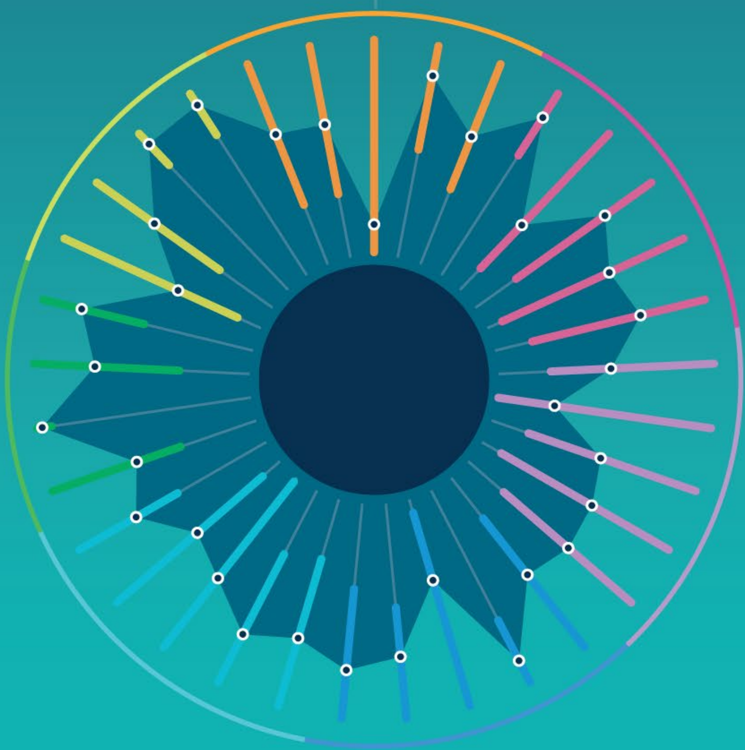
OECD GOING DIGITAL TOOLKIT

MEASUREMENT NOTE

Measuring the economic value of data



OECD
BETTER POLICIES FOR BETTER LIVES

going digital

This Toolkit note was written by John Mitchell, Daniel Ker and Molly Lesher. It was reviewed by the Working Party on the Measurement and Analysis of the Digital Economy and the Working Party on National Accounts, and it was declassified by the Committee on Digital Economy Policy on 2 December 2021. The note was prepared for publication by the OECD Secretariat.

This Toolkit note is a contribution to the OECD Going Digital project, which aims to provide policy makers with the tools they need to help their economies and societies thrive in an increasingly digital and data-driven world.

For more information, visit www.oecd.org/going-digital.

#GoingDigital

Please cite this publication as:

Mitchell, J., D. Ker and M. Lesher (2021), "Measuring the economic value of data", *Going Digital Toolkit Note*, No. 20, https://goingdigital.oecd.org/data/notes/No20_ToolkitNote_MeasuringtheValueofData.pdf.

Note to Delegations:

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

DSTI/CDEP/GD(2021)2/FINAL.

Table of Contents

Tables

Figures

Boxes

Measuring the economic value of data

As data have become a social and economic resource, including for value creation, decision-making, innovation and production, policy makers are facing a number of challenges. Among the most important issues – but also one that is particularly complex – is how to measure the economic value of data to provide a solid evidence base for policymaking. This Going Digital Toolkit note brings clarity about what is meant by the term "data" in the context of efforts to conceptualise and measure the value of data from a statistical perspective. The note also highlights why estimating the value of data is increasingly important, identifies the conceptual and practical measurement challenges faced, and catalogues various innovative initiatives underway across countries in the context of the forthcoming revision of the System of National Accounts and beyond.

The digital technology ecosystem relies on data, which can have both positive and negative impacts on people, firms and governments. From an economic perspective, data underpin digital transformation and have become an important source of value, including for decision-making and production (OECD, 2019[1]). Data are an increasingly valuable asset for firms, governments, individuals and society-at-large. The availability and prevalence of data has given rise to new or significantly improved products, services and business models, and has contributed to enhancing productivity.

This phenomenon, sometimes referred to as "data-driven innovation", is growing in importance as digital transformation progresses. Across all sectors, from agriculture to energy to transportation, data are contributing to more efficient uses of resources. Data are also helping to address societal challenges, ranging from climate change to the management of natural disasters to health crises. As a result, data and data flows are of ever-increasing economic and wider policy interest. Measures of the value of data are needed to address these policy needs.

Despite its importance, however, attempts to conceptualise and measure the value of data within economic statistics remain relatively underdeveloped, within both the System of National Accounts (SNA) and other statistical approaches. The need to make progress in measuring the value of data was recognised in the Going Digital Measurement Roadmap (OECD, 2019[2]). In particular, the Roadmap highlights the need to develop taxonomies and classifications of data for statistical purposes, further study the role of data in business models and processes, and improve the measurement of knowledge-based assets including data and their role in production, productivity and competitiveness.

This Going Digital Toolkit note sets out why finding measures for the value of data is increasingly important, identifies the conceptual and measurement challenges faced, and catalogues various initiatives underway across countries. This note also highlights innovative, practical approaches to measuring the value of data within the SNA and beyond. Such efforts hold much promise to provide a better evidence base on which policy makers can formulate data policies going forward. This note does not discuss the "value" of data – both positive and negative – from a social welfare perspective.

What is "data"?

The term "data" is used flexibly both in policy discussions and in academic literature, often without a clear definition of what precisely is being talked about. In the context of the SNA, the term is sometimes used in reference to individual records of the most basic facts or *observable phenomena*[1] (e.g. e-commerce shopping information on purchases made by an individual on a given day). It may also be used to refer to wider sets of similar data (e.g. e-commerce shopping information in a specific country over time or in multiple countries). In addition while analogue and quantum representations of data exist, it is usually implicit (if not explicit) that the data of interest are in digital form. This is the case for purposes of measuring data to assess its economic value (such as within the SNA), where analogue and quantum representations are too small to matter statistically.

References to "data" are also often used with respect to vast datasets containing not only information on a large number of observations (e.g. e-commerce shopping information collected every day in countries around the globe), but also statistics and indicators derived from these data. While these examples all involve facts and statistics collected together, they differ considerably in important ways: 1) the volume and nature of productive activities involved in their creation, 2) the amount of information that they contain, and 3) their usefulness to the broader community and potential monetary value to their owner.

"Data" may even be used as a shorthand to refer to an organisation's entire business model. For example, companies in the USD 7 billion private weather forecasting industry (Forrester, 2020[3]) might be said to be in the business of providing "data" (e.g. on temperatures) to their clients. However, a significant part of the "data products" they produce is likely to consist of forecasts and accompanying commentary that are derived – but far removed – from the individual observable phenomena. The firm could not exist without data, but that does not mean that the firm's entire "value" (e.g. market capitalisation) is equal to the value of the underlying data.

The lack of a single definition of data and its scope is one of the key challenges in determining a framework through which to value "data"[2]. This Toolkit note uses an OECD working definition of data proposed to the Inter-Secretariat Working Group on National Accounts (ISWGNA) Advisory Expert Group (AEG): *Information content that is produced by accessing observable phenomena and recording, organising and storing relevant information elements from these in a digital format, which can be accessed electronically for reference or processing"*

[1] Observable phenomena is defined in this note as "a fact or situation, whose characteristics and attributes may be recorded".

[2] A recently agreed upon definition of data in the policy context is: "Data refers to recorded information in structured or unstructured formats, including text, images, sound, and video" (OECD, 2021[27]).

(OECD, 2021[4]) [3] Under this definition – which was created explicitly for the purpose of including the production and value of data in economic statistics – data is delineated from observable phenomena (i.e. the facts and situations whose attributes can be observed and recorded digitally, and that underpin the production of data).

Care should be taken to avoid equating "data" with "bandwidth", which is the volume of information that can be sent over an online network in a given amount of time (megabits per second (Mbps)). Bandwidth does not relate to the amount or value of data produced or consumed; data and bandwidth exist at different and largely unrelated layers.

Data underpins value creation in many business models

Many of the world's most well-known companies today were born digital – bringing together inputs that include data analytics, digital marketing, and increasingly artificial intelligence (AI) – to develop business models that are entirely "data-enabled" (Nguyen and Paczos, 2020[5]). This approach has helped some of these companies and their digital products to achieve vast and even global scale and become "household names" – or indeed widely understood verbs (e.g. "to Google", "to Photoshop", "to Skype"). These firms' business models are reliant on data[4]" and are now generally accepted as viable concepts. Indeed, many of these "data-driven companies" listed on U.S. stock exchanges – including these well-known firms – appear to grow in market value more rapidly than the wider population of firms (Ker and Mazzini, 2020[6])

Online platforms are perhaps the typical example of firms and business models that are centred around digital transactions and heavily reliant on data. These platforms, defined as "digital services that facilitate interactions between two or more distinct but interdependent sets of users (whether firms or individuals) who interact through the service via the Internet"[5] (OECD, 2019[7]), have facilitated an ever wider scope of economic activity, with households joining businesses in engaging in productive activities. These include creating digital content and providing other services that were previously the preserve of

[3] This definition expands on a previous one provided in a draft ISWGNA guidance note which defined data as "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" (ISWGNA Advisory Expert Group on National Accounts, 2020[22]).

[4] In this Toolkit note, "data" is used to refer to data in digital form. Digital data differ from analogue data insofar that they can be used, re-used, copied, moved, and processed cheaply, without degradation, and very fast (OECD, 2019[1]).

[5] This definition excludes businesses such as direct business-to-consumer (B2C) e-commerce and ad-free content streaming, as those serve only one set of customers. It does, however, include businesses such as third party B2C e-commerce and ad-supported content streaming, because those services involve two separate sets of users.

"professionals" or "specialists" (e.g. courier services). Indeed, Li and Chi (Li and Chi, 2021[8]) finds that the entry of online platforms (with their data-driven business models) in the U.S. hospitality and transportation industries was most disruptive for incumbents that were less digitalised and had comparatively limited access to data.

Beyond these "digital natives", many existing firms are embracing the use of data-based inputs, databases, data analysis and online platforms to improve their existing products and business processes (OECD, 2019[9]), (OECD, 2021[9]). Such firms have been coined "data-enhanced businesses" (Nguyen and Paczos, 2020[5]). For example, financial services firms have adopted electronic ledgers, databases, secure networks, digital delivery (i.e. online banking), and the like to enhance their products and processes almost beyond recognition – unlocking efficiencies, spurring innovation, and driving profits (OECD, 2018[10]), (OECD, 2020[12]). More universal examples across industries include:

- Databases, which are routinely used as an efficient means to store and manage information that is essential to a business' activities on customers, suppliers, transactions and personnel, among others.
- Phone or Internet services, which involve the creation and transfer of data packets, are crucial tools for almost all businesses, regardless of size and industry.
- E-commerce, enabled by flows of data between buyers and sellers providing product details, order information, etc., underpins an ever-increasing share of transactions. Furthermore, online platforms are leveraging e-commerce, along with data-driven business models, to find new ways of generating economic value (OECD, 2019[12]).

Data are thus vital to creating, accessing, disrupting, increasing and shaping markets, and maximising economic value generation. However, hand in hand with any additional value generation comes the possibility that future fiscal policies and frameworks may include the source of data generation as one component in determining any final tax imposition.

Furthermore, the value of data goes beyond the relatively narrow perspective of influencing firms' profitability, market valuations or tax burden. The "data revolution" is not limited to businesses – public sector and non-profit organisations are also embracing data-based inputs, databases, data analysis, and online platforms to reduce costs, improve efficiency, and find innovative, data-driven solutions to a range of societal challenges. Indeed, data can contribute to "social welfare" in a broad sense and in a wide variety of ways – encompassing not only the profitability of businesses, but also how individuals' needs are met (including but not limited to the need for income), and non-monetary benefits such as convenience or health (Coyle et al., 2020[13]).

This multi-faceted socio-economic role of data, and the need to make informed policy decisions that touch upon many different policy domains (e.g. privacy, security, labour markets, tax, and trade, among many others), is a key reason why conceptualising and measuring the value of data is of great interest. Nevertheless, the nature and features of data, as well as the relative scarcity of "data about data", do not make this straightforward. The following sections consider conceptual and practical challenges of assigning an economic value to data, before various efforts to measure and value data and data flows are presented.

Conceptual challenges around valuing data

It is unlikely that the *value* of data will be directly related to its *volume* in any routine or systematic way. At a very detailed level, two records in a company's customer database may well take up the same amount of file storage space. However, the contact details for the purchasing manager at a firm's biggest customer (i.e. making a big contribution to their bottom line) is likely to be of more "value" to the firm than the fax number for a potential customer who was approached but declined (i.e. contributing only cost to the bottom line).

At a much broader level, a vast database coupled with AI to track the movements of a particular group of people within a country or region might create debatable economic or social value, while a smaller database coupled with AI to search for treatments for COVID-19 might yield tremendous economic and social value (OECD, 2020[15]). Meanwhile, the data flow arising from two 10-minute phone calls may be identical, but one call might clinch a deal between small and medium sized enterprises (SMEs) while the other might avert a war between states. As such, the *volume* of data – which is at least theoretically straightforward to measure as it at least comes in well-defined units – is unlikely to provide a good basis for inferring the *value* of data holdings and flows.

Nevertheless, the above suggests two crucial components of data value:

- *Content* – the information the data embodies, and
- *Context* – the context in which the data were gathered and are (or could be) held, analysed, and used.

The information *content* of the data will ultimately determine the gamut of what it can potentially be used for, and thereby the economic and social value it could generate. For example, a dataset of star locations and sizes does not have much potential to be of use in planning and optimising suburban bus routes to maximise bus company profits and minimise local air pollution. As such, the information *content* strongly influences the economic and social value that could arise from that data. Meanwhile, the *context* in which the data were collected, in which they are stored, and in which they are used will be important in determining how much of this potential value is ultimately realised.

Content and *context* are inextricably linked – especially as most data are gathered or generated and stored for an entity's own use (i.e. internally to a given business or other organisation) and so tailored to use by that organisation (Coyle and Li, 2021[15]). The information content of the data in many cases is likely to result from the context of its intended use, including what the organisation obtaining the data intends to use them for, the tools, skills, and expertise available to analyse and interpret the data, and how resulting findings and knowledge will be applied.

The context of collection will also affect the information content and quality of the data. For example, measurements and observations may have been recorded by specialist personnel or instruments, or have been reported by private individuals. Depending on the phenomena being recorded (e.g. precise measurements of physical characteristics or personal opinions), one or another of these may result in "better" data.

Content and *context* are also crucial to measuring value at a more practical level. Additionally, the context within which a certain datum or dataset exists can evolve over time. Consider the aforementioned entry in a company's client database: The information content is designed to facilitate the firm's business with a particular client. It is understandable to the company because it is adapted to the context in which it is used – from being in a format that works with the firm's technology systems to being structured according to the firm's processes. In other words, the data are shaped by, and an important component of, the firm's accumulated information, or organisational capital (Li and Chi, 2021[8]) – the knowledge, know-how, and business practices embedded in the firm (including in its managers and employees) (Squicciarini and Le Mouel, 2012[16]).

While such a client record is likely to be especially adapted and relevant to the context of its intended use, it may nevertheless be of use to others. For example, the information in that client record may also have an obvious potential value to a competitor firm, or to other agents such as tax authorities, provided it is in (or can be converted to) a format that they can read and interpret.

Timeliness – and the time-period covered by a dataset – is also a factor affecting both content and context and therefore value. If the record relates not to a present-day client but one that went out of business a month, a year, or a decade ago, this will again affect its value (or potential value) to the aforementioned various actors. Such "old" information may nevertheless be of value to a researcher looking at the dynamics of firm closures, or as an entry in a large dataset being used to train AI.

Two of the economic features of data (Box 1) have especially important implications for the context in which they should be viewed and their potential to generate economic and social value, as well as the extent to which this is realised. Firstly, data are *non-rivalrous*; that is, the use of data by one agent does not reduce the quantity of data available for potential use by others. As such, various uses for the client record mentioned above need not be mutually exclusive.

Box 1. Economic characteristics of data

In addition to the information content of data and the context around it, a number of economic characteristics are also relevant:

Data are non-rival. Many agents can make use of the same data at the same time without it being "used up" or degraded. This means that the "ownership" of data and data-related transactions exhibit important differences compared to those for typical goods and services.

Data can be excludable. It is likely hard to exclude agents from data that is relatively easy to collect, such as data scraped from websites, whereas client or administrative data is likely to have tightly restricted access.

Data involve externalities. Positive externalities can arise when datasets are combined, making the "sum of greater value than the parts". By contrast, negative externalities arise when the collection or use of data leads to harm (e.g. exclusion from healthcare coverage). Strong incentives to use data intensively can undermining privacy, for example. At the same time, weak incentives to use data can lead to missed opportunities for generating economic and social value. Data can also arise as an externality, often as a by-product of the standard production process.

Data may have increasing or decreasing returns to scale. Sometimes, collecting more data can provide additional insights, but at other times more data adds little extra value (and likely some cost) but some organisations continue to accumulate and potentially hoard it.

Data has a large option value. It is hard to predict how the value of data might change. New data, new technologies or algorithms, and new measurement methodologies mean existing data can have unpredictable future importance. Organisations may keep data for their potential rather than current value.

Data collection may have high up-front cost and low marginal cost. Collecting data can entail investment in hardware (such as sensors) and software, among other costs, but on-going collection can be cheap. This can create barriers to beginning to collect data that would be useful or potentially financially lucrative. Data use requires complementary and often on-going expenditures. Organisations may need to invest in complementary hardware and software to be able to process and analyse the data. These may also rely on complementary specialist skills. These can create further disincentives to collecting and using data.

Data about people have particular features. Information about one person will often have no meaningful value to an organisation, but may well be highly valued in principle by the individual. People create positive and negative incentives and impacts for others as they share information about themselves (e.g. through social networks or genetic analysis services). Furthermore, personal data are generally subject to additional legal controls, with compliance costs acting as another potential barrier to data collection.

Several of the features above point to costs and risks associated with collection and use of data. Often, these will be relatively straightforward to understand and quantify. By contrast, the benefits of having and using the data may well be less clear.

***Source*:** (Coyle et al., 2020[13]) (OECD, 2019[1]).

However, economic rents can be generated through holding and exploiting data – and these may be finite (e.g. limited due to market-size). This can create strong incentives to exploit another feature of data – its potential *excludability* – in order to keep others from accessing it. Furthermore, as most data are gathered on own account (i.e. internally to a given business or other organisation) (Coyle and Li, 2021[15]), even if a data holder would be willing to agree terms for access, parties that would be interested may often not even know that the data exist. Legislation also sets limits on data sharing. The barriers to demand- and supply- driven data sharing can therefore be extremely high and, as a result, opportunities for innovation and growth are sometimes limited by a lack of access to data (Furman, 2019[17]), even though some of those restrictions have public policy objectives (e.g. privacy protection).

The importance of the specific *content* of each individual dataset and the *context* around it – including how the data were gathered, how they are stored, the formats they are in, who can access them and under what terms, etc. – means that data cannot be considered as a monolithic mass. Indeed, the opposite is true. This feeds into the practical measurement challenges as well as the conceptual challenges faced when attempting to incorporate a data product into established statistical classifications and frameworks – notably the SNA.

Valuing data in the SNA

Within the SNA framework, there are limited examples of data being specifically represented in any balance sheet or estimate of capital stock, despite data being generally perceived as a fundamental driver of production and growth for many businesses. These businesses include both those with data-driven business models ("data-enabled businesses"), as well traditional businesses that leverage data to improve efficiency or lower costs ("data-enhanced businesses") (Nguyen and Paczos, 2020[5]).

This absence in itself is surprising, as data appears very much to be an asset, both from a broad business accounting perspective and on a more specific 2008 SNA basis. Within the 2008 SNA, as well as other macro-economic statistical frameworks, data would appear to meet the clear definition of assets 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" (European Commission, International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), United Nations and World Bank, 2009[18]).[6]

Additionally, for reasons elaborated upon below, it is important to note that these assets are separated into "produced" and "non-produced" assets. That is, those "that have come into existence as outputs from production processes that fall within the production boundary of the SNA" and those "that have come into existence in ways other than through processes of production" (European Commission, International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), United Nations and World Bank, 2009[18]) These "other ways" generally mean "naturally occurring". Land is the classic example of a non-produced asset.

While the SNA includes inventories and valuables[7] as categories of produced assets, the vast majority of produced assets used in the economy are fixed assets. They are *produced* once and then used repeatedly as an input into production, through which they steadily decline in value over an extended period. Machinery and equipment, buildings, and computer hardware are all examples of fixed assets. All of these items provide repeated benefit to the economic owner from their use. Data would also appear to fit into this category.

However, data are different. While they can act like a fixed asset in repeatedly contributing to production over time, the unique characteristics of data mean that the conventional analysis used to assess the value of physical assets and commodities cannot effectively capture the value of data. Perhaps the two most significant characteristics are the unique way that the production of data is often never fully "completed", and that there may also be a blurring of the

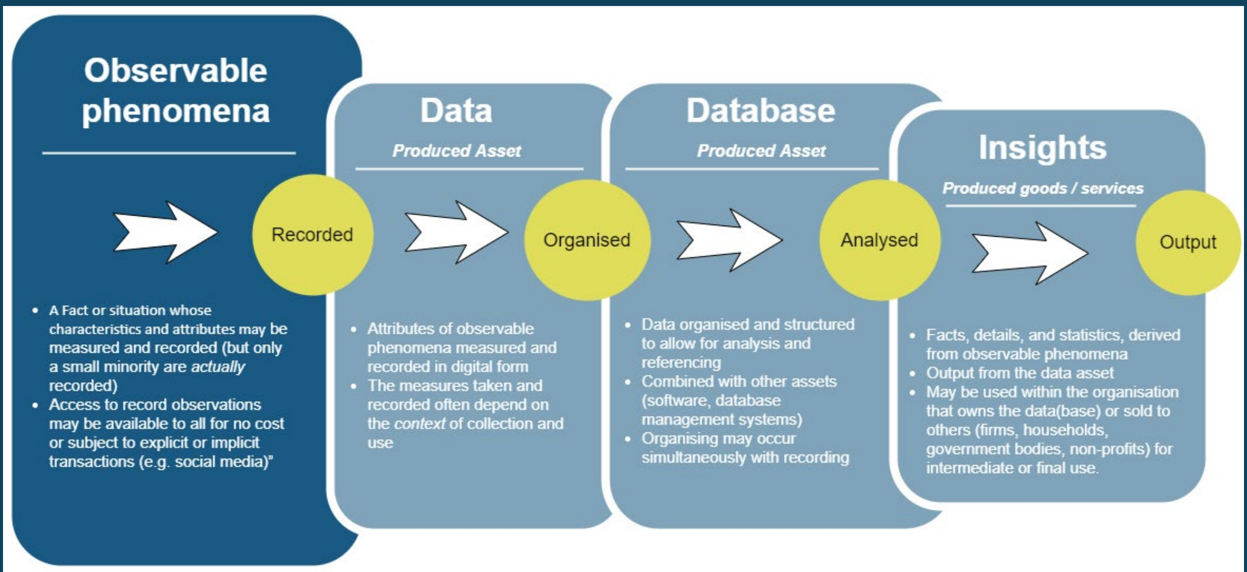
[6] "Over a period of time" is usually considered to mean at least one year.

[7] Valuables are a specific category of produced assets that are "acquired as stores of value: They are not used up in production and do not deteriorate physically over time." (European Commission, International Monetary Fund (IMF), Organisation for Economic Co-operation and Development (OECD), United Nations and World Bank, 2009[18] §6.214). Examples include precious jewellery, works of art, memorabilia, etc.

line between the value created as a result of production and value stemming from the observable phenomena that underpin data products.

The idea that observable phenomena underpin data is fundamental to understanding not just the conceptual challenges, but also the practical challenges of measurement and valuation. The data-information chain (Figure 1), inspired by a conceptualisation produced by Statistics Canada, sets out how agents accumulate elements of information by obtaining access to observable phenomena in order to measure and record observations about attributes of interest (Statistics Canada, 2019[19]). This act of measuring and recording (in digital form) "generates" data describing the observable phenomena. These observations are recorded, organised into databases, and subsequently analysed to gain insight that can be exploited by businesses or other agents.

Figure 1. Data-information chain from a SNA perspective



Source: Authors.
Note: For pragmatic reasons, analogue data is excluded from the national accounts asset boundary.

In Figure 1, *observable phenomena* are a crucial foundation for data by providing a subject for observation and measurement. Observable phenomena can be "naturally occurring", such as the temperatures at a given outdoor location at a certain time, or occur as a corollary of productive activity, such as the operating temperature of an aircraft engine during flight or the existence of each car coming off a production line. These latter items *relate to* productive activity – and each car or flight is already reflected in measures of output and GDP – but the production gives rise to new observable phenomena (facts and situations, i.e., the characteristics of the car coming from the factory, the aircraft engine temperature) of which there is an *opportunity* to take observations and measurements.

Taking and recording (in digital form) measurements about the attributes of observable phenomena requires specific action. The result is *data* – the attributes of observable phenomena measured and recorded in digital form. To be clear, "data" as conceptualised in Figure 1 meets the definition of "*information content that is produced…in a digital format, for reference or processing"* but arguably refers to data in its most "basic" form. At this point, it has yet to be linked with other data and organised in a database; actions which enhance the overall information content (and value) of the data asset. Nevertheless, due to the initial active productive step involved in measuring and recording observations about attributes of observable phenomena, data would be considered as a form of produced product (and assets if they meet the SNA asset criteria).

The attributes of observable phenomena may be measured and recorded in various ways; often machines (such as sensors) are used for this purpose. In other cases, people take and record observations. In yet another case, people may themselves indicate that they are an entity that exists (i.e. an observable phenomena) and report, for example, the date of the event of their birth (which was an observable phenomena in itself, as recorded on their birth certificate, and also gave rise to an attribute of the person – their birthday). Such actions are common when signing up to social networks or applying for financial services products, for example. Importantly, measurement and recording (in digital form) cannot happen "naturally", it requires some sort of action (e.g. installing a sensor). Furthermore, a specific decision has to be taken about what information should be measured and recorded (e.g. what type of sensor to install and where). This adds further weight to the assertion that data are the result of productive activities.

The recorded measures and observations comprising data are organised and structured within *databases* to enable analysis and referencing. This relies on other forms of assets including database management software. Often, the recording of an observation and organising it into a database structure occur at the same time, such as when a person creates an account on a social networking site. They may be decoupled though – for example, smart electricity meters constantly record energy usage but store these locally prior to sending them to a central system.

Through analysis of these organised and structured data, *insights*, in the form of facts, details, and statistics about the attribute of observable phenomena represented by the data can be derived. These insights, considered as output derived from the data asset, may be used within the organisation that gathered the data for monitoring and optimising business processes, or could be a direct input to the firm's main production (e.g. targeted advertising services).

Data or information may also be provided to others, for example through a market transaction such as when a company pays for access to a credit records for the purpose of assessing potential customers. This output, which could equally be provided to others at no cost, may take the form of summarised information produced for sale or simple datasets, allowing for specific information to be derived by the consumer themselves.

From this perspective, observable phenomena are considered facts or situations (such as people's personal information or characteristics of a machine) whose various attributes can be measured and recorded.[8] While some observable phenomena are "naturally occurring", others arise as a corollary of productive activity. Nevertheless, information about them is required to be obtained, recorded and organised, in order to create any data assets and allow for the subsequent outputs to be produced. Importantly, for many datasets, information about the underlying observable phenomena or additional observable phenomena, may continue to feed into the dataset on a regular basis.

Due to this cycle of periodic data updates, for many datasets it is impossible to state that the investment in them is ever truly "complete"[9] – rather, it is simply waiting for the next "tranche" of observable phenomena to occur and the information about them to be recorded. This is one reason why firms may be willing to invest in the high initial costs of starting to collect data, as they are aware that it will continue to be generated on an ongoing basis that has the potential to deliver ongoing value. However, even in its current "incomplete" state, the data already compiled are likely to be feeding into production, thereby generating benefits to their owner and meeting the definition of an asset.

Does the information obtained from accessing each additional recorded observable phenomena, when added into the data asset, create a brand new different data asset? The reasonable answer is likely no, but at the same time, the value of the data asset can change (potentially considerably) with the addition of each newly recorded observable phenomena. While in many cases the most recent addition may well be the most useful or "valuable" of all the observations (on account of its recentness), in many instances a portion of the overall data(set) value is likely to arise from preceding observations. In this

[8] This definition provides additional guidance and context to that provided in the draft SNA research guidance note on data presented at the 14th meeting of the ISWGNA Advisory Expert Group on National Accounts, which defined OP as *"the occurrence of a singular event or piece of information"*.

[9] This idea of incomplete data should not be thought of from only a chronological perspective, where the next data point is simply the most recent. Data can be improved (or made more complete) by improving the "resolution" of the data, that is bringing in more disaggregated observations or observations that provide greater content or context to the original data.

way, a dataset could be considered as a never-ending work in progress, despite the pre-existing data already contributing to production.

Due to this, it is very conceivable that data assets may continue to become more useful, and thus more valuable, with each observable phenomena (or tranche of observable phenomena) that is added to, even after already being used in production for a period. Unlike a traditional fixed asset, the value of which declines in a constant, although not linear, way, due to wear and tear, as well as obsolescence. While data are not susceptible to wear and tear, by contrast, obsolescence can affect some, but not all, data. This obsolescence may depend heavily on the context of use. [10] While the firm gathering the data may mainly find value in the latest observation (e.g. the latest daily output figure), which would imply a strong rate of depreciation for earlier observations, a business operations consultant advising the firm on finding efficiencies may place much more value on the time-series and complementary data items (implying a different depreciation profile for the data asset).

Therefore, a data asset may simply continue to increase in value over time as long as further observations about the relevant observable phenomena are continually fed into it. For example, a dataset containing daily temperatures recorded in Paris will theoretically increase in value (though likely very slightly) with each additional day of information that is added to it. Unless this information is deemed no longer relevant (hence why obsolescence might occur), there is no reason to think that this dataset will ever decline in value, rather its economic and social value will continue to grow as the time series is extended.

As a result, it is possible that an increase in value of a data asset may *not* only be due to new investment being higher than the decline in depreciation as is the case for traditional assets. Moreover, such a possibility may reflect the need for a new type of asset classification, separately delineated from other assets, which can appropriately show the value of data assets used in production.

The second characteristic that clearly separates data from other fixed assets and causes conceptual concerns regards the unique nature of most data assets arising due to the individual observable phenomena that underpins it. Most fixed asset categories, such as buildings, plants and machinery, and computer software, share various unifying characteristics. While the parameters governing a data asset may be similar, two pieces of data are almost never identical as observable phenomena on which the data are based usually have unique attributes. While the practical concern that it is impossible to have an accurate market rate when no two assets are alike will be addressed in the

[10] Since conceivably the change in context could result in an increase in value as well as a decrease, further guidance is likely required to delineate between traditional depreciation/obsolescence and when the asset undergoes a revaluation.

following section, the essence of this lack of homogeneity is what calls into question whether the entire value of the asset is solely the result of a productive activity. To use the terminology provided earlier, while the *context* of the data asset, including its collection, and application in production, is the result of capital and labour, its *content*, the second critical feature on which it is based, is usually independent of, or only a corollary of, productive activity.

Observable phenomena may have attributes that, when measured and recorded, are more valuable than those of other observable phenomena. When observed attributes of these two different observable phenomena are combined with others to make a data asset, its value may vary greatly based on the information about observable phenomena, its *content*, regardless of identical inputs of capital and labour having been used to measure and record information on its attributes, resulting in the asset. Due to this, is it appropriate to record additional value added (and thus production) for one data asset purely due to the fact that the attributes of observable phenomena underpinning it give rise to more valuable information rather than being the result of a different (and more valuable) production process?

To illustrate this point, and following on from the earlier example, information regarding the temperature over a period of time at the highest point in specific city is unlikely to be of value to everyone, but (assuming that it had not been made available for free), it would likely be possible to find some party willing to pay for that information (such as the local news channel). Measures of the dew point might also find a willing buyer, but as a much more specific and less widely understood observable phenomena, there is likely to be fewer potential buyers and/or buyers would offer a lower price. The labour and capital cost as well as the cost of inputs required to produce data on the air temperature and the dew point are likely to be very similar, but one is ultimately likely to be "worth" more. Since this additional value would not appear to be the result of production, it is arguable that at least a component of the assets' value could be considered non-produced.

Some similarities can be drawn with the measurement of established asset classes already defined in the national accounts – research and development and artistic originals – both of which contain a large amount of unique human information. A key difference is that for these asset classes, there is no certainty that the information contribution will actually add to the value of the asset. For every literary classic or "successful" research project, there is also a published book that sells very few copies or a research result that is not of use. By contrast, data collections are usually designed specifically to meet the needs of (and be of value to) the collecting organisation it is much more likely that the data collected have some value. In addition, valuable markets already exist for some specific data containing certain information (Coyle and Li, 2021[15]), (Ahmad and van de Ven, 2018[20]).

For data to be incorporated into the SNA in an accurate and appropriate manner, it must be made very clear exactly what is creating the value of the data assets that are becoming so fundamental to the production of output and value in the modern economy. Does the value of a data asset arise only as a result of inputs, including labour and capital? If so, should it be considered an act of production and contribute to GDP? Alternatively, should part of the asset be considered non-produced if some of the additional income being generated is due to the *content* observed from observable phenomena? The difficulties in answering this question definitively is a key reason why guidance on the matter has not already been developed.

Practical challenges around valuing data

The conceptual challenges of valuing data are also associated with various practical challenges. As most data are gathered on own account and highly integrated with a firm's organisational capital, markets for data and datasets are relatively underdeveloped. This limits the number of data sales and purchases that occur and prices that could be observed. Furthermore, the *content* and *context* of each transaction (i.e. the information the data embodies, the parties involved, how the data will be transferred or shared) would likely result in a price that is highly specific to that transaction. As such, such prices are unlikely to be representative of the value data holdings at-large.

There are very few types of data for which there exist standard, widely accepted valuations. In addition, there is no universal standard for categorising data into "types" for statistical purposes. Various categorisations exist, but they tend to be adapted to different analytical questions (Nguyen and Paczos, 2020[5]) – another manifestation of the importance of *context*. Indeed, it can be argued that there is a tension between the unique *content* and *context* of data and the statistical necessity of finding common features to compile groups that gloss over such detail. Any statistical categorisation of data types will need to successfully manage this tension in order to define a grouping that is both meaningful and operable in practice.

As a result, a more promising way forward is likely to involve finding indirect means of valuation. In particular, analysing the amounts spent on gathering, storing, maintaining, analysing, and transferring data, on labour and other inputs. Business expenditures on labour and intermediate consumption are routinely gathered as part of the structural business surveys underpinning economic statistics. However, the way in which these expenditures are aggregated and published – according to the International Standards for classifying products, industries, and occupations – poses challenges for focusing in on expenditures and activities related to data and data flows.

For one, identifying which classes within a given classification relate to data (or are most likely to) is usually difficult, as is finding a basis for deciding what portion of categories that combine data-related items with others should be taken (Ker and

Mazzini, 2020[6]). In addition, avoiding double counting of the same inputs for multiple related activities (e.g. software, data, research and development) and related concepts (e.g. human capital, organisational capital) is likely to pose considerable challenges. Nevertheless, various efforts are being made to begin to grapple with these issues.

From a SNA perspective, there are extensive practical issues that create challenges for national statistics organisations as they try to estimate both the level of production involved in creating new data assets, as well as the value of the data assets that are already in use across the economy. It is important to note that there is no definition of "data" within the 2008 SNA. A recent draft issues paper prepared for the ISWGNA Advisory Expert Group (AEG) defines data as "information content that is produced by accessing observable phenomena and recording, organising and storing relevant information elements from these in a digital format, which can be accessed electronically for reference or processing" (OECD, 2021[4]). This definition is useful in describing a concept, but as of yet, it does not include practical guidance on data measurement.

The overriding challenge faced by compilers is the very limited amount of information on the market price of data assets. The most common method in the national accounts and broader macro-economic statistics for generating estimates of capital investment is by recording purchases of such fixed assets. This not only provides a measure of investment, but also a market price for the good or service. However, data assets are usually manufactured for own account purposes, (Coyle and Li, 2021[15]). Even if data were sold often enough that prices could be recorded, it is likely that because of the lack of homogeneity across data assets (i.e. the *content* and the *context* of each item of data varies considerably), finding an appropriate value for each different data may not be feasible (Nguyen and Paczos, 2020[5]).

In addition to a lack of transactions in data assets making market prices difficult to ascertain, it is also important to consider that data assets created by the private sector for market use comprise only a portion of the data assets that exist. The government and non-profit sector make widespread use of digital information, creating data assets that are then used repeatedly in their (productive) activities.

Because of the large amount of investment in "data" undertaken on an own account basis or by non-market economic units, a substitute to market prices is required as a basis for the estimation of the value of that investment and the capital stock of data assets. Data is not unique in this respect, and the 2008 SNA makes provisions for such situations by recommending that the value of the capital investment is estimated based on the sum-of-costs of production for both non-market producers (§6.130), e.g. public roads and schools, and for own account capital investment with limited market transactions (§13.36), e.g. software and research and development (European Commission, International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), United Nations and World Bank, 2009[18]).

National statistical offices' widespread use of sum-of-costs estimation for various intangible assets was highlighted by the final report on intellectual property products by the joint EUROSTAT – OECD Task Force on land and other non-financial assets. This provides the most detailed guidance to-date on the specific cost elements to include for each asset. Importantly, for databases they include:

- The cost of preparing data in the appropriate format;
- Staff time spent on developing the database;
- Capital services of the assets used in developing the database; and
- Costs of items used as intermediate consumption (Eurostat-OECD, 2020[21]).

This breakdown for databases is important, as the first two items above describe costs associated with tasks that are very similar to those described in the definition of data provided earlier in the paper, that is that *"information content that is produced by accessing observable phenomena and recording, organising and storing relevant information elements from these in a digital format".* Organising and storing information elements from observable phenomena are arguably the same as "preparing data and developing the database". This is not to say that the definition of databases could not also be altered to better distinguish between the production of databases and of data. However, it does show that if an additional and separate data asset class is to be identified in macro-economic statistics, guidance will be required to delineate which costs be apportioned to data and which to databases.

Rather than separating them, the best solution may be to instead create some form of joint asset class, which brings together both databases and data (ISWGNA Advisory Expert Group on National Accounts, 2020[22]). While keeping the two components together is likely to be more feasible for compilers, having separate categories would likely be the superior option from a user perspective provided that sufficient quality could be attained. An important challenge of combining the asset categories, though, is the practical difference between valuing a database – the specific parameters of which are defined once, paid for and then utilised, similar to other fixed assets – and data which changes over time with the addition of new observations of observable phenomena.

This kind of continual augmentation of the data asset through the constant addition of data could be considered as similar to a renovation done to a building (albeit at a much higher frequency), where additional investment adds to the overall value of the asset. Conceivably, an estimate of this investment, added to the overall value of the asset, could be generated through the use of averages and a large dose of assumptions. In this way, a value could be imputed for every new item of data added to the data asset. However, the use of such imputations based on unitary values without any consideration for the information that the data contains could potentially distort important macro-economic indicators (Ahmad and van de Ven, 2018[20]).

In addition to the practical challenge of identifying specific costs to allocate to data as opposed to other classes of assets (i.e. software, databases etc.), there is the fundamental question of when to begin recording data production costs. That is, at what point does the process of producing data assets actually start? Many data assets are created based on observable phenomena that were only accessed due to additional expenditure by the organisation. For example, access to observable phenomena is often obtained in exchange for the receipt of a "free" (zero price) digital service. Examples include an algorithm-based search engine, a free phone application, or a website that aggregates publically available information in an easy to use format. These free services are provided only when the right to observe and record information elements is given to the firm by the consumer. These type of free services require production (input of labour and capital) by the producer, even though this output is generally provided free of charge to the consumer.

Such expenditure does not include the recording, processing and storing of information from observable phenomena, which is considered fundamental to the production of data. However, it is arguable that in some cases it is the preliminary step in accessing and recording information *about* observable phenomena. Therefore, this production expenditure could also be included in any sum-of-cost calculation.

However, this kind of expenditure may serve an additional purpose to providing an opportunity to access observable phenomena. Often it also provides the firm an opportunity to produce their output. For example, social media usually contains advertising, which is the primary way that the firms generate revenue. The social media platform allows for this to occur, and provides an opportunity to obtain access to sought after observable phenomena. Should the expenditure on the platform be considered an investment in a data asset, or an investment in producing the advertising services?

Alternatively, on some occasions the unit may simply pay an amount to purchase access to observable phenomena in order to record the information contained in it. In this case, where there has been no production, just a monetary transaction, should this additional expenditure contribute to the overall production assigned to the data asset?

These practical challenges are non-trivial, and it has been well established that the existing national accounting framework is not well equipped to reveal the current data revolution that is driving many new digitally related business models (Ahmad and van de Ven, 2018[20]). However, any changes to the framework must be made in a way that maintain the integrity and usefulness of the framework. Due to the sheer number of firms now utilising data to generate revenue directly or through efficiencies within their business, any decision regarding data measurement – conceptual or practical – will affect not only the information that statistical offices need to gather in order to make the required compilations, but more importantly a large number of the outputs they produce.

Approaches to measuring the economic value of data

A range of approaches have been undertaken to try to find ways of valuing data and data flows, both in the context of the SNA and more broadly. Recent work explored four different perspectives from which the value of data, databases, and data flows can be conceptualised and associated with measures of economic value (Ker and Mazzini, 2020[6]).

The first perspective took data from use tables (part of the supply-use framework) and business statistics databases (expenditures on data storage products, i.e. hardware, software, and services) under the assumption that economic agents will only pay for these if they expect to reap economic (or other) benefits of equal or greater value. It is estimated that firms in the United States (U.S.) spent over USD 36 billion on storing data in 2017 (comprising 0.25% of total intermediate consumption), and that overall expenditure on data storage is increasing over time. These estimates are possible because the U.S. publishes detailed expenditure by product line data, as well as supply-side product detail; extending the approach to other countries will rely on additional statistical details becoming available.

A second perspective uses business statistics to look at the revenues generated by firms creating explicit value from data (i.e. those collecting, compiling, and selling databases and associated products). These activities are estimated to have been worth over USD 60 billion in the U.S. in 2017, around USD 1.4 billion in Canada, and around EUR 19-50 billion in the European Union in 2016. The wide range of the latter arises from the relatively less detailed information available for European countries and the different empirically-based estimation parameters that can be used.

A third perspective considers whether data holdings and analysis may be contributing to firm valuations. "Data-driven firms" are identified from web-scraped information, matching such data to stock market tickers used to web-scrape market capitalisation data for those firms. On average, the value of the identified firms has grown faster than other firms listed on the NASDAQ and NYSE over the period 1986-2020. These firms had a combined market capitalisation of USD 5 trillion in 2020. Further work is needed to extend this approach to firms listed on other stock markets and to improve the identification of "data-driven firms". Other measures of firm performance could also be analysed including enterprise value, profitability and productivity.

Finally, the fourth perspective looks at potential links between trade flows and data flows, comparing different measures of the trade in "digitally deliverable products" as defined from several perspectives. Overall, the evidence suggests that there are a number of global hubs for trade in these products. The U.S. appears to be the largest by quite some margin, but others include Germany, India, Ireland, France, the Netherlands, and the United Kingdom (U.K.). These countries' digitally deliverable exports are estimated to be worth between USD 36 billion and 1.2 trillion, with this wide variation depending on the product classes used to operationalise the definition of "digitally deliverable products".

Li and Chi (Li and Chi, 2021[8]) adopt another approach which relates commercial estimates of global data flows to "Big Tech" firms' organisational capital (proxied by sales as well as general and administrative expenditures). Findings suggest that a five-fold increase in data flows is associated with a doubling of Big Tech firms' organisational capital. In addition, using firm-level data for U.S. hospitality and transportation businesses, it also finds that the market entry of an online platform – with their intensively "data-driven" business models – increases the rate at which less digitalised incumbents' organisational capital depreciates. However, no immediate impact on incumbents' output, employment, or total factor productivity is detected.

In further work, Coyle and Li (Coyle and Li, 2021[15]) leverage the link between data and organisational capital to estimate the economic value of markets for data with an initial focus on the hospitality sector. This results in a "conservative" estimate that the market size for data in the global hospitality industry was USD 43.2 billion in 2018, doubling in size every three years over the preceding period. They note that this method could be applied in future work at both industry and country levels.

While they are not inherently incompatible with the established frameworks used for compiling economic statistics such as the SNA, the concepts and measures set out above, such as organisational capital and firm market valuation growth, are not fully aligned and integrated with them. Therefore, while such exercises can allow for useful analysis and insights, at this point they do not represent established and tested methods that are replicable across all countries or that align with fundamental principles of the SNA such as volume based outputs, estimates based on market prices, etc.

Due to the practical and conceptual challenges mentioned earlier in the note, as well as the need for any method to be accurate, replicable and transparent in its calculation there is, as of yet, no consensus regarding best practice for valuing data in the SNA context. Theoretical approaches to estimating the value of data have largely focused on three methods:

- Market-based: Value is determined based on the market price of comparable products on the market.
- Cost-based: Value is determined by the cost of producing the information/know-how derived from data.
- Income-based: Value is determined by estimating the future cash flows that can be derived from the data (OECD, 2019[25]).

While markets for data do exist (e.g. data brokers), they tend to focus on a very specific type of data, usually personal information that can be leveraged for advertising purposes. A large proportion of data used in the economy is created and used by the same firm focusing exclusively on their own production. This type of data, while highly valuable to the firm, is both rarely sold and extremely heterogeneous. As a result, it is unlikely that this can be used as a proxy for the value of other data not the subject of market transactions.

An income-based valuation – that is, a valuation based on the expected future income from the asset – is often very hard to calculate and thus not used regularly by statistical offices. While theoretically the income approach is logical, businesses often value their productive assets based on tax minimisation, not potential future earnings.[11] Due to this, valuation methods by business vary and are not directly comparable to the valuation of assets for national accounts purposes. For this reason, statistical offices rarely ask directly for valuations of assets, preferring to build up estimates of capital stock using the initial purchase cost and the perpetual inventory method (OECD, 2009[23]).

Alternatively, national statistical organisations (NSOs) could compile an estimate of the stock based on potential future earnings. Encouragingly, this is a widespread practice for measuring natural resources. However, since data can have so many context-dependent uses, including the possibility of the same data being used multiple times and by multiple agents (i.e. its non-rivalry), the potential revenue stream is essentially limitless. In the case of a natural resource the stock of the resource, the uses, the pattern of use, the price, and the amount of time until the known stock is depleted, are broadly understood. In the case of data, with the industry in such infancy, how long they will be used, (which theoretically could be forever), is unknown, as is the price (since it depends on the use) and the potential uses (OECD, 2021[4]). For these reasons, an approach based on future income is considered extremely difficult in practice.

Therefore, a cost-based approach has proven the most conducive to the development of experimental estimates by countries. This approach takes the standard SNA sum-of-cost approach incorporating the value of the inputs used in production and wage costs of production, as well as a return on capital for the use of any fixed asset in the production. Even with this pre-existing approach, there have only been limited attempts by statistical offices at placing an experimental monetary value on data assets used in the economy.

The most comprehensive effort so far was by Statistics Canada in 2019. In this experimental work, wage and employment information was used as the basis for a preliminary set of Gross Fixed Capital Formation (GFCF) estimates

[11] These different valuation approaches may often be due to both tax minimisation strategies as well as accounting regulations aimed at limiting inflated asset prices being claimed.

representing investment in data assets. These were then combined with certain assumptions to create estimates of capital stock using the perpetual inventory method.[12] This allowed for a time series of GFCF and capital stock to be estimated from 1990 to 2018 (Statistics Canada, 2019[19]).

This work estimated that investment in Canada to produce data assets was between CAD 29.5 billion and CAD 40.1 billion in 2018, having grown consistently since the earliest estimate in 1990 when it was estimated at between CAD 14.6 billion and CAD 20.0 billion (Figure 2). Overall, this continual rise in investment in data assets has resulted in the total value of data assets in the Canadian economy being estimated to be worth at least CAD 104.8 billion and potentially as high as CAD 150.9 billion as of 2018.

Figure 2. Range of investments in data products, Canada
1990-2018

Source: (Statistics Canada, 2019[19]).

A key advantage of attempting to measure data assets using a sum-of-cost approach is that it can be applied to all sectors of the economy, including those sectors not attempting to derive a profit such as the Government and Non-Profit Institutions Serving Households sectors. As part of its publication, Statistics Canada showed in combination, the government sector and non-profit institutions made up slightly more than 20% of the investment in data, databases and data science in 2018 (Table 1).

[12] The perpetual inventory method outlines how capital stock estimates can be created using only estimates of capital investment. For more information, see the *OECD Manual on Measuring Capital* (OECD, 2009[23]).

Table 1. Investment and capital stock of data, databases, and data science in Canada

By sector, 2018

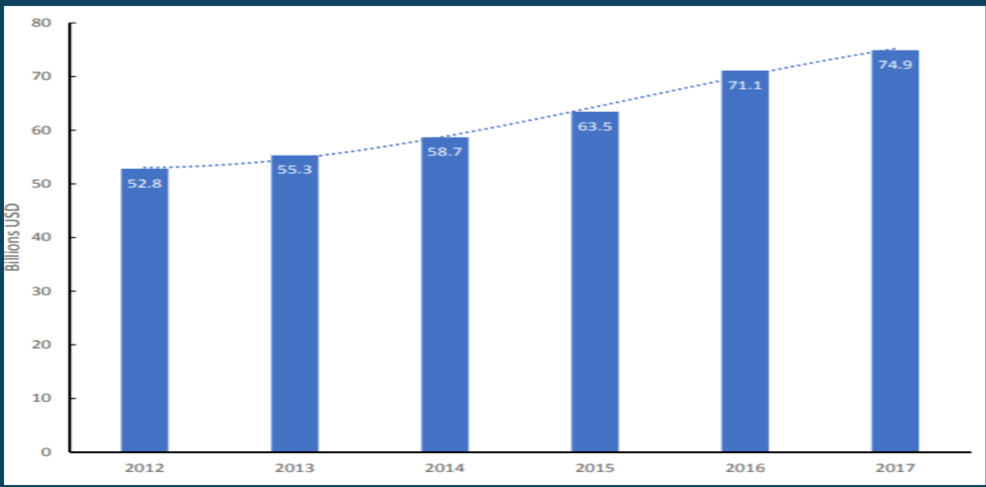
	Total	Non-financial corporations		Financial corporations		Government		Non-profit institutions serving households	
	millions of dollars	millions of dollars	percent of total	millions of dollars	percent of total	millions of dollars	percent of total	millions of dollars	percent of total
Investment									
lower range value	29,455	13,676	46.4	9,327	31.7	6,027	20.5	425	1.4
upper range value	40,025	19,403	48.5	12,224	30.5	7,842	19.6	556	1.4
Capital stock									
lower range value	157,067	80,875	51.5	38,835	24.7	34,834	22.2	2,524	1.6
upper range value	217,659	114,562	52.6	54,097	24.9	45,646	21.0	3,354	1.5

Source: (Statistics Canada, 2019[19]).

The U.S. Bureau of Economic Analysis (BEA) has also undertaken similar work producing estimates of investment in data using a sum-of-cost approach. Basic preliminary estimates used as a "proof of concept" were generated for the BEA advisory committee. This work took wage costs of certain occupations identified as data-intensive and combined them with other production expenses, based on a ratio of employee to non-employee costs of similar industries. This basic approach found that growth in data-related production costs within private industries, ranged from 4.7% in 2013 to 12.1% in 2016, culminating in a total estimate of USD 74.9 billion in in 2017(Figure 3) (BEA, 2019[24]).

Figure 3. Data related production costs for private industries, United States

2012-2017



Source: (BEA, 2019[24]).

The BEA is now undertaking a more comprehensive approach for estimating the labour costs of data collection, storage, and analysis by using an unsupervised machine-learning algorithm to produce a better estimate of which occupations are involved in data-related tasks and the proportion of time spent on these activities (BEA, 2020[25]). This is combined with a traditional labour costs estimation framework to provide experimental estimates for the value of the data economy.

The U.K. Office of National Statistics (ONS) undertook a similar process using a sum-of-cost approach to improve GFCF estimates of own account software and databases. This approach included using occupation data to derive estimates that were incorporated into the 2019 annual national accounts publication (ONS, 2019[26]). By including this sum-of-cost approach in its published estimates, the ONS illustrated the feasibility of using such a method similar to the experimental estimates created by the BEA and Statistics Canada. Conversely, it also showed the practical difficulty in separating investment in databases (already an asset in SNA 2008) with data (potentially an asset in the next iteration of the SNA) because the occupations used in the estimation are likely to be producing both types of asset with no simple way to delineate the expenditure associated with each.

Academics have also attempted to generate estimates of data using similar methodologies. For example, Goodridge et al. (Goodridge, Haskel and Edquist, 2021[27]) use statistics from the European labour force survey to estimate own account investment in data. However unlike the work of the national statistical offices, who simply compare the derived level of investment with that already in the core accounts, this work inserts the new investment figures into the established estimates to observe the impact that this expanded asset boundary might have on productivity.

Looking ahead, data will continue to be a social and economic resource. While there is no one established approach to measuring the economic value of data, this Toolkit note provides insights into how different communities define the term “data”, as well as innovative approaches countries are taking to measuring the economic value of data. These initiatives, many of which are listed in the Annex, are an important part of the statistical and research community’s efforts to measure data. It is important that national statistical offices and other research bodies continue to take practical steps that not only add to the discussion, but also allows them to implement the latest methods and classifications in a timely, accurate, transparent way and cross-country comparable way.

Annex. A selection of approaches conceptualising and valuing data

Treatment of data in the national accounts

Responsible entity: U.S. Bureau of Economic Analysis

Description: The paper outlines preliminary thoughts and considerations for the inclusion of data stocks and flows in a national accounts framework. It summarises the current treatment of data in the SNA and presents considerations that serve as a foundation for how data may fit into the SNA framework. The paper includes some cursory estimates of data-related flows based on official statistical sources although there are extremely experimental and should not be used to compare against published measures in the U.S. national accounts.

Read more: *https://www.bea.gov/system/files/2019-05/Paper-on-Treatment-of-Data-BEA-ACM.pdf*.

Measuring investment in data, databases and data science: Conceptual framework

Responsible entity: Statistics Canada

Description: This paper addresses the lack of visibility in the modern national accounting framework afforded to data by expanding current national accounting concepts and statistical methods for measuring data. It sheds light on these highly consequential changes in society that are related to the rising usage of data. The paper includes examples of some of the new ways data are being used by businesses and households in order to contextualise the discussion. This is further elaborated upon by describing the examples in the context of an information chain that exist in the creation and use of data. The paper then discusses the topic of ownership before concluding by discussing possible methods that can be used to assign an economic value to the various elements in the information chain.

Read more: *https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00008-eng.htm*.

The value of data in Canada: Experimental estimates Statistics Canada

Responsible entity: Statistics Canada

Description: The paper extends, and to a certain extent tests, a statistical framework created to provide an estimate of the value of data used in production in Canada. It presents a preliminary set of statistical estimates of the amounts invested to produce Canadian data, databases and data science in recent years. The estimates are calculated from employment and wage information collected by the quinquennial Census of Population and the monthly Labour Force Survey, combined with a number of important, but as yet largely untested, assumptions. The results indicate rapid growth in investment in data, databases and data science and a significant accumulation of these kinds of capital over time.

Read more: *https://www150.statcan.gc.ca/n1/en/pub/13-605-x/2019001/article/00009-eng.pdf?st=Wzd1A5d8*.

Estimating national investment in data assets in European countries using Labour Force Survey data.

Authors: Jonathan Haskel and Harald Edquist

Description: This paper uses labour force survey data for European countries to estimate national investment in data assets where the asset boundary is extended beyond that for software and databases as currently defined in the System of National Accounts. In this way the methodology is similar to those undertaken by statistical offices but applies the approach to all European countries. The paper estimates how much this extension of the asset boundary impact estimates of capital formation and productivity. **Read more:** *https://onlinelibrary.wiley.com/doi/full/10.1111/roiw.12542*.

How much is UK business investing in big data?

Authors: Peter Goodridge and Jonathan Haskel

Description: This paper attempts to document the contribution that data and data-based assets are making to UK growth. It presents a conceptual framework based around publically available labour market statistics to understand and measure the production of transformed data and data-based knowledge While acknowledging that a majority of the investment is already captured in software and database assets, it proposes that there is an additional amount of investment currently unreported in official statistics.

Read more: *https://ideas.repec.org/p/imp/wpaper/25159.html*.

The Data Economy: Market Size and Global Trade

Authors: Diane Coyle and Wendy Li

Description: This paper provides a novel sectoral methodology that is used to estimate the economic value of markets for data, providing industry-level and country-level information on data markets. This methodology results in a conservative estimate of the market size for data in the global hospitality industry of USD 43.2 billion in 2018, and it has been doubling its size every three years. With many jurisdictions introducing different data protection and trade regimes, affecting the data gap and data access by market participants, the paper presents a trade typology of countries and discuss their ability to benefit from data value creation.

Read more: *https://www.escoe.ac.uk/publications/the-data-economy-market-size-and-global-trade/*.

Online Platforms' Creative "Disruption" in Organizational Capital

Authors: Wendy Li and P.J. Chi

Description: This paper presents a new methodology, focusing on the concept of organisational capital, to examine how the entry of an online platform, a new data-driven business innovation, affects an existing firm's value of organisational capital and investment in organisational capital. An online platform's key disruption in its sector is traditional firms' knowledge derived from their relatively limited amounts of data. This disruption can hence be measured by a firm's organisational capital, the accumulated information of the firm. The approach is supported by findings that the organisational capital of dominant online platforms is highly correlated with rising global data flows, the first empirical evidence that successfully links the explosive global data flow to an economic value. Moreover, when the global data flow increases five-fold, Big Tech's organisational capital stock doubles. The paper also uses firm-level data for the U.S. hospitality and transportation industries during the period 2002 to 2018. This is the first empirical evidence of the anticipated effect of new business innovations on the depreciation rate of organisational capital. However, there is no immediate impact on output, employment, or the total factor productivity of existing incumbents. In the increasingly digitally and physically inter-connected world, new online platforms' disruptions in traditional industries will be significant, fast, and on a massive scale. This paper provides a new methodology to measure online platforms' disruptions in traditional brick-and-mortar firms in a timely manner.

Read more: *https://iariw.org/wp-content/uploads/2021/08/LiChi_paper.pdf.*

Recording and measuring data in the System of National Accounts

Responsible entity: OECD

Description: The 1993 SNA introduced the notion of databases, with further clarifications provided in the 2008 SNA that specified that databases should reflect only the value of the underlying database management systems and the costs associated with the digitisation of data. This recommendation reflected the view that the underlying value (information content) associated with the data itself was de facto a non-produced asset (because to do otherwise would indirectly open the door to the capitalisation of knowledge), with outright purchases of databases recognised in the accounts as goodwill, and as such, their contribution, as a factor of production, is de facto invisible in the accounts. Recent years have seen an explosion in the generation of data, and the use of data, notably in advertising based business models, raising questions about the 'invisibility' of data in the accounts. This paper and presentation attempts to address these issues as a way of encouraging further debate, both conceptually and in the field of actual measurement.

Read more:
https://unstats.un.org/unsd/nationalaccount/aeg/2018/M12_3c1_Data_SNA_asset_boundary.pdf.

An update on recording and measuring data in the system of national accounts: An issues paper, utilising theoretical scenarios

Responsible entity: OECD

Description: This paper expands on previous discussion on how best to conceptually and practically record data in the System of National Accounts. It focuses on the issue that information, which underpins the creation of data, is obtained by firms in different ways, the result of implicit and explicit transactions as well as a by-product of conventional production. In order to appropriately record data in the National Accounts a framework must be created that allows for not only the representation of these assets as they are, a crucial contributor to production but also that maintains the usefulness and integrity of national account indicators.

Read more:
https://unstats.un.org/unsd/nationalaccount/aeg/2021/M15_7_4_Recording_Data.pdf.

U.S. App Economy Update

Responsible entity: Progressive Policy Institute

Description: This paper examines the number of jobs created by the "app economy" in the United States. The paper, part of a larger research project examining app economy employment in different countries and regions, attempts to improve the visibility of the app economy, including the employment impact of the Internet and the "new economy". While not extending the link between wages of these occupations and the value of the asset they are creating, it provides useful guidance on the use of occupations to value elements of the economy.

Read more: *https://www.progressivepolicy.org/wp-content/uploads/2017/05/PPI_USAppEconomy.pdf*.

Changes to gross fixed capital formation and business investment: Blue Book 2019

Responsible entity: U.K. Office of National Statistics

Description: This paper outlines improvements made to estimates of gross fixed capital formation (GFCF) and business investment in the national accounts for the Blue Book 2019. The improvements follow a range of new research that ensures the estimates better reflect activity in the economy today. This includes the use of labour costs of relevant workers input into a multiplicative model. Adjustments are made for the different kinds of workers involved, non-wage labour costs and non-labour costs (such as intermediate inputs, overheads and consumption of fixed capital), it includes a mark-up for operating surplus. The paper explains how additional research was used to identify relevant workers with job titles that are not easily identifiable in occupation classifications.

Read more:

https://www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/articles/nationalaccountsarticles/impactofbluebook2019changesongrossfixedcapitalformationandbusinessinvestment.

Perspectives on the value of data and data flows

Responsible entity: OECD

Description: Data and databases are increasingly central to business activities today, with even relatively small data flows having the potential to create considerable economic value. Despite this, attempts to conceptualise and measure the value of data remain underdeveloped. This paper explores four different perspectives from which the value of data, databases, and data flows can be conceptualised and measured: 1) how much businesses spend on storing data; 2) how much money businesses make from selling data-based products; 3) how the market valuation of "data-driven firms" compares to that of other firms; and 4) the value of trade flows in digitally deliverable products.

Read more: *https://doi.org/10.1787/a2216bc1-en.*

The value of data: how is the value of data created, captured, and distributed?

Responsible entity: Bennett Institute + ODI

Description: This paper reviews different approaches to measuring the value of data and examines how sharing data affects what value is unlocked and how that value is distributed. This work shows that the economic characteristics of data and the data economy mean the market alone will not unlock data's full potential value, but that it is possible to gain more from data with the right data policies, and an institutional framework that supports trustworthy access to data.

Read more:
https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_summary_report_26_Feb.pdf.

References

Ahmad, N. and P. van de Ven (2018), *Recording and measuring data in the national accounts*, https://unstats.un.org/unsd/nationalaccount/aeg/2018/M12_3c1_Data_SNA_asset_boundary.pdf. [20]

BEA (2020), *A Labor Costs Approach Using Unsupervised Machine Learning*, OECD Publishing, Paris, http://dx.doi.org/doi:https://community.oecd.org/servlet/JiveServlet/downloadBody/176318-102-1-311572/AG%20GDP%20Digital%20eco%20-%20Item%204.d%20-%20BEA%20-%20Blackburn.pdf. [25]

BEA (2019), *Treatment of Data in National Accounts*, https://www.bea.gov/system/files/2019-05/Paper-on-Treatment-of-Data-BEA-ACM.pdf. [24]

Coyle, D. et al. (2020), *The value of data: Policy Implications*, OECD Publishing, Paris, https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_Policy_Implications_Report_26_Feb_ok4noWn.pdf. [13]

Coyle, D. and W. Li (2021), *The data economy: Market size and global trade*, OECD Publishing, Paris, https://www.escoe.ac.uk/publications/the-data-economy-market-size-and-global-trade. [15]

European Commission, International Monetary Fund (IMF), Organisation for Economic Cooperation and Development (OECD), United Nations and World Bank (2009), *System of National Accounts, 2008*, United Nations, New York, https://doi.org/10.5089/9789211615227.071. [18]

Eurostat-OECD (2020), *Joint Eurostat-OECD Task Force on Land and Other Non-financial Assets*, OECD Publishing, Paris, https://ec.europa.eu/eurostat/documents/24987/725066/Eurostat-OECD+Report+on+Intellectual+Property+Products.pdf. [21]

Forrester, J. (2020), *What's The Difference Between Private Weather Companies And The National Weather Service?*, Forbes, https://www.forbes.com/sites/jimfoerster/2020/01/24/whats-the-difference-between-private-weather-companies-and-the-national-weather-service/. [3]

Furman, J. (2019), *Unlocking Digital Competition*, OECD Publishing, Paris, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf. [17]

Goodridge, P., J. Haskel and H. Edquist (2021), "We see data everywhere except in the productivity statistics", *The Review of Income and Wealth*, https://doi.org/10.1111/roiw.12542. [27]

ISWGNA Advisory Expert Group on National Accounts (2020), *Recording and Valuation of Data in National Accounts*, OECD Publishing, Paris, https://unstats.un.org/unsd/nationalaccount/aeg/2020/M14_5_3_1_Recording_of_Data_in_SNA.pdf. [22]

Ker, D. and E. Mazzini (2020), "Perspectives on the value of data and data flows", *OECD Digital Economy Papers*, No. 299, OECD Publishing, Paris, https://dx.doi.org/10.1787/a2216bc1-en. [6]

Li, W. and P. Chi (2021), *Online platforms' creative "disruption" in organizational capital - The cumulated information of the firm*, OECD Publishing, Paris, https://iariw.org/wp-content/uploads/2021/08/LiChi_paper.pdf. [8]

Nguyen, D. and M. Paczos (2020), "Measuring the economic value of data and cross-border data flows: A business perspective", *OECD Digital Economy Papers*, No. 297, OECD Publishing, Paris, https://dx.doi.org/10.1787/6345995e-en. [5]

OECD (2021), *Draft Recommendation of the Council on Enhanced Access to and Sharing of Data*, internal document [C(2021)96]. [28]

OECD (2021), *Issues paper: Recording observable phenomena and data in the national accounts*, OECD Publishing, Paris, https://unstats.un.org/unsd/nationalaccount/aeg/2021/M15_7_4_Recording_Data.pdf. [4]

OECD (2021), *The Digital Transformation of SMEs*, OECD Studies on SMEs and Entrepreneurship, OECD Publishing, Paris, https://dx.doi.org/10.1787/bdb9256a-en. [9]

OECD (2020), *The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector*, OECD Publishing, Paris, https://www.oecd.org/finance/The-Impact-Big-Data-AI-Insurance-Sector.pdf. [11]

OECD (2020), *Using artificial intelligence to help combat COVID-19*, OECD Publishing, Paris, http://www.oecd.org/coronavirus/policy-responses/using-artificial-intelligence-to-help-combat-covid-19-ae4c5c21. [14]

OECD (2019), *An Introduction to Online Platforms and Their Role in the Digital Transformation*, OECD Publishing, Paris, https://dx.doi.org/10.1787/53e5f593-en. [7]

OECD (2019), *Going Digital: Shaping Policies, Improving Lives*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264312012-en. [1]

OECD (2019), *Measuring the Digital Transformation: A Roadmap for the Future*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264311992-en. [2]

OECD (2019), *Unpacking E-commerce: Business Models, Trends and Policies*, OECD Publishing, Paris, https://dx.doi.org/10.1787/23561431-en. [12]

OECD (2018), *Digitalisation and Finance*, OECD Publishing, Paris, http://www.oecd.org/finance/Financial-markets-insurance-pensions-digitalisation-and-finance.pdf. [10]

OECD (2009), *Measuring Capital - OECD Manual 2009: Second edition*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264068476-en. [23]

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

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

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