IARIW – CIGI 2023

IARIW – CIGI 2023

Thursday, November 2 – Friday, November 3

Toward Data as an Asset in the System of National Accounts

John Mitchell
(OECD)

Nicola Massarelli
(Eurostat)

Richard Heys
(ONS)

Erich Strassner
(IMF)

Paper prepared for the Conference on The Valuation of Data November 2 – November 3, 2023
Session 1: Methodologies of Data Valuation: National Accounts Approaches
Time: Thursday, November 2, 2023 [10:00-11:30AM EST]

Toward Data as an Asset in the System of National Accounts.

John Mitchell[1] , Richard Heys (U.K. ONS), Nicola Massarelli (Eurostat), and Erich H. Strassner (IMF)[2]

The availability and prevalence of data has given rise to new or significantly improved products, services, and business models. Across all industries and sectors, from agriculture to transport, government and private enterprise, data contributes to more efficient uses of resources. However, despite the omnipresence of data within the economy, the System of National Accounts, 2008 (SNA2008) does not explicitly identify data as a standalone asset. Since the relevance of the SNA depends on its ability to accurately measure the economy including its evolution, the absence of an explicit data asset in the accounts, and its effect on estimates of output and capital stocks is becoming increasingly untenable. However, while there are many different ways that one can assign a value to data, not all of them are compatible with the overriding valuation principles used within the System of National Accounts. The SNA is currently under revision, with an updated standard to be approved by 2025. A part of this revision includes the incorporation of data into the production and asset boundaries of the national accounts, thereby acknowledging that data is a produced asset. Initial feedback from countries has shown that there is strong support for this position. A task team, comprising members from both international and national statistical offices, have developed a potential data measurement framework as part of an SNA guidance note on the subject. The guidance note includes concepts and definitions that would allow for the recording of data in a way that is not only comparable across countries but importantly is consistent with the concepts and accounting rules that currently exist in the national accounts. This paper summarises the recommendations contained within the guidance note as well as highlights the specific areas requiring further research. Furthermore, it provides an overview of the initial attempts by specific countries to compile estimates consistent with the proposed framework.

[1] Consultant to the Joint Eurostat – IMF Task Team on Measuring Data as an Asset in National Accounts, formerly of the OECD and ABS.

[2] The views expressed herein are those of the author(s) and should not be attributed to the IMF, its Executive Board, or its management.

Introduction

Data[3] has become fundamental to businesses across the economy. The ability to collect and organise huge amounts of information, resulting in the production of massive amounts of data is not only a fundamental aspect of many business models but is also a quintessential example of digitalisation impact on the economy. While the practice of collecting and organising information is not new, the digitisation of so many facets of the economy has made the collection of facts and scenarios (labelled in this paper as observable phenomena) much easier and cheaper than ever before. This in turn has resulted in the creation of brand-new business models reliant on data, while increasing the risk that many traditional enterprises will be left behind if they are not creating data about their customers or production processes.

Despite the omnipresent impact of data on the production occurring in the economy there is currently a 'data shaped' gap in GDP and the national accounts more broadly. While the 2008 System of National Accounts (SNA) (UNSD, 2009) currently hints at the presence of data, acknowledging its existence in several roundabout ways, it currently falls short of acknowledging data as an explicit input into production and asset on its own. Considering that GDP is often used as a synonym for the economy itself, this has the potential to erode the value and trust that users place in these figures if they feel that they are not properly reflecting the modern economy.

The SNA is currently being revised, with publication of an updated SNA set for 2025. The revision encompasses solutions to a broad range of conceptual challenges including, but not limited to, better recognition of environmental degradation, more emphasis on the distribution of income and general conceptual changes required to match an economy that has moved on from 2009, when the most recent version of the SNA was published. An important part of this revision is the inclusion of data as a stand-alone produced asset.

This paper begins by explaining the current treatment of data in the SNA and why this has become untenable, before moving on to a range of recommendations that are being put forward to incorporate data into the SNA production and asset boundary. It finishes with an inventory of initial estimates of data investment produced by national statistical offices and academia, which are broadly consistent with the proposed recommendations and can be used as a starting point for other countries to compile their own estimates of data.

The concept of measuring data and why it is needed in the System of National Accounts

Rightly or wrongly, Gross Domestic Product (GDP) is the most heavily used and sighted indicator of economic activity. GDP is of course just a single indicator of the much broader and comprehensive System of National Accounts (SNA). This framework has been updated on several occasions to reflect

[3] For this paper, Data will be used for both the singular and plural. While there exists perfectly good arguments for why data should only be considered as a plural, the lead author of this paper has long argued that the since the English language is continually evolving, both are acceptable, a viewpoint pleasingly supported recently by the Economist (The Economist, 2022).

developments in the economy[4]. These updates have assisted GDP in maintaining its relevance by continually altering the asset and production boundary to reflect changes in the economy. The changes are a fundamental part of the SNA's quest to appropriately measure the input into, and output from, production as undertaken today. In the forthcoming change, Data will be added as an explicit asset class.

In some ways the SNA is perhaps playing "catch up" to other efforts to measure the value of data. Since data has become so important to the economy, it is logical that businesses, government and academics have undertaken efforts to better understand it, including its value to firms and society. Just a few examples include estimating data based on the market capitalisation of the producer of the data (Ker & Mazzini, 2020), the consumer surplus created by its use (Soloveichik, 2023), the value of venture capital in data intensive businesses (Ker & Mazzini, 2020), tracking actual transactions in personal data, and revenue from business that sell data (OECD, 2022). Additionally, there is a large amount of material that creates estimates based on willing-to-accept (WTA) or willing-to-pay (WTP) surveys. While these surveys are usually undertaken to value the free services that people receive from data gathering platforms, the free service is often in return for the information elements contained within the observation phenomena they provide, allowing for a quasi-value to be placed on the data collected.

Finally, there are a range of ways that data can provide value to society. These benefits are often valued via impact-based methods. As outlined in (Slotin, 2018), these typically exploit natural experiments, and as such provide information for policy makers but are of limited use to economic statistics since they are usually far removed from "market prices" and are limited to data used for a specific purpose in a narrow context rather than all data in the economy[5].

Much of this work has been undertaken without the restrictions of the SNA framework, allowing for different valuation techniques and definitions on the actual data they are measuring. While these many different efforts to value data are useful for assisting with policy development, improving insight, and informing enquiring minds on what choices to make, they are usually unable to be transferred and used within the SNA. A cornerstone of the SNA is that it is undertaken based on specific guidelines and accounting rules. If a value is to be included within the SNA production or asset boundary, it must be created in adherence to these rules. Valuations of fixed assets created via subjective valuations, equity prices, or consumer surplus are not consistent with the established SNA guidelines and are not compatible with the established values placed on other fixed assets used in production.

This compatibility with other fixed assets and inclusion in the SNA is important to produce estimates of capital stock and the productivity estimates that follow. So, while the benefits of including data within the SNA asset boundary extend well beyond productivity statistics, this is one very clear and tangible benefit. That said, early investigations on the impact of a potential data asset into the productivity statistics have shown it to have a relatively minor impact (Goodridge, Haskel, & Edquist, 2022), although some authors find that a focus on proprietary big data has contributed a negative effect on overall productivity (Corrado, Haskel, Iommi, Jona-Lasinio, & Bontadini, 2023). Overall, its very inclusion will assist in removing potential ambiguity caused by its exclusion.

[4] As pointed out by (Oulton, 2018) while the original version published in 1953 was just 48 pages, revisions in 1968, 1993 and 2008 have increased its length to 662 pages, which is reflective of the complexities of the modern economy as well as improvement to the accuracy of the framework.

[5] An excellent summary of the many different ways that data can be valued is provided by (Coyle & Manley, 2022).

To a certain extent investment in data is somewhat already in the SNA framework, however its current treatment contains some inconsistencies and slight vagueness. During the drafting of the current 2008 version of the SNA, there was careful consideration on how far to include data in the database category of Gross Fixed Capital Formation (GFCF) (Ahmad, 2005). Ultimately, in order to avoid the possibility of implicitly "capitalising knowledge" the final recommendation limited capital formation to databases which would include the cost of preparing data in a format that conforms to the "database management system (DBMS)" but excludes the cost of acquiring or producing the data (SNA §10.113) (UNSD, 2009). At the same time however, when recording databases purchased via a market transaction the SNA suggested that the value should be determined by a market price, which would implicitly include the value of the information content. Thus, the result, whether by accident or design, is that the SNA recommends a different treatment for data in capital formation depending on whether a database is developed for own use or for sale or license[6].

This decision to exclude the inherent value of the data might be viewed as a pragmatic decision as separating out the value of the database management system from the inherent value of the information would have been arduous at best and impossible at worse. However, as pointed out by Ahmad and van de Ven (2018) this also reflected the view that "the underlying value (information content) associated with the data itself was de facto a non-produced asset".

While only discussed 18 years ago, this pragmatic decision was taken at a time when the use of data in the production process was not as omnipresent as now. This last point is important, the motivation to introduce data as an asset in the SNA is not due to the sudden realisation by the international statistical community that data is now used in production. Numerous examples exist of collected and organised data (albeit non-digital data) being fundamental to production throughout history[7]. However, most producers who used data did so in a rudimentary manner or by purchasing the service from a specialist. Rather, as the value of data insights have become more obvious, and the cost of collecting, storing and analysing data has declined, its use has increased exponentially[8] necessitating some form of change to the SNA framework.

Today, many firms undertake some form of data analytics, importantly however, on most occasions it is not a case of firms buying a data product that suits them, rather they are creating data themselves to increase revenue or improve efficiency and productivity. This anecdotal observation now has empirical backing with data from Eurostat showing that while 36.9% of large businesses in Europe undertake some form of data analytics only 4.6% of business are purchasing data and even less are selling it (Eurostat, 2022). Such results show that the vast majority of data used repeatedly during production is created on an own account basis.

It is useful to also specifically note that the proposed change in conceptual treatment is not as simple as including something that was previously excluded, rather it is an important change in definition of data

[6] Several recent papers provide a slightly longer discussion on the measurement and classification of data within the current SNA, this includes (Ahmad & van de Ven, Recording and measuring data in the System of National Accounts, 2018) (Rassier, Kornfeld, & Strassner, 2019) (Goodridge, Haskel, & Edquist, 2022).

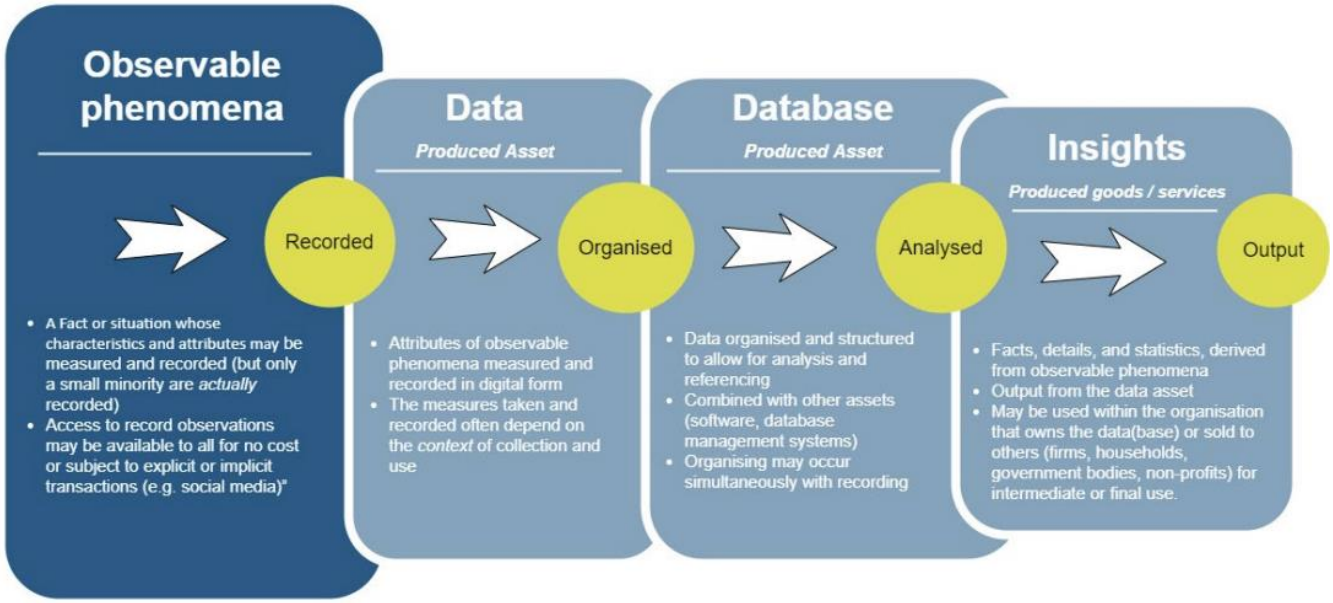
[7] Examples include the formation of the first modern insurance fund in 1740's (Ferguson, 2009) to the Feist publications court decision in 1991 (Supreme Court of the United States, 1991).

[8] There is a range of figures (of varying quality) that demonstrate the ever-increasing amount of data potentially generated. Most are a variance of the 2010 quote by then Google CEO, Eric Schmidt, who stated that every two days we create as much information as we did from the dawn of civilization up until 2003. (Seagate, 2023)

used previously in the SNA and that proposed for statistical measurement in the forthcoming SNA. Previously, when the SNA implicitly took the view that data had value but was non-produced, it viewed 'data' as the embodied information content of what is now typically referred to in the new lexicon of data value chains as the information content of 'observations' or 'observable phenomena.' In simple terms this is information content that had not yet been recorded[9].

For the upcoming SNA, the concept of data has been moved slightly along the data value chain (see Figure 1). Data is now considered as *"information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities."* (ISWGNA, 2023) Importantly there are productive activities such as accessing, observing, recording, organizing, and storing that must occur in order for the data to be recognised as produced. The information element (or knowledge) is not enough by itself to be considered a produced asset.

Figure 1: Data-information chain from a SNA perspective



Source: (Mitchell, Ker, & Lesher, 2022)

Estimates of data produced using the same standards and conventions as other capital inputs within the SNA results in outputs that are easier to compile, are more comparable, and can be used with greater confidence than is currently the case. Importantly, while compared with the methods used to measure the investment and stock of other fixed assets in the national accounts, the methodology and practical experience for data is still in its infancy, however, improvements are occurring quickly. As can be observed in some of the cases studies presented later in the paper, many important steps, which improve the accuracy and dependability of the valuations, have already occurred in only a few years. As all countries begin to create their own estimates of data assets, for inclusion within the SNA production and asset

[9] In (Mitchell, Ker, & Lesher, 2022) Observable Phenomena are defined as "a fact or situation, whose characteristics and attributes may be recorded".

boundary, compilation methods will continue to build of one another, improving the overall quality of the estimate.

Recommendations for putting data into the 2025 SNA.

The basis for the incorporation of data into the System of National Accounts (SNA) has been laid out in the formal guidance note "DZ.6 – data in the national accounts" (ISWGNA, 2023)[10]. This section of the paper summarises the more salient recommendations included in the final version of the guidance note. These recommendations are predominately conceptual in nature. The SNA is not a compilation guide and does not contain specific guidance regarding measurement methodology. Guidance and information on specific compilation practice are usually saved for companion handbooks and manuals[11]. A similar publication focusing on data is currently being worked on via a joint Eurostat-IMF task team and will be published in unison with the updated SNA (ISWGNA, 2023).

When attempting to measure something, a clear definition is a very important first step. The term data is used flexibly in both in policy discussions and in academia, often without a clear definition on what is being presented. Data is used to represent both a single point of information (i.e., the rainfall in London for a 24 hour period) to specific datasets (i.e., the data on daily rainfall in London for an entire year)) to much larger all-encompassing datasets while also including the statistics and indicators derived from this dataset. (i.e., Meteorological data for the United Kingdom, which would usually include the forecasts as well as the observations.)

For the purpose of its inclusion in the SNA, **Data is defined as "information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities"** (ISWGNA, 2023).

It is worth reminding that this definition is specifically for economic statistics and not for everyday usage. Different definitions will still be required for scientific pursuits and other reasons, but for the purpose of measuring value added, it is felt that this definition finds a balance between aligning with SNA principles, being prescriptive enough to be followed consistently, while still being open enough to include all the different types of data used in the economy.

A clear example of this specific purpose of economic measurement within the SNA is the prerequisite that **data exists in a "digital format" only**. Data certainly exists in non-digital form, as mentioned earlier, the use of data in production predates digitalisation, however, for the purpose of the SNA only digitised data is considered the result of production and therefore contributing to GDP. This is a pragmatic decision, based, not only on avoiding the almost impossible task of deriving a value for all non-digital data but

[10] Guidance notes were created for all research topics, with each following the same process of consultation with the ISWGNA AEG before being sent for global consultation. Edits and amendments were then made before final endorsement by the AEG.

[11] Examples of these include measuring capital, informal economy, IPP products, back casting, Price and volumes, quarterly NA, global value chain, find some others.

predominately on the reasonable proposal that almost all data *actually* used in production is digitalised and therefore the exclusion results in only a trivial amount of data assets not being counted.

In addition to non-digital data, **the creation of any data (digital or non-digital) that is not directly used in the productive activities of the economic unit that owns it will be considered outside of the SNA production and asset boundary.** This caveat is not restricted solely to data. It is an expectation that all assets considered as fixed assets in the SNA will be "produced assets that are used repeatedly or continuously in production processes for more than one year."[12] This caveat removes the possibility of gross value added being created by externality/accident. As has been well documented, much data can be created as a by-product of conventional productive activities. If a business assigns capital and labour to try and "access, observe, record, organize and store" the information elements coming as a by-product then it is considered that an asset is being created. Conversely, if another business decides to let these observable phenomena lapse and remain uncollected, then this no value added has occurred and no asset produced. Theoretically this places the choice of capitalisation somewhat on the economic unit, resources must be spent producing the data asset for it to come into existence. This is not dissimilar to the current standard applied to Research and Development, which requires work to be "undertaken on a systematic basis." From a R & D perspective, this therefore excludes the immediate capitalisation of the bright idea had by someone while walking along a beach or watching the clouds. Rather, once they use this idea to increase the stock of knowledge, in a systematic way, it becomes an asset. From a Data perspective, this would exclude the information content created as a by-product (records of retail sales, personal information captured at point of sale) unless the resources are committed to turn this information content created as an externality into data for re-use in production.

Data is the result of production; it is an intangible product considered part of the intellectual property product suite of assets. It is produced with the input of labour and capital; therefore, when capitalised in the National Accounts it is classified as a produced asset. As outlined in figure 1, Data is distinct from the information elements of 'observable phenomena' (OP), which are inputs for data but are regarded as non-produced. The OP is considered to have no value unless explicitly purchased, therefore, the value creation (a.k.a. value added) comes once the information element of the OP is recorded and stored in a digital format, at which point it becomes data[13]. Due to this the data producer and owner is the economic unit that records, organises, and stores the information elements in a digital format.

Data that is produced and used in production for more than one year meets the SNA characteristics of an asset and, as such, should be capitalised in the National Accounts. **Potentially all own account production of data may be considered capital formation.** Although it is likely that some own account data may be fully consumed within one year, due to practical limitations on delineating this data from data used repeatedly for more than year, countries may decide to capitalise all own account production of data or only a portion of it. This specific compilation decision will be determined by each NSO based on guidance and recommendations still being developed. Taking computer software as an example, many countries calculate an estimate of their own account computer software, via a sum of cost approach. They then remove a percentage of this total on the assumption that a certain amount will not be used for more than

[12] While the produced assets category also includes inventories and valuables, both of which do not have this 'production process' requirements, as detailed more thoroughly in the guidance note, data is not considered either of these.

[13] (Goodridge, Haskel, & Edquist, 2022) does a nice job of summarising this important distinction in the framework they built where they suggest that "data does not constitute an asset until it is transformed into a usable, analyzable, long-lived form that can repeatedly be used in the extraction of knowledge and contribute to production".

a year. This step is reasonable and has usually been introduced following some form of specific investigation by the NSO. Such steps should be encouraged if supported by appropriate research and are explained in a transparent manner for the user. **Data purchased via market transaction is treated as per other products,** that is, it is capitalised if intended to be used for more than one year or considered as intermediate consumption if consumed during production (e.g., no longer usable after one year).

Like other assets in the National Accounts, data is subject to economic ownership, valuation (and re-valuation) and depreciation - although depreciation occurs not due to wear and tear but rather due to obsolescence. Currently information on the appropriate service life for data assets is scarce. As will be observed later in the paper, countries that have already produced capital stocks of data have used services lives ranging from 3 years to 25. Due to the unique characteristics of data, its heterogeneous content, and the highly contextual ways it is used, data likely has a larger range of service lives than almost any other asset. While this problem of different services lives applies to all assets recorded in the SNA, it is magnified for data. Despite this, it is considered that applying an estimated average to all data produced is still the most practical approach going forward[14]. Currently a general consensus is building that while the average life might be quite short, the retirement curve is likely asymmetrical, with a very long tail to account for certain very long used data. However, as mentioned, NSO's should be encouraged to apply service lives and depreciation schedules that they consider most appropriate for their country and economy.

While most data are produced on an own account basis and used in house, data can be and certainly is sold via market transactions. **Data that is sold in a market transaction with no exclusive rights connected to the data is considered a sale of a copy of an original** (See 2008 SNA §6.208 to 6.212). If exclusive rights are granted, it is recorded as a sale of a produced asset, in line with the sale of other fixed assets that provide exclusive rights to use. In many cases, data itself will not be sold; rather it will be used by the data producer as input in the production and provision of specific goods and services.

Own-account production of data is recommended to be valued at the sum of costs. This is consistent with existing guidance in the 2008 System of National Accounts for other own-account output including intellectual property product (see 2008 SNA §6.114). Encouragingly, the work undertaken by NSO and academics so far to try and estimate a value of data assets in a manner consistent with the overall SNA framework have all approached valuation via sum of costs, this somewhat validates this recommendation. While valuations via market transactions or estimating the net present value of the asset based on future income is conceptually possible, to encourage compilation by all countries as well as provide comparability in the results, sum of costs was chosen. This important decision was strongly supported in the global consultation and is discussed more thoroughly in the ISWGNA guidance note.

Expenditure undertaken to access, record and organise additional information elements from OPs that are then added to an established data asset is considered new gross fixed capital formation not repair and maintenance. Such an action prolongs the life of the data asset rather than just ensuring that it maintains its current asset life. This recommendation follows a similar one made regarding databases prior to the production of the 2008 SNA.

[14] This need to apply an average set of assumptions (retirement profile, asset life etc.) to a cohort, even though these assumptions cannot actually be represented by a single individual asset is a fundamental aspect of the PIM. This point as well as why it is still considered appropriate for the national accounts is explained in greater length in (OECD, 2009).

Initial work to estimate the value of data in a manner consistent with the SNA.

An indication of the general demand for information on data valued from a SNA perspective is that even before the formal endorsement to have data assets included in the forthcoming updated SNA, several countries and academics undertook work to produce estimates. This work has been important for confirming the feasibility of proposed recommendations. All the following examples use the same fundamental methodology that can be explained in four steps.

The first is to estimate the value for the labour cost of involved in producing data. This is done by identifying specific occupations involved in data creation and multiplying the wage rate of these occupations by the number of hours spent *actually* producing data. The next step is to add expenditure related to the non-labour costs of producing data. This includes the intermediate consumption of other goods and services, as well as a return on the fixed capital involved in production. For non-market producers this would be simply the consumption of fixed capital involved in the production, while for market producers an additional mark-up is required to represent the operating surplus. The third step is to then deflate this nominal estimate using a price index to provide the estimates on a volume basis and removing any growth due to inflation. Finally capital stock estimates are created using a PIM.

Each example described below took slightly different variations on this fundamental methodology. A non-exhaustive list of these variations includes the occupations chosen, the method to deduce the concentration/intensity of data production to these specific occupations, the price index or depreciation schedule applied, and the mark up used to represent non-labour costs.

In July 2019, Statistics Canada was the first NSO to produce an estimate of data assets consistent with the SNA (Statistics Canada, 2019). They used the population census and labour force data to derive a nominal estimate based on eight Occupational groups, "selected from among those in the National Occupational Classification (NOC) that are generally associated with converting observations into digital format (the process of digitization)" (Statistics Canada, 2019). An additional three groups were selected for the estimate of databases and another five for the estimate for data science. Due to a lack of information on non-labour costs, a general mark-up of 50% was applied to the labour costs to represent all non-labour costs with a further 3% added to represent the return on capital services.

In 2021, Statistics Netherlands (De bondt & Mushkudiani, 2021) published their version of the framework put forward by Statistics Canada. While making some slight changes around the proportion being allotted to non labour costs (60% instead of 50%) and using a weighted price index drawn from price changes in all three input categories (labour/capital/other inputs) instead of just labour inputs as was the case with Canada. A final difference in the Dutch paper was the exclusion of the government sector. Statistics Netherlands deviates slightly from the subsequent recommendations in the formal guidance note put forward by the digitalisation task team which states that all sectors of the economy can produce data assets. Statistics Netherlands takes the view point that since "a lot of government administered data is freely available, e.g. through NSIs [and] for something to qualify as an asset, economic ownership needs to be asserted" they feel therefore, that "it is difficult to regard government data as a government asset"

(De bondt & Mushkudiani, 2021). Such a viewpoint would appear to be at odds with other fixed assets currently recorded on the SNA balance sheet of the government sector such as roads and bridges, both of which are also publicly available for use at no cost. Instead, the digitalisation task team takes the viewpoint that these assets are the capital inputs to providing various administrative services (i.e., tax databases for administrating tax policy). Despite the absence of data assets from the education and public administration industries, the Dutch estimate was slightly higher than the Canadian when viewed from a proportion of GDP (see Figure 2)

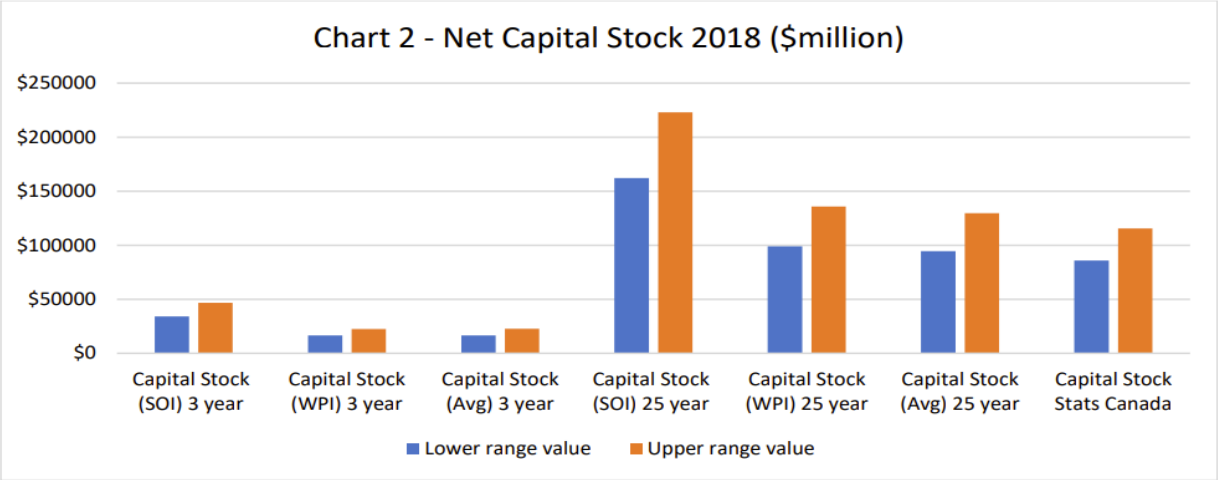
Figure 2: Initial estimates of the value of data assets, selected countries.

Country	Year	Value of data asset, % of total GDP	PPT difference in total GDP growth for year	PPT difference in GCF for the year
Australia	2016	2.7%	0.016%	0.57%
Canada	2018	1.8%	-0.037%	-0.09%
Netherlands	2017	2.7%	-0.012%	-0.12%
India	2019	1.0%	0.000%	0.14%
USA	2020	0.8%	0.047%	0.26%

Source (ISWGNA, 2023)

In March 2022, the Australian Bureau of Statistics presented estimates of data within the Australian Economy (Smedes, Nguyen, & Tenburren, 2022). They also broadly followed Statistics Canada's method, using the same 50% and 3% mark up for non-labour costs and capital services. Due to this their estimate of GFCF of data is broadly like Canada's and Netherlands when considered as a share of GDP (See Figure 2). Two areas where the ABS deviated from the Canadian work was in their choice of price index and asset lives when creating estimates of capital stock. Australia preferred a weighted price index reflecting the different inputs used in production. This contrasted with Statistics Canada's choice of a price index based entirely on labour costs. This has a small but non-trivial impact on capital stock estimates; however, Australia also applied a mean asset life of 3 years and a maximum asset life of 5, this is in contrast with the 25-year life applied in Canada. As would be expected this change produce a net capital stock estimate significantly lower than that published by Statistics Canada (Figure 3).

Figure 3: Variations in Capital Stock of Data in Australia, based on index and asset life applied.



Source: (Smedes, Nguyen, & Tenburren, 2022)

Later in 2022, the United States Bureau of Economic Analysis published estimates of the data economy (Calderón & Rassier, 2022) and while they also used sum-of-costs as the method, they applied a much more systematic way to determine both the occupations from which labour costs should be taken but also in estimating the time use / level of data intensiveness that should be applied to each occupation. The previous three studies all used somewhat arbitrary estimations of the level of time different occupations were spending producing data. In contrast, the BEA uses information on job ads to determine which jobs involved data related tasks and depending on the concentration of certain tasks and required skills in the job ad, also the level of time use that should be applied to each job. The overall estimate of data put forward by the BEA is lower than those previously published by other countries but still recorded an average annual growth rate higher than the overall rate of growth for business activity.

This idea of using a more systematic approach to estimate the data intensity of different occupations was also picked up by the OECD, who used job advertisements from Lightcast data to score the data intensity of occupations based on natural language processing (Schmidt, Pilgrim, & Mourougane, 2023). This information can be broken down by industry or individual occupation level and can then be used as a starting point for determining the occupations contributing to data production as well as the time share of each occupation. Importantly this approach can be applied consistently across countries where the data is available providing potentially a more objective approach to determining the parameters of the sum of cost approach. This method is generally considered to be more exhaustive and as such the overall aggregate amount of data investment produced using this approach is slightly higher than those put out by the NSOs of Canada and the United States (See Figure 4).

Figure 4: Data assets, Comparison between different methods, USA & Canada

Country	Year	Value of data asset, % of total GDP	
		NSO Estimate	Estimated via natural language processing
Canada	2018	1.8%	3.1%
USA	2020	0.8%	4.4%

Source: (Schmidt, Pilgrim, & Mourougane, 2023)

The desire to compare across countries was picked up by (Goodridge, Haskel, & Edquist, 2022) who used a sum of cost approach to derive estimates of data investment using European labour force data. This enabled them to produce estimates for a majority of European countries and found that expanding the production boundary of Database and Software[15] to include data assets would add around 60% to the EU-16 estimate of this assets class while growing at a rate twice as fast as that currently recorded for simply software and databases.

Conclusion

This paper has provided an update on how Data will be included in the updated SNA as a new produced asset class, including examples by countries who have published estimates broadly in line with the proposed recommendations. While conceptual recommendations have generally been consulted and agreed on, specific compilation methods will continue to evolve as more countries produce estimates of data assets in their economy and best practice is shared across NSOs. Pleasingly, the early results from initial estimates are that the incorporation of data will not swamp or "break" GDP or other produced indicators. Rather it will more accurately reflect the capital inputs that are being used in the production of other goods and services.

That said, as highlighted in section 2, there are still some compilation decisions for which countries have deviated from one another, on some occasions by a fair margin. While compilation methods are a decision for individual countries, more work is required to better define the international best practice on certain aspects of data compilation, in particular on assumptions used to create capital stock estimates.

Inclusion in the SNA is just one component of recording data in economic statistics. Updates to the international classifications that cover products, activities and international transactions are also underway to complement the existence of data in the national accounts. This is important, as producing a

[15] As noted in the paper, this asset class is the logical one where data would be already included or likely added. While the SNA delineates databases as a separate asset class, almost all countries who present GFCF estimates broken up by asset type, combine software and databases into a single category.

consistent and replicable framework for compiling estimates of data investment is only useful if NSO's subsequently classify these estimates in a consistent manner.

To maintain the relevance and accuracy of the SNA including its preeminent indicator, GDP, data assets must be included in the revised framework. Its explicit absence in the current framework continues to raise questions and concerns that are becoming harder to ignore or write off as trivial. While this paper and the work described in it will not be the final and definitive solution for statistical offices, the work to data has advanced the concept of measuring data in the national accounts providing a base for statistical offices to build on.

References

Ahmad, N. (2005). The measurement of databases in the National Accounts. https://unstats.un.org/unsd/nationalaccount/aeg/papers/m3Databases.PDF.

Ahmad, N., & van de Ven, P. (2018). Recording and measuring data in the System of National Accounts. https://unstats.un.org/unsd/nationalaccount/aeg/2018/M12_3c1_Data_SNA_asset_boundary.pdf.

Calderón, J., & Rassier, D. (2022). *Valuing the U.S. Data Economy Using Machine Learning and Online Job Postings.* https://www.bea.gov/system/files/papers/BEA-WP2022-13.pdf.

Corrado, C., Haskel, J., Iommi, M., Jona-Lasinio, C., & Bontadini, F. (2023). *Data, Intangible Capital, and Productivity.* https://www.nber.org/system/files/chapters/c14737/c14737.pdf.

Coyle, D., & Manley, A. (2022). *What is the Value of Data? A review of empirical methods.* https://www.bennettinstitute.cam.ac.uk/wp-content/uploads/2022/07/policy-brief_what-is-the-value-of-data.pdf.

De bondt, H., & Mushkudiani. (2021). *Estimating the Value of Data in the Netherlands.* https://iariw.org/wp-content/uploads/2021/10/bondt_paper.pdf.

Eurostat. (2022). *Digital Economy and Society (database).* Retrieved from https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database

Ferguson, N. (2009). *The ascent of money : a financial history of the world.* New York: Penguin Books.

Goodridge, P., Haskel, J., & Edquist, H. (2022). We see data everywhere except in the productivity statistics. *Review of Income and Wealth*, https://www.ifn.se/media/u0tjjvjn/2022-goodridge-haskel-edquist-we-see-data-everywhere-except-in-the-productivity-statistics.pdf.

ISWGNA. (2023, July 5). *Minutes from Meeting of the ISWGNA and Task Team leads.* Retrieved from https://unstats.un.org/unsd/nationalaccount/iswgnaMinutes/iswgna189.pdf

ISWGNA. (2023). *SNA Guidance note, DZ.6: Recording of data in the National Accounts .* https://unstats.un.org/unsd/nationalaccount/RADOCS/ENDORSED_DZ6_Recording_of_Data_in_NA.pdf.

Ker, D., & Mazzini. (2020). *Perspectives on the value of data and data flows.* https://www.oecd-ilibrary.org/science-and-technology/perspectives-on-the-value-of-data-and-data-flows_a2216bc1-en.

Mitchell, J., Ker, D., & Lesher, M. (2022). *Measuring the economic value of data.* https://www.oecd-ilibrary.org/science-and-technology/measuring-the-economic-value-of-data_f46b3691-en: OECD Publishing.

OECD. (2009). *Measuring Capital.* https://www.oecd.org/sdd/productivity-stats/43734711.pdf.

OECD. (2022). *Measuring the value of data and data flows.* https://www.oecd-ilibrary.org/docserver/923230a6-en.pdf?expires=1695240160&id=id&accname=guest&checksum=A23446F17CC73C2BAA1DEEAADB8AE8F3.

Oulton, N. (2018). GDP and the System of National Accounts: A past, present and future. https://eprints.lse.ac.uk/87178/1/__lse.ac.uk_storage_LIBRARY_Secondary_libfile_shared_repository_Content_Centre%20For%20Macroeconomics%20discussion%20papers_GDP%20and%20the%20system%20of%20national%20accounts.pdf.

Rassier, D., Kornfeld, R., & Strassner, E. (2019). *Treatment of Data in National Accounts.* https://www.bea.gov/system/files/2019-05/Paper-on-Treatment-of-Data-BEA-ACM.pdf.

Schmidt, J., Pilgrim, G., & Mourougane, A. (2023). *"What is the role of data in jobs in the United Kingdom, Canada, and the United States?: A natural language processing approach".* https://doi.org/10.1787/fa65d29e-en.

Schmidt, J., Pilgrim, G., & Mourougane, A. (2023). What is the role of data in jobs in Canada, the United Kingdom and the United States –A natural language processing approach. *ESCoE Conference of Economic Measurement.* https://escoe-website.s3.amazonaws.com/wp-content/uploads/2023/05/23160724/Schmidt-Estimating-investment-in-data-in-the-United-Kingdom-and-selected-trade-partners-using-online-job-advertisements.pdf.

Slotin, J. (2018). *What do we know about the value of data.* https://www.data4sdgs.org/blog/what-do-we-know-about-value-data.

Smedes, M., Nguyen, T., & Tenburren, B. (2022). *"Valuing data as an asset, implications for economic measurement.* https://www.abs.gov.au/system/files/documents/7bfccb4ddb8aded818330bebe6b76b14/Smedes%20-%20Valuing%20data%20as%20an%20asset.pdf.

Soloveichik, R. (2023). *Capitalizing Data: Case Studies of Tax Forms and Individual Credit Reports.* https://www.bea.gov/system/files/papers/BEA-WP2023-7.pdf.

Statistics Canada. (2019). *The value of data in Canada: Experimental estimates".* Retrieved from https://www150.statcan.gc.ca/n1/en/pub/13-605-x/2019001/article/00009-eng.pdf?st=ifEOEPUK

Supreme Court of the United States. (1991). Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340. https://supreme.justia.com/cases/federal/us/499/340/.

The Economist. (2022). Should "data" be singular or plural? https://www.economist.com/culture/2022/08/11/should-data-be-singular-or-plural.

UNSD. (2009). *2008 System of National Accounts.*