistical agencies in Canada, the Netherlands and the United States, each of which tries to value data-related assets using a cost-based methodology. While each of these studies finds that data is a significant asset — investment in data-related assets is estimated at anywhere from one to three percent of output — there are significant differences in the results, which are, in turn, driven by some arbitrary assumptions that the authors were required to make.

Going forward, a key question is whether data will be treated as an asset in the 2025 System of National Accounts (SNA). The authors argue that it likely will, but that more work will need to be done before statistical agencies will start including data as an asset in individual countries' national account systems.

The authors then go on to examine two other ways of valuing data: the income-based method and the market-based method, as well as a hybrid approach. The authors find that while conceptually superior, these methods are hard to implement in practice.

Finally, the authors conclude with suggestions for further work. The scope of efforts to value data could be broadened to include the public sector, which has significant holdings of data. Also, the social and not just the private value of data needs to be calculated — this will mean considering both the positive and negative externalities of data-related investment and use. Finally, given these broader social implications of data acquisition and use, more work needs to be done on appropriate data governance models to ensure that data is being used for the benefit of all.

Introduction

In 2017, *The Economist* (2017) announced that “the world’s most valuable resource is no longer oil, but data,” arguing that data is fuelling the modern economy just as oil did a century earlier. In the same year, Jonathan Haskel and Stian Westlake (2017) published *Capitalism without Capital: The Rise of the Intangible Economy*, arguing that intangible capital, which includes data but also research and development (R&D), patents and other intellectual property (IP), is now the main driver of advanced economies, rather than physical capital such as machinery and buildings. Since then, the perceived importance of data to business and the economy at large has only grown, especially given the importance of data as the raw material in artificial intelligence (AI) systems, which promise to be a new “general purpose technology” as consequential and wide ranging as computers or electricity.

The importance of data to the economy, and particularly its increasing importance as an asset, poses a challenge to statistical agencies. While the current SNA standard recognizes the costs of database management software, it does not treat data as an asset, and does not count the costs of acquiring the data. This absence reflects some of the very real problems that exist in valuing a commodity that is very situation-specific, is usually not bought and sold in transparent markets (or at all), and, consequently, is not easily obtainable from firms' balance sheets.

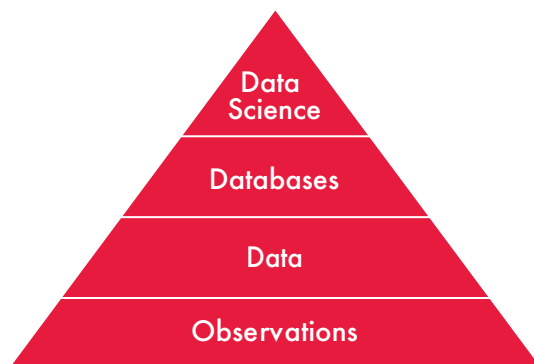
Despite these problems, several national statistical offices have risen to the challenge of attempting to measure the value of data in a way that is consistent with national accounts concepts for capital. In this paper, the authors will take stock of these attempts, assess how plausible their estimates are, examine other possible methods of valuation and suggest potential ways forward for the valuation of data, both in an SNA context and more broadly.

Current Treatment of Data in the National Accounts

Fundamental Concepts and Definitions

The authors begin by exploring the key concepts associated with valuing data as an asset in the SNA. The first is the definition of data, and how it relates to adjacent concepts such as databases and data science. One useful way to address this issue is to situate data in what Statistics Canada (2019a) calls the “information value chain,” shown in Figure 1.

Figure 1: Information Value Chain



Source: Statistics Canada (2019a).

“Observations” (or “observed phenomena”) are the first step in the information value chain. Observations can be practically anything. For example, observations include what an individual ate for lunch, the weather, or the time someone gets home from work. They are emitted constantly by people, objects and the environment, and are usually transitory and intangible. However, observations are not inherently recorded and stored. In most cases, someone must decide to capture and digitize them. Only when observations are recorded in a digital format do they become “data,” the second step of the information value pyramid.

However, data is merely stored and digitized observations. It is not organized or processed; it is still essentially raw material. It cannot yet be interpreted or used to gain insights. To be useful, it needs to be entered into the third tier of the

information value chain: a database, defined as “an organized store of data that can be readily retrieved and manipulated” (ibid., 8).¹

The highest tier of the pyramid is “data science.” This tier includes the analysis done to gain knowledge from the data to inform future activities. Data science looks at the data set as a whole to reveal trends and patterns that cannot be gleaned from individual data points.

National Accounts Concepts

Appropriately categorizing the elements of the information value chain in the SNA requires assessment along two different dimensions. The first is whether the element is an asset or not. If it is, then any expenditures count as an investment. The SNA defines assets as “entities that must be owned by some unit...and from which economic benefits are derived by their owner(s) by holding or using them over a period of time” (European Commission et al. 2009, 7). Thus, a factory is an asset because it is durable and provides benefits over many years, whereas the car parts it produces are intermediate consumption because they are quickly transformed into something else and so are not durable. Similarly, an oil field for which rights have been assigned is an asset because it is owned and provides benefits over time, whereas once the oil is pumped out, it is subsequently refined and so would be counted as intermediate consumption. Assets, of course, need not last forever, and if something is determined to be an asset, then a view needs to be taken on an appropriate depreciation rate, which would need to consider both physical decay and reduction in economic value.

The second dimension along which elements of the information chain need to be assessed is whether the element is produced or not. In the SNA, “Production is understood to be a physical process, carried out under the responsibility, control and management of an institutional unit, in which labour and assets are used to transform inputs of goods and services into outputs of other goods and services. All goods and services produced as outputs must be such that they can be sold on markets or at least be capable of being provided by one unit to another, with or without charge” (ibid., 6).

1 This distinction between databases and raw data would disappear to the extent that AI applications are able to draw insights from raw data without the need for these data to be “organized” in databases.

In the examples above, a factory is a produced asset, whereas car parts and barrels of oil are produced goods; however, the oil field is a non-produced asset because it has not been created by human agency.

Current Treatment of Data in the SNA

While the concepts of asset and produced good are relatively clear and well accepted, their application to the information chain is not, and discussions on how to value data-related assets have been ongoing for more than 30 years. Originally, expenditures on collecting and producing data, entering it into a database and analyzing the results, were treated as intermediate expenditures and so not as investment. Despite these expenditures, data, databases and data science were essentially treated as a non-produced asset, along with other forms of IP. This changed partially with the 1993 SNA, which recommended the inclusion of large databases as produced assets (Commission of the European Communities et al. 1993). However, this change was largely made to ensure consistency with the proposed inclusion of software in the SNA: it was (and remains) impractical to separate database management software from the database itself (Statistics Canada 2001). Thus, the purchase and development costs of database software were capitalized, but not database content, creation or updating. Data and data science therefore remained “outside the asset boundary” in statisticians’ parlance — not treated as produced assets.

A key innovation of the 2008 SNA was the creation of an asset class of IP products, which includes the existing asset classes of software and databases, but also R&D. In the buildup to the 2008 SNA, the Canberra II Group re-examined the issue of how best to include data and databases in the SNA (Rassier, Kornfeld and Strassner 2019). This group considered two definitions for databases. The first included “the value of the information... stored on the databases” (the value of the data) and the second did not (Ahmad 2005, 2). The group ultimately recommended the second definition, excluding the value of data from the database. This was done to not risk “the capitalization of knowledge” (ibid.): the fear was that by including

the value of data, one would ultimately be driven to include the value of all knowledge, which would risk overwhelming the rest of the national accounts (Ahmad and van de Ven 2018).

Ultimately, the SNA 2008 does not come down completely on one side or the other; instead, whether data should be included as an asset or not depends on the practical issue of how the value of the database is determined. For those databases that are developed for in-house purposes and not intended for sale, which is most databases, “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” (European Commission et al. 2009, 10.113). However, those databases intended for sale “should be valued at their market price, which includes the value of the information content” (ibid., 10.114). Essentially, the authors of the SNA prefer to exclude data from the list of assets but recognize that this would be impractical when databases are valued at market price, as it would be very difficult to disentangle the value of the data from the rest of the value of the database.

This approach is somewhat different from the approach used to capitalize R&D in the 2008 SNA where R&D, unless its market price is observed directly, is valued as the sum of all costs associated with activity — acquisition and production costs are not excluded.

Where does this leave data science? In principle, data science meets the criterion for R&D: “Research and [experimental] development consists of the value of expenditures on creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and use of this stock of knowledge to devise new applications” (ibid., 10.103). However, Peter Goodridge, Jonathan Haskel and Harald Edquist (2022) argue that, in practice, R&D statistical agencies generally measure only traditional science laboratories, which, by and large, excludes data science. The latter might get categorized under software development but would then be assigned to software. Data science therefore remains effectively uncaptured in the current version of the SNA.

Experimental Estimates of the Value of Data by Statistical Agencies

The increasing importance and prominence of data in the economy, as well as the above-mentioned inconsistencies in the current treatment of data, have led a number of researchers to consider how data could be capitalized in the SNA, and to develop experimental estimates of the asset value of data in advance of the next version of the SNA, which is expected in 2025. In this section, the authors will evaluate the three studies, all from national statistical agencies, that develop experimental estimates of the value of data: Statistics Canada (2019b); Hugo de Bondt and Nino Mushkudiani (2021) from Statistics Netherlands; and José Bayoán Santiago Calderón and Dylan G. Rassier (2022) from the Bureau of Economic Analysis (BEA) in the United States.

Concepts and Methods

The three papers all use essentially the same approach. In each case, estimates are presented for data, databases and data science (the US study aggregates the three, whereas the other two present aggregates for the three different concepts). Estimates are for own account data; data for sale is excluded. Databases already included in existing measures of software are also excluded.

To compute the value of the stock and flow of data-related assets, all three studies used the sum-of-costs approach, whereby the value of assets is measured by the labour, capital and costs incurred in production. This is a common approach in the SNA when no market price is observed — it is also used for valuing own account software and R&D, as well as many government activities. Since businesses are unlikely to pay more for producing an asset than what it is worth to them, this method at least provides a lower bound on the value of the asset to a company.

The value of data according to this approach is calculated as follows: For each occupation where there is data production, one calculates the proportion of labour time devoted to data production and multiplies this by the average

wage.² Summing this over all relevant occupations gives the total labour cost, which is then marked up to account for capital and intermediate consumption costs to provide an overall value for production costs of data for a given year. This gives a value for nominal investment in data. To calculate real investment, a price deflator is constructed using wages, the price of intermediate consumption and capital (if available), and estimates of the rate of technological change.

To calculate the stock of data assets, the Canadian and US studies (the Dutch study did not estimate the value of the stock) both use the perpetual inventory method (PIM). The stock of data is calculated by cumulating over time the flows of investment minus the flows of depreciation and discards. This method requires assumptions about the useful life of data and its depreciation rate but, given these considerations, is straightforward to implement. The PIM is a standard method used by statistical agencies when direct measures of the value of a capital asset are not available.

Differences in Methodology

Despite a common methodology, there are some important differences between the three studies, largely driven by data availability.

- **Aggregation:** The Canadian and Dutch studies both disaggregate data-related assets into data, databases and data science in their estimates; the US study does not.
- **Occupations:** Each study uses its own method for determining relevant occupations. The Dutch study tries to be as close as possible to the Canadian categorization; however, the US study uses machine learning applied to job descriptions from online job advertisements to determine the data content of an occupation.
- **Sectoral scope:** The US and Dutch studies only cover the business sector, whereas the Canadian study covers government data assets as well.
- **Time period:** The Canadian study covers the period 2005–2018; the Dutch study 2001–2017; and the US study 2002–2021.

2 This approach assumes that wages are equal to the worker's marginal product. To the extent that workers earn a wage premium, this approach will overestimate that value of data. Large firms often pay significant wage premia in order to motivate and retain workers.

→ **Adjustment for overlap with existing measures of capital formation:** The US study makes two adjustments for potential overlap between estimates of production of data-related assets and existing measures of production of R&D and software, which are already incorporated into measures of investment in the national accounts. The study reduces the estimates of labour income by the proportion of employees engaged in R&D, removes occupations such as computer programmer and software developer that are likely engaged in developing software, and further assumes that 50 percent of labour input is already included in capital formation. The Canadian and Dutch studies note the problem but do not address it.

→ **Adjustment for own account data:** The Canadian and Dutch studies do not have an explicit adjustment factor for the portion of labour input that is used to produce own account data rather than data for sale. Instead, the subjective estimates of data-related input for each occupation are meant to exclude data for sale. The US study assumes that only 50 percent of the activities of one of the key occupations (data processing and hosting) is for own account data.

→ **Markup:** The Canadian study assumes a markup on wage costs of 50 percent to cover “non-direct salary and other costs”; the Dutch study assumes a markup of 60 percent. Both assume an additional markup of three percent to cover capital costs. The US study assumes a markup of 153 percent — much higher than the other two studies.

→ **Depreciation rates and useful lives:** The Canadian study assumes a useful life for data of 25 years, databases of five years (to be consistent with software in the Canadian SNA) and data science of six years. A geometric depreciation rate is assumed. The US study, which does not make a distinction between the three categories of information, assumes a service life of five years and a geometric depreciation rate of 0.33 (this is the same treatment as software in the US SNA). There is therefore a very large difference between the US and Canadian treatment of data: 25 years versus five years.

→ **Price indices:** The Canadian study bases its estimates of prices solely on labour costs, adjusted down by one percent per year to

account for assumed productivity growth. The Dutch study uses labour costs, consumption costs and capital costs to produce a price index, which is then (as with the Canadian study) adjusted down by one percent per year to reflect productivity growth. The US study uses labour costs and intermediate consumption costs (but not capital costs) to construct an input price index that is then adjusted down by estimates of total factor productivity (TFP). This index is then combined with the index for products of data-related industries.

Comparison of Results

Table 1 compares estimates of the share of GDP and of gross fixed capital formation (GFCF) represented by data-related investment and its three components from the three studies. For Canada and the Netherlands, both upper and lower bounds are shown.

By and large, the estimates are somewhat different across the three countries. Data-related assets are smallest as a share of output in the United States (1.1 percent) and largest in the Netherlands (2.1–2.7 percent), with Canada in the middle (1.3–1.8 percent). Similarly, data-related assets as a percentage of GFCF range from 5.1 percent in the United States to 5.8–7.9 percent in Canada to 12.7–16.3 percent in the Netherlands.

Looking at the three different subcategories of data-related assets for the two studies that provided this breakdown, these statistics were of similar magnitude, with data science having the largest share of GDP (0.5–0.6 percent), closely followed by data (0.4–0.6 percent) and databases a little further behind (0.4–0.5 percent). In the Netherlands, the data category (1.0–1.4 percent) was well ahead of data science (0.6–0.7 percent) and databases (0.5–0.6 percent). This significantly higher contribution of data can explain most of the discrepancy between the Dutch and Canadian estimates.

Table 2 compares nominal growth rates for data-related investment for the three countries. Investment in data-related assets grew fastest in Canada (6.5 percent) and slowest in the Netherlands (3.5–3.6 percent), with the United States in the middle (4.7 percent).

Looking at components, growth in both countries was fastest in data science (8.9 percent in Canada

Table 1: Share of Nominal GDP and Nominal GFCF Represented by Data-Related Investment

Share of GDP and GFCF (%)						
		Canada 2018		Netherlands 2017		United States 2020
		Lower	Upper	Lower	Upper	
Share of GDP/BVA	Data-related	1.3	1.8	2.1	2.7	1.1
	Data	0.4	0.6	1.0	1.4	-
	Database	0.4	0.5	0.5	0.6	-
	Data science	0.5	0.6	0.6	0.7	-
		Total		Private		Private
Share of GFCF	Data-related	5.8	7.9	12.7	16.3	5.1
	Data	1.9	2.8	5.9	8.5	-
	Database	1.6	2.3	3.2	3.6	-
	Data science	2.4	2.8	3.6	4.2	-

Sources: Statistics Canada (2019b) (Canada); de Bondt and Mushkudiani (2021) (Netherlands); Calderón and Rassier (2022) (United States); Centre for the Study of Living Standards (CSLS) calculations.

Notes: BVA = business value added. US GDP share is a share of BVA; US and Dutch GFCF is private GFCF.

Table 2: Growth of Investment in Data-Related Assets in Current Prices

Compound Average Growth Rates (%)					
Investment	Canada		Netherlands		United States
	Lower	Upper	Lower	Upper	
	2010-2018		2010-2017		2012-2021
Data-related	6.5	6.5	3.6	3.5	4.7
Data	2.8	3.5	2.7	2.4	-
Database	8.7	8.4	-3.4	-3.3	-
Data science	8.9	8.9	21.4	21.2	-

Sources: Statistics Canada (2019b) (Canada); de Bondt and Mushkudiani (2021) (Netherlands); Calderón and Rassier (2022) (United States); CSLS calculations.

and 21.2–21.4 percent in the Netherlands). Growth in data was similar (2.8–3.5 percent in Canada and 2.4–2.7 percent in the Netherlands). However, a very different pattern emerges for databases. In Canada, database investment grew 8.4–8.7 percent, almost as fast as investment in data science; however, in the Netherlands, database investment fell by an average of 3.3–3.4 percent annually. The authors of the Dutch study speculate that this could be due to offshoring by Dutch firms, or an underestimate of productivity growth in this sector.

We can compare growth in the net stock of data-related assets in current prices for the Canadian and US studies. The Canadian study finds a net stock of data-related assets of between \$157 billion and \$217 billion — 5.7–7.7 percent of total non-residential fixed assets. (For context, IP products are about nine percent of total assets.) Two-thirds of this stock is in data, 12 percent in databases and about 20 percent in data science. This preponderance of data is very much driven by assumed service lives: 25 years for data as opposed to five years for databases and six years for data science. In the US study, data-related assets are calculated at only 1.3 percent of private fixed assets — a very large difference. (IP products were 14 percent of total US non-residential fixed assets in 2021.)

Each of the three studies calculates a price index for data-related investment, using input costs (particularly wages) and assumptions about productivity growth. The Canadian study finds an annual average increase of 0.9 percent for data and databases and 0.6 percent for data science (2010–2018). Overall, the price increase is 1.8 percent; presumably, this overall price growth is higher than that of its individual components because of compositional shifts (recall data grew more slowly than databases and data science over this period). The Dutch study calculates an increase of zero percent for data and databases and 0.4 percent for data science (2010–2017), while the US study finds a decline in the price of data-related assets of more than one percent (2012–2020).

Finally, the Canadian and US results permit some sectoral breakdowns. The Canadian study finds that in 2018, non-financial corporations accounted for about half of all investment in data-related assets, with financial corporations accounting for a third of investment and government about a fifth. The US study breaks down investment by industry sector and finds that for the period 2002–2021, the industries with the largest investments were

professional, scientific and technical services (25 percent), followed by manufacturing (14 percent), and finance and insurance (13 percent).

Reasons for the Disparities

It is notable that, according to the three studies, the relative importance of data assets is much lower in the United States, which seems to have much lower rates of investment and a much lower stock of assets. Given the importance of the information economy in the United States, which domiciles information powerhouses such as Google and Meta, this is very unlikely. The more likely explanation is differences in methodology. The first candidate is differences in the adjusted markup. The Canadian and Dutch studies multiply the wage data by a ratio of 1.53–1.63 percent to reflect non-wage costs. The United States assumes a markup ratio of 2.52 but halves this number to 1.26 to adjust for overlap with existing measures of capital formation, and halves it again to 0.63 for industries in North American Industry Classification System (NAICS) code 518 (data processing, hosting and related services) to adjust for overlap with data-related products that are sold rather than used for a firm's own account (this is the industry where data sales largely occur). As a result, the adjusted wage data is anywhere from 25 to 60 percent lower for the United States, depending on whether the industry is in NAICS code 518. It is likely that many industries do, in fact, fall into the NAICS code as it includes industries that have data as a primary business activity.

The other factor that will likely have an important impact on the difference between the two sets of results is the method the US study uses to allocate occupations and the proportion of time dedicated within occupations to data-related production, which is much more sophisticated than the method adopted by the other two studies. Assessing the extent of this difference is challenging, as the machine-learning process by which the US study assigns labour input to occupations is, by its very nature, opaque; nonetheless, it is likely to be significant given how very different it is from the manual approach adopted by the Canadian and Dutch studies.

For prices, a key difference is assumptions about the pace of productivity growth and the mix of inputs. As noted above, the Canadian study simply assumes a productivity growth rate of one percent annually and uses only labour to calculate input prices. The Dutch study makes

the same assumption about productivity growth, but allows for intermediate inputs and capital as well as labour, and thus uses price data on these two inputs as well. This would explain the somewhat lower price growth than the Canadian study, as wages tend to grow faster than intermediate input costs or capital costs.

The US study uses actual estimates of TFP growth from the two relevant industries (data processing and hosting, and computer systems design), along with input cost data to construct an input cost index. This is then combined in a simple average with an average industry price index. According to published data from the Bureau of Labor Statistics, TFP growth in data processing and hosting was 3.3 percent annually over the 2010–2020 period, whereas TFP growth in computer systems design was -0.9 percent. Using, as the US study does, industry gross output to weight these two indices, we arrive at a TFP growth for the two industries combined of 1.5 percent annually, half a percentage point higher than the one percent assumed by the other two studies. This difference is not enough to explain most of the difference in price growth between the Dutch and US studies. The other difference between the two studies is that the US study combines the input index with an average price index. However, as industry prices were growing (at least for data processing — about one percent per year),³ it seems that the key reason why prices for data decline in the United States but not in the Netherlands would be faster price declines in the United States for intermediate inputs.

Weakness of the Studies

Interesting as the three studies are, there are some clear challenges in moving forward with any of these methodologies to produce estimates for the SNA. The most obvious are in assessing the proportion of own account production versus production for sale, and in assessing the amounts of non-labour inputs, particularly capital, in the production process. The studies either sidestep the issue or make very arbitrary assumptions. As we have seen, those assumptions seem to contribute to quite large differences in the share of

data-related investment in GDP, and so statistical agencies would need to survey producers to get a better sense of what these proportions should be.

Another area where quite arbitrary assumptions are made is the asset life of data. As we have seen, the Canadian study assumes 25 years for data, whereas the US study assumes five years. The Canadian study bases the 25-year assumption on the idea that data would be useful for about a generation, and this is how long firms would store the data. However, it is likely that a lot of data is quickly replaced by more timely data, which would argue for a much shorter service life. It is plausible, therefore, that asset lives for data are quite short, which would significantly reduce the estimated size of the net stock of data assets from what is estimated in the Canadian study.

Finally, there are some fairly arbitrary assumptions made to calculate the price data. Both the Canadian and Dutch studies use an assumed productivity growth rate; using an actual productivity growth series, as the US study does, would obviously be better if the data is available. Furthermore, basing output prices on input prices, as the Canadian and Dutch studies do, is not ideal; if the data were sold, the firm would presumably make some return. The US approach, as noted above, is to combine an input price index with an output price index from the industry that is sold: this seems a better approach, even if the data produced for own account is likely different from that produced for sale. However, the equal weight of the two indices seems quite arbitrary, and more work should ideally be done to determine the appropriate weights.

Next Steps for National Accounts

Given these experimental studies, how likely is it that data will indeed be included in the 2025 SNA? A good indication is given in a guidance note prepared for the United Nations Economic and Social Council in 2022 by its Digitalization Task Team (United Nations Economic and Social Council 2022). The note covers the main issues raised in recording data in the SNA.

3 The authors were not able to find producer price data for the computer systems design industry, but as this industry is part of professional services, the other components of which did see significant price increases, it is implausible that there were price decreases in this industry large enough to lead to a significant price decrease for the two industries combined.

Data as a Produced Good

The task team essentially accepts the taxonomy of the information chain outlined above, with observations (following John Mitchell, Molly Leshner and Marion Barberis [2022], it uses the term “observed phenomena”) outside the production boundary, and data, databases and data science (it uses the term “insights,” again following Mitchell, Leshner and Barberis [2022]) within the production boundary. The rationale for including data is that it is the result of production, with labour and capital being used to transform a non-produced good (observations) into a digital format, which is the produced good.

Data as an Asset

The SNA recommends that a produced good that is used in production but is consumed within a year be treated as intermediate consumption, whereas a good that is used in production for longer than a year be treated as capital. The guidance note recommends this approach for data that is purchased: if the data is intended to be used for more than a year, it should be capitalized and the resulting asset form part of the IP class of assets. For own account data, the guidance note argues that, in practice, it will be difficult for statistical agencies to get information on the expected service life of a data set when it is produced; the note therefore recommends that all data be capitalized in order to be consistent with the way some other IP assets are treated in the SNA, particularly mineral exploration. The note recognizes that because of its time-sensitive nature, a lot of data is in fact used within a year; however, the note recommends that this issue be dealt with by assuming appropriately short service lives.

Valuation of Data and Data Science

The guidance note recommends that own account data be valued according to the sum-of-costs approach, which is consistent with how other own-account IP assets are valued. While the guidance note recognizes data science as part of the production chain, downstream from data and databases, it does not argue for its inclusion in the asset boundary.

Given these conclusions, which build upon the work done by the three experimental studies, it seems likely that the 2025 SNA will recommend

the inclusion of data, but not data science, in the asset boundary (unless it is purchased data intended to be used within a year). In many ways, data-related assets are following the same well-trodden path as other IP assets such as software and R&D, which are now capitalized in the SNA. The main issues at this point are practical: how best to measure service lives and prices, and how to avoid duplication with other IP assets. On the former, it seems likely that the guidance will be for a much shorter service life than the 25 years contemplated by the Canadian experimental study, and perhaps less than the five years in the US study, in order to offset the inclusion of data with a relatively short life in the asset boundary. The note recognizes that a significant amount of testing may be required before an acceptable approach to determining service life is established. It could therefore be several years after the release of the 2025 SNA before statistical agencies start capitalizing data in the national accounts.

Alternate Methods for Data Valuation

The cost-based methodology used in the three studies examined above has numerous advantages that explain its popularity in data-valuation exercises. Costs usually occur in the past and are easy to measure: for data, most of the costs are salary costs for employees engaged in data acquisition, manipulation and analysis, which are picked up in regular surveys of firms and workers conducted by statistical agencies. The cost-based method is already used by statistical agencies to value intangibles such as software and R&D expenditures that are produced by firms as intermediate inputs into the production process and are not sold to third parties. It is also used in the academic literature: Goodridge, Haskel and Edquist (2022) use this method to estimate the impact of data on productivity growth in the European Union.

However, while conceptually straightforward, the cost-based method is, at best, only a lower bound on the value of data. Given how profitable data-driven firms are, it is likely that the returns on investments in data-related assets are well

above the actual costs of the investments. Also, as we have seen, assumptions about service lives and depreciation rates need to be made to arrive at a capitalized value for data.

These problems have led researchers to consider other approaches to valuing data. These are the income-based method, which attempts to value the flow of revenues from investing in data; the market-based method, which uses market prices for data or for the value of the firm; and a hybrid model, which combines the income-based and cost-based approaches.

Income-Based Methods

The income-based method uses direct estimates of expected revenue streams that derive from the data asset. If reliable estimates of revenue streams are available, then this is a straightforward method to implement as long as one is willing to specify an appropriate discount rate. This method is a type of “value in use” method, which is already familiar to most accountants as it is defined in the International Accounting Standard 36. Value-in-use methods are any methods where the asset’s value is derived as “a function of the future value streams the asset creates or enables” (Girard, Lionais and McLean 2021, 6). An example of the use of this approach for other types of intangible assets is for patents and trademarks, where streams of payments are directly observable over a significant period (Coyle and Manley 2022).

However, in practice, this method has only limited application for valuing data. First, it is only really feasible when data is being produced for sale, so that the income stream can be directly observed, rather than being used as an input further along the information value chain to produce insights and analytics that are used elsewhere in the business or sold to clients. In the latter case, it is difficult — if not impossible — to disentangle the contribution of the data itself from other inputs to the production process.

Second, even if data is produced for sale, identifying the revenue from this activity is not straightforward. Daniel Ker and Emanuele Mazzini (2020) attempt to use business survey data from a number of Organisation for Economic Co-operation and Development countries to measure the output of activities related to compiling and selling databases. They find that for many industries, sales of data are not the main product, and

so data on revenue by industry significantly overestimates revenues from data sales. Unfortunately, the detailed product line data, which would be required for more reliable estimates, is simply not available for most countries.

Finally, even if reliable revenue data were available, there would still be the problem of valuing future flows in order to arrive at a valuation of the data asset. While for some assets, such as natural resources, one might be able to forecast future revenues with some confidence, the pace of technological change in applications of data, such as AI, makes forecasting revenue streams from current revenues a very difficult endeavour.

Market-Based Methods

An alternative to the income-based method of valuing data is the market-based method, which uses the current observed price of the asset, rather than the revenue stream, to make the valuation. In principle, this is an ideal solution as there is no need to make assumptions about discount rates or depreciation: the market takes care of all this.

However, for this approach to work in practice, one requires transparent markets with enough transactions to establish the market price for a specific type of data at any given time. Unfortunately, there is very little observable information on prices for buying and selling data sets (Coyle and Manley 2022). This is partly because, as discussed above, firms usually produce data for their own use rather than sell it directly on the open market. However, even for those firms that are engaged in selling data directly, the transactions are kept private. Part of the challenge, as noted by Pantelis Koutroumpis, Aija Leiponen and Llewellyn D. W. Thomas (2020), is that data is an “experience good”: its value is only known once it is purchased. This implies an inherent asymmetry of information between buyer and seller, which makes the market prone to failure. The private nature of transactions also makes it easier for sellers to price discriminate, so that different buyers may pay a very different price depending on their willingness to pay, with the result that there is no one market price for a given data set.

One way around the lack of observable transactions in the market for data is to turn to another market, the stock market. In principle, if a company is primarily engaged in selling data, its market valuation should reflect the net present value of

that data, along with the other assets it owns, both tangible (buildings and equipment) and intangible (organizational strengths, customer knowledge, patents, proprietary research and so forth). The attraction of this method is that, at least for publicly traded companies, market valuations are readily available. Furthermore, the value of a company's assets is available from the information companies are obliged to make public (even if data is not mentioned explicitly, other more conventional assets are).

Ker and Mazzini (2020) look at stock market valuations for 64 data-driven firms whose business model relies primarily on data collecting and data analytics (for example, Amazon, Cognisant, Microsoft and Oracle). They find that the market capitalization of these firms in 2020 was more than US\$5 trillion (about 12 percent of total US market capitalization) and equivalent to about one-quarter of US GDP.

There are numerous challenges to this approach, as Ker and Mazzini admit. First, not all data-driven firms are listed on public markets: social media company X (formerly known as Twitter) is one example (Murphy 2022). Second, it is very difficult to separate out the value of data from the value of other intangible assets such as proprietary research or organizational know-how. Companies such as Amazon or Microsoft may be data-driven, but their primary product is not data, and there are many assets, both tangible and intangible, that contribute to the production process. Attributing all or even most of these companies' stock market valuations to data is likely to produce a large overestimate, even when the value of tangible assets such as information technology infrastructure and buildings is netted out.

Hybrid Method

Wendy C. Y. Li, Makoto Nirei and Kazufumi Yamana (2019) use a hybrid of the cost-based and revenue-based methods to value data, using a method developed by Wendy Li and Bronwyn H. Hall (2018) for valuing R&D expenditures. They argue that the value of data for data-driven firms can be calculated by measuring the value of these firms' business models, or what is referred to in the literature as "organization capital" (originally coined by Edward C. Prescott and Michael Visscher [1980]).

Following earlier studies, Li, Nirei and Yamana (2019) reported selling, general and administrative

expenses from annual income statements for four online platform companies as a proxy for those firms' investment in organization capital. These expenses include "employee training costs, brand enhancement costs, consulting fees, and the installation and management costs of supply chains" (ibid., 22–23). The authors then use this investment data along with revenue data (also from annual income statements) to estimate the capital value of organization capital using a perpetual inventory model. The key identifying assumption is that the (unobserved) rate of return on organization capital is the same as the firm's overall (observed) rate of return — which, in a standard model, would be so as the firm would seek to ensure that the marginal rate of return on all its investments was the same.

It should be noted that, like the market capitalization-based method, this hybrid method can only really be used for firms whose business model is focused on using data. To the extent that, say, Amazon, has organizational knowledge about logistics that is not simply a result of data holdings, this method, which ascribes all expenditures that are not spent on production or R&D to data, will overestimate the value of data per se.

That said, it is interesting to note that the estimated values for data are well below what would be obtained from a market capitalization approach. The authors calculate that Amazon, for example, had organization capital of US\$125 billion in 2017, about 16 percent of its total valuation. For Google, the figure was US\$48.2 billion, about eight percent of its total valuation.

One way to refine these figures is to estimate the impact of a data-driven firm that enters a market where incumbents are not data-driven. The decline in the value of organization capital of these incumbent firms then becomes a proxy for their willingness to pay for the data that the new entrant possesses. This method is adopted by Diane Coyle and Wendy Li (2021), who look at the impact of the arrival of Airbnb on the hotel market and, in particular, the impact on the Marriott hotel chain. The authors find that there was a significant drop in the value of Marriott's organization capital following Airbnb's entry, and based on that drop, they estimate the market size for data in the global hospitality sector to have been US\$43 billion in 2017, about seven percent of the approximately US\$600 billion global hospitality

market in 2018. Of course, this approach relies on the natural experiment of a data-driven new entrant, and so has limited wider applicability.

Conclusions and Directions for Future Research

Since the publication of *The Economist* article referenced at the beginning of this paper, there has been considerable work on valuing data, both quantitative and theoretical. National statistical agencies, propelled in part by the forthcoming revision of the SNA in 2025, have done a significant amount of experimental work, using a traditional cost-based methodology that may well enable data (if not data science) to follow other intangibles, such as databases, software and R&D spending, and be capitalized in the SNA.

Nonetheless, the relatively wide variation in estimates of the value of data flows and stocks shows that there are still some significant problems to be sorted out. First, national statistical agencies will need to decide on an appropriate methodology for identifying data-related occupations: the machine-learning approach of the BEA, while likely more accurate than conventional methods of determining data-related occupations, may prove too opaque for some agencies to be comfortable with. Second, there will need to be agreement on appropriate depreciation rates and service lives, and the proportion of production that is for own account rather than for sale: this will require actual surveys of firms rather than the arbitrary assumptions used by the studies outlined in this paper. Third, more work will need to be done on developing appropriate price indices, particularly agreeing on the right measure of productivity growth by which to deflate price indices; two of the three studies use an entirely arbitrary approach to estimating productivity growth.

What will be the likely magnitude of estimates of data-related assets? At this point, the authors of this paper regard the US study as the most reliable guide, given that it has the most sophisticated approach to determining occupations, and more conservative assumptions for depreciation rates.

This study calculates the value of data-related assets in 2021 at US\$421 billion, equivalent to 1.3 percent of total private fixed assets.

Substantial though this number is, it is significantly less than the value of total IP products (US\$4.18 billion in 2021), and very significantly below the combined market capitalization of just two major data-related companies (Alphabet and Meta), which was US\$2.9 billion in 2020. This discrepancy arises, in part, from the cost-based methodology used by statistical agencies: given that data-related companies may be earning very large rents from their ownership of data, costs and revenues of data assets will not be as closely aligned as they would be in a competitive market where competition would eliminate these rents over time. (Part of this rent comes from the non-rival nature of data, which means it can be sold to many customers simultaneously.)

Unfortunately, alternative and conceptually superior methods of calculating the value of data are difficult to implement in practice. One problem is the absence of open and transparent markets for data. While it is possible to get around this by estimating measures of organization capital, the problem is that this is very much a residual category, which includes not just data but other forms of hard-to-measure intangible capital such as management expertise and accumulated knowledge. While one could argue that the latter is ultimately founded on data, or even data science, these forms of capital are still not what the authors mean by data-related assets and are conceptually separate. Ultimately, it may well be the case that the bulk of the value of a company such as Alphabet or Meta is bound up in what one might call “intangible intangibles” rather than “tangible intangibles” such as data and data science, or patents and software.

What are the next steps for research in this area? From an SNA perspective, the most pressing is probably to establish reasonable depreciation rates and service lives for data. Without international agreement on this issue, we are likely to get dramatically different capital stock estimates for different countries.

The next issue is one of scope. Measures of data-related assets will need to be developed for the public sector, as they have been for other IP products. This could be done, as it was in the Canadian study, using the cost-based method: as

most public data is free, the revenue or market-based approaches are unsuitable. However, as with private sector data, the challenge is that the value of public sector data is likely well above what it costs to produce. One way of getting this information would be through survey or experimental data on users' willingness to pay: the UK Office for National Statistics has done significant work along these lines (see Cope, n.d., quoted in Slotin [2018], and Office for National Statistics [2021]), and this approach could be taken further.

This leads to the broader issue of the social valuation of data: Are there positive or negative externalities associated with data that are not accounted for by private sector actors? On the positive side, investment in data and data science, such as R&D, may generate new insights and ways of doing things that cannot be entirely hidden from competitors, thereby generating positive spillovers to the rest of the economy. In this case, the social value of data would exceed the private value. It would be useful to estimate a model of economic growth that would capture the extent of these spillovers by estimating the impact of investment in data by one firm on the productivity of other firms in the industry, just as studies have attempted to do for R&D.

On the other hand, there may also be negative spillovers associated with data. When an individual's personal data is correlated with other individuals' personal data, then if the individual shares their data, they compromise not only their own privacy but also the privacy of all the other individuals with correlated data. Daron Acemoglu et al. (2019) argue that this negative externality leads to excessive data sharing (individuals are less conservative with allowing their personal data to be collected) and results in an underpricing of personal data. They explain that this externality is why individuals are often willing to divulge their data without any sort of financial remuneration (ibid.). They also argue that this makes individuals more willing to share their data: if their personal information can already be assembled from other people's data, what incentive do they have to protect their own privacy? Assessing the costs of this negative externality would be another direction for future research.

Discussion of externalities raises the question of data governance. As Diane Coyle and Stephanie Diepeveen (2021) note, if the private and public valuations of data differ, then there is a potential

role for government in trying to influence the current "market" for data (for individuals, this market is largely implicit: they give up their data in exchange for services provided by free apps). It is not sufficient to simply try to create a market for data by, for example, clearly establishing individuals' property rights over data, because the resulting prices will be suboptimal if they do not take into account externalities. This is an area where more policy work is clearly required: to quote Dan Ciuriak (2019), "there is no good historical analogue for a regulatory framework to control the negative externalities to which [data] gives rise or to equitably share the benefits that it generates." Depending on the nature of the externalities, regulatory tools could include restricting the kinds of data individuals can share, requiring open data access by companies that collect data, and creating ways for individuals to band together in order to bargain on an equal footing with firms that collect their data. Assessing the social value of data will be an essential underpinning of the policy work that will be required to effectively govern the data space in the years to come.

Acknowledgement

The authors would like to thank CIGI Managing Director of Digital Economy Robert Fay, CSLS Executive Director Andrew Sharpe and two anonymous reviewers for helpful comments. Comments are welcome and may be sent to tim.sargent@csls.ca.

Works Cited

- Acemoglu, Daron, Ali Makhdomi, Azarakhsh Malekian and Asuman Ozdaglar. 2019. "Too Much Data: Prices and Inefficiencies in Data Markets." National Bureau of Economic Research Working Paper Series. <https://doi.org/10.3386/w26296>.
- Ahmad, Nadim. 2005. "Follow-Up to the Measurement of Databases in the National Accounts." Issues paper for the AEG July 2005 – SNA Update Issue 12. <https://unstats.un.org/unsd/nationalaccount/aeg/papers/m3Databases.PDF>.
- Ahmad, Nadim and Peter van de Ven. 2018. "Recording and measuring data in the System of National Accounts." Paper prepared for the Meeting of the OECD Informal Advisory Group on measuring GDP in a digitalised economy, November 9. SDD/CSSP/WPNA(2018)5. https://unstats.un.org/unsd/nationalaccount/aeg/2018/M12_3c1_Data_SNA_asset_boundary.pdf.
- Calderón, José Bayoán Santiago and Dylan G. Rassier. 2022. "Valuing the U.S. Data Economy Using Machine Learning and Online Job Postings." BEA Working Paper Series, WP2022-13.
- Ciuriak, Dan. 2019. "Data as a Contested Economic Resource: Framing the Issues." Paper prepared for Data and Digital Intelligence as People's Resources: Reclaiming Freedom and Control in a Data-based Society, Berlin, Germany, November 23–24. <https://doi.org/10.2139/ssrn.3496281>.
- Commission of the European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank. 1993. *System of National Accounts 1993*. New York, NY: United Nations. <https://unstats.un.org/unsd/nationalaccount/docs/1993sna.pdf>.
- Cope, Ian. n.d. "The value of census statistics." Office for National Statistics. <https://dam.ukdataservice.ac.uk/media/455474/cope.pdf>.
- Coyle, Diane and Stephanie Diepeveen. 2021. "Creating and governing social value from data." Paper prepared for the IARIW-ESCoE Conference, Measuring Intangible Capitals and their Contribution to Growth, London, UK, November 11–12. <https://doi.org/10.2139/ssrn.3973034>.
- Coyle, Diane and Wendy Li. 2021. "The Data Economy: Market Size and Global Trade." Paper prepared for the 36th IARIW Virtual General Conference, August 23–27. <https://doi.org/10.2139/ssrn.3973028>.
- Coyle, Diane and Annabel Manley. 2022. "What is the value of data? A review of empirical methods." Policy Brief. Bennett Institute for Public Policy Cambridge. July 17. www.bennettinstitute.cam.ac.uk/publications/value-of-data/.
- de Bondt, Hugo and Nino Mushkudiani. 2021. "Estimating the Value of Data in the Netherlands." Paper prepared for the IARIW-ESCoE Conference, Measuring Intangible Capitals and their Contribution to Growth, London, UK, November 11–12. https://iariw.org/wp-content/uploads/2021/10/bondt_paper.pdf.
- European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank. 2009. *System of National Accounts 2008*. New York, NY: United Nations. <https://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf>.
- Girard, Michel, Michael Lionais and Rob McLean. 2021. "What Is Your Data Worth? Insights for CPAs." Toronto, ON: Chartered Professional Accountants Canada. March. www.cpacanada.ca/en/foresight-initiative/value-creation/data-value-primer-whats-your-data-worth.
- Goodridge, Peter, Jonathan Haskel and Harald Edquist. 2022. "We See Data Everywhere Except in the Productivity Statistics." *Review of Income and Wealth* 68 (4): 862–94. <https://onlinelibrary.wiley.com/doi/full/10.1111/roiw.12542>.
- Haskel, Jonathan and Stian Westlake. 2017. *Capitalism Without Capital: The Rise of the Intangible Economy*. Princeton, NJ: Princeton University Press.
- Ker, Daniel and Emanuele Mazzini. 2020. "Perspectives on the Value of Data and Data Flows." OECD Digital Economy Papers No. 299. December. Paris, France: OECD Publishing. <https://doi.org/10.1787/a2216bc1-en>.
- Koutroumpis, Pantelis, Aija Leiponen and Llewellyn D. W. Thomas. 2020. "Markets for data." *Industrial and Corporate Change* 29 (3): 645–60. <https://academic.oup.com/icc/article/29/3/645/5804957>.
- Li, Wendy C. Y. and Bronwyn H. Hall. 2018. "Depreciation of Business R&D Capital." National Bureau of Economic Research Working Paper No. 22473. www.nber.org/papers/w22473.
- Li, Wendy C. Y., Makoto Nirei and Kazufumi Yamana. 2019. "Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy." BEA Working Paper. www.bea.gov/research/papers/2018/value-data-theres-no-such-thing-free-lunch-digital-economy.

- Mitchell, John, Molly Leshner and Marion Barberis. 2022. "Going Digital Toolkit Note: Measuring the economic value of data." Organisation for Economic Co-operation and Development. January 3. DSTI/CDEP/GD(2021)2/FINAL. [https://one.oecd.org/document/DSTI/CDEP/GD\(2021\)2/FINAL/en/pdf](https://one.oecd.org/document/DSTI/CDEP/GD(2021)2/FINAL/en/pdf).
- Murphy, Andrea. 2022. "America's Largest Private Companies 2022: Twitter And Continental Resources Join The Ranks." *Forbes*, December 1. www.forbes.com/sites/andreamurphy/2022/12/01/americas-largest-private-companies-2022-twitter-and-continental-resources-join-the-ranks/?sh=7cbf006b34c7.
- Office for National Statistics. 2021. "Valuing official statistics with conjoint analysis: April 2021." www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/valuingofficialstatisticswithconjointanalysisapril2021.
- Prescott, Edward C. and Michael Visscher. 1980. "Organization Capital." *Journal of Political Economy* 88 (3): 446–61.
- Rassier, Dylan G., Robert J. Kornfeld and Erick H. Strassner. 2019. "Treatment of Data in National Accounts." Paper prepared for the BEA Advisory Committee. May. www.bea.gov/system/files/2019-05/Paper-on-Treatment-of-Data-BEA-ACM.pdf.
- Slotin, Jenna. 2018. "What do we know about the value of data?" *Global Partnership for Sustainable Development Data Blog*, May 3. www.data4sdgs.org/blog/what-do-we-know-about-value-data.
- Statistics Canada. 2001. "Capitalization of Software in the National Accounts." www150.statcan.gc.ca/n1/pub/13-605-x/2003001/concept/2001software-2001logiciel/index-eng.htm.
- . 2019a. "Measuring investment in data, databases and data science: Conceptual framework." June 24. www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00008-eng.htm.
- . 2019b. "The value of data in Canada: Experimental estimates." July 10. www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00009-eng.htm.
- The Economist*. 2017. "The world's most valuable resource is no longer oil, but data." *The Economist*, May 6. www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.
- United Nations Economic and Social Council. 2022. "DZ. 6 – Recording of data in the National Accounts." https://unstats.un.org/unsd/nationalaccount/RAdocs/DZ6_GN_Recording_of_Data_in_NA.pdf.

**Centre for International
Governance Innovation**

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

🐦 @cigionline