United Nations ECE/CES/GE.20/2024/12

Economic and Social Council

Distr.: General
7 March 2024

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on National Accounts

Twenty-third session
Geneva, 23-25 April 2024
Item 2 (b) of the provisional agenda
Towards the 2025 System of National Accounts:
Measuring intangible assets and natural capital in 2025 System of National Accounts

Handbook on the Capitalisation of Data in the System National Accounts - Annotated outline

Prepared by Eurostat and International Monetary Fund

Summary

The inclusion of Data as a new asset class is one of the main innovations planned for the System of National Accounts 2025 (2025 SNA). While concepts are well defined, the compilation of data assets is only in its initial stages for many countries. To improve international comparability, Eurostat and the IMF have set up an international Task Team that is developing a handbook with practical recommendations for a harmonized compilation of this important component of modern economies. This document presents the proposed outline of the Handbook on Capitalization of Data in the SNA.

I. Executive summary

A. An executive summary (or some other nomenclature)

1. *This section would include the specific high-level recommendations pertaining to the conceptual and practical compilation of data as a form of summary of recommendations. It would replicate those recommendations included later in the handbook.*

II. Introduction

Introducing the broader context for this work

2. *This section will include several paragraphs setting the scene for the work of the task team and the need for this handbook beyond a desire to have better knowledge of data's value. It will include references to the decision to revise the System of National Accounts 2008 (2008 SNA), how this is regular (although infrequent occurrence) and the impact this has on indicators such as Gross Domestic Product (GDP). It will discuss that the incorporation of data into the production and asset boundary is only one change as part of many for the planned 2025 SNA. These paragraphs will include a discussion on the how the decision to revise the 2008 SNA was made, the process involved in having changes approved (i.e., guidance notes, global consultation, and approval by Advisory groups) and the expected timeline for implementation.*

Why this handbook has been written and what it covers

3. Data has become a fundamental pillar of the economy. Advances in technology have increased the ability for corporations to produce and process massive amounts of data. This data creates more insights, which further incentivises more investment in data assets. While the practice of collecting and organising information is not new, the digitisation of so many facets of the economy has made the production of data easier and cheaper than ever before. This in turn has resulted in the creation of brand-new business models reliant on data, while increasing the risk that many traditional enterprises will be left behind if they are not creating data about their customers or production processes.

4. Digitalisation is one of the priority areas of the SNA research agenda. In this context, the absence of an explicit data asset within the 2008 SNA production boundary, despite its increasing importance to both new and existing business models has become untenable. While it is arguable that the value of data can be somewhat implied in the value assigned to other assets such as Databases and Research & Development (R&D), as will be explained in chapter 1, this current treatment is vague and inconsistent in regard to purchased data assets and those produced on one's own account.

5. As such, following extensive discussions in several international forums and a global consultation with statistical compilers and other stakeholders, a decision was made to explicitly incorporate data into the 2025 SNA production and asset boundary.

6. Conceptually, the estimation of data as an asset in the 2025 SNA follows existing methods used for other Intellectual Property Product (IPP) assets such as R&D and Computer Software. However, the unique characteristics of data, both in the way that it is produced and used, have required some additional discussions to reach consensus on how to treat certain aspects of data production in the 2025 SNA. These topics are discussed throughout the handbook with final recommendations presented so that countries are able to easily and consistently explain to users the conceptual boundary of data within the 2025 SNA and how it impacts important indicators such as GDP.

7. The practical interpretation and implementation of these conceptual recommendations is a larger, but not insurmountable, challenge for countries. Most countries have not produced estimates of data investment before so while most have experience in estimating own account

capital formation of various IPP assets, the intricacies of data require a review of certain assumptions and modelling undertaken.

8. As such, the handbook will, at times, provide two sets of compilation recommendations that countries can follow. The first will include a broader aspirational approach to which countries can aim towards as their data availability improves. The second set of recommendations will be more basic step-by-step guide, directed towards (groups of) countries for whom data on certain aspects of data production is not yet available. The handbook also includes information and case studies from those countries that have already compiled estimates of data, which can guide and inform other countries in their work.

9. By including a basic set of recommendations on certain aspects of data compilation, the handbook is striving to maximise methodological consistency across countries. This is vital for any change to the GDP production boundary as the international comparability of the System of National Accounts, and GDP in particular, is a fundamental pillar of the framework.

10. The compilation of data requires countries to make many decisions on how to appropriately reflect the level of data output and investment as well as the changing value of data assets over time. These measurement challenges are not unique to data assets, however since most countries are producing estimates of data for the first time, there is a strong desire for compilers to apply similar assumptions when data gaps exist in an effort to ensure that differences between estimates reflect variations in the real economy rather than alternative methodology.

A summary of the different chapters of the handbook

11. The handbook is written broadly in the order that countries will attempt to compile estimates of data.

12. Chapter 1 defines data for economic statistical purposes and covers how, for the purpose of the 2025 SNA, data are being considered as an explicit output of production and if capitalised in the accounts will be classified as a standalone produced asset (combined with databases). It will explore how this differs from the previous treatment of data in the 2008 SNA. It also covers some broad underlying measurement questions such as how to distinguish intermediate consumption of data compared to Gross Fixed Capital Formation (GFCF) of data (or if this is even feasible), the treatment of ancillary data – which is not considered a productive output – and the classification of data across the economic statistics landscape.

13. Chapter 2 covers the compilation of estimates of data on a nominal basis. While the chapter will touch on the different approaches theoretically possible, it will focus on the sum of costs method as this is seen as the most practical for countries to implement. It discusses the treatment of government data, as well as expanding on the treatment of data consumed immediately. It contains specific recommendations regarding the choice of occupations in the initial compilation of labour costs, how time factors, or data intensiveness of these occupations may be represented as well as where information and ratios may be sourced which represent other costs involved in the production of data. Helpfully it includes case studies from countries showcasing how they have produced these estimates.

14. Chapter 3 focusses on converting nominal estimates of data output and GFCF to volume estimates, a fundamental requirement for incorporation into headline GDP measures. This is done through the use of price indexes, as such the chapter explores the different price indices that countries should aim to apply to data. The chapter includes a discussion on potential of applying adjustments to the price index so that any quality changes in the data asset can be represented.

15. Chapter 4 covers the compilation of capital stock estimates through the use of the Perpetual Inventory Method (PIM). This includes examples from countries on the assumptions they are applying to the PIM to estimate the capital stock of data, including the asset lives applied.

16. Chapter 5 discusses the remaining conceptual and practical elements of including data into the 2025 SNA. This includes recording the sale of data assets and services related to data assets, including those transactions that take place across international borders. This chapter also presents recommendations on creating a back series for data including examples of how countries have already approached this challenge.

What this handbook does not cover

17. While it is envisioned that this handbook contains all the information required for countries to compile estimates of data output including GFCF of data, the handbook does not cover every aspect of statistical collection and production. It assumes a certain level of knowledge on various aspects required to compile these estimates of data. For example, the handbook does not cover the fundamental background to the Perpetual Inventory Method, the production of price indexes or the data collection methods used to source the required information discussed in the handbook.

18. It is assumed that countries are already sufficiently knowledgeable in these areas, or if additional information is needed, that this can be sourced from the alternative reference material included as links within the handbook and which focuses specifically on these topics.

III. Chapter 1 – Defining the conceptual boundary of data for inclusion in the 2025 SNA.

This section opens with a clear definition of data for the purpose of the 2025 SNA

19. Due to the many possible understandings of what data is and is not and potential for misinterpretation, a detailed and extensive definition is required. Within this handbook and the proposed 2025 SNA data is considered as "***Information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities".***

20. This technical definition differs greatly from that likely thought of by the proverbial "person in the street". For many, data is a simpler concept even though it can refer to many different things. In fact, when used by most people, the term data is broadly indistinguishable from Information and can cover a single fact or point of knowledge up to large datasets from which numerous insights can be drawn from. This is not wrong; data can exist on a single item (the personal information of an individual) or on whole economies (the GDP for an entire country). Additionally, as pointed out by the OECD, data can take on a quantity perspective when it refers to "*Internet Protocol (IP) traffic or the volume of digitised information stored on servers and other hardware"* (OECD, 2022).

21. *Due to the obvious differences between the technical definition for use in the SNA and the more mainstream understanding of what data is, the handbook will discuss various ways that statistical offices may be able to describe data in order to collect information on it, including via business surveys.*

What are the specific characteristics that data must exhibit to be considered a produced asset from SNA point of view, including paragraphs on excluding non-digital and ancillary data

22. Primarily it was considered necessary to add additional caveats regarding how the data has been created and used to ensure that the data which would be capitalised from the perspective of the national accounts was consistent with other produced fixed assets within the SNA. The 2008 SNA considers produced fixed assets to be "*assets that have come into existence as outputs from production processes*" (2008 SNA §10.9) (EC et al., 2009) and that

"*are used repeatedly or continuously in production processes for more than one year. The distinguishing feature of a fixed asset is not that it is durable in some physical sense, but that it may be used repeatedly or continuously in production over a long period of time, which is taken to be more than one year.*" (2008 SNA 10.11) (EC et al., 2009). It has been broadly agreed that in today's economy, there are countless examples of data being created as an output of a production process and subsequently being used in business processes repeatedly over a period of one year. Simple examples include sales data to assist with forecasting demand, customer information as part of loyalty programs, cookies collected from websites and used to personalize an experience. As such, the suggestion that data should be considered a produced asset has been widely supported, however, it has also been acknowledged that the broadly accepted concept of data used in the economy can extend beyond that which is the specific output of production.

23. To compensate for this, the 2025 SNA definition of data brings in several aspects that must be fulfilled for data to be considered an output of production. This includes being on a digital format and providing an economic benefit to the owner.

24. While it is well established that non-digital data exists and can theoretically be used in production, it is considered that this would make up a very small and inconsequential amount of the data used in production. The inclusion of non-digital data is considered a significant measurement burden for countries and not commensurate with the influence of this data on the economy. Therefore, while acknowledging that non-digital data exists, **for the purpose of the 2025 SNA, only digital data is considered within the 2025 SNA production boundary.**

25. In addition, most businesses generate data that is not directly relevant to the production of the business. This data may be captured digitally, however, **if the data is not providing a direct economic benefit to the business, it is considered outside of the 2025 SNA production and asset boundary**. It is not practical to explicitly list what type of data this might entail as it will be different from business to business. However if the data is collected solely for record keeping or to facilitate the internal running of the business and it is not used to derive insights or information which may further production (a.k.a., providing an economic benefit), then any costs associated with it should be considered as a current input cost of production of the underlying output (intermediate consumption) rather than a capital cost (GFCF of data).

26. Theoretically, a distinction between the production of these different types of data (data that used in production and data that is considered ancillary) can be implied through the occupations chosen. Occupations that are not involved in the pro-active creation of value of data should not be included in the initial labor costs assigned to data production. As mentioned, this delineation is not clear cut. In many instances, data may have been kept for record keeping originally, only for business to later realize the value of this data and begin to analyze the data in order for it to provide economic value to the business through its insights.

27. While not controversial and in fact, merely brings data in line with other assets already included in the 2008 SNA asset boundary, this concept of not including expenses related to the production of data that is not providing an economic benefit to the owner is a fundamental and important one. *As such this section will elaborate further on this topic to ensure that countries clearly understand the concept, especially for the non-market sector.*

28. **Data from the perspective of the SNA does not automatically include everything that was saved digitally and stored on a computer.** While modern lexicon often used the term data for anything saved/stored on a computer, (i.e., how much data does your computer hold?). A digital file may contain data (as defined by the SNA) or it may contain a video, photo, software, old emails, etc. These things are not information content, produced by accessing and observing phenomena, which provide an economic benefit. *This distinction between data as defined in this handbook that makes up the production of data and data assets and data that underpins other goods and services (telecommunication services, artistic originals, publishing services) will be further elaborated on as this is an important point for users to understand.*

A discussion on the treatment of short-lived data, a.k.a, data that is consumed within one year

29. The SNA is very clear that expenses should only be capitalized if involved in production for more than one year. As such several countries in their initial estimates of data GFCF have made an adjustment to the final nominal estimate of own account output of data to represent the data which is consumed within one year and thus should not be capitalized.

30. However, as noted by the countries who have made these adjustments, they are at the moment considered quite arbitrary as information on the percentage of data used within one year is so far not readily available from businesses or any other source.

31. Practically, the compilation of economic statistics relies greatly on business accounting. Currently most assets in the 2008 SNA are also considered assets for the purpose of financial accounting. As such, when asked, businesses are able to separate capital expenditure from current expenditure relatively easily. This is not the case for data expenditure which, while not explicitly excluded, is not (yet) explicitly considered an asset within the international accounting standard. Making it difficult for business to distinguish between current and capital expenditure on data. *This link between the SNA framework and the International Financial Reporting Standards (IFRS) is an important concept requiring further exploration.*

32. Additionally, it is possible to incorporate the retirement of a large cohort of data assets within the first year of existence through PIM – See chapter 4. This would alleviate the need for an adjustment based on unavailable data. This treatment would also align with the proposal that was supported by the global consultation and expert group when considering the incorporation of data as an asset in the 2025 SNA (ISWGNA, 2023).

33. *Overall, this is an important concept that needs to be discussed and explored further in the handbook. Both the concept and the subsequent practical implementation are still being discussed by the task team; however, it is agreed that a final recommendation should be included in the handbook in order to maximize consistency across countries.*

Explanation on how data in the revised 2025 SNA differs from the decision to not include data in the 2008 SNA.

34. This concept of data needing to be digital and providing an economic benefit to the owner also solidifies what data is in the 2025 SNA compared to the 2008 SNA. Previously, when the SNA implicitly took the view that data had value but was non-produced[1], it viewed 'data' as the embodied information content of what is now typically referred to in the new lexicon of data value chains as the information content of 'observations' or 'observable phenomena.' In simple terms this is information content that had not yet been recorded[2]. This consideration of data as embedded information is one reason why the authors of 2008 SNA, in an attempt to limit the possibility of implicitly "capitalising knowledge" (Ahmad & van de Ven, 2018), chose to limit the value of databases to only include the cost of preparing data in a format that conforms to the "database management system (DBMS)" while excluding the cost of acquiring or producing the data (2008 SNA §10.113) (EC et al., 2009).

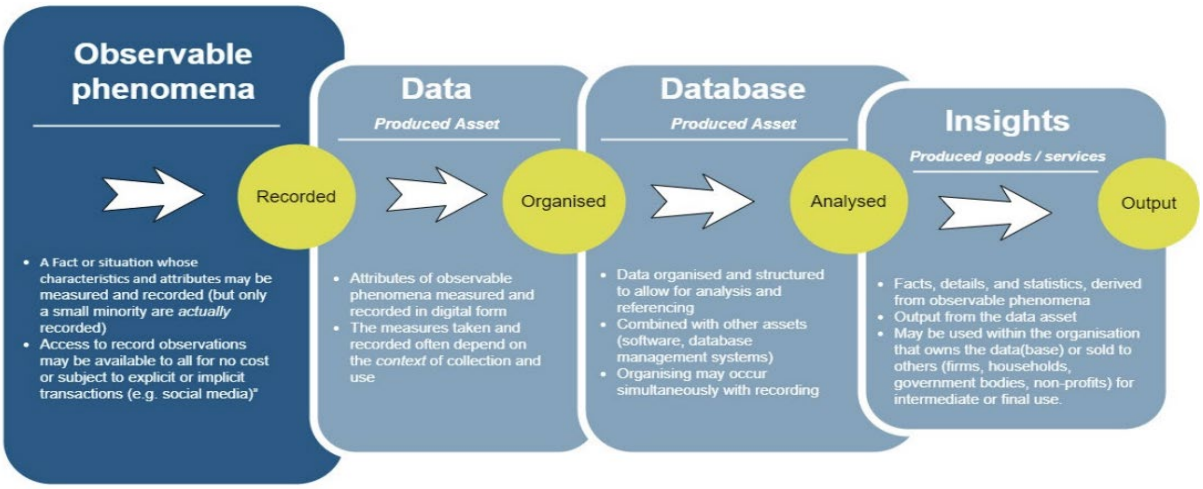
35. With data now considered the information content that comes from accessing and observing phenomena; and recording, organizing, and storing information elements from these phenomena rather than the embedded information contained in the phenomena themselves, the concept of data, for the purpose of economic measurement and analysis has moved along the data value chain as represented in Figure 1. This movement introduces a

[1] It is considered "implicitly" as data was not explicitly considered a non-produced asset. Rather when an explicit transaction in goodwill was made, it was considered that some of the value of the goodwill (a non-produced asset) was derived from the data contained within business.

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

clear element of production (as defined in the SNA) to the process and creates additional value to the output.

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



(Mitchell, Ker, & Lesher, 2022)

Where data is currently classified in various classifications used in economic statistics

36. *This section includes several paragraphs detailing how data is (and is not) currently classified in the international statistical classification. Both the Central Product Classification (CPC) CPC and International Standard Industrial Classification (ISIC) are undergoing revisions to coincide with the revision to the SNA and Balance of Payments Manual. Data has been explicitly included in both classifications; therefore, this section includes how the explicit identification of data in the 2025 SNA matches with the work that has been done by the classification task team.*

Data can be produced by all sectors of the economy

37. Like all other assets in the economy, **data can be produced and used in production by all sectors of the economy including the government sector** (or non-market). Occupations listed as part of the sum of costs calculation would include those working for the general government and non-profit institutions serving households (NPISH) sectors, their output would be recorded as production of data and if capitalized, GFCF for their respective sectors. This may include both publicly available and non-publicly available data assets as both are considered as contributing to the production of the government sector. For example, data compiled by security forces, that assist in public safety or by government agencies that assist in the efficient delivery of government services (i.e., tax, social welfare) are clearly investments (GFCF in SNA terms) made by government that provide an economic benefit to its owner (the government) over future periods.

38. Like other assets that are publicly owned and made available to use, with no direct charge to the users, the services produced by these data assets are consumed collectively and theoretically the value that the public places on these assets may extend well beyond the sum of costs it took to produce them. The consumer surplus in this situation is no different from that applied to other services that the government provides but which consumers would happily pay more for (i.e. running water). This type of consumer surplus is just one of many different perspectives that can be used to estimate the "public value" of data and data sharing, however, the values attached are not consistent with the valuation concepts used within the national accounts (more discussion of this is in chapter 2) and so should only be used for academic purposes.

39. Theoretically, data can be produced by the household sector. However, since data for the purpose of SNA involves information content obtained through accessing and observing phenomena rather than simply anything saved digitally, this excludes videos, photos, blogs, and other self-published material from being considered as data in the SNA. That said, these activities may well constitute production of a good or service other than data. As such, it is thought that the contribution of the household sector to the overall amount of data production would be minor compared to other sectors.

40. *These paragraphs will benefit from some numerical examples, which show the change in treatment whereby some current government consumption, becomes government investment, such examples will be placed in the Annex.*

IV. Chapter 2 - Creating a nominal estimate of GFCF through the sum of costs method.

A. Chapter 2.1: The Sum-of-cost is the recommended method for measurement.

41. **It is recommended that data produced on an own account basis is valued using the sum-of-cost method as already established in the 2008 SNA.** The reason for this will be explained in the upcoming paragraphs. Other valuation methods beyond those outlined in the SNA have previously been tested by academia or statistical offices and will also be briefly discussed. However, these are considered either not consistent with the overall SNA framework or not practical for the consistent implementation across countries, a major requirement for the purpose of national accounting.

How the SNA values output

42. The 2008 SNA is not prescriptive in its choice of valuation methods for measuring output or asset, only on the valuation principle. That is, in order to make the SNA the powerful analytical tool that it is, it uses a single accounting unit, money terms (EC et al., 2009). Therefore, while the 2008 SNA does suggests using "values at the actual price agreed upon by the transactors" (2008 SNA §2.59) (EC et al., 2009) which makes exchanged prices the basic reference for valuation in the SNA, several other valuation approaches are mentioned as appropriate or acceptable in certain circumstances.

43. One of those circumstances is when there is an absence of market prices, a situation that is certainly applicable for data assets which are not sold. One of the central pillars of data's value is exclusivity: an organisation having data that its competitors do not possess provides a clear point of value. While publicly available data or data that is not exclusive can still be used in production, its potential for value adding is greatly diminished. For this reason, it has been consistently observed that business models are becoming more dependent on proprietary data (Nguyen & Paczosi, 2020), a trend that is likely to continue with increases in legislation aimed at protecting consumer privacy that prohibits or limits the sale of third party data (Corrado, Haskel, Iommi, Jona-Lasinio, & Bontadini, 2023). As such, and as shown in results published by Eurostat, the vast majority of data that is used in production is produced by organisations themselves resulting in only a small amount of market transactions in data relative to the amount being used in production (see Box 1).

Box 1: Information and figures from Eurostat relating to businesses buying, selling, and using big data

44. Importantly, even if a greater number of market transactions existed, the extreme heterogeneity of data presents another issue for their use in the accounts. The highly contextual and independent nature of data means that the collected prices are not nearly as representative of other transactions as is the normal case for many other goods and services.

45. This absence of market prices and the difficulty in using them when they are available, means that compilers must search for alternative valuation methods. The 2008 SNA provides several options for valuing output or assets where market prices are absent. These include.

- Estimating a value according to costs incurred (2008 SNA §2.59)
- by referring to market prices for analogous goods or services (2008 SNA §2.59)
- estimating a discounted present value of future returns expected from a given asset (2008 SNA §2.60)

46. The very low level of market transaction in data combined with the heterogeneous nature of data, make the second option unfeasible. The third option is often used within the national accounts, for example when valuing natural resources. However, while data is often referred to as the "new oil", from a valuation perspective there are clear differences between the characteristics of data and natural resources that impact the ability to accurately forecast future returns. These include the near limitless stock of future data, the lack of homogeneity in data products as well as the highly contextual nature that data is used in production. As such, the production of forecast that would allow for an accurate estimate of the value of all data assets within the economy was considered unrealistic.

47. Due to data characteristics the 2nd and 3rd valuation options are considered untenable. It is recommended that data is estimated based on the sum of costs involved in its production.

How the sum of costs is put together in practice

48. The 2008 SNA provides a simple description of the sum of costs approach. It lists output as the sum of the following items: intermediate consumption, compensation of employees, consumption of fixed capital and other taxes on production less other subsidies on production. For market producers, it is necessary to also include an estimated mark-up to account for its net operating surplus. (EC et al., 2009)[3]

49. All countries that have produced estimates of data assets have used this sum of costs approach when constructing estimates of data GFCF. When applied for the purposes of data it can be shown in the following manner as was done by the Bureau of Economic Analysis (see Figure 2), where for each occupation ω, industry i, and year t, the wage bill is calculated by multiplying the annual number of employees ($H\omega,i,t$) by the average annual wage ($W\omega,i,t$) and an occupation-specific time-use factor ($\tau\omega$) that reflects the time-effort that the occupation allocates to data-related activities. The parameter α is a markup that reflects other costs (not included in the wage bill) including capital costs, intermediate consumption, and an operating surplus (Calderón & Rassier, 2022). This final markup is covered much more extensively in late in the section.

Figure 2: BEA Sum of costs approach presented mathematically.

$$Ci,t = \alpha \sum \tau\omega W\omega,i,t H\omega,i,t$$

(Calderón & Rassier, 2022)

Detailing aspects of the sum of costs formula in more detail

50. The remaining portion of this section covers each component of this formula with chapter 2.2 covering the creation of a wage bill. This includes the occupations chosen and the labour outputs used. Chapter 2.3. covers the occupation specific time-use factor. All countries have applied some form of time use adjustment to acknowledge that employees are unlikely to spend 100% of their time producing data assets, this adjustment seeks to appropriate capture the proportion of their labour that is *actually* contributing to producing capital assets. Chapter 2.4 covers the non-labour markup, including what this represents economically and how countries have obtained estimates for this. Chapter 2.5 covers some

[3] For non-market producers net operating surplus is zero by convention

additional adjustment that countries may choose to make to their nominal estimate as well as a discussion on indicators that can be used to produce quarterly estimates of data.

51. The recommendations in these chapters involve two components. Due to the infancy of data compilation, there is a shortage of information on various assumptions used to compile the nominal estimate. As such, countries are encouraged to continue to research various aspects in order to improve the quality and robustness of the output. In this vein the recommendations labelled as desirable will likely involve the incorporation of data that is not yet available for many countries.

52. At the same time, more defined recommendations, provided in order to promote international comparability estimates of data are required in the interim. These recommendations labelled as acceptable are also to be included in these sections and should be viewed as a basic methodology that countries can use to compile initial estimates.

B. Chapter 2.2: The choice of occupations when producing data GFCF using the sum of costs approach

53. The choice of occupations is an important foundation stone in the construction of nominal estimates of GFCF of data. However, unlike other IPP assets, it has been observed that the production of data can spread more broadly across a wider range of occupations. While there are some obvious occupations that have been included in all studies completed to date, some work has also included other occupations, often with a small time share or intensity ratio.

54. All countries that have produced estimates have published their list of occupations chosen, these are provided in Annex 1. The manner in which these occupations have been chosen can be split into two methods. The first is a selection through expert knowledge of the analyst or by using some key words to identify occupations within a statistical classification. The alternative method is to nominate specific task that are associated with the production of data and then use machine learning to review job advertisements and identify advertisements for occupations that include these specific tasks. This process not only identifies data producing occupations but is able to also provide a systematic measure of the occupations' level of data intensity or time factor spent producing data, such information is also required (see chapter 2.3). This method is outlined in more detail in Box 2 that presents the work done by the BEA. However the use of machine learning also creates additional challenges that needs addressing and is also quite resource intensive which may make it unavailable for many countries. *As such, this section will outline both a desirable approach involving a systematic approach to selecting occupations which countries may aim for as well as an acceptable approach to selecting occupations without the use of machine learning.*

Box 2: A summary of the process by which occupations are selected (To be completed by the BEA)

55. When choosing occupations, it is important to consider how an occupation fits into the data values chain presented in Chapter 1 (See figure 1). The SNA research guidance note endorsed by the AEG as part of the revision to the SNA outlined expenditure on the following tasks as possibly contributing to the production of data.

- Planning, preparing, and developing a data production strategy,
- accessing, recording, and storing information embedded in observable phenomena,
- processing, cleaning, and organising the data to allow for use in productive activities.

56. The tasks listed may entail both labour or non-labour costs, however, they can be considered a starting point in the selection of occupations.

57. *This section will expand to better explain the factors that countries should consider when choosing the actual occupation categories (and classification level) to use including data availability in order to ensure that outputs are able to be created on a regular basis. A core list of occupations that can form the genesis of an occupation list used for the construction of the labour cost by countries will be provided.*

58. *This will also cover the topic of double counting when compared with Computer Software and R&D.*

A discussion on when occupations are actively producing data compared to those that are passively involved in data production

59. Labour cost component used as part of the formula displayed earlier (See Figure 2) should involve the labour cost of employers and employees whose *primary* role is related to completing one or more of the previously mentioned tasks. However, this is not always an easy distinction to make. For example, the concept of which occupations are involved in accessing, recording, and storing information elements could arguably be considered extremely broad if followed scrupulously.

60. Today many electronic goods that contain internet connectivity include data tracking software that feeds information elements from observable phenomena associated with the product, back to the producer who use this to create data[4]. In this situation it could be argued that the salesperson who facilitated the sale of the product has assisted in the producer being able to access and record observable phenomenon. As such a (very small) portion of their wage could be considered expenditure related to the production of data. On this basis many retail workers, including cashiers in supermarkets might be considered as producers of data. This is not the intention of the measurement framework and an interpretation such as this should not be followed. Rather, **occupations should be included if their job involves tasks (such as those mentioned previously) that pro-actively contribute to the production of data as opposed to occupations that play a passive role in data production** such as the cashier and retail worker.

61. *Although this final point is generally agreed upon by the task team, further nuancing is expected before the finalisation of the handbook which will be included here.*

C. Chapter 2.3: Applying time factors to occupational labour costs

62. It is well accepted when compiling estimates of own account capital formation that the entire output from a single worker is unlikely to contribute entirely to the production of an asset. This may be due to the specific requirements of their role, their skill limitations or possibly the characteristics of the industry or organizations that the occupation is placed in. These time factors can be applied at the occupation / industry or even aggregate level.

63. The application of time factors in the compilation of own account computer software is recommended with the final Eurostat-OECD report on land and other non-financial assets. It specifically recommend making the adjustment at the most granular level possible since "workers in specific industries may spend more time on own-account software and database production, and workers in larger enterprises may be able to spend more time on own-account software and database production than those in smaller firms" (Eurostat-OECD, 2019).

64. A similar phenomenon is almost certainly occurring for the production of data and therefore **It is recommended that, if possible, a time share proportion should be applied at either the occupation or industry level rather than at an aggregate whole-of-economy level.**

65. That said, data on these proportions is very difficult to capture, with many of the countries using best guesses or applying upper and lower bound time factors resulting in upper and lower bound estimates of data GFCF. Since such a range estimation is not be suitable for the inclusion of data GFCF in the core national accounts as countries should produce a single estimate using the best available information.

66. While the data has been hard to find, there are several examples of work that has been undertaken to inform compilers. The Japanese cabinet office uses information from the

[4] *Often this is information on the manner that the products are used. Clients often accept this retrieval of information as part of commencing use of the product and connecting it to the internet.*

special internet survey to estimate how much time workers are spending on data related work, more information on this is presented in Box 3.

67. Alternatively, the OECD has used natural language processing (NLP) on job advertisements to estimate the time factors/ data intensity of occupations/sectors as well as identifying the occupations themselves. This work has tended to produce slightly lower time share ratios; however, these lower rates tend to be applied to a larger number of occupations and sectors, thereby producing similar overall estimates of Labour costs. Work is ongoing about how easily this work can be replicated across countries or can be combined with previously chosen occupations to make more feasible for a broader range of statistical offices. Box 4 provides more information on this process.

68. *The section will further discuss the appropriateness and feasibility of using time factors generated in country A to create estimates of data in country B. The handbook will contain an inventory of time factors used by countries to assist in countries determining their own, as well as containing more basic guidance that countries can follow in the absence of any country specific data.*

Box 3: The Japanese "special internet survey". How the survey is run & some preliminary results. To be completed by the Japanese cabinet office.

Box 4: OECD "summary of natural language processing (NLP) on job advertisements in the United Kingdom, Canada and the United States." To be completed by OECD

D. Chapter 2.4: How countries may derive the mark-up applied to labour costs to represent intermediate consumption, consumption of fixed capital (CFC) and operating surplus. Referred to as 'non-labour expenditure'.

69. There is no single correct way to estimate the non-labour component of investment in data assets. Conceptually, as long as a component of the final nominal estimate of GFCF contains expenditure representing intermediate consumption, consumption of fixed capital used in production of the asset and for market producers, an amount covering the producers net operating surplus then it can be considered an appropriate estimate of own account GFCF. However, countries are unlikely to have this level of specific information for the production of data.

70. No countries have surveys only targeting data producers which might capture this specific information, therefore, to date, all countries have estimated this expenditure by applying a mark up to the labour cost to estimate the expenditure covering intermediate consumption, depreciation and GOS. This main assumption associated with the method is that there is a consistent production function for the data asset, in so much that a consistent amount of non-labour input is required for each unit of labour input. This is not a heroic assumption and is considered quite acceptable in lieu of actual data on expenditure of non-labour inputs, the means that the focus becomes how such a mark-up should be derived?

71. Ideally it would be desirable to have country and asset specific information being applied to specific occupations and industries, in order to represent the real-world differences occurring across different economies. Due to this, **countries are encouraged to investigate potential data sources that may provide accurate estimates of the non-labour expenses involved in the production of data assets.**

72. In more recent work, efforts have been made to have the non-labour mark up more reflective of the industries likely producing the majority of the data assets. Using industry estimates of intermediate consumption, consumption of fixed capital and operating surplus taken from Supply-Use or Input-Output tables can be applied against the overall output for that industry to create various ratios which can be used as a proxy mark-up and applied to the labor cost estimate to create a final estimate of total expenditure. These ratios, created on an industry by industry basis is an improvement on the initial rudimentary adjustments with a single ratio (representing all non-labor expenditure) applied to the entire labor cost estimate.

73. **It is desirable for any ratios used to mark-up the labour cost component to be calculated from a wide range of industries reflecting the various occupations that have been chosen as contributing to the investment in data. The industry's contribution should be weighted based on the proportion of overall labour costs derived from that industry.**

74. **An acceptable alternative would be to simply select one or two industries that contain a large amount of output from occupations that have been chosen**. While the industry classification varies across countries, of the work published so far, countries have used derivatives of "Computer programming, consultancy and related activities" and "Information service activities". These equate to industries 62 and 63 respectively, of the international standard of industrial classification (ISIC Rev. 4). Ratios for intermediate consumption, consumption of fixed capital and operating surplus could be taken from these industries in the absence of more detailed or accurate information on non-labour expenses.

75. If specific information relating to the production of data on an industry basis is not available, using the same mark up as that applied to similar IPP assets in existing compilation is also considered acceptable.

76. *This section could also include a discussion on how to estimate an appropriate rate of return applied to market producers.*

77. *A numerical example showing how non-labour mark ups are calculated and applied to the labour estimates is provided in Annex 2.*

78. *Case studies summarizing how certain countries have compiled nominal estimates of data investment will be provided in Annex 3.*

E. Chapter 2.5: Potential additional adjustment covering a range of concepts.

This section covers the following points.

79. *That* ***no adjustment should be made to represent expenditure on repair and maintenance of the data assets. All expenditure is considered to extend the asset life and thus should be considered GFCF.*** *This would include the clear guidance already provided by the 2008 SNA separating out the difference between new GFCF investment and maintenance expenditure treated as a current expense and how this applies to the production of data assets.*

80. ***Adjustment may be made to the nominal estimate to represent data that is consumed within one year****. Following on from the conceptual point raised in chapter 1, the task team may consider such an adjustment is appropriate. If so, this paragraph would outline how it has been applied by several countries already and provide information to make the adjustment as consistent as possible across countries.*

81. ***The possibility of capturing information on market transaction on data sales.*** *Some countries have discussed adding additional questions to business surveys to obtain additional information on expenditure related to purchasing established data assets. If these transactions are recorded, then conceptually the amounts from these can be used to shift GFCF in data from the industry that has undertaken its production, and captured via the sum-of-cost methodology, to the industry that has subsequently purchased the asset, captured via the business survey.*

82. ***The production of quarterly estimates of data output.*** *So far all estimates of data produced by countries have been on an annual basis. A majority of countries produce estimates of GDP on a quarterly or even monthly basis. Depending on how these outputs are produced and disseminated, it is likely that most countries will need to produce an estimate of data output and GFCF on a quarterly basis. This section would discuss how countries might approach this need. It is unlikely that all of the source data discussed in the previous section is available on a quarterly basis, therefore some form of indicator will be required to move the annual estimate forward. This is a standard procedure in compiling the national*

accounts. This section will discuss the various options available to countries and make some recommendations for countries to follow.

V. Chapter 3 – Creating volume estimates of data.

Why deflation using price indexes is seen as the most pragmatic route to deriving volume estimates

83. In the System of National Accounts, certain high profile indicators including estimates of production (i.e., GDP) are presented in volume terms as well as nominal terms. Until now, this handbook has focused only on the production of a nominal estimate of data investment. A standard way for nominal estimates to be represented on a volume basis is to be deflated by taking into account the change in the price of the output. This change is usually calculated by recording the difference in the market price in the current period to the previous period. When estimates are calculated as the sum of the costs, as is recommended for data, the overall change in price may be calculated as the weighted change in price of the inputs used in production[5].

84. Within the national accounts, volume estimates are occasionally calculated based on an output indicator which often represents a quantity good or service produced. This is usually for estimates of production and trade which involve natural minerals or agriculture products since they are relatively homogeneous and quantity counts are relatively easy to obtain.

85. In one regard the quantity of data is relatively easy to measure. The bits and bytes that make up data when saved to a computer take up a specific amount of memory. Due to this, it should be, theoretically, possible to measure the additional quantity of data produced each period when compared to the previous period. In fact, this undertaking has already been done by several organisations who estimate that around 2.5 quintillion bytes are created every day with the overall amount of data doubling every two years[6]. However, despite the presence of this estimate there are several reasons why such an estimate of quantity cannot be used for compiling volume estimates of data in the SNA.

86. The first and most important reason is that this incredible number includes a large amount of data that is *not* data as defined within this handbook and the 2025 SNA. Rather it is closer to the more specific definition of data as Internet Protocol (IP) traffic or the volume of digitised information stored on servers and other hardware. A large amount of this data includes photos, text messages, email and other communications that fail the 2025 SNA data definition, as they are often not produced by accessing and observing phenomena, are not used in productive activities or are in fact an externality of digital service delivery. This distinction was discussed previously in chapter 2.

87. The other reason data quantity is not able to be used within the 2025 SNA is due to the lack of consistent relationship between the quantity of data within data assets and their subsequent value. The majority of the data value comes from the content of the information and the context that it has been gathered or could be used. Both these factors are often unrelated to the size of the data. While it is true that data that contains more information is likely to be worth more than data with less information, the relationship is not consistent enough to create any form of reliable value based solely on quantity. Proof of this is the evidence that the huge increase in data production observed in the economy is driven more by the declining cost and increasing efficiency of data storage than by a positive liner relationship between the amount of data produced and its value.

[5] Theoretically, a price index based on changes in the price of the output produced can also be applied to the overall estimate, however estimates compiled using the sum of costs approach are usually done this way due to the absence of output prices.

[6] A quintillion is 10 raised to the power of 18, that is, a 1 followed by 18 zeros. See https://www.the-next-tech.com/blockchain-technology/how-much-data-is-produced-every-day-2019/

88. Overall, while a quantity estimate of data production may be achievable, **the use of a direct volume measure within the National account is deemed inappropriate due to the heterogeneous nature of data as well as the volatility and treatment of prices applying in different markets.** Interestingly, data is not the only good that falls into this category, with the 2008 SNA pointing out that the volume estimates of electricity (as well as other utilities) should not be derived through quantity, even though it appears relatively feasible, due to the difficulty in capturing a single representative price. (SNA §15.103 (EC et al., 2009)

A discussion on the type of price indexes that might be considered

89. Even though the nominal estimate of data investment has been constructed via a sum-of-cost approach, the price indexes used to deflate this nominal estimate do not *need* to reflect these input costs. The SNA is open to the idea that the deflation of output compiled via the sum of costs can be undertaken using a pseudo-output price index (2008 SNA §15.117) (EC et al., 2009). A clear benefit of this approach is that when it is compared to the aggregate input price index the difference reflects the productivity growth thought to be occurring in the production process price index based on changing price of the final output.

90. That said, since there are few market prices for the sale of data, a traditional price index calculated based on recording the change in transaction prices paid for data is likely not obtainable. Therefore, it is expected that the price indexes used will either be based on a similar asset or reflect the inputs that were used.

91. In preliminary work around data, countries often created volume estimates by deflating the nominal estimates with the established price index for Computer Software, Computer Hardware, or Research & Development. This in itself is not an outrageous proxy index, data is an intellectual property product and so in the absence of an alternative, using an available price index of an asset with similar characteristics can be considered a suitable option.

92. Importantly however, the choice of the price index can have a significant impact on the volume estimates, especially if the price indexes being considered are trending is opposite directions.

93. Often the price indexes representing the market costs of computer software and hardware, and used in initial work, showed strong deflations, due to the overall cost of computer software and hardware declining in value over the past 5-10 years as technology has advanced. Such a price movement would appear to be at odds with the production of data considering that such a large input cost is labour costs, which have grown consistently through this period. The ABS demonstrated this difference in their initial work (See Box 5), which persuaded them to use a mixture of labour and intermediate cost in the production of their price index.

94. The use of a weighted price index which reflects the input costs contributing to the final estimate of data GFCF, would provide a better reflection of the change in costs associated with producing data assets. Such an approach received wide support from the Joint Eurostat – IMF Task Team on Measuring Data as an Asset in National Accounts, Therefore, where possible, **this handbook recommends deflating nominal estimates of Data GFCF using a weighted price index, based on the specific ratio of labour and non-labour inputs used in the production of data GFCF.**

95. Depending on data sources available, both price information and data used in the production of the nominal estimate, it may only be possible to use a set ratio of labour and non-labour in the price index. This may be because of the lack not only of data source but also of clear set of inputs representing the non-labour component, as may be the case if the non-labour component is created via ratio (see chapter 2). Regardless, it is still preferable to use a mixed price index, even with a set ratio, rather than a price index wholly influenced by one or the other or based on the market price of a similar asset (i.e., software R & D). However, it is important that the ratio is reviewed regularly enough to ensure that it remains relevant to the current inputs being used.

BOX 5: Demonstration of impact on volume GFCF estimates based on using divergent price indexes - to be completed by ABS.

BOX 6: Showing the difference in European price indexes for the different inputs used for data (i.e., labour vs software / hardware / R & D output) – to be completed by Eurostat.

BOX 7: Case study of the how the BEA creates weighted price indexes to deflate nominal data estimates. - to be completed by the BEA.

The inclusion of quality adjustments in the price index used

96. The concept of including an adjustment to the price index to represent quality improvements to the finished product has been discussed by both the data task team and other organisations who have created estimates of data assets. Interestingly, in the final report of the joint Eurostat – OECD task force on land and other non-financial assets, the prospect of including quality adjustments to the price indexes used was largely absent, perhaps an acknowledgement of the conflict that exist between their conceptual reasoning with practical implementation. Despite this, there has been several requests for the handbook to cover this topic.

97. *The following paragraphs would cover both the advantages of including quality adjustments as well as discussing the concerns that such a recommendation would create. The includes the possible contradiction with the current guidance contained within the European system of national accounts that forbids adjustments related to quality when calculating output via a sum of costs approach.*

VI. Chapter 4 – Creating Capital Stock estimates.

A summary of the PIM, explaining why it is the recommended approach for compiling estimates of consumption of fixed capital and capital stock within the 2025 SNA

98. The SNA discusses the creation of balance sheets and therefore includes guidance on compiling estimates of capital stock and consumption of fixed capital. On several occasions it explains the challenges caused by obtaining estimates of capital stock and consumption of fixed capital direct from business in the same manner as the estimates for expenditure are obtained.

99. Therefore, while the 2008 SNA does not explicitly recommend the use of the PIM, it does state that "*consumption of fixed capital must be valued with reference to the same overall set of current prices as that used to value output and intermediate consumption*"(2008 SNA §6.248) (EC et al., 2009). In simple terms this suggests that the same information used to derive output (such as GFCF) should be considered when deriving estimates of consumption of fixed capital.

100. Furthermore the 2008 SNA recommends that '*independent estimates of consumption of fixed capital should be compiled in conjunction with estimates of the capital stock. These can be built up from data on gross fixed capital formation in the past combined with estimates of the rates at which the efficiency of fixed assets declines over their service lives'*2008 SNA §6.249) (EC et al., 2009). This is in essence describing the Perpetual Inventory Method. As such, the PIM has become the standard method used by all statistical officers to compile estimates of Consumption of Fixed Capital and capital stock in their National Accounts.

101. This handbook does not cover the specific concepts and practical implementation of the PIM as this is a considerably detailed and technical endeavour. Furthermore, existing works such as the OECD manual on measuring capital are already in place that cover these topics at length. Finally, since the PIM is used by all countries, including by all those that created already created capital stock estimate of data, it is seen as redundant to discuss the generic concepts. Rather this section focusses on the specific assumptions that countries have

used within the PIM, when calculating consumption of fixed capital and capital stock estimates of data.

Most countries have applied the same assumptions for data as they have for other produced fixed assets

102. *There would be several paragraphs detailing how those countries that have created data estimates via the PIM have applied the same Age-Price profiles, Age-efficiency profiles, and retirement profiles as other assets.*

103. *This section would discuss the advantages of being consistent with these assumptions across the different assets as mentioned below.*

104. *There are benefits to applying consistent and transparent assumptions when real world source data on the subject is hard to come by. Users are then able to understand how the estimates are derived, removing the idea that the PIM is a form of black box. Additionally, since these estimates are highly modelled, it can be hard for a statistical office to demonstrate why assumptions for one asset should be different to another. Finally, the difference in the final output cause by altering the age-price, age-efficiency and retirement profiles are relatively minor when compared with difference cause by the application of different asset lives. This was again shown in the specific case of data by recent work by Destatis that showed that 'the assumed asset service life has the biggest impact on [capital stock] results.*

What is an appropriate asset service life for data assets, the section would include recommendations

105. Work by Destatis, Statistics Canada as well as by the ABS has suggested that the single most impactful assumption applied in the compilation of capital stock estimates of data is the asset service life (Smedes, Nguyen, & Tenburren, 2022; Statistics Canada, 2019). *This section would detail the work by these three countries and the various results achieved based on applying different asset service lives.*

106. When determining an appropriate asset life for data there is several important things to consider. First is that with the exception of the odd data source (See Box 8) information on the use of assets in production is extremely difficult to obtain and is usually subject to various biases. The OECD manual on measuring capital recommends various approaches for obtaining information on asset lives, this includes those prescribed by tax authorities, company accounts, statistical surveys, administrative records, expert advice, and other countries' estimates (OECD, 2009). Importantly, not all of these are applicable for data. For example, currently no government recognises data as an asset for which depreciation can be claimed to reduce a tax liability, therefore immediately ruling out one potential source.

107. Another consideration is that the service life applied does not represent a single asset but rather the average of a cohort of assets. Within the PIM it is not feasible to apply specific asset lives to individual assets, rather an average is applied to a cohort of assets with similar characteristics. Since data is extremely heterogeneous, deciding on this average for data is perhaps more challenging than for other fixed assets. It is relatively easy to find examples of data that are used instantly by businesses and are likely to have little value past the immediate future (i.e., consumer searching habits online) as well as find data that contains value for the medium to long term, (i.e., business transaction data, or data on natural occurrences such as rainfall or temperature). The introduction of AI also brings an additional perspective as seemingly worthless data can be re-used in production for the purpose of machine learning.

108. In an attempt to make these cohorts as similar as possible, assets are often broken down below their aggregate level based on specific characteristics of the asset which may impact the asset life, (e.g., dwellings are often broken up into wood houses vs concrete houses, public infrastructure is broken up between roads and railways). A similar disaggregation is conceptually possible for data. There are a large number of taxonomies that already exist for classifying different types of data, and while conceptually accurate to do this, as discussed in Box 9, practical limitations likely result in only a single type of data

being represented in the PIM. However, a more rudimentary version of this delineation based on industry may be possible. Although it is important to note that any delineation between data types has to be supported by the ability to compile commensurate price indexes and GFCF estimates.

109. *This section will conclude with a discussion on the table displaying the different asset service lives applied by the countries who have already produced estimates of data assets (table 1).* It will present that while the very first estimate of data published by Statistics Canada included an asset life of 25 years all subsequent estimates, including a subsequent project by Statistics Canada have included a much lower estimate of between 3 – 8 years. This length is supported by discussions within the task team that favour an asset life of this length. *Additionally, a final decision on the service life will also be impacted by the decision to recommend any adjustments to the nominal estimate of data investment to remove data that is used within one year.*

110. Therefore, while the task team has not finalised guidance on this matter, **it is the current view that the handbook provides a specific recommendation of an asset live within a relatively narrow range of 5 years.**

Table 1: Table of current service lives applied by countries (to be completed by task team consultant)

Box 8: Japanese investigation into the length of use of data assets by businesses (to be completed by Japanese cabinet office)

Box 9: Discussion on breaking up data into different types to apply different asset lives (to be completed by task team consultant)

VII. Chapter 5 – Overarching measurement questions

How is the sale of data represented in the accounts, including international transactions

111. *So far, the handbook has covered almost exclusively the compilation of estimates associated with the investment of data, that is GFCF, depreciation and capital stock. However, while it is a minority, data is sold as well as the sale of services associated with data assets. These paragraphs would cover under what circumstances the sale of data may be considered as the sale of an asset and recorded in the capital account, compared to the sale of a copy, sale of a license or simply a different service connected to the data asset, all of which would be considered as output, and recorded in the production account. This would also cover the scenario of data produced for sale and how recorded selling/purchase transactions should be considered in unison with the overall sum-of-cost methodology. This section would include some numerical examples of these scenarios included in an annex.*

112. *Several paragraphs would be devoted to transactions in data that cross international boundaries. Classification covering the balance of payments and trade has been updated so recommendations would be included on when transactions should be included in the capital and current account.*

How data can be back cast for implementation into Aggregate GFCF estimates and the PIM

113. *Back series will be needed, both for time series representing final demand but also for use in the PIM so that capital stocks and CFC can be generated. There would be several paragraphs providing guidance on how best to back cast GFCF in data, this would include both preferable and acceptable options for countries to follow.*

114. There is firm support from the task team that the back casting of data should be done to produce as realistic time series as possible. So far this has been approached in several different ways by those countries who have already produced estimates of data. This includes using a variety of business indicators not specifically related to data, to move data back, as

well as tapering the impact by using an evolving time factor to increase the intensity of the occupations.

115. *The paragraphs will discuss the advantages and disadvantages of various approaches to back casting. The handbook will recommend an ideal approach but also list some acceptable methods since it is unlikely that source data will be available for all countries.*

VIII. Chapter 6 – Conclusion and annex

A. Annex 1

List of occupations used by countries in initial estimation of data.

B. Annex 2

Numerical example showing basic construction of sum of costs approach, involving calculation of labor cost and then mark-up based on industry ratios applied. This would include some real-world case studies from countries that have already completed this work, e.g., Pakistan, Canada, Germany.

C. Annex 3

National Accounting T diagrams showing the sale of data, either as a sale of a produced non-financial asset or as an output within the production account.

Bibliography

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

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

Corrado, C., Haskel, J., Iommi, M., Jona-Lasinio, C., & Bontadini, F. (2023). *Data, Intangible Capital, and Productivity.*

EC et al. (2009). *2008 System of National Accounts.*

Eurostat-OECD. (2019). *Final report from task force on Land and other Non-Financial assets - Intellectual Property Products.* https://ec.europa.eu/eurostat/documents/24987/725066/Eurostat-OECD+Report+on+Intellectual+Property+Products.pdf.

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

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

Nguyen, D., & Paczosi, M. (2020). *Measuring the economic value of data and cross-border data flows: A business perspective".* https://doi.org/10.1787/6345995e-en.

OECD. (2009). *Measuring Capital, OECD Manual.* https://unstats.un.org/unsd/nationalaccount/docs/oecd-capital-e.pdf.

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

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

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