Working paper 3.4

8 July 2020

Group of Experts on National Accounts

September 2020

Webinar on Digitalization

14 September 2020, 13:30 - 16:30

Organizers: United States and Eurostat

RECORDING AND VALUATION OF DATA IN NATIONAL ACCOUNTS

Recording and Valuation of Data in National Accounts

ISWGNA – Subgroup on digitalization[1]

Issue note – June 2020

1. Executive summary

1.1. Introduction

The System of National Accounts (SNA) is the core statistical framework that supports policymaking and other purposes. Digitalization is transforming production processes and products for businesses and creating new consumption options for households. As digitalization becomes more common in economic activity, the relevance of macroeconomic statistics depends on the ability to adapt the SNA and develop supplemental frameworks to meet the evolving needs of policymakers and other users. From a macroeconomic measurement perspective, there are three developments from digitalization that require careful consideration for future valuation and treatment in statistical frameworks: 1) data as an asset, 2) 'free' assets, and 3) 'free' services. Each of these is generally considered to be outside the current scope of the 2008 SNA boundaries. A number of questions come to mind for each:

- **a.** How should they be defined and classified for statistical purposes?
- **b.** How should they be valued?
- **c.** How should their flows (and stocks) be recorded in a national accounting framework?
- **d.** Should they be included in the core accounts or satellite accounts?
- **e.** How can we track their cross-border flows?
- **f.** Who is the economic owner of data and 'free' assets?

The ISWGNA 'Subgroup on digitalization' (hereafter 'the subgroup') is trying to answer these questions and make recommendations that allow an accurate measurement of GDP and productivity and a correct allocation of production and expenditure across institutional sectors. This issue note presents the views and the proposals of the subgroup on the *Recording and valuation of data in national accounts* so far. Since its setup in 2019, the subgroup has reviewed the existing literature and has identified and analysed possible options to tackle the issues at stake, considering both conceptual soundness and practical aspects. This has allowed to already identify possible solutions for several of them.

The subgroup is addressing in parallel issues related to 'free assets' and 'free services', taking into account the strong links between data and these topics. 'Free assets' and 'free services' will be the

[1] This issue note has been prepared by the ISWGNA Subgroup on digitalization, based on an issue paper drafted by the U.S. Bureau of Economic Analysis in March 2020 and incorporating inputs from the Subgroup members. Contributors to this paper are Andreas Dollt and Nicola Massarelli (Eurostat), Dylan Rassier and Rachel Soloveichik (BEA), John Mitchell (OECD), Marshall Reinsdorf, Jennifer Ribarsky, Jim Tebrake, Margarida Martins and Silvia Matei (IMF), Richard Heys (ONS), Ziad Ghanem (Statistics Canada), Sri Soelistyowati (BPS Indonesia), Stanimira Kosekova (ECB), Benson Sim (UNSD) and Kevin Fox (UNSW).

subject of a separate note. At present, work is more advanced on 'data' than on the other two topics.

The subgroup shares the opinion that 'data' is produced and should be included as such in the 'core' part of the SNA. The subgroup has not yet formed an agreed view on the treatment of free assets and free services, in particular on whether, and to what extent, these free products should affect the core accounts and/or if they should be addressed in detail in satellite accounts.

The main findings of the subgroup on 'data' so far, including recommended options when already available, are presented below. Chapter 2 shows an overview of the existing literature. Chapter 3 presents a broader view on the considered options, including pros and cons and reasons for recommending them or not. Chapter 4 includes remaining questions on the recording and valuation of data.

1.2. Recommended options

Definition of data

In essence, the subgroup recommends updating the 2008 SNA to include *data* in the production and asset boundaries. The subgroup also considers that the underlying *observable phenomena* (which lead to data) are non-produced and do not have the characteristics of an asset. As such, they are considered neither produced nor non-produced assets and should not be recorded at all in the SNA.

The main valuation approach for data would be as for an asset produced on 'own-account', and forming an asset which should be depreciated using a worldwide standard method. Data are to be seen in strict relation to databases, which are already covered in the 2008 SNA. The recording of data as an asset may be made together with databases, thus increasing the value of the latter, or as a separate asset. Both approaches have merits and downsides. More details are provided below.

The 2008 SNA (10.112) provides a definition of databases that refers to data, without however specifying what data is: ***Databases*** *consist of files of* ***data*** *organized in such a way as to permit resource-effective access and use of the data. Databases may be developed exclusively for own use or for sale as an entity or for sale by means of a license to access the information contained.*

2008 SNA para 10.113 specifies: *The cost of preparing data in the appropriate format is included in the cost of the database <u>but not the cost of acquiring or producing the data</u>.*

It is therefore the cost of acquiring and producing the data that the next SNA update should cover, to complement that of preparing the data in the appropriate format. The subgroup proposes the following statistical definition for data:

> **Data** is information content that is produced by collecting, recording, organising and storing *observable phenomena* in a digital format, which can be accessed electronically for reference or processing. Data from which its owner(s) derive economic benefits by using it in production for at least one year is an asset.

According to this definition, *data* is included in the SNA production boundary. When it produces an economic benefit to its owner(s) by using it in production for at least one year (2008 SNA 10.33), termed 'long-lived data' in this paper, then data is also to be included in the SNA asset boundary. Short-lived data (i.e. useful for less than one year) is instead to be considered as intermediate consumption when it is purchased from third parties, or as the product of an ancillary activity, i.e. an integral part of the primary activity, when the production is taking place within the same unit. In this latter case, while it should not be separately recorded as output or intermediate consumption, it should be reflected in the value of output it contributes to.

Data can be stored in a digital as well as in a non-digital format. However, the subgroup recommends the exclusion of non-digitalised data from the SNA production and asset boundaries for practical reasons. It only makes up a small amount of the data within the economy and its monetary value to production is considerably lower than digitalised data. The subgroup thus recommends to focus on digital data only[2], which has the potential to generate much larger valuations and greatly influence the efficiency and output of production due to its ability to be easily processed and to be sold/leased/purchased.

Another crucial distinction is between produced *data* and the underlying non-produced observable phenomena[3]. *An **observable phenomenon** is the occurrence of a singular event or piece of information*[4]. Observable phenomena simply exist, independently of if they are observed, recorded and used for economic purposes. They are non-produced. They do not meet the characteristics of an asset and should not be included in the SNA asset boundary. Further considerations on observable phenomena are provided in section 3.6.

The above definition identifies 'data' as the result of a production process. It also implies that data is subject to economic ownership and has the characteristics of an asset, and as such it is subject to valuation and depreciation.

Economic ownership

[2] A general definition of data, including digital and non-digital data, is: 'Data is information content, which results from collecting, recording, organizing, and storing observable phenomena that can be accessed for reference or processing and from which economic benefits are derived by their owner(s) by holding or using it'.

[3] The term 'observations' has also been considered by the subgroup. It should be noted however, that this term is already loosely used in the SNA, although not defined. For example, in paragraph 18.39: '*Simply benchmarking four quarterly observations to the eventual annual figure, though, may give unexpected and implausible changes from the last revised quarter to the next quarter (a 'step') unless techniques are used that address this problem*'. To avoid confusion and make clear that these are states of nature or events and not a produced output, the subgroup recommends the use of the term 'observable phenomena'.

[4] Adapted from OECD (2020), 'Measuring data products'.

Based on the definition above, the economic owner of data is the owner of the intellectual property rights over the data. This is (at least initially) the entity that produces the data and not the person or entity the underlying observable phenomena refer to. For example, when someone searches on his/her computer, search engines record all those 'clicks' and digitise those observable phenomena into data, producing them and thus becoming their owner.

In general, the SNA definition of economic ownership applies. In case of own account production, the economic owner is the producer of the data and not necessarily the owner of the place where the data is stored (e.g. if the server that hosts the data belongs to a different unit). In general, a license to use data is not to be considered a change in economic ownership. However, when a copy of data is made available under a license to be used in production for more than one year, the licensee assumes all the risks and rewards of ownership (2008 SNA 10.100). Multiple copies of data can be licensed to several licensees at the same time, which opens the possibility of multiple owners of the same data. A second case where the producer is not the owner is when an outright sale of data (i.e. an explicit transaction) occurs.

Determining the economic owner in the case of a multi-national enterprise group (MNE) may not be straightforward, e.g. when a digital platform is headquartered in one country, records its IPP in another country, stores data in the cloud and/or operates data centres around the world. A joint ISWGNA-BOPCOM Task Force on IPPs is addressing these issues. Their guidance will apply also to data owned by MNEs.

Data as an asset

The subgroup agrees that long-lived data should be included in the **fixed asset category**, be this own-produced or purchased. The inclusion of data as a produced asset has the effect to increase output, value added and GDP. For market producers, this stems from the fact that the new data assets will be the result of productive activities. For non-market producers, this increase derives from the additional consumption of fixed capital of the data assets.

The usefulness of data in production widely differs, with some data being useful for a long period and others becoming soon obsolete. Of course, only data with a long-lasting (i.e. one year or more) use in production qualifies as an asset, whereas short-lived data does not.

Data cannot be classified as inventories, because it cannot exit the capital stock like other inventories, nor as valuables, as it is not acquired with the main intention to store value over time.

Several options have been considered on how to record long-lived data within fixed assets. From a conceptual point of view, the subgroup's favoured option is to create a new specific category of fixed assets for data, under 'Computer software, *data* and databases (AN1173)'. This option, which keeps data separate from databases, would give more relevance to data as a separate fixed asset, thus highlighting its importance in today's economy. Measuring it separately from databases may however be challenging in practice.

An alternative option to expand the scope of databases in capital formation, to include the costs to produce or acquire data, may be considered. The rationale for this option is mainly practical: national accountants may not always be able to separate the costs of database structure from that

of the database content (i.e. data). As long as own-account estimates will be based on occupational data, it may be hard to make a clear distinction between these categories. In practice, the publication of more detailed breakdowns might be restricted to growth rates, if estimating actual nominal values is not feasible. This approach is in line with the current 2008 SNA approach for software and databases, which would ideally be separately estimated but in practice are often combined. This approach would just require an extension of the paragraphs on software and databases to also cover data.

Practical tests of the two options will be crucial to make a final recommendation. Whatever the chosen option, the conceptual distinction between data and databases is useful to bring clarity and should be included in the next version of the SNA.

Valuation and depreciation of data

Data assets can be produced on own-account or purchased. This distinction is relevant for the *valuation* of data, which should use the sum of costs approach for own-produced data and market prices for purchased data.

The cost of producing own-account data should include the costs of collecting or acquiring data. This includes surveying, locating and capturing the underlying observable phenomena, including through providing free services or discounts. The preparation of the database structure, the cost of preparing data in appropriate format and storage costs (including cloud storage) are already included in the value of own-produced databases in the 2008 SNA. However, it is possible to conceive of storage, processing and distribution costs specific to data production, which should be included in the value of own-account data. Disentangling the part of costs specific to data from that related to databases is an issue that requires practical considerations, not addressed in this paper.

Purchased data are part of pre-existing datasets and are bought and sold in the commercial market as part of a database.

Depreciation of data as a fixed asset should be estimated through the PIM approach, using appropriate service lives and depreciation patterns. If data assets are valued and recorded separately from databases, services lives could be estimated empirically where feasible. If instead data assets are combined with databases and software, then the same service lives could be assumed for data.

Finally, the value of some data may increase over time. In general, the increase would be treated as a revaluation. However, if the source of the change is a discovery of a new use for the data, then this should be recorded as an 'other change in volume'.

1.3. Overview of main definitions

Database [from 2008 SNA (10.112)]: *Databases consist of files of data organized in such a way as to permit resource-effective access and use of the data. Databases may be developed exclusively for own use or for sale as an entity or for sale by means of a license to access the information contained.*

Data: *Data is information content that is produced by collecting, recording, organising and storing observable phenomena in a digital format, which can be accessed electronically for reference or processing and from which their owner(s) derive economic benefits by holding or using it.*

Data as an asset: *Data from which their owner(s) derive economic benefits by using them in production for at least one year (long-lived data) are fixed assets.*

Short-lived data: *Data from which their owner(s) derive economic benefits by using them in production for less than one year is to be considered as intermediate consumption when it is purchased from third parties, or as the product of an ancillary activity, i.e. an integral part of the primary activity, when the production is taking place within the same unit.*

Observable phenomenon: *An observable phenomenon is the occurrence of a singular event or piece of information.*

2. Existing Materials

Leading up to the 2008 version of the SNA, the Canberra II Group carefully considered the inclusion of embedded data in capital formation (Ahmad 2004, Ahmad 2005). The recommendation that was ultimately written into the SNA included databases combined with computer software as a separate category of intellectual property products (IPPs) in capital formation (2008 SNA para. 10.109-10.114). If a database is developed for own use, the SNA recommends a sum of costs approach to value the database. The sum of costs includes the cost of preparing data in a format that conforms to the database but excludes the cost of acquiring or producing the data. In addition, the sum of costs excludes the value of the database management system (DBMS), which is included instead with computer software.

In essence, the SNA implicitly took the view that 'data' had value but was actually non-produced. It is important to note however, what the SNA meant by 'data' as this will be helpful in the rest of this note. Essentially the SNA viewed 'data' as the embodied information content of what is now typically referred to in the new lexicon of data value chains as 'observations' or 'observable phenomena'.

Recognising that transactions related to these databases often included the value of the observations, the SNA took a pragmatic view that if a database is developed for sale or for license, its value should be determined by a market price, which includes the value of the information content. Thus, *de facto* 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.

More recent conversations have a renewed focus on the treatment of the information content of databases – i.e., the embedded *data* – in response to the rapid increase in the collection and use of data among businesses, governments, non-profits, and households over the last couple decades (Ahmad and Van de Ven 2018, Nijmeijer 2018, OECD 2020, Rassier et al. 2019, Reinsdorf and Ribarsky 2019, Statistics Canada 2019).

The Joint Eurostat-OECD Task Force on Land and Other Non-Financial Assets did not address data as an asset in their final report on IPPs because it was considered out of scope for the Task Force (see box 3.1 in the report). [5] The recommendations outlined for databases are consistent with those already in the SNA. The existing work by national accountants on the treatment of data stocks and flows in national accounts leave open the likely possibility that data may be a produced asset as a result of the data value chain (OECD 2020, Rassier et al. 2019, Reinsdorf and Ribarsky 2019, Statistics Canada 2019). These studies also discuss a number of considerations that need to be settled in the decision to include data stocks and flows in national accounts—such as economic ownership, multiple counting, and valuation methods. The most assertive study to date of data as a produced asset offers estimates of investment flows and stocks for three data-related categories in the Canadian economy: data, databases, and data science (Statistics Canada 2019).

3. Subgroup's considerations

3.1. Definition of Data

There is widespread agreement among statistical offices and international organizations that a definition of the word *'data'* is required before recommendations can be made on valuation and recording in national accounts (e.g. OECD 2020, Statistics Canada 2019). A few definitions have been proposed that the subgroup used as a starting point.

Statistics Canada defines *data* as 'observations[6] that have been converted into a digital form that can be stored, transmitted, or processed and from which knowledge can be drawn.' Under this definition, 'observations' are naturally occurring and do not become *data* until they have been digitally recorded. For example, the outside temperature exists as an observation whether or not it is recorded. The 'observation' only becomes *data* once it is recorded. Moreover, under Statistics Canada's definition, the recording must be digital for it to be considered *data*. In other words, observations are non-produced because they simply appear, but *data* result from activities that can be identified as a production process.

[5] A link to the report is available here: https://ec.europa.eu/eurostat/documents/24987/725066/Eurostat-OECD+Report+on+Intellectual+Property+Products.pdf.

[6] Here 'observations' is used as a synonym of 'observable phenomena'. See footnote 3 on this terminology.

The OECD (2020) offers a definition that omits use of the word *data* to avoid ambiguity. Their definition includes three terms illustrated in figure 1a: database structure, observations, and databases. Databases are composed of a database structure and observations[7].

- The database structure 'includes the general parameters of the database but excludes the content, i.e., files of data'.
- Observations are 'the occurrence and recording of a singular event or piece of information.'
- Databases 'consist of files of data organized in such a way as to permit resource-effective access and use of the data.'

The latter definition is from the 2008 SNA. The OECD considers the database structure to be produced and the observations to be non-produced. However, they also distinguish observations that have been 'collected, structured, and packaged' as 'information content' that has been produced and should be capitalized along with the database structure as the value of the database. In addition, the OECD definition does not require information content to be digital in order to have value. However, the OECD does acknowledge that 'digital databases have the potential to generate much larger valuations due to its ability to be easily processed and to be sold/leased/purchased.' Likewise, the SNA definition of databases used by the OECD implies the information content of databases to be digital, because databases are grouped with software as a type of IPP.

Reinsdorf and Ribarsky (2019) utilize a dictionary definition of *data* that includes 'facts and statistics collected together for reference or analysis' or 'the quantities, characters, or symbols on which operations are performed by a computer, being stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.' They also recognize that *data* can be analog form or digital form, but the latter form is what gives *data* the most value for their use in processing. They also distinguish between 'raw data' as events and conditions that are observed and can be considered non-produced assets and 'information assets' that can be considered produced through digitalization, processing, and analysis.

The three definitions summarized above have four shared features. Firstly, they each consider *data* to be information content that is the result of a production process of transforming observable phenomena into a digitalised form. Secondly, they each make a distinction between *data* produced as an asset and naturally occurring *observable phenomena* that are non-produced. Any intrinsic value embodied in the underlying observable phenomena (e.g. single observable phenomena with a high value) is not the outcome of a production process and should not be included in the value of digitalised data. Thirdly, they each outline uses of *data* that include reference or processing. Fourthly, they each agree that without digitalization, *data* would not be as valuable in economic activity. Thus, the definition of data proposed in chapter 1 is intended to reflect these shared features. Additional elements that the definition takes into account are the fact that data is subject to economic ownership and that it can provide economic benefit to its owner(s) for one year or more.

[7] This definition does not consider the database structure as a separate fixed asset. A database is considered a single fixed asset where two components can be logically identified.

3.2. Data as an asset in the SNA

An asset is defined in paragraph 10.8 of the 2008 SNA as a *'…store of value representing a benefit or series of benefits accruing to the economic owner by holding or using the entity over a period of time. It is a means of carrying forward value from one accounting period to another.'* All SNA assets must be subject to economic ownership. Assets may be either produced or non-produced. The definition and categorization as produced or non-produced implies that only produced assets show as GFCF and consumption of fixed capital, while income from non-produced assets show as property income.

As argued above, long-lived data is considered as a produced fixed asset. Examples are data on household browsing and consumption patterns, which are used repeatedly in machine learning algorithms designed to formulate recommendations by online retailers such as Amazon or online content providers such as Spotify.

As explained in Chapter 1 and more extensively later in the paper, two options may be considered for recording data as a fixed asset: either (i) expanding the database category to also include data, mainly for the practical reason that producing distinct estimates of the two elements may not always be feasible or reliable; or (ii) creating a new category within *intellectual property products*, to give relevance to the production process of digitalizing observable phenomena that we are trying to capture and value.

Inventories, and valuables have also been considered as alternative options for data within produced assets. However, as mentioned in Chapter 1, the subgroup thinks that neither of these options is suitable.

Paragraph 10.12 of the 2008 SNA defines *inventories* as '…produced assets that consist of goods and services, which came into existence in the current period or in an earlier period, and that are held for sale, use in production, or other use at a later date.' Inventories exit the capital stock by being removed and used, not via depreciation like fixed assets. The qualification of data as SNA inventories is complicated by the unique features of data. In the case of data generated solely to train a specific AI algorithm, for example, the data would appear to be inventory whose value will eventually be subsumed into the finished software. Nevertheless, the data in principle can likely be used repeatedly, and the assumption that inventories are used up does not fit to the distinctive features of data. Dropping the assumption would have unacceptable consequences because counting both the gross production of inventories of short-lived data and the full price of products embodying the data will lead to double counting.

Valuables are defined in paragraph 10.13 of the 2208 SNA as '…produced goods of considerable value that are not used primarily for purposes of production or consumption but are held as stores of value over time. Valuables are expected to appreciate or at least not to decline in real value, nor to deteriorate over time under normal conditions. They consist of precious metals and stones, jewelry, works of art, etc.' Clearly data are not used as a store of value but for use in production, and therefore do not qualify as valuables.

3.3. Economic Ownership

The SNA definition of economic ownership in paragraph 10.5 states '…the economic owner of entities such as goods and services, natural resources, financial assets and liabilities is the institutional unit entitled to claim the benefits associated with the use of the entity in question in the course of an economic activity by virtue of accepting the associated risks.'

Based on this definition, observable phenomena that have been collected, recorded, organized, and stored (i.e., data) for own account use, for license, or for sale seem to be subject to economic ownership. In the case of own-account data, the producer is the user who bears the risks and is the economic owner of the data asset. In the case of licensed data, the producer normally retains economic ownership of the data asset while granting access to the user in exchange of a fee. In this case, data remain in the balance sheet of the producer as a fixed asset, while their use by the lessee is to be considered as intermediate consumption. However, when a copy of data is made available under a license to be used in production for more than one year, the licensee assumes all the risks and rewards of ownership (2008 SNA 10.100). In the case of a sale of data, economic ownership is transferred from the producer to the user upon the sale.

3.4. Recording data as an asset in the core SNA

Based on the literature generated by statistical offices and international organizations so far, there is general agreement that data share characteristics with other IPPs. Based on that literature, the subgroup has considered three options for the recording of data as an asset in the SNA.

One option is to retain the current SNA recommendations on software and databases. This option means that software and databases would continue to be grouped together as a single category of IPPs, with countries providing separate estimates of databases when possible. Own-account databases would generally be valued using a sum-of-costs approach that includes the cost of preparing data in an appropriate format and excludes the cost of acquiring or producing the embedded data. Purchased databases would generally be valued using market prices, which include the value of any embedded data. Likewise, the value of data that is reflected in goodwill with the acquisition of a firm would be included in measures of non-produced assets. Under this option, own-account data and data reflected in goodwill would not affect measures of production and income. At the same time it would not be possible to purchase data, as such a category would not exist separate from databases.

A second option is to expand the scope of databases in capital formation to include select costs associated with acquiring or producing data. Based on the proposed definition for data, production costs could include the costs associated with surveying, locating, collecting, recording, organizing, and storing observable phenomena. In this case, estimates of databases would include costs that are deemed to be within scope for national accounts and are not already included in other national accounts measures. Under this option, production and income measures would be affected by data production activities and acquisitions of data, which are currently excluded from the scope of database assets. There would also be an impact on subsequent flows of consumption of fixed capital as databases depreciate and an impact on subsequent production and income flows to reflect payments for the right to use database assets (i.e., intellectual property products).

A third option is to include data as a distinct category of IPPs. In this case, the value of data would be recorded separately from the value of databases (and software). This option would only be advisable in the core SNA accounts if an accurate and reliable estimate of data as an asset is possible. Under this option, production and income measures would be affected by data production activities and acquisitions of data. In addition, the treatment would generate subsequent flows for consumption of fixed capital as data depreciates and may generate production flows to reflect payments for the right to use data assets (i.e., intellectual property products).

The subgroup considers the first option overall problematic and has a preference for the second or third options. In particular, the first option does not appropriately classify the additional expenditure and production that occurs in the generation of 'digitalized data', an asset that is subsequently used in production. In addition, it does not clarify if purchased data are to be recorded as GFCF or as purchases of non-produced assets. Moreover, paragraph 10.114 in the 2008 SNA implies that part of data is valued (if subject to a monetary transaction) and part is not (if produced on own-account), which is not consistent. Finally, the current recommendation clearly doesn't meet the users' needs, which is why this issue is a priority in the SNA research agenda.

Options 2 and 3 both have pros and cons. Keeping *data* together with *databases* (option 2) has practical advantages, as countries may not always be able to separately estimate their values and depreciation. This approach is in line with the current 2008 SNA indications for software and databases, which would ideally be separately estimated but in practice are often combined. Furthermore, it may be hard to describe to users what a database is if it does not include the data (or the cost of digitizing the data which is currently included in databases) and the software. This approach would require an extension of the paragraphs on software and databases to also cover data. The asset category could be called 'Data and databases' in order to clearly show that it contains two distinct but aggregated assets used in production. On the negative side, this option may be perceived as not putting sufficient emphasis on data as an additional asset. This may reduce countries' efforts to properly measure it. In addition, users may not have separate estimates for data. Finally, this approach may require significant revisions of the database series for many countries.

Computer software, *data* and databases (AN1173)
- Computer software (AN11731)
- *Data* and databases (AN11732)
  - *Data (AN117321)*
  - Databases (*AN117322*)

Creating a separate asset category for data (option 3) is appealing from a user perspective as it provides a more refined breakdown. In addition, it permits the occurrence where the same data may be fed into multiple databases, sold or used to develop new products (although this requires to consider how to measure this). Whilst this option may cause some complexities in terms of splitting the value of databases, the flexibility it preserves in terms of data having multiple uses may well justify the cost of this.

Computer software, *data* and databases (AN1173)

Computer software (AN11731)
Databases (AN11732)
Data (AN11733)

A crucial factor for the choice between options 2 and 3 will be the possibility to estimate the cost of producing data separately from those of producing the other database components. Most of data production is done on own-account and would be valued at a sum of cost approach (see next section). This requires to determine the appropriate occupational groups and associated non-salary costs involved in the process. While this should be possible for countries with advanced statistical systems, other countries may have difficulties with implementing option 3.

3.5. Valuation of Data as an Asset

In general, the preferred method of valuation in the SNA is market prices for actual or comparable transactions. When active markets do not exist for a given product, which is often the case for IPPs, the SNA generally recommends a sum of costs method. The sum of costs includes labour compensation, intermediate consumption, consumption of fixed capital, other taxes less subsidies on production, and a net return to fixed capital for market producers (2008 SNA paragraph 6.125). When measured correctly, these components should result in a value that is comparable to a market value. Compensation, intermediate consumption, and other taxes less subsidies are less vulnerable to mismeasurement because they are market-based. Consumption of fixed capital and net return to fixed capital are generally not market-based and more prone to mismeasurement.

Purchased Data

If data are exchanged in an active market, the transaction should be valued at the market price. In the case of a sale, the value of data reflects the producer's costs and profit for collecting, recording, organizing, processing and distributing data. In the case of a license, storage costs would also be included. In each case, the recording would be consistent with SNA recommendations. If data are initially valued using a sum of costs method and then subsequently sold, the difference in valuation may be treated as a revaluation. Purchased data are normally part of pre-existing datasets and are bought and sold in the commercial market as part of a database. The value of data is normally included in the price of the database. Separating the price of data from that of the other database components will likely not be straightforward.

Own-Account Data

Own-account IPPs are generally valued using a sum of costs method. For data as an asset, the data value chain can be referenced to determine which costs are within scope for national accounts and not already included in other national accounts measures. The data value chain is presented in figure 2. The data value chain demonstrates a five-stage production process from an unstructured form – i.e., *observable phenomena* – that has very little value, to a structured form – i.e., *data* – that can be leveraged in a business model or other usage. When high volumes of observable phenomena are collected in the first stage, they may be unstructured such as those collected via

electronic payment systems (e.g., credit card purchases and Venmo), internet-connected machines and devices (e.g., smart phones and Internet of Things - IoT), or other methods. In this case, their accuracy has not been validated and they are not ready for use, so value is low. Likewise, observable phenomena may be collected and accessed in the first stage from sources such as regulatory filings (e.g., tax returns and business financial reports), surveys, and other methods that require less validation and are much closer to usage in the last stage.

In principle, the valuation of own-account data should cover the cost of collection and storage (first two stages of the data value chain). The cost of processing (third stage) should also be included to the extent that processing activities are required to validate and affirm the accuracy and reliability of collecting, recording, organizing, and storing data. However, some of these costs may already be included in the valuation of the databases that include the data, and double counting should of course be avoided.

Figure 3 breaks down direct costs associated with database development, including the embedded data. The costs included in figure 3 can be matched to the first three stages of the data value chain. The last two stages of the data value chain – i.e., distribution and usage – are not considered part of the value. The first entry in green in figure 3, corresponding to the first stage of the data value chain in figure 2, is currently excluded from the value of SNA databases. Entries 2, 3, 4, and 5 in orange are included in either database or software measures. The cost of storing data includes payments for cloud storage, which are included in the sum of costs when calculating the database asset. The last entry in purple may be included in part in own-account R&D measures.

In practice, statistical compilers often estimate own-account software and databases based on labour costs for relevant occupations plus a markup for other expenses. Ideally, double counting would be avoided by assuring that labour classes are not included in more than one asset. This however may not always be feasible. If occupations relevant for software and databases overlap with occupations relevant for data as an asset, then own-account data may already be included to some extent in existing measures of software and databases. Likewise, if surveys used to measure own-account R&D include expenditures on workers that perform activities on the data value chain – such as data scientists – then own-account data may already be included to some extent in existing measures of R&D.

In the end, data will be part of a database. Pragmatic considerations will be needed when estimating the separate values of data and databases to avoid double counting, possibly based on empirical evidence. Table 1 provides an overview of the specific costs attributable to data, databases and software.

Table 1: Overview of costs attributable to data, database, database software

Data (digitalised information content)	**Database**	**Database Software**
Own account production		
Costs of collecting or acquiring observable phenomena, storing and processing data.	Costs for work on manipulating or analysing existing data in the database; this includes the cost of preparing data in a format that conforms to the database; cost of storing data, including payments for cloud storage.	Costs for developing and programming the database management system. Costs of additional tools used to analyse data (e.g. software algorithms).
Sale/Purchase		
Acquisition of a whole database including data.		Acquisition of standard/readymade database management software. Acquisition of individual/customised database management software.
Payments for license to access		
Access to the data in the database. Payments are output, or final or intermediate use, respectively		

Economic depreciation and revaluation

The perpetual inventory method (PIM) is recommended in the SNA for stocks of assets. If data are to be recorded as a separate asset, using this method for data would allow consistency with the other IPP stock estimation[8]. The PIM requires appropriate service lives and depreciation profiles. The service life of data clearly varies substantially, with some data having a much longer useful life span than other data. Therefore, it is conceptually possible for different data to have different service lives and depreciation schedules. If data as an asset is valued and recorded separately from databases, ideally national statistical institutes should strive to obtain detailed information on

[8] Caution should be used in the case of short time-series, for which the reliability of the PIM results may be questionable.

service lives and depreciation patterns from data-owner companies, if they capitalise data in their financial statements, which often may not be the case. As an alternative, the feasibility of empirically estimating services lives could be investigated. In the end, for practical reasons, data will likely have to be aggregated and treated in an 'average' way. In the worst case scenario where no better information is available, a geometric depreciation function with a common estimated service life (e.g. 10 years, as Eurostat and the OECD's approach to R&D) should be used by all countries, as a way to ensure consistency.

If instead data assets are combined with databases (and software), then the same service lives could be assumed for data. The Joint Eurostat-OECD Task Force on Land and Other Non-Financial Assets found in a survey that many countries use geometric depreciation for R&D and software, which is also an option for data as an asset. Across a broad cohort of data, most data will likely experience positive depreciation due to obsolescence. In concept, data assets (as well as other assets) could increase in value. Like other assets, e.g. cultivated biological assets that often enter service before they are fully mature, or ships that often need a 'shakedown' period to identify early flaws before they are fully reliable, data may experience increases in value early in the service lives. One way to handle this type of early negative depreciation is by averaging the depreciation rate over the entire lifespan and then using the average depreciation rate in calculation.

Increases in the value of a data asset may also be caused by events such as the discovery of a new application for data, better technology for processing data, or a change in data prices. The discovery of a new use for the data should be treated as an 'other change in volume'. This is similar to the treatment of land that is re-zoned for a new use in current SNA (e.g., farmland that is re-zoned as residential). Changes in data prices, by contrast, should be treated in the SNA as a revaluation because they do not arise from production.

3.6. Treatment of 'observable phenomena'

The definition of data in Chapter 1 refers to 'observable phenomena' as the non-produced input for the production of data. *An **observable phenomenon** is the occurrence of a singular event or piece of information.*[9]

Observable phenomena are ubiquitous and can result from events, interactions and participation by actors in the economy. Importantly, they can come into existence through the actions of, or information about, a single person; as the direct result of an interaction between two parties; or as a by-product of production. Some observable phenomena such as telephone numbers or social security numbers may be considered as produced. However, for such items the relevant production is already included in the market or non-market output of the producing unit (the telephone company or social security institution). Once they have been produced, telephone numbers or social security numbers simply exist and become observable phenomena that can be captured to produce data once or several times. Therefore, in this context they are to be treated as non-produced observable phenomena.

[9] See footnote 3.

While a certain single observable phenomenon may be valuable, the vast majority are of zero or minimal value. Observable phenomena have two important characteristics that are useful for classifying them within the national accounts:

a. Ignoring the very rare occurrences of a single valuable observable phenomenon, most individual observable phenomena by themselves do not represent a 'store of value representing a benefit or series of benefits accruing to the economic owner by holding or using the entity over a period of time'. It should also be borne in mind that observable phenomena are usually available in very large quantities, hence the value of one of them individually considered becomes virtually zero. Furthermore, many observable phenomena may be interchangeable, as a consequence of which one additional piece of information does not add much value, unless it contains very scarce information. For all these reasons, single observable phenomena do not meet the definition of an asset.

b. A single specific observable phenomenon is not the direct result of a production process. As mentioned, many observable phenomena are simply facts about a person, including e.g. data on age, occupation, family situation, etc. The individual does not engage in a productive activity to generate these facts. Other observable phenomena can come into existence as a by-product of production; however, since the purpose of the production was not to produce them, these should be considered an externality and therefore 'no values are imputed for them in the SNA' (SNA 2008, para. 6.47).

This means that observable phenomena, as defined here, fall outside the SNA production and asset boundaries[10].

While observable phenomena are considered outside the scope of the SNA production and asset boundaries (except for the exceptional cases of valuable individual observations), their possible treatment in a satellite account goes hand in hand with the treatment of free digital products.

4. Practical considerations

This paper proposes solutions for the recording and valuation of data in national accounts that are considered the most appropriate from a conceptual point of view, in line with the indication of the Advisory Expert Group on National Accounts in October 2019. However, some of them raise potential issues regarding their practical implementation and should be subject to test against possible fallback solutions, also already considered in this paper.

As a summary of the argumentation in the previous chapters of this paper, practical considerations are needed at least for:

a) The valuation of own-account data. Ideally, all the cost components should be considered. These include collecting and recording the underlying observable phenomena and storing and processing data. Tests should explore if the costs of recording, storing and processing

[10] An exception are observable phenomena which are explicitly exchanged on the market thus becoming an asset.

data can be separately estimated from the costs of the same operations for the production of the databases to which the corresponding data belong. The fallback option is to value data only for the costs of acquisition of observable phenomena, currently not included in databases. This may however lead to a considerable underestimation of the value of data (but to a more limited impact on overall production as long as the estimation of databases covers those costs);

b) The recording of data as an asset. The favoured option to record data separate from databases depends on the possibility to obtain reliable distinct estimates. The fallback solution is to enlarge the database asset category to also include the (additional) value of data.

c) The choice of depreciation patterns. Ideally, specific service lives should be identified for each data asset. Empirical tests should provide evidence on the feasibility to identify them. A fallback solution is to adopt a common service life for data assets in general (e.g. 10 years). However, should data be included in the accounts together with databases, the ultimate option may be to adopt for data the same service life used for databases.

5. Open questions

1. Do you agree with the proposed distinction between produced data and non-produced observable phenomena?
2. Do you agree to limit the focus to digital data?
3. Does the proposed definition of data omit any relevant element?
4. Do you agree that data belong to their producer (unless they are sold/licensed) and not to the persons/households to which the underlying observable phenomena refer to?
5. Do you agree that long-lived data should ideally be recorded as an asset category separate from databases, subject to its feasibility?
6. Do you agree that the cost of producing own-account data should ideally include the cost of acquiring data as well as the costs of storing and processing data?
7. Do you agree with using the PIM to estimate the depreciation of data assets?

References

Ahmad, Nadim. 2004. "The Measurement of Databases in the National Accounts." Paper prepared for the December 2004 Meeting of the Advisory Expert Group on National Accounts.

Ahmad, Nadim. 2005. "Follow-Up to the Measurement of Databases in the National Accounts." Paper prepared for the July 2005 SNA Update Issue 12.

Ahmad, Nadim and Paul Schreyer. 2016. "Measuring GDP in a Digitalised Economy." OECD Statistics Working Papers, 2016/07, OECD Publishing, Paris.

Ahmad, Nadim and Peter van de Ven. 2018. "Recording and Measuring Data in the System of National Accounts." Paper prepared for the OECD Meeting of the Informal Advisory Group on Measuring GDP in a Digitalized Economy, Paris France. November.

Brynjolfsson, Erik, Avinash Collis, W. Erwin Diewert, Felix Eggers, and Kevin J. Fox. 2020. "Measuring the Impact of Free Goods on Real Household Consumption."

Brynjolffson, Erik, Avinash Collis, W. Erwin Diewert, Fliex Eggers, and Kevin J. Fox. 2019. "GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy." NBER working paper number 25695.

Corrado, Carol A. and Janet X. Hao. 2014. "Brands as Productive Assets: Concepts, Measurement, and Global Trends." WIPO Economic Research Working Papers 13, World Intellectual Property Organization – Economics and Statistics Division.

European Commission. 2020. "Final Report on Intellectual Property Products." Paper prepared by the Joint Eurostat-OECD Task Force on Land and Other Non-Financial Assets for the 20th Meeting of the Directors of Macroeconomic Statistics, Luxembourg.

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.

Heys, Richard. 2020. "The Impact of Digitalization on the National Accounts and the Satellite Accounts." Paper prepared for the SNA Subgroup on Digitalization.

Li, Wendy C.Y., Makoto Nirei, and Kazufumi Yamana. 2018. "Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy." BEA working paper.

Nakamura, Leonard, Jon D. Samuels, and Rachel Soloveichik. 2017. "Measuring the 'Free' Digital Economy within the GDP and Productivity Accounts." BEA working paper.

Nijmeijer, Henk. 2018. "Issue Paper on Databases." Paper prepared for the Joint Eurostat-OECD Task Force on Land and Other Non-Financial Assets.

Organization for Economic Cooperation and Development. 2019. "Guidelines for Supply-Use Tables for the Digital Economy." Paper prepared for the Informal Advisory Group on Measuring GDP in a Digitalized Economy.

Organization for Economic Cooperation and Development. 2020. "Measuring Data Products." Paper prepared for the *SNA* Subgroup on Digitalization.

Rassier, Dylan G., Robert J. Kornfeld, and Erich H. Strassner. 2019. "Treatment of Data in National Accounts." Paper prepared for the BEA Advisory Committee, Washington DC. May.

Reinsdorf, Marshal and Jennifer Ribarsky. 2019. "Measuring the Digital Economy in Macroeconomic Statistics: The Role of Data." Paper presented at the ASSA 2020 Annual Meeting, San Diego, CA.

Rincon-Aznar, Ana, Rebecca Riley and Garry Young. 2017. "Academic Review of Asset Lives in the UK - Report to the Office for National Statistics". National Institute of Economic and Social Research, NIESR Discussion Paper No. 474.

Robbins, Carol A., Gizem Korkmaz, José Bayoán Santiago Calderón, Claire Kelling, Stephanie Shipp, and Sallie Keller. 2018. "The Scope and Impact of Open Source Software: A Framework for Analysis and Preliminary Cost Estimates." Paper prepared for the 35th IARIW General Conference, Copenhagen.

Soloveichik, Rachel. 2020. "'Free' Content in the National Accounts and the Satellite Accounts." Paper prepared for the SNA Subgroup on Digitalization.

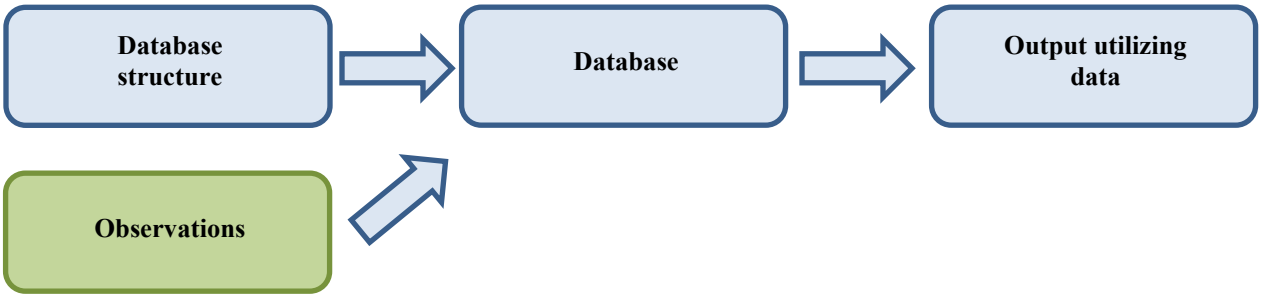
Statistics Canada. 2019. "Measuring Investment in Data, Databases, and Data Science: Conceptual Framework." *Latest Developments in the Canadian National Accounts*.

U.S. Bureau of Economic Analysis. 2020. "Issue Paper on Recording and Valuation of Data, Free Assets, and Free Services in National Accounts". Paper prepared for the SNA Subgroup on Digitalization.

Van Elp, Martin and Nino Mushkudiani. 2019. "Free Services." Presentation for the OECD Joint Meeting of the WPFS and WPNA, Paris France. November.

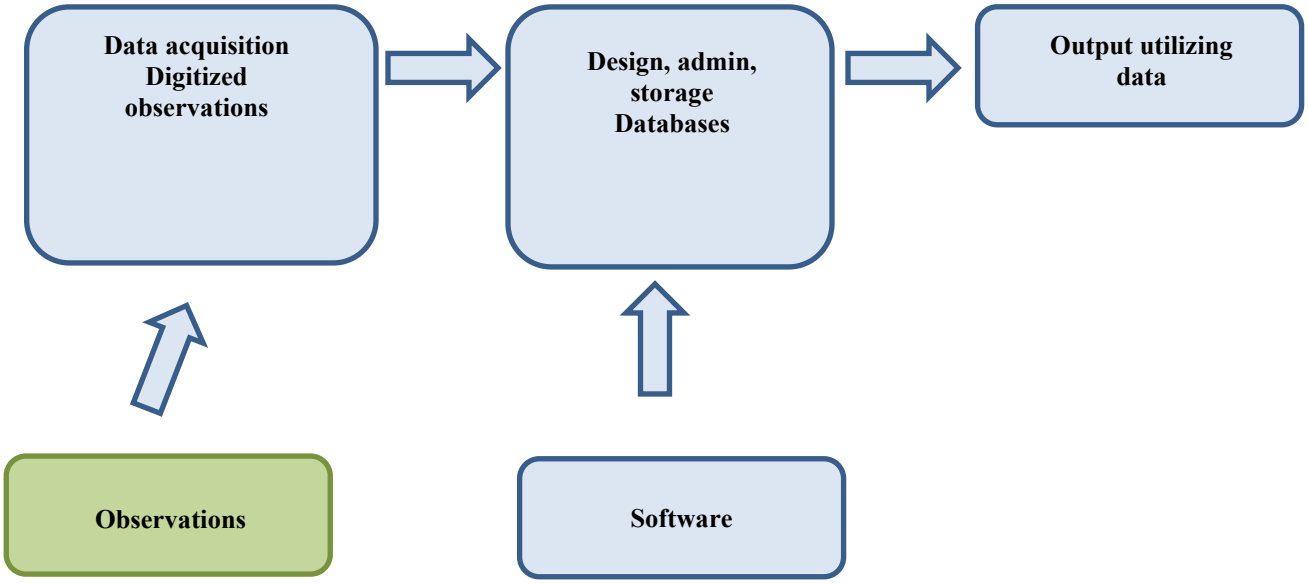
Figures

Figure 1a: Database Structure, Observations, and Databases



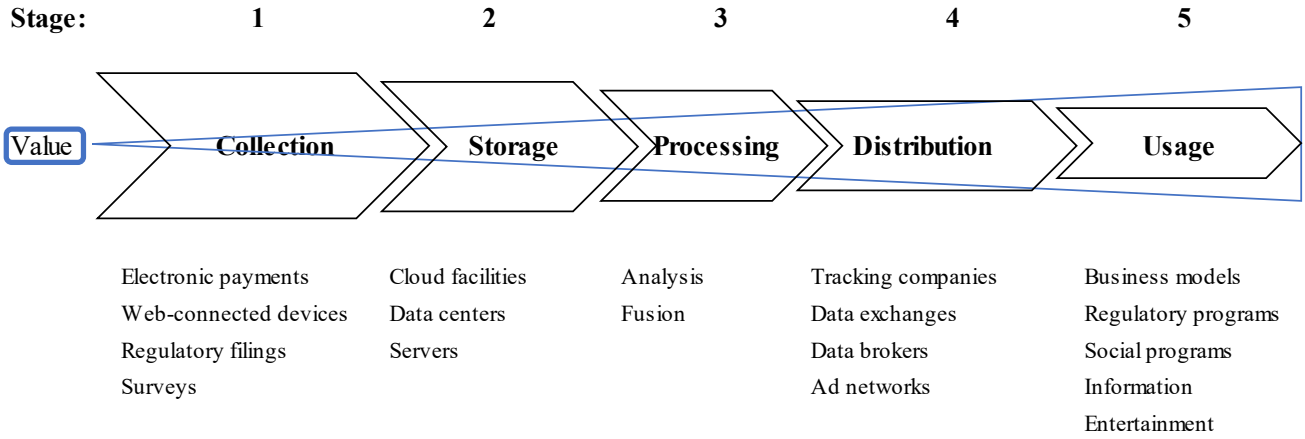
Source: OECD (2020).

Figure 1b: Observations, Data, Databases, and software



Source: IMF (2020) based on OECD (2020).

Figure 2: Data Value Chain



Source: Adapted from OECD (2013) and Visconti et al. (2017).

Figure 3: Direct Costs Associated with Developing Databases

	Cost	*2008 SNA Treatment*
1	Costs of collecting or acquiring data (survey, locate, capture, provide free services or discounts, purchase)	Excluded from database assets
2	Costs of preparing data in appropriate format for storage	Included in database assets
3	Costs of storing data	Included in database assets
4	Costs of designing the DBMS or purchasing database management services	Included in software assets
5	Costs of tools used to analyze data (e.g., software, algorithms)	Included in software assets
6	Costs of analyzing data (including data validation, cleaning, contextualizing)	May be partly included in R&D assets

Source: Adapted from the International Monetary Fund.