



Economic and Social Council

Distr.: General
7 March 2022

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on National Accounts

Twenty-first session

Online, 17, 19, 23 and 25 May 2022

Item 3 of the provisional agenda

New developments in communication, globalization and digitalization

Recording of data in the National Accounts

Prepared by the Digitalisation Task Team

Summary

Data is a crucial input into productive activities across almost all facets of the economy. However, within the System of National Accounts 2008 (2008 SNA), data is not explicitly identified as a standalone asset. Instead, some aspects of data investment are included in the estimation of database assets or implicitly recorded as goodwill when an entire economic unit is sold, or if an explicit transaction in databases occurs. Since the relevance of macroeconomic statistics depends on the ability to adapt the SNA to meet the evolving needs of policymakers and respond to a changing economy, the absence of an explicit data asset in the accounts is becoming increasingly untenable. The guidance note Recording of data in national accounts developed by the Digitalization Task Team provides an overview of available options for recording data in the National Accounts, with each discussion ending with a recommendation. It also presents some practical considerations related to the proposed recommendations, most of which will need to be addressed during a testing and experimentation phase before inclusion in the revised SNA.

I. Executive summary – list of recommendations

1. In order to incorporate the production of data into the core National Accounts, as well as provide a framework for consistent valuation, the guidance note recommends the following;

- **Data is defined as “information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities”.** While data certainly exists in non-digital form, for the purpose of the System of National Accounts (SNA) only digitised data is considered the result of production. Additionally, the creation of data that is not directly used in the productive activities of the economic unit will also be considered outside of the SNA production boundary.
- **Data is the result of production**, it is an intangible product, produced with the input of labour and capital; therefore, when capitalised in the National Accounts it is classified as a produced asset. Data is considered part of the intellectual property product suite of assets.
- **Data is distinct from the information elements of ‘observable phenomena’ (OP), which are inputs for data. Observable phenomena are “a fact or situation whose characteristics or attributes may be recorded”.** Data consists of recorded and organised information elements of OPs, i.e., once the information element of the OP is recorded and stored in a digital format it becomes data. Observable phenomena are regarded as non-produced¹ and as in general having no value, except if they are purchased. These **explicit purchases are considered as either purchase of a service (output) or payment of rent for providing access to specific observable phenomenon**, from which their information elements can be recorded. For pragmatic reasons, the information elements of the OPs, which may contain asset-like qualities, are excluded from the asset boundary of the SNA.²
- **Data that is produced and used in production for more than one year meets the SNA characteristics of an asset and, as such, should be capitalised in the National Accounts.** While in general they should be treated as per other capital formation in the National Accounts, due to the unique characteristics of data, additional guidance on certain transactions is required.
- **Like other assets in the National Accounts, data is subject to economic ownership, valuation (and re-valuation) and depreciation** - although depreciation occurs not due to wear and tear but rather due to obsolescence.
- **All own account production of data is considered capital formation.** Although it is likely that some own account data may be fully consumed within one year, due to practical limitations on delineating this data from data used repeatedly for more than year, it is recommended to capitalise all own account production of data. **Data purchased via market transaction is treated as per other products, that is capitalised if intended to be used for more than one year or as intermediate consumption if consumed during production.**
- **Own-account production of data is valued at the sum of costs.** This is consistent with existing guidance in the National Accounts for other own-account produced intellectual property products. A complete list of the relevant costs is provided in Section IV.
- **Data that is sold in a market transaction with no exclusive rights connected to the data is considered a sale of a copy of an original. If exclusive rights are**

¹ While regarded as non-produced, they can be an externality of productive activities.

² The exact asset that is being excluded from the balance sheet it not explicitly defined, the asset could be considered as a licence to observe, or access to certain information or even ones individual privacy. However, just because there is no desire to include this hard-to-quantify asset on any balance sheet this need not be a reason not to treat the transaction associated with it as a rent payment.

granted, it is recorded as a sale of a produced asset, in line with the sale of other fixed assets. In many cases, data itself will not be sold; rather it will be used by the data producer as input in the production and provision of specific goods and services, such as advertising services.

- **The data producer is the economic unit that collects, records, organises and stores information elements of observable phenomena in a digital format, thereby creating data.** They are the initial owner of the data, not the persons or entities to which the underlying observable phenomena may refer.
- **Expenditure undertaken to access and record OPs that are added to an established data asset is considered new gross fixed capital formation**, as it prolongs the life of the data asset.
- **If practically possible, data would be classified to a newly-created asset category called “data” which would include the current output associated with the production of databases and be separate to the current category of computer software.** Such a recommendation is dependent on testing by national statistical offices (NSOs).
- **Only data that provides an economic benefit when used in the productive activities of its owner is included in the SNA asset boundary.** Other data, referred to as ancillary data, likely incorporates data created as a by-product of operating the business and which is not used as a direct input into production and as such, it is excluded from the asset boundary.
- **While no specific recommendations are provided in this guidance note, it is expected that the service life applied to data assets will be quite short.** In part, this reflects the relatively large proportion of own account produced data that is not used for longer than one year. Appropriate guidance on this will rely on research and testing by NSOs.

II. Introduction

2. Data is a crucial input into productive activities across almost all facets of the economy. Data is both consumed (as intermediate consumption) and used repeatedly (as an asset) in production. However, within the 2008 SNA, data is not explicitly identified as a standalone asset. Instead, some aspects of data investment are included in the estimation of database assets when calculating own account creation of databases, additionally, the value of data is likely implicitly recorded as goodwill when an entire economic unit is sold, or if an explicit transaction in databases occurs. Since the relevance of macroeconomic statistics depends on the ability to adapt the SNA to meet the evolving needs of policymakers and respond to a changing economy, the absence of an explicit data asset in the accounts is becoming increasingly untenable. Digitalisation and the creation of data continues to become more common in economic activity and additional information is sought on data’s impact on productivity. Based on this, there is consensus that data should be incorporated into the core accounts of the revised SNA.

3. As part of the SNA research agenda, a task team was created to study if the production and use of data could be explicitly incorporated into the National Accounts. Based on discussion with statistical offices and users, it appears clear to the task team that data is the result of production and on many occasions the data produced is used for more than a year, therefore meeting the asset definition in the National Accounts. However, due to the unusual characteristics exhibited by data, its valuation and inclusion in the SNA is not a case of simply including an additional class to the capital account.³

³ For a full discussion of the differences between data and other assets and the difficulties these cause in attempting to value data see Nguyen and Paczos (2020) <https://www.oecd-ilibrary.org/docserver/6345995e-en.pdf?expires=1643032809&id=id&accname=ocid84004878&checksum=4A2B19E0B6EA15874A4573DACD4A6B5D>

4. Some of the specific questions that required guidance from the task team included;
 - What is the most appropriate method of valuation?
 - What costs should be included as contributing to the production of data when applying a sum-of-cost approach?
 - What is the life length of data and would all data meet the asset definition?
 - How to record the expenditure associated with the continual updating of databases?
 - How to record the different ways of producing data and of using data?
5. In Section IV, the guidance note (GN) discusses these and other considerations, providing an overview of available options for recording data in the National Accounts, with each discussion ending with a recommendation. Section V presents some practical considerations related to the proposed recommendations, most of which will need to be addressed during a testing and experimentation phase before inclusion in the revised SNA.
6. In many cases (although importantly, not all), the production of data is intrinsically tied to the consumption of “free digital services”. The digitalisation task team is, in parallel, also addressing issues related to the creation and consumption of these free services, both digital and non-digital. While there is no suggestion that consumption of these services should be incorporated into the core SNA accounts, a separate GN is being prepared on how they might be incorporated in supplementary tables. This GN focuses on determining a market value for the final data product than can be used within the core National Accounts, rather than any imputed theoretical values that might be assigned to non-monetary transactions that can occur with respect to free services.

III. Existing material

7. Leading up to the 2008 SNA, the Canberra II Group carefully considered the inclusion of embedded data in capital formation (Ahmad 2004; Ahmad 2005). The recommendation that was ultimately included in 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 2008 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 in computer software.
8. 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 2008 SNA viewed ‘data’ as the embodied information content of what is now typically referred to in the new lexicon of data value chains as the information content of ‘observations’ or ‘observable phenomena’, that is information that had not yet been recorded.
9. Recognising that transactions related to databases often included the value of the information content, 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.
10. The Joint Eurostat-OECD Task Force on Land and Other Non-Financial Assets did not explicitly address data as an asset in their final report on IPPs, rather focussing on compilation guidance for databases, in which their recommendations are consistent with

those already in the SNA (see Box 3.1 in the report).⁴ In response to user demands, some National Statistical Offices have begun to further explore the issue, trying to come up with an estimate of the value of data. This work often attempts to derive a value on the basis of a sum of costs approach, classifying certain expenditure as directly involved in the creation of data. This often provided experimental estimates of flows and stocks for data related investments, usually derived based on the costs of labour inputs. This includes work by Statistics Canada (Statistics Canada, 2019), Statistics Netherlands (De Bondt and Mushkudiani, 2021) as well as for many different European countries by Goodridge, Haskel and Edquist (Goodridge et al, 2021). While this work is useful in providing rough estimates and guiding the practical ability for NSOs to implement the task teams' recommendations, they do not explicitly address some of the trickier conceptual questions mentioned in the previous section.

IV. Options considered and recommended by the task team

11. This section examines several challenges, presenting multiple options on how data could be valued and included in the accounts. On the basis of these options, the task team came to specific recommendations which are presented following each discussion.

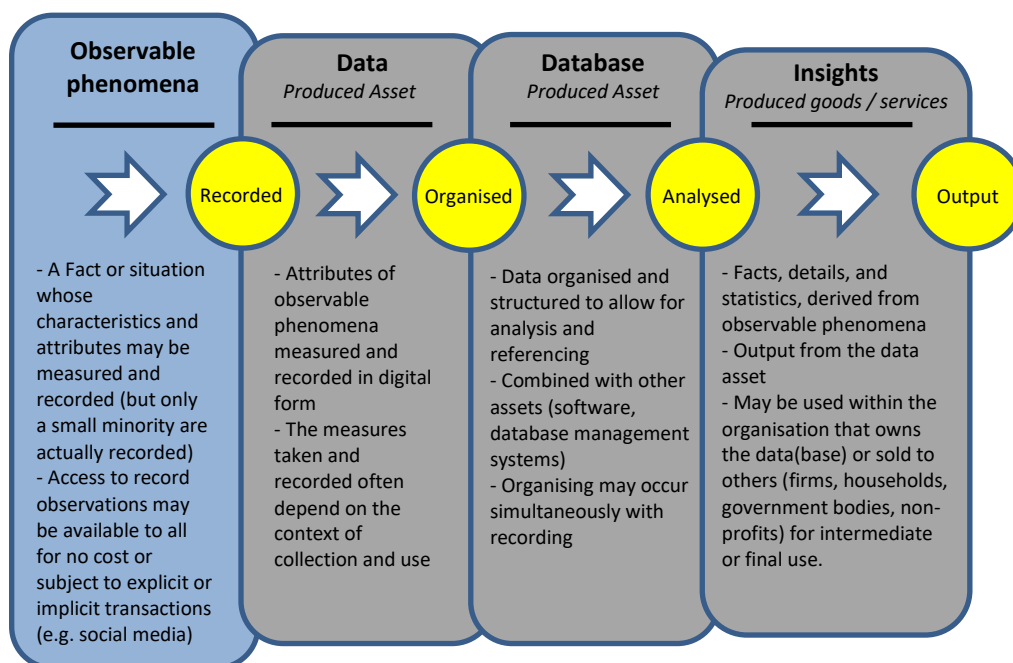
A. Definition of data

12. There is widespread agreement among statistical offices and international organisations that a clear definition of the word '*data*' for economic measurement purposes is required before recommendations can be made on valuation and recording in the National Accounts. While diverse definitions of data exist, many are drafted for specific policy purposes, so may not be of direct use for a discussion of the recording of data in macroeconomic statistics.

13. While data is broadly used as a term to describe recorded information, the specificities of this information are extremely varied. Currently, data might be used in reference to an individual record of the most basic fact or single data point on a specific observable phenomenon (E-commerce record of one individual's purchase). It may also be used to describe a broader set of similar data (E-commerce record for one store for the entire day) or even huge datasets containing a vast number of records on the information content of millions of observations/observable phenomena (the E-commerce information across multiple countries).

⁴ A link to the report is available here:
<https://ec.europa.eu/eurostat/documents/24987/725066/Eurostat-OECD+Report+on+Intellectual+Property+Products.pdf>.

Figure 1
Step by step process of data production chain



Source: Mitchell, Ker & Leshner (2021)

14. The point that information elements of OPs are recorded would appear the logical step as to where any potential data asset comes into existence. It is at this point that the OP is transformed into something that can be “provided from one unit to another” (SNA 1.39) a key characteristic of output included in the SNA production boundary. While the point of recording is the moment that data is created, importantly and as outlined in the definition, this production process may have begun prior to the recording when inputs of labour and capital are used to decide how the chosen OPs are to be accessed and recorded.

15. As such, and as shown in the data production chain, **this GN considers that data is produced at the point that the information content of the OP is digitalised.**⁵ Therefore the definition used for data in this GN is “*information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities*”.

B. Is data produced or a mixture of produced and non-produced?

16. Two sets of data can cost the same amount to produce; however, due to the information contained within the data it is possible that one of the two data sets could be significantly more valuable than the other. If the cost of production is the same, but the value different, should compilers consider some of the assets’ value as resulting from the intrinsic value of the information embedded in the dataset (which could then be regarded as non-produced asset). With this in mind, the task team faced two choices:

- Recommend the creation of a new hybrid data asset category, consisting of produced and non-produced parts. The produced part could be derived on the basis of the costs of the productive activities involved in the creation of the data asset. Any remaining value

⁵ Importantly, while this is the specific point that data (from the SNA perspective) is produced, just because information is recorded digitally does not make that content data for the purpose of the SNA. For instance, music, podcasts, books may be considered as digitised information content but they are not derived by observing phenomena or by recording, organizing and storing information elements from these phenomena.

could then be recorded as the value of the non-produced part, with its emergence recorded in the other volume changes account.

b) Recommend that data is entirely the result of production.

17. In regards to option 1, there are currently no assets within the SNA that are considered both produced and non-produced.⁶ The impact of such a change would be considerable, raising several questions, including how to present a single asset under two categories within the capital account and balance sheet, how to divide expenditure between the production account and capital account, as well as considerations regarding the calculation of consumption of fixed capital (COFC). This also appears as a poor compromise if it is only created for a single type of asset, and may introduce a risk that other assets are possibly reconsidered as being a mixture of produced and non-produced elements (Research and development (R&D), organisational capital, etc.). Additionally, but importantly, such a change would require considerable “re-education” of users (and likely compilers) who are not familiar with the other changes in the volume of assets account and are comfortable with the basic concept that the sum of gross fixed capital formation (GFCF) and COFC (along with any revaluations) is equal to the change in value of assets.

18. An important consideration for Option 2 is that it would be, by far, the most practical for NSOs to implement, as implementation is similar for other IPP products such as own account software and R&D.

19. From a conceptual perspective, it is undeniable that data relies on the information embedded within OPs to provide it with its value, and that this value might not be fully attributable to the production of the data asset. However, any difference is often caused by the context that the data is used, as such the explicit inclusion of this residual value as non-produced would appear incorrect when compared to other non-produced assets.

20. Equally, data that is worth different amounts upon creation does not unequivocally imply some additional non-produced component contributing to the assets value. The higher value assigned to a piece of data may often reflect the additional labour, innovative new process, creative thinking or other inputs of production that has contributed to the ability to access the more valuable OPs that are contributing to the higher valued data.

21. Considering the relevant points, **the task team recommends that the data assets be considered entirely the result of production.**

C. Valuation methodology – Sum of costs vs Net present value

22. Once confirmed that data is entirely produced, a recommendation is required on the most appropriate methodology to measure its value. Like most other assets in the SNA, the most obvious valuation method would be to collect prices from market transactions. However, it is well established that the majority of data used in production is constructed on an own account basis and is not actively traded on the market; therefore, an alternative valuation method to market price should be considered. Two options recommended in the SNA in this case include (i) calculating estimates of GFCF based on the sum-of-costs of production or (ii) on the basis of the net present value (NPV) of the asset.

23. A clear advantage exists in advocating for a sum-of-cost approach due to the familiarity for statistical offices with this approach and the similarity between the proposed measurement of data and other own account intellectual property products. Importantly, any sum of costs for creation of data by the private sector would need to incorporate an additional mark-up for net operating surplus for market producers. The exact costs that should be included and excluded are discussed below.

⁶ The closest that might exist is the non-produced land underneath buildings, which when sold are considered an entirely produced asset. However when placed on a balance sheet, land and structures are separated, and for many (although likely not all) transactions in GFCF across sectors in dwellings or buildings the land component is considered separately. They are certainly not jointly represented in sectoral balance sheets.

24. The NPV is based on potential future revenues that may be derived from the asset; this is also already in use in some areas of the National Accounts, for example when valuing natural resources. However, since data can have so many context-dependent uses, including the possibility of the same data being used multiple times, the potential revenue stream is limitless. Additionally, in the case of a natural resource, the stock of the resource, its use, the pattern of use, the price and the amount of time until the known stock may be depleted are broadly understood. In the case of data, with the industry in its infancy, a lot of this information is unknown.⁷

25. **Due to the clear practical advantages that come with the sum of costs, this approach is recommended.** Such a decision would appear to be supported by the fact that, so far, all NSOs and academics that have attempted to create estimates of the value of data consistent with the SNA have used the sum-of-costs methodology.

D. Which costs to include in sum of cost calculation?

26. The 2008 SNA makes the following statement in regard to the cost that should contribute to the sum of cost compilation of databases. “*The cost of preparing data in the appropriate format is included in the cost of the database but not the cost of acquiring or producing the data*” (SNA 10.113). There is general agreement that there should be a change in the final portion of this statement so that there is an expansion in the costs contributing to a data and database asset. This would relate most notably to the costs of accessing OPs and collecting the relevant information elements.

27. The production of data has changed significantly since discussions were held prior to the finalisation of the 2008 SNA. At that time, a large driver of the value of data was not just the information it contained but that it was information in digitised form. This digitisation allowed for easy analysis, sharing and manipulation compared with data in non-digitised form. This digitised data spawned new ways that data could be used in the production process creating different opportunities to add value to the production process. It made sense then that a fundamental cost involved in creating this asset was “converting” or “preparing” the data (information content of OP’s in the current typology). To some extent, this reflected not just the way data was used in production at the time but also the technology of 15-20 years ago.

28. Data production today exhibits greatly reduced costs in relation to “preparation”. Accessing, collecting and recording usually occur simultaneously, often as a by-product of standard processes taking place. Due to this, the value of the data has shifted from being based on its interpretability and usability (although that still matters) that previously came with being digitised, to being fundamentally based on the information that it contains.

29. The value of the information embedded in data will always have a contextual component to it, related to the relevance, scarcity and timeliness of the information. With this in mind, the ability to produce productive (and therefore valuable) data is not dependent on the ability of an economic unit to “prepare” the data in an appropriate format, rather it is dependent on the ability of producers to access useful and timely OPs preferably in an exclusive manner. To do this, producers must come up with creative, efficient and legal ways to gain access to these useful and timely OPs, access which takes resources to plan and undertake.

30. With this in mind, it was proposed that the inclusion of data into the production boundary could be as simple as reversing the final exclusion in the existing SNA definition of databases and incorporating the costs of “acquisition and production of data”. However, it has been generally agreed that more specific guidance is required, as the exact meaning of

⁷ Furthermore, if data asset values and the GFCF undertaken to produce them were measured based on future earnings, the valuation could bring in additional unrelated external non-produced effects such as potential monopolistic network affects, rent seeking due to market power and contributions from other unknown capital. Alternatively, such a methodology would likely involve assumptions that, as pointed out by Reinsdorf and Ribarsky (2019) would be “unacceptable for national accounts purposes”.

“acquire and produce” might incorporate a large range of input costs that may also be involved in the creation of other assets.

31. With these considerations the task team recommends to expand the costs involved in producing data and databases to include specific costs involved in planning and implementing how and which OPs are accessed. **Therefore, own account production of data should be valued at the sum of costs, which would include;**

- costs of planning, preparing and developing a data production strategy,
- costs associated with accessing, recording and storing information embedded in OPs, which may include, but is not limited to, the explicit purchases of OPs or already produced data,
- costs associated with designing, organising, testing and analysing the data in order to draw information and conclusions from it.

32. These costs include staff time and costs of items used as intermediate consumption. **In addition an estimate for consumption of the fixed capital used in the own account production of data and a mark-up for net operating surplus for market producers should be included.**

E. How to record explicit purchases of OPs that are used in data

33. There is broad agreement that one of the expanded costs involved in producing data is the explicit monetary payments made in exchange for the right to access OPs and record their information elements. However, there are several options on how these purchases might be recorded in the National Accounts. Importantly, there are two characteristics regarding these transactions that might be considered when selecting the appropriate option.

34. While examples do exist and are relatively easy to find, purchases related to accessing OPs is by far *the least utilised way that information elements of OPs are recorded*. Only a few specific business models provide monetary payment (or even payment in kind, such as vouchers) in exchange for providing access to the OPs. That said, it is foreseeable that only a slight change in business models might make these transactions much more common. For instance, if free service providers switch to some form of freemium model whereby people had to pay for a service unless they would provide access to OPs, this could also be regarded as obtaining access to OPs in exchange for monetary or in kind payments.

35. While they all involve the provision of information in exchange for money, the exact nature of what is being purchased may take many different forms. This ranges from placing a tracker on a phone to record additional information about the phones user, to uploading a photo of a recent purchase receipt, to the situation where people undertake some form of scientific testing, whereby they must eat, live, and act a certain way for a period of time in exchange for payment. Each of these examples require different levels of time and effort from the person the OPs refer to. The task team has continually stressed the need to not transform everyday activities like participating in social media or signing up to a mailing list into productive activities. However, there would appear to be a threshold when the activities change from regular, day-to-day activities for no monetary payment to activities that must meet certain requirements as prescribed by the payer/data producer. The latter would certainly appear more like productive output than the former.

36. With these two considerations in mind, four options have been discussed for how this transaction might be recorded.

- Option 1: Purchase of a non-produced asset.
- Option 2: Rent payment for accessing a non-produced asset.
- Option 3: Payment for a service, considered output of the household sector.
- Option 4: No specific treatment assigned, continue current way of recording.

1. Option 1: Purchase of non-produced asset.

37. Option 1, whereby the information content of the OP is considered a non-produced asset appears to be the only option that would be considered entirely inappropriate, both conceptually and practically. In the vast majority of cases, the information contained within the OP cannot be held exclusively by the data producer. Therefore the data producer is not able to establish true ownership rights over the OP. Furthermore, it has been agreed by the task team that while OPs have “asset like qualities”, for a variety of practical and conceptual reasons it is recommended that they are not recorded on the balance sheet of the sector that ‘produces’ them or the sector that subsequently obtains them. Due to this, the implementation of option 1 would first require an other change in the volume of assets in the initial sector to allow the asset to be exchanged, in order to then appear on the balance sheet of the data producer as a non-produced asset. Overall, due to its exclusion from balance sheets it seems counterintuitive to record the transaction as a purchase / sale of an asset.

2. Option 2: recorded as a rent payment for accessing a non-produced asset

38. If the payment is regarded as a rent payment, then unlike option 1 no asset will be purchased, rather it could be viewed as a payment for granting access to a person’s (or economic unit’s) non-produced asset (which is still excluded from the balance sheet for pragmatic reasons)⁸. Since no asset then changes hands, it negates the issues raised by the purchase of a non-produced asset approach. This treatment would also have the benefit of not introducing productive output (value added) for the payee, which would appear appropriate when creating the OPs involves minimal to no effort for the payee.

39. Such a treatment would require a change in the SNA, so that the method of deriving the sum of costs could incorporate rent payments on the use of non-produced assets. Currently, the sum of costs is limited to just; “*Intermediate consumption; compensation of employees; consumption of fixed capital; a net return to fixed capital, and other taxes (less subsidies) on production*” (SNA 6.125). This does not include rent payments, although the expenses clearly contribute to the creation and value of the final data asset. This issue of including rent on non-produced assets in the sum-of-costs valuations impacts areas of the National Accounts beyond the explicit purchases made to access and record the information elements of OPs and is being taken up in a separate guidance note. From the perspective of data, such an option will be recommended if the general guidance is changed to include such as costs.

3. Option 3: Payment for a service, considered output of the payee sector.

4. Option 4: No specific treatment assigned, continue to be recorded as it is now.

40. Option 3 of considering the payment an exchange for a service and option 4, the continuation of the current practice of not treating the payment differently to any other business expense, somewhat co-exist. If the payment is ignored and treated as an other operating expense, as is likely the current practice, it is implicitly recorded as output of the sector to which the payment is made. While this treatment would appear appropriate when significant efforts are undertaken by the payee to create the OPs⁹, it appears less appropriate if the payee makes little or no recognisable effort. The largest benefit of option 3 & 4 is a practical one, since for compilers it would likely require no change to their current methods.

41. Overall, the GN acknowledges that conceptually the most appropriate treatment would be to delineate transactions, depending on the level of recognisable effort undertaken by the payee sector; with some being recorded as output and some as a rent payment. However, it is recognised that this is highly unlikely to be possible for NSO’s. As such, **the GN recommends recording all transactions directly related to obtaining access to OPs as a rent payment. The benefit of applying a single consistent treatment outweighs the**

⁸ This undefined asset relates to the right to withhold one’s information from others. While not as specific as privacy, the asset is the ability to grant or withhold the right to observe oneself creating OPs.

⁹ While it appears that output is occurring and would need to be recorded, it is not immediately clear what product is being produced. Such an issue will be taken up by the Central Product Classification (CPC) revision task team.

risk that a small amount of transactions made in repose to productive activities are classified incorrectly.

F. How to record ongoing expenditure related to established data assets

42. Most data assets continually receive new inputs in the form of new information elements from various OPs. It is important that the guidance note clearly recommends how this output should be recorded. Most assets, after being created, receive maintenance, a current expense that is not capitalised or may undergo “major improvements”, which are counted as investment and recorded as GFCF.

43. Alternatively, a data asset could be regarded as never actually being completed, in which case, (additional) production could be considered as “work in progress” and the data asset being considered as an inventory. However, data in any form is able to continually be sold or contribute to production. Therefore, it would appear incorrect to ever label data as a work-in-progress and place it within the inventories classification. Such a classification includes “output that is not sufficiently processed to be in a state in which it can easily be supplied or sold to other institutional units” (SNA 6.110). Data can certainly be supplied to other units or used in production even if not “finished”. Therefore, the ongoing expenditure does not meet the definition of a work in progress.

44. The expected service life of data varies greatly depending on the content of the information embedded and the context this data is used. However, all data assets will begin to decline in value if no additional expenditure is undertaken and no information content from new OPs are added. Therefore while each addition of an individual information element may not seem to constitute a “major improvements, addition or extension” (SNA 1.156), when considered in their entirety, they certainly do have a material impact on the fixed asset. This expenditure improves the relevance of the asset which in turn “improves the assets performance, increase[s] their capacity or prolong their expected working lives” (SNA 1.156).

45. For these reasons, **this GN recommends that expenditure undertaken to update a data asset with new OPs should be considered as investment rather than repair and maintenance.** This recommendation does not mean that a separate new data asset is created (or not created)¹⁰ but the output related to this expenditure should be considered GFCF rather than intermediate consumption.

G. How to record the sale of data?

46. A large amount of data used in production will concern own account data used exclusively and repeatedly by the economic unit that produces it. This type of data is an asset to the unit and should be included as economic output and on the balance sheet of the units.

47. Data can also be purchased in a market transaction from a data producer, with the purchaser then using the information in the data to improve their own production. This transaction could be recorded as a sale of a produced asset, as an operating lease of a fixed asset, as the sale of a copy, or as output, where the purchaser simply buys a service from the data producer. As many different scenarios may exist regarding the sale of data, a recommendation should be provided to assist compilers to treat similar transactions similarly.

¹⁰ This guidance note does not try to articulate the boundaries of one data asset, delineating where one finishes and another begins. Across the National Accounts, such a process is not undertaken with other assets. The specific numerical number of cars, computers and new buildings are not recorded in the accounts. Even on survey forms, business are not asked how many computer software packages they purchased, instead, they are asked about their expenditure on assets. This guidance note believes that there is no need to hold data to a higher standard than done for other assets. Theoretically, it does not matter if Google considers their database on one French person a single data asset or their database on all people in France a single data asset. What is important for the National Accounts is the value of the expenditure required to create this data.

48. An important aspect regarding the sale of data is distinguishing between transactions involving data and transactions involving outputs from data, which may be very similar but should reflect the product the data is contributing too. In this regard, information from data can be used as intermediate consumption to produce a range of other products which are subsequently sold; in these circumstances the transaction does not explicitly involve a data asset.

49. When a transaction in data is undertaken, how such a transaction is recorded depends if this was done on an exclusive arrangement, where the seller has given up rights to the data or if the data seller has provided equivalent access to multiple users.

50. **If the data producer reproduces and makes available several (any number more than 1) copies of the same data asset, then this is considered similar to the SNA concept of originals and copies**, where it is considered GFCF to produce the original data asset, but then a current expense to produce the copies. These sales are treated as output of the data producer; however, their treatment by the data users depend on their use. They may be considered GFCF by the purchasing unit if they fulfil the conventional SNA asset test, i.e., “a copy sold outright may be treated as a fixed asset if it satisfies the necessary conditions, that is, it will be used in production for a period in excess of one year” (SNA 10.100). In this case, investment may have occurred for the buyer, but the seller has not recorded a corresponding capital transaction (outside of the original GFCF to create the original). Alternatively, the buyer may purchase the copy but then uses it for less than a year. As such, this would be treated as intermediate consumption.

51. **If exclusive rights to the asset are sold (including the right to reproduce the asset), this would be considered the sale of a produced asset.** Thus, it would be investment on the purchaser side, and potentially output on the producers side if the data asset was produced specifically for sale. If the data asset had been used in production for a while before being sold, the sale would be considered similar to other second hand assets. Such a transaction would involve a capital transaction on the seller’s side rather than simple output. Following the sale, the data producer would no longer have access to the data so it would need to leave the data producers balance sheet.

52. Often data might be made available to users for no cost. This may concern data produced by both market and non-market data producers. The treatment would be similar regardless of the sector. That is, the creation of the asset is recorded as production and recorded on the balance sheet of the relevant entity. **Since the data is provided for free, no transaction takes place vis-à-vis any users.**

53. In this circumstance, data should still be recorded as an asset as the data producers’ are still receiving an economic benefit from the data, even if it’s not a direct monetary benefit. For example, the general government often produces assets that facilitate utility across society when they are provided for free to consumers, such as parks and roads. At the same time, as has been discussed in the guidance note addressing free services, market producers often provide free services (or assets) for a variety of reasons. These may be done in order to build goodwill in the business or for other marketing reasons, such as a pseudo “loss-leader”. Ultimately, the data producer could decide to invoke their ownership rights at any time and begin to charge users.

54. **This GN recommends that when data is sold on a non-exclusive basis, this should be considered as the production and sale of a copy of an original, and be treated in line with paragraph 10.100 of the existing SNA. If data is sold on an exclusive basis, this should be treated as a sale of a produced asset. Publicly available data or data accessed free of charge is still considered an asset for the producer but does not create a transaction and as such no asset is recorded on the data user’s balance sheet.**

H. How to treat data that is essentially consumed within a year.

55. Since a large amount of data are extremely time sensitive, on many occasions data is used repeatedly in production but for a period that is less than one year, therefore failing to meet the asset test within the SNA. The treatment of own account data if used within a year

is relatively clear; the SNA states that own account output that is fully consumed in the purpose of producing another output should be considered as intermediate consumption. Additionally, paragraph 6.120 notes that it is unusual to record this intermediate consumption within the same economic unit. However, much data is used in production for much longer. Additionally, the continually evolving contextual nature of data combined with its physical lack of wear and tear mean that data is likely to be kept (even if not explicitly used in production) for longer than a year.

56. Since it is well established that it is unlikely that data producers and statistical compilers will be able to delineate, at the point of production, data that is used immediately from data that is used repeatedly for more than a year, two simplified options were considered by the task team.

a) A set percentage of total expenditure could be capitalised in line with the other recommendations in this GN, while the remaining amount of expenditure is considered own account intermediate consumption.

b) Capitalise all expenditure but apply a very short asset life and highly skewed retirement distribution to reflect the high proportion of data that is expected to retire early.

57. Option 1 is considered the easier for NSOs to implement. Expenditure on own account production of data could be calculated, before a certain percentage is removed and treated as intermediate consumption. In many cases, this would better reflect the process undertaken by economic units, whereby a percentage of the information content compiled is extremely time sensitive and thus not used after a very short period. Once this time sensitive information is used in production the information is essentially “consumed” and no longer used again.

58. In many cases data is stored for a longer period of time. Additionally, at the time of production, all data producers hope that the data produced will be useful as an input into production for extended periods. Although this will not always be the case, it is important to consider that since it is not subject to wear and tear, data can be, and likely is, used many years after production, potentially for a use that was never originally envisioned. This may not happen for all data but the fact that it does suggests that arbitrarily removing a set amount of data from being considered an asset may be premature.

59. Importantly, this “premature” exit from production for some assets is normal. COFC represents “the decline, during the course of the accounting period, in the current value of the stock of fixed assets owned and used by a producer as a result of physical deterioration, normal obsolescence or normal accidental damage.” (SNA 6.240). As outlined in the OECD handbook on measuring capital, the “decision to retire is taken because a new and possibly more productive and/or cheaper asset appears, rendering the old model obsolete” (P.106, Handbook on Measuring capital, OECD). In simple words, obsolescence in data assets occurs when a new and more productive piece of information contained in another data asset appears rendering the previous data asset obsolete. This normal obsolescence (as applicable to data) should be incorporated into the calculation of the stock of data assets in line with that of other assets. Computers and cars experience a much quicker rate of obsolescence than buildings or dwellings, which is reflected in their asset lives and retirement profiles; data and databases likely experience an even quicker rate of obsolescence¹¹.

60. While it may appear counterintuitive to record 100% of data as asset when evidence suggests that a much lower percentage is still utilised after a year, there are parallels with the way that mineral and petroleum exploration is recorded. As currently treated in the National Accounts there is an implied recognition that a firm must experience some misses in order to get some hits. Such a viewpoint is outlined in paragraph 6.231 which states regarding mineral exploration “whether successful or not, they are needed to acquire new reserves and so are all classified as gross fixed capital formation” (SNA 6.132). Just as exploration is undertaken

¹¹ This incorporation of a higher rate of obsolescence with data assets need not be done simply with short asset lives. Different retirement profiles can be implemented that acknowledge the much larger obsolescence occurring for data during year 1, so unlike normal asset live distributions, which are usually bell shaped which imply the even distribution of retirements, it can be acknowledged that for many data assets (perhaps even a majority of data assets) a larger percentage are removed from production earlier.

in order to gain greater information about what lies below the surface, the production of data is undertaken to gain greater information about various elements of the production process. Sometimes this information will be forthcoming and useful, sometimes not, but the production of all of it is needed to acquire the new information.

61. A small but important consideration regards data that is only stored for a set amount of time due to commercial or regulatory reason, where it is impossible to use the data again. Expenditure on this 'transitory' data should not be capitalised.

62. **The GN notes that due to the impact on important aggregates, costs should, if possible, be recorded as intermediate consumption or capitalised in line with the current guidance of the SNA. If this delineation is not possible, The GN recommends all costs included in the sum-of-cost calculation are capitalised (option 2), but that an appropriate asset life and retirement distribution profiles are applied to the capital formation expenditure to reflect the significant amount of data that is consumed or becomes obsolete within a year.** While it was considered that Option 1 would be easier to implement and to a certain extent represents the real world economy, in so far that only a percentage of data produced is still used after one year, Option 2 was considered more consistent with current methodologies within the SNA.

V. Practical considerations in the classification of data

63. 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 while also being mindful of some of the practical limitations that NSOs might face in trying to derive estimates of data.

64. Certain recommendations will still require testing by compilers to help determine if the recommendations are practically feasible for implementation in a revised SNA. These include,

- The ability of NSOs to clearly determine which costs are contributing to the production of data.
- Whether expenditure on data can be separated from expenditure on computer software and whether information on data can be presented as a standalone asset within the National Accounts.
- Whether compilers are able to delineate and exclude costs associated with data that is not providing a clear economic benefit to the economic unit.
- The ability for compilers to obtain the information needed to apply appropriate assumptions in the creation of capital stock and how much assumptions may impact high profile indicators such as gross domestic product (GDP) and GFCF.

65. Some of the practical considerations of these choices are discussed in more detail below.

66. The joint use of other intellectual property products (and possibly other assets) in the production of data while at the same time contributing to other production (often the primary production of the unit) might cause situations where it may be difficult to determine which part of these assets have been used in the creation of data. For example, certain software may be used in the production of the primary output of the organisation (i.e. advertising) while also being used to access OPs and record the information elements of them, thereby contributing to the production of data. While a single asset is often used to produce multiple products, it is unusual for the products to then be capitalised requiring a more detailed audit of the costs involved in producing it.

67. Additionally, there is currently a large amount of trial and error in the production of data than there is in the production of other assets. Units are likely undertaking specific research and development to try to improve the OPs that they have access too as well as the information elements they are recording. The thin line between research and development

and production of the underlying asset may well be harder to judge than for data than for existing assets.

68. Overall, the purpose of bringing data into the SNA asset boundary is not to usurp the recording or importance of other assets. Instead, the purpose is to try and correctly record certain expenditure, that may currently be recorded as intermediate consumption or as input in the creation of other assets, as input in the creation of a data asset.

69. This intertwining nature of expenditure on certain aspects of the data will also challenge the recommendation to present data as a standalone asset within the National Accounts. While databases have been included in the SNA since the last revision, almost all countries present estimates of databases as a combined asset with computer software. This was seen as a pragmatic decision due to the difficulty of separating the two. The GN strongly recommends that data and databases are compiled and presented separately from software within the revised SNA, as the addition of data creates further differences in the asset characteristics between the two. This is not just a decision driven by user demands, rather, the addition of data further differentiates the various characteristics between the assets, and as such, the accuracy of the capital stock measures will improve if they are separated. This recommendation to present data and databases together but separate from software is a clear difference from the current presentation of the accounts. The ability for NSO to separate this established relationship is a clear question to be answered as part of the testing and experimentation phase.

70. An important recommendation that will require testing to ensure its feasibility, is the exclusion of data that is not providing a clear economic benefit to the economic unit. This type of data may be described as ancillary data, non-tradable data, or generic and universal data. It consists of digitalised information that is not used as direct input into the primary productive activities of the economic unit. This type of data exist in almost all firms and economic units and although not identical between firms has very little to no proprietary value. While theoretically the data has some value to the firm in so much that it ensures a smoother running of the business, since it is not considered a store of value and does not directly contribute repeatedly to production, it remains outside of the asset boundary. That said, the practical ability for NSOs to exclude this type of data may depend on how the estimates are compiled.

71. The recommendation to capitalise all expenditure made in the production of data will require substantial research and testing to try to determine the appropriate service lives and retirement distribution. Before it is formally incorporated into the SNA, it is important to try to quantify the level of impact that this change would have on some of the key indicators in the National Accounts such as GDP, GFCF and capital stocks.

72. Although their impact on estimates will potentially be much smaller, other recommendations will also require testing to ensure that they can be followed. This includes the ability for NSO to separately identify explicit payments made to access OPs. While the payment will be captured as an expense in usual business surveys, it may be difficult to separate any payment made in exchange for “minimal effort or output” from those made for other intermediate inputs into the production. This would limit the ability of NSOs to separate these payments as rent payments rather than payments that correspond to output of another sector. For this reason, the recommendation may be reviewed to record all payments as relating to output if NSO’s are not able to separately identify them.

73. Overall, there remains a significant amount of further research into how these recommendations could or should be implemented by NSO’s. **Importantly, the requirement for additional research should not preclude a revised SNA from incorporating the inclusion of data production into the SNA production boundary.** Such a precedent has previously been followed with R&D¹², which was incorporated into the revised 2008 SNA with only a broad definition. Concurrent stand-alone research was

¹² An important difference, that should be acknowledged, when comparing the implementation of R&D into the SNA with the proposed implementation of data is the large amount of research and classifications that had already taken place for R&D as part of the various iterations of the Frascati manuals.

undertaken that helped develop more detailed guidelines and practical recommendations. It is envisioned that the inclusion of data will follow a similar path.

VI. Changes to SNA Required

74. While some fundamental changes to the SNA will be required to accommodate the inclusion of data production, additional changes will depend on the specific recommendations chosen.

75. The most obvious fundamental change is the creation of an explicit asset category for “Data”. As proposed, this category will be considered a produced fixed asset and exist alongside other intellectual property products such as research and development and computer software. **It is recommended that this category would include expenditure currently classified to “databases” and be separately compiled and presented to computer software.**

76. If it is decided that explicit monetary transactions in the purchase of OPs should be considered as a rent payment made from the data producer to the unit creating the OP, then two more additional changes would be required, both of which are discussed in other guidance notes.

77. The Guidance note G.9 “Payments for Non-produced Knowledge-Based Capital (Marketing Assets)” discusses the creation of an alternative rent category termed “*rent on other non-produced nonfinancial assets*” which would sit alongside the current category of rent on natural resources. Payment for OPs would appear more closely aligned with the more generic title as while some OPs are naturally occurring, the vast majority are due to some form of human intervention. As mentioned previously, for reasons of practicality, OPs are excluded from the SNA asset boundary. However, this does not negate the ability to incorporate a flow connected to this excluded asset.

78. The inclusion of these explicit payments, recorded as rent paid into the cost of production for own account capital formation, would necessitate another change within the SNA. Currently, any rent payments are excluded from the listed sum-of-costs formula in the SNA. In regards to the production of data, it would appear logical that such a cost is a fundamental cost of production and as such should be included in the calculation of the asset value. Since this issue has broader implications, especially in regards to the use of land and agriculture in production, a guidance note is planned on the subject.

VII. Terminology

79. This section provides a definition of the main concepts used in this guidance note.

- **Observable Phenomenon** are a fact or situation, whose characteristics and attributes may be recorded.
- **Data** is information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which can be accessed electronically for reference or processing.
- **Data assets** are data that is intended to be used in production for more than one year. All data when produced are capitalised, thereby becoming data assets. A data asset may be produced as an original, with copies subsequently sold which appear on multiple balance sheets; the original recorded on the data producers balance sheet and the copy, if used by the data users for more than a year, recorded on the balance sheet of the data user as a produced asset.
- **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.

- **The data producer** is the entity that collects, records, organises and stores observable phenomena in a digital format, the production associated with this production is output for the data producer economic unit. They are the initial owner of the data asset.
 - **The data user** is the economic unit who is accessing the data for use in its own production. It accesses the data either because it is publically available or through a monetary payment. In many cases, the data producer and the data user may be the same unit.
 - **Government data assets** are data assets produced by the general government sector. These include data assets that may or may not be publically available. These data assets are recorded on the balance sheet of the general government.
 - **Public data asset or publically available data asset** is a data asset produced by any sector of the economy and made available for use in production at no monetary cost to the data user. These data assets are recorded on the balance sheets of the data producers sector only.
 - **Productive data** is data that is contributing to the primary or secondary activity of the economic unit and provides a clear economic benefit to the owner. Transactions in productive data (both own account creation or the purchase of a copy) are recorded as GFCF and the subsequent assets appear on the balance sheet of the economic unit that has undertaken the GFCF
 - **Ancillary data** is data that is generated as an externality of production undertaken in order to create the conditions within which the principal or secondary activities can be carried out. The generation of ancillary data is not considered GFCF and it is not recorded on the balance sheet of the data producer. Any expenditure on storing and organising ancillary data is considered intermediate consumption towards the primary output of the economic unit.
-