Australian Bureau of Statistics

Economic Implications of the Digital Economy Conference

9-10 March 2022

Valuing data as an asset, implications for economic measurement

Michael Smedes

Thai Nguyen

Bode Tenburren

1. Introduction

1. Intangible assets are increasingly viewed as providing a strong competitive advantage for business and as being central to creating consumer value. Federal Reserve Bank of Philadelphia calculations put a value on intangible assets held by US corporations as being equivalent to roughly half the market capitalisation of the S&P 500 index (Nakamura 2008). Despite this importance, intangible assets are largely absent from company balance sheets due to the difficultly in determining 'fair' value for these assets.

2. Data is one type of intangible asset whose use has increased exponentially due to the ease with which it can be captured, stored, and analysed thanks to the digital revolution. Yet despite this apparent trend, data have little visibility in business accounts and economic statistics. This is because data usage is to a large extent unpriced in the modern economy.

3. This paper aims to help improve this situation by expanding current national accounting concepts and statistical methods for measuring the value of data as an asset. The paper does this by testings a range of assumptions and providing indicative estimates on the value of data assets in the Australian economy. This work forms part of international efforts by statistical offices to better inform the digital economy in the lead-up to a proposed 2025 revision to the System of National Accounts.

2. What is Data?

4. The term "data" has become ubiquitous in discussions across all fields - including in policy debates, business performance, and academic literature. The term is used flexibly often without a clear understanding of what is being discussed.

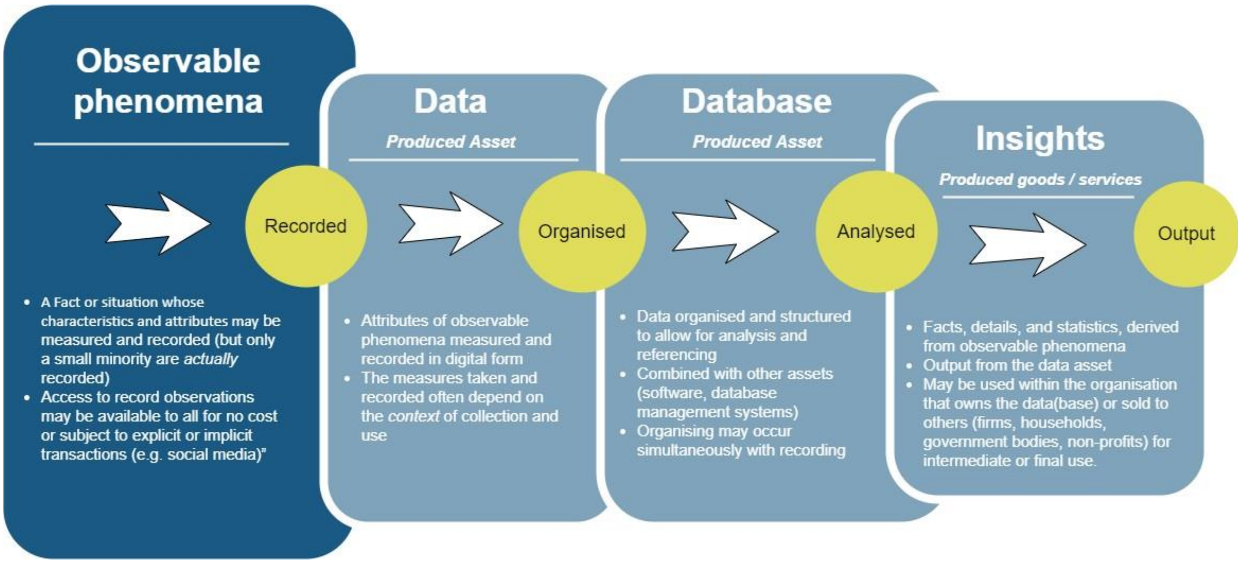
5. References to "data" can be made with respect to, amongst other things: vast datasets containing large numbers of observations; statistics and indicators derived from these observations; scientific studies of a specific phenomena; or curated observations collated to drive decision making processes.

6. While these examples all involve, at some level, facts collected together, they can differ in important ways: 1) the volume and nature of productive activities involved in their creation, 2) the amount of information that they contain, and 3) their usefulness to the broader community and potential monetary value to their owner. (OECD, 2021)

7. The lack of an agreed definition of data has been one of the key challenges in determining an approach for valuing "data". This paper uses a working definition of data proposed by the Advisory Expert Group on National Accounts explicitly for the purpose of including the production and value of data in economic statistics:

Information content that is produced by accessing observable phenomena and recording, organising and storing relevant information elements from these in a digital format, which can be accessed electronically for reference or processing" (OECD, 2021)*.*

Figure 1. Data-information chain from a System of National Accounts perspective



Source: OECD 2021.

8. Data, like most intangible assets, have economic characteristics which can differ from tangible assets, and which have implications for measurement in economic statistics. A range of these characteristics (taken from the work of *Coyle et al 2020* and *OECD 2021*) are set out below:

- **Data are non-rival.** Many agents can make use of the same data at the same time without it being "used up" or degraded. This means the ownership of data and data-related transactions exhibit important differences compared to those for typical goods and services.
- **Data can be excludable.** It is likely hard to exclude agents from data that is relatively easy to collect, such as data scraped from websites, whereas client or administrative data is likely to have tightly restricted access.
- **Data involve externalities.** Positive externalities can arise when datasets are combined, making the "sum of greater value than the parts". By contrast, negative externalities arise when the collection or use of data leads to harm (e.g. exclusion from healthcare coverage). Strong incentives to use data intensively can undermine privacy, for example. At the same time, weak incentives to use data can lead to missed opportunities for generating economic and social value. Data can also arise as an externality, often as a by-product of the standard production process.
- **Data may have increasing or decreasing returns to scale.** Sometimes collecting more data can provide additional insights, at other times more data adds little extra value (and likely some cost)
- **Data has a large option value.** It is hard to predict how the value of data might change. New data, new technologies or algorithms, and new measurement methodologies mean existing data can have unpredictable future importance. Organisations may keep data for their potential rather than current value.
- **Data collection may have high up-front cost and low marginal cost**. Collecting data can entail investment in hardware (such as sensors) and software, among other costs, but on-going collection can be cheap. This can create barriers to beginning to collect data that

would be useful or potentially financially lucrative. Data use requires complementary and often on-going expenditures. Organisations may need to invest in complementary hardware and software to be able to process and analyse the data.

- **Data about people have particular features.** Information about one person will often have nomeaningful value to an organisation but may well be highly valued in principle by the individual. People create positive and negative incentives and impacts for others as they share information about themselves (e.g. through social networks or genetic analysis services). Furthermore, personal data are generally subject to additional legal controls, with compliance costs acting as another potential barrier to data collection.

9. Taken together these characteristics typically lead to a situation where the costs and risks associated with collection and use of data are much clearer than are the benefits of having and using the data. This situation has major implications for the approach to valuing data and for the measurement options available. Based on these characteristics it would appear more practical to value data based on costs of production rather than anticipated future revenue for example.

3. The Role of Data in the Production Process

10. An assumption that data is an economic asset underpins this paper, however this position is far from settled within the economic measurement community and there remain a range of threshold issues under discussion.

11. The role data plays within the production functions of business is perhaps the most fundamental of these issues. To be considered an economic (fixed) asset data must be used repeatedly in the process of production for more than a year. However engagement with companies that extract value from consumer behaviour data indicate that the utility of data declines rapidly and that it becomes effectively obsolete in an extremely very short timeframe, unless it is 'replenished' with more recent observations.

12. This need to constantly 'replenish' data may call into question its classification as an economic asset providing a repeated flow of capital services into an entities' production function. Instead, some argue that data is more like electricity, a new input revolutionising production processes but an input which is inherently consumption rather than capital by nature.

13. Alternatively, this challenge in classifying data as an economic asset may indicate that current definitions and typology underpinning economic theory and measurement have failed to keep pace with technology change, the growth of the knowledge economy, and the rise of intangible products. See for example *Hill 1999.*

14. The ABS will continue to engage with academics and the international statistical community in discussions on the appropriate treatment and is keen to work with private sector businesses to better understand practical applications of data in their business models and production functions.

15. In the meantime, the calculations that follow assume data is an economic asset providing a flow of services which are used (across a period greater than one year) in production functions.

4. Valuing Data-Related Capital Formation

16. Placing a price on products within the System of National Accounts typically relies on the existence of a market established price for that product. Such a market price is not observable for data. Most commercial data is internally generated, and where data is involved in a monetary transaction it is usually a data derived insight which is sold rather than the data itself, for example advertising targeting. There are non-monetary data transactions which occur (for example someone trading their personal location information in exchange for mapping and other services), but it is extremely challenging to place a monetary value on these transactions.

17. The International Accounting Standards Board have similarly found it challenging to determine an approach for valuing data which meets their 'fair value' test, and as a result the value of data is almost never recorded on company balance sheets.

18. A standard valuation technique used in economic statistics when a product is not sold on the market is to employ a sum of costs approach. Under this approach the value of the asset is represented by the sum of the costs of the inputs used to create the asset plus an appropriate rate of return. Statistics Canada have pioneered a sum of costs approach to valuing data which the calculations below initially replicate, and then build on.

19. In the approach first set out by Statistics Canada (2019) the value of data is estimated with reference to the labour costs incurred in their production plus associated non-direct labour and other costs, such as the costs of the associated software, hardware, electricity, building and telecommunications services.

20. Occupational groups were selected from among those in the occupational classification that are generally associated with converting observations into (useable) digital stores. Employees working in these categories are unlikely to devote all their work time to producing data so subjective assumptions on the proportion of time spent producing data were made. In view of the uncertainty associated with these assumptions, two alternatives were considered. They are labelled 'lower' and 'upper' in the tables below.

21. The five yearly Census of Population and Housing, provides good quality employment and earnings statistics by occupation and was used for these calculations. This information is used for each of the census years 2006, 2011 and 2016.

22. It is assumed that non-direct salary and other intermediate input costs represent 50% of the salary costs. Another markup of 3% is added to this margin for the cost of capital. This combination of labour costs, intermediate inputs, and capital costs provide a sum of costs estimate of the value of data-related activities.

23. Total gross fixed capital formation in each component of the information chain is presented in Table 1. Total investment is between $35 billion and $47 billion in 2016 at current prices. The growth between 2006 and 2016 is 80 - 85%, or 6.5 – 6.8% on an average annual basis.

24. Please note these amounts cannot be directly added to existing estimates of gross domestic product, since they overlap to a degree with components in the already the published estimates. However it does provide an indicative estimate of the increase to GDP that may result from the inclusion of data assets within the National Accounts. Currently published estimate for GDP (current prices) in 2016 is approximately $1.65 trillion, the lower and upper values calculated in this exercise work out at around 2.2% and 2.8% of GDP in 2016.

Table 1. Estimates of Data related Capital Formation (current price)

	2006	2011	2016
	millions of AUD		
Total of all data-related categories			
lower range value	**19,790**	**25,951**	**35,512**
upper range value	**25,543**	**33,682**	**46,728**
Data			
lower range value	8,755	11,339	14,193
upper range value	11,942	15,533	19,851
Databases			
lower range value	3,323	4,549	6,698
upper range value	4,448	6,188	9,486
Data Science			
lower range value	7,712	10,064	14,622
upper range value	9,153	11,961	17,390

Table 2. Annual growth rates for Data related Capital Formation (current price)

	2006	2011	2016
Annual growth rate (all data-related categories)			
lower range value		5.6%	6.5%
upper range value		5.7%	6.8%
Annual growth rate (Data)			
lower range value		5.3%	4.6%
upper range value		5.4%	5.0%
Annual growth rate (Databases)			
lower range value		6.5%	8.0%
upper range value		6.8%	8.9%
Annual growth rate (Data science)			
lower range value		5.5%	7.8%
upper range value		5.5%	7.8%

Table 3. Share of total Gross Fixed Capital Formation (current prices)

	2006	2011	2016
Total of all data-related categories			
lower range value	**7.1%**	**7.0%**	**8.5%**
upper range value	**9.2%**	**9.1%**	**11.1%**
Data			
lower range value	3.2%	3.1%	3.4%
upper range value	4.3%	4.2%	4.7%
Databases			
lower range value	1.2%	1.2%	1.6%
upper range value	1.6%	1.7%	2.3%
Data Science			
lower range value	2.8%	2.7%	3.5%
upper range value	3.3%	3.2%	4.1%

5. Valuing Data-Related Capital Stock

25. Taking the next step and calculating capital stock estimates requires:

- annual estimates of capital formation in current prices
- estimates of price change to calculate constant price estimates
- estimates of asset life to depreciate asset values over time

Annual capital formation estimates

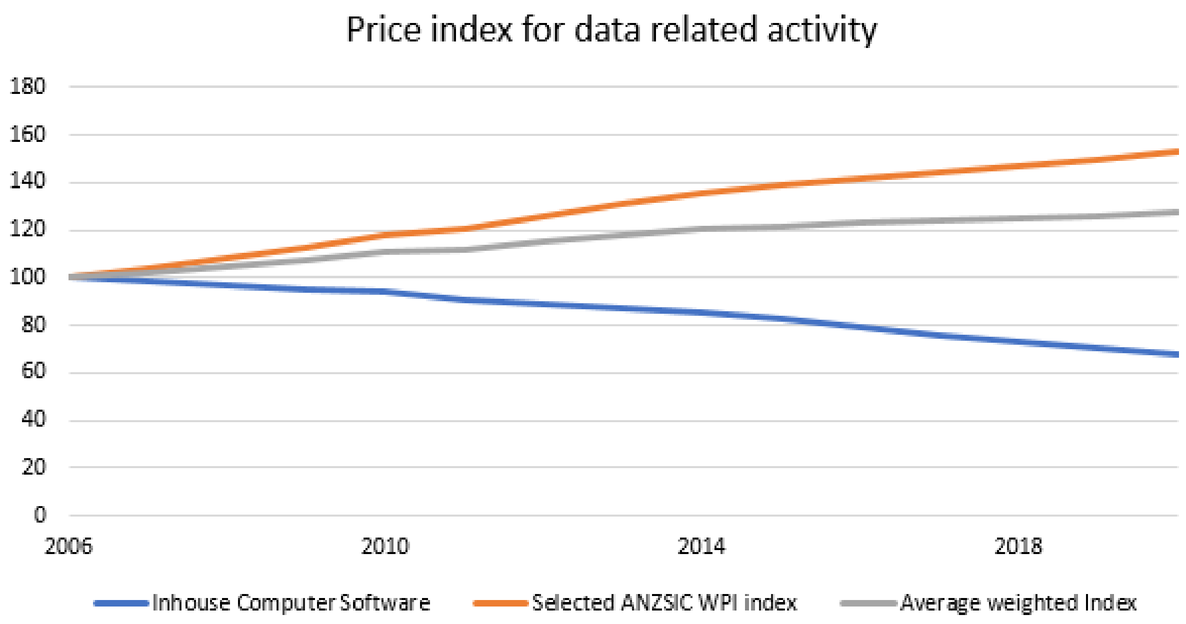
26. Labour Force Survey and Survey of Employee Earnings data were used as an indicator to interpolate between the five-yearly Census benchmarks, providing an annual estimates of capital formation. This was in line with the approach taken by Statistics Canada.

Data asset price change

27. Statistics Canada applied a labour price index to deflate their current price values and calculate constate price estimates. However, the usual practice when deflating a value calculated using the sum of costs approach is to weight together price indexes for the component cost series. Given the increasing wage rates for data professionals and the falling prices for software and other ICT equipment used the data production process this provides a quite different result shown in graph 1 below:

- The orange line uses the Wage Price Index for selected industries to capture the wage growth paid to employees working in this space as demand for employees with these skills increases to be able to “accessing and observe phenomena; and recording, organizing and storing information elements”
- The blue line uses software prices to capture the technology improvements in the non-wage costs e.g. it is cheaper for a company to record, organize and store information elements due technology improvements.
- The grey line is a cost weighted combination of the two

Chart 1 – Price Index Sensitivity



28. Deflating using the cost weighted price index results in an average annual growth rate for constant price data assets of just over 5% a year across the time series.

Asset Life

29. Statistics Canada assumed data have a useful life of 25 years, "since much data that are currently used are behavioural, it can be assumed such data will retain their value for a generation" (Statistics Canada 2019).

30. While agreeing that much (most) data currently used are for the purposes of understanding behaviour, the authors understand that most of the commercial value being extracted from data is linked to economic behaviour and consumer preferences. And that for this type of use a 25-year data asset life is too long.

31. To understand the useful life of data related to consumer preferences we looked to work published by the ABS in the *Information Paper: Increasing the Frequency of CPI Expenditure Class Weight Updates, July 2016*. To reduce the bias in the CPI arising from consumers substituting between different combinations of products the frequency of reweighting the CPI was increased from from every six years to annually. This change was supported by the empirical evidence in the paper and suggested a more appropriate mean asset life of 3 years and maximum asset life of 5 years.

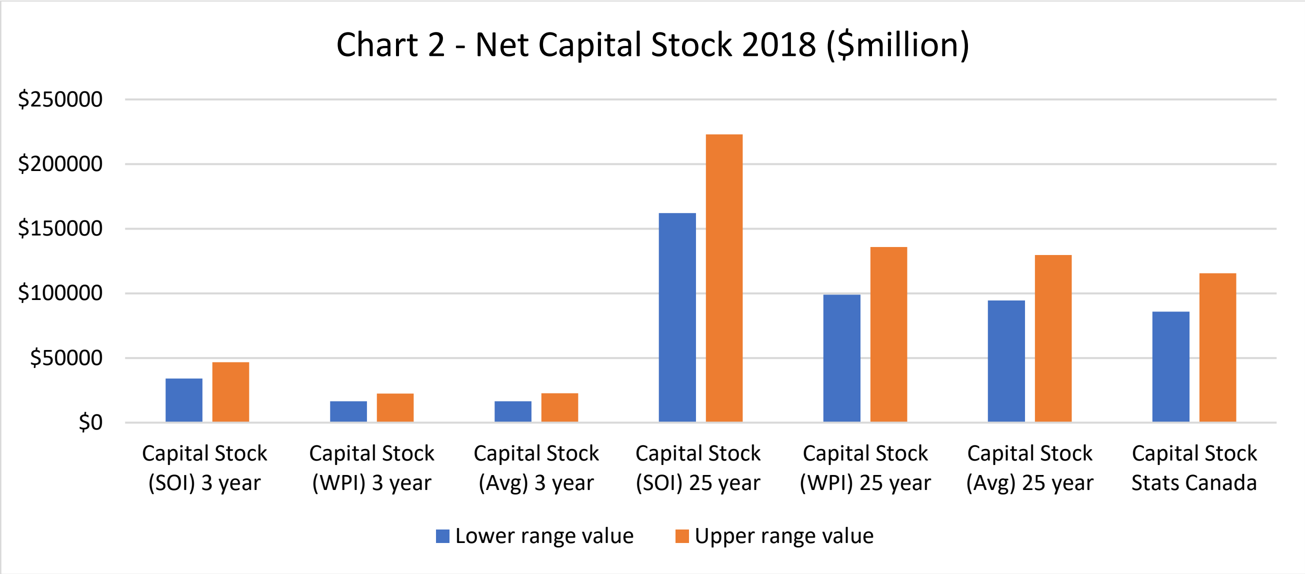
32. To give some context for these asset lives, the mean asset lives used for other intangible assets are: research and development 11.3 years; purchased software 4.5 years; music originals 3 years; and film and TV originals 3.5 years.

33. To understand the impact on capital stock estimates we tested the two different assumptions:

- The first using the Statistics Canada assumption that a firm is expected to draw upon data to gain 'generational' insights for 25 years
- The second based on the assumption most value in data exploited now is linked to consumer preferences. Giving us a mean asset life of 3 years and a maximum asset life of 5

years

Chart 2 – Asset Life and Price Index Sensitivity Testing Results



34. Clearly the large difference in the assumed asset lives (3 years vs 25 years) lead to very different capital stock values, but the choice of price index between software (SOI), wages (WPI) and the weighted average (Avg) is also material to the outcome.

35. The value of intangible assets currently on the national accounts balance sheet (intangible assets currently included consist of R&D, software, mineral & petroleum expenditure, and artistic originals) was roughly $250b in 2018. The upper bound of this sensitivity test (25 year asset life, deflated by a software index) would increase that value by around 90%. The lower bound (3 year asset life, deflated with a wage index) would increase that value by only 6.6%.

36. The inclusion of these data assets values would result in only a small increase in the value of total net capital stock which in 2018 was valued at approximately $6.2 trillion, if a mean asset life of 3 years is used the increase in net capital stock would be a fraction of a percent.

6. Conclusion

37. This paper provides an estimate for the value of data assets using, in the absence of market prices or business accounting values, a sum of costs approach. Given the uncertainties implicit in this valuation, several assumptions and ranges are tested.

38. The results of this exercise indicate that valuing data as an asset within the national accounts may increase GDP in the order of 2% and would have little impact on the value of capital stock (dependent on the asset life chosen).

39. These estimates have only been calculated at the total economy level and there are as yet no industry or sector level estimates available. However, based on these results it would appear unlikely that including data assets within the accounts would explain the much-discussed productivity ‘puzzle’.

40. The ABS intends to engage with private sector data companies to sense check the assumptions and the approach used in these calculations. Further refinement of the approach will also lead to industry level estimates and, eventually, estimates of the impact of productivity.

References

BEA (2019), *Treatment of Data in National Accounts*, https://www.bea.gov/system/files/2019-05/Paper-on-Treatment-of-Data-BEA-ACM.pdf.

Coyle, D. et al. (2020), *The value of data: Policy Implications*, OECD Publishing, Paris, https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_Policy_Implications_Report_26_Feb_ok4noWn.pdf.

Coyle, D. and W. Li (2021), *The data economy: Market size and global trade*, OECD Publishing,Paris, https://www.escoe.ac.uk/publications/the-data-economy-market-size-and-global-trade.

European Commission, International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), United Nations and World Bank (2009), *System ofNational Accounts, 2008*, United Nations, New York, https://doi.org/10.5089/9789211615227.071.

Eurostat-OECD (2020), *Joint Eurostat-OECD Task Force on Land and Other Non-financialAssets*, OECD Publishing, Paris, https://ec.europa.eu/eurostat/documents/24987/725066/Eurostat-OECD+Report+on+Intellectual+Property+Products.pdf.

Forrester, J. (2020), *What's The Difference Between Private Weather Companies And The National Weather Service?*, Forbes, https://www.forbes.com/sites/jimfoerster/2020/01/24/whats-the-difference-between-private- weather-companies-and-the-national-weather-service/.

Furman, J. (2019), *Unlocking Digital Competition*, OECD Publishing, Paris, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf.

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

Hill, P. (1999), "Tangibles, intangibles and services: a new taxonomy for the classification of output" Economic Commission for Europe

ISWGNA Advisory Expert Group on National Accounts (2020), *Recording and Valuation of Datain National Accounts*, OECD Publishing, Paris, https://unstats.un.org/unsd/nationalaccount/aeg/2020/M14_5_3_1_Recording_of_Data_in_SNA.pdf.

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

Li, W. and P. Chi (2021), *Online platforms' creative "disruption" in organizational capital - The cumulated information of the firm*, OECD Publishing, Paris, https://iariw.org/wp-content/uploads/2021/08/LiChi_paper.pdf.

Nguyen, D. and M. Paczos (2020), "Measuring the economic value of data and cross-borderdata flows: A business perspective", *OECD Digital Economy Papers*, No. 297, OECD Publishing, Paris, https://dx.doi.org/10.1787/6345995e-en.

OECD (2021), *Draft Recommendation of the Council on Enhanced Access to and Sharing ofData*, internal document [C(2021)96].

OECD (2021), *Issues paper: Recording observable phenomena and data in the nationalaccounts*, OECD Publishing, Paris, https://unstats.un.org/unsd/nationalaccount/aeg/2021/M15_7_4_Recording_Data.pdf.

OECD (2019), *An Introduction to Online Platforms and Their Role in the Digital Transformation*, OECD Publishing, Paris, https://dx.doi.org/10.1787/53e5f593-en.

OECD (2019), *Going Digital: Shaping Policies, Improving Lives*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264312012-en.

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

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

Statistics Canada (2019), *The Value of Data in Canada: Experimental Estimates*

Statistics Canada (2019), *Measuring Investment in Data, Databases and Data Science: Conceptual Framework*

US Federal Reserve Bank of Philadelphia (2008) WORKING PAPER NO. 08-23 INTANGIBLE ASSETS AND NATIONAL INCOME ACCOUNTING Leonard I. Nakamura

ONS (2019), *Impact of Blue Book 2019 changes on gross fixed capital formation and business investment*, OECD Publishing, Paris, https://www.ons.gov.uk/releases/nationalaccountsarticlesimpactofbluebook2019changesongrossfixedcapitalformationandbusinessinvestment. [26]

Squicciarini, M. and M. Le Mouel (2012), "Defining and Measuring Investment in Organisational Capital: Using US Microdata to Develop a Task-based Approach", *OECD Science, Technology and Industry Working Papers*, No. 2012/5, OECD Publishing, Paris, https://dx.doi.org/10.1787/5k92n2t3045b-en. [16]

Statistics Canada (2019), *The value of data in Canada: Experimental estimates*, OECDPublishing, Paris, https://dx.doi.org/10.1787/5k92n2t3045b-en. [19]