Twenty-seventh Meeting of the Advisory Expert Group on National Accounts

Inter-secretariat Working Group on National Accounts

Washington, D.C

28 – 30 October 2024

SNA/M3.24/03

Cover Note

Handbook on measuring data in the System of National Accounts

Handbook on measuring data in the System of National Accounts

Cover note for AEG

While the introduction of data into the SNA production and asset boundary was endorsed at the 2024 UNSC, the commission also "emphasized the importance of addressing conceptual uncertainties, and stressed the importance of continuing to develop implementation guidance on the new recommendations to facilitate the implementation of the 2025 SNA in an internationally comparable way". While not explicitly mentioned, it was inferred that guidance on the compilation of data output and GFCF was one of several changes that this statement referred too.

In response the ISWGNA endorsed the creation of the Eurostat-IMF task team on Measuring Data as an Asset in National Accounts, which was tasked with providing "recommendations on how to estimate data as an asset in the national accounts by exploring data sources and methods. The TT should develop concrete guidance for countries on how to compile the requested data in line with the 2025 SNA" (Task force on measuring data in the National Accounts, 2023). The task team met frequently (seven meetings since August 2023) to discuss conceptual challenges, exchange compilation best practice and agree on a standard set of concepts and recommendations that can be used as a starting point for the compilation of data output and data gross fixed capital formation. The task team has produced "the handbook on measuring data in the national accounts". The handbook is envisioned to be made available publicly at the same time as the official endorsement of the revised SNA.

In chapter 1, the handbook begins by providing clarity around the definition of data for the purpose of the SNA. This includes clarifications on which types of data should be considered as an output of production, who is likely to be producing it, where the line exists between the production of data and the production of other goods and services (including other IPPs) and how it should be reported in the accounts. The clarifications are offered to provide insight to users to better comprehend the concept behind the data estimates, while also helping to guide compilers to better understand what kind of information and data sources are likely to be useful in the compilation of estimates of data output and GFCF .

Chapter 2 details the methodology involved in compiling nominal estimates of data output and GFCF using the sum of cost methodology. It includes the default source material and recommendations that have been gradually refined following testing by the task force. Unlike the clarifications included in chapter 1, the recommendations in chapter 2 (as well as the remainder of the handbook) offer both a default and aspirational recommendations. The task force agreed that a default set of compilation recommendations would greatly assist in promoting international comparability as countries began the task of compiling estimates of data output and data GFCF for the first time. At the same time, the inclusion of aspirational recommendations reflects the task force desire to see the methodology continually developed and improved, so that users are able to have greater confidence in the accuracy of the estimates. Aspirational recommendations also provide countries the ability to use already available source material and compilation practices, including those beyond what is required in the default recommendations. At the heart of the methodology is a list of occupations contributing to production of data assets and related involvement rates. When developing this list,

attention has been given to avoid double counting, through defining appropriate involvement rates. However, the handbook does not provide occupation lists or involvement rates for the production of other intellectual property products, which is out of its scope.

Chapters 3-5 cover compilation requirements such as the deflation of nominal estimates, the creation of depreciation and capital stock estimates through the PIM, recommendations on back casting and other compilation challenges. Similar to chapter 2, these recommendations usually contain both a default and aspirational variety.

Prior to the formation of the task team, several countries had already compiled experimental estimates of data investment. While this work formed the basis of the discussion, additional testing, source data gathering and research by task team members refined the recommendations to arrive at a final set of conceptual clarifications and compilation recommendations. A summary of these are presented below.

The handbook also contains case studies from countries that participated in the task team which often provided the basis for the aspirational recommendations. As well as offering real world reference material which other countries can learn from, these case studies provide a foundation for the overall continual improvement of data estimates.

Overall, this handbook has been written to serve two main purposes. Firstly, as a compilation guide to assist economic statisticians compile outputs consistent with the newly updated 2025 SNA. While also providing information to users who seek to understand what is represented in the estimates of data output and GFCF of data being produced by statistical offices.

Conceptual clarifications regarding the measurement of data production[1].

Definition of data: *Information content that is produced by accessing and observing phenomena, and recording and storing information elements from these phenomena in a digital format and that provides an economic benefit when used in productive activities.*

Treatment of non-digital data: *For the purpose of the 2025 SNA, only digital data is considered within the 2025 SNA production and asset boundary.*

Treatment of Auxiliary data: *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.*

Reporting of combined data and databases asset: *While conceptually different assets, both assets are usually produced or purchased together, therefore it is practically difficult to compile separate estimates for each. As such, for reporting purposes, data and databases are combined into a single detailed intellectual property (IP) product called data and databases.*

Separately identifying data even when used in the production of other IPP: *It is recommended that the value of produced data should be separately identified when capitalized. The value should be excluded from the cost of own account production of other goods and services regardless of how data dependent the final output is.*

[1] These clarifications are in line with the guidance note DZ.6 "The Recording of Data in the National Accounts" endorsed by the AEG and with the draft 2025 SNA.

The production of data across all sectors of the economy: *Data can be produced and used in production by all sectors of the economy including the government sector.*

The limit to expenditure associated with data production: *The incorporation of additional information content or improving the data's quality at either a granular or aggregate level is considered expenditure on data production and is considered GFCF. Analysis of the data to obtain insights or using the information contained in the data in productive activities is considered the production of a good or service other than data.*

Potential adjustments to account for short lived data : *All expenditure on production of data on an own account basis is regarded as a capitalised expense and should be classified as GFCF, with no adjustment made to represent data consumed within one year. However, if countries have obtained statistically appropriate information that can provide guidance on the proportion of data that is consumed within one year, they are encouraged to make such an adjustment. If such an adjustment is made, countries are recommended to publish the accompanying microdata to improve the comparability across countries. If an adjustment is undertaken, in order to follow convention and be consistent with other own account output consumed internally, the proportion of data produced on an own account basis but consumed within a year should not be explicitly identified, rather it is considered embedded into the output of the subsequent product.*

Recommendations associated with the compilation of estimates of data output and GFCF.

Valuation approach: *Data produced on an own account basis is valued using the already established sum-of-cost methodology.*

Choice of occupations: *In the absence of other sources, NSAs should use the list of occupations provided in this handbook for the compilation of data output. However, if possible, NSAs are encouraged to derive a list of occupations through an objective and systematic approach to better determine which occupations are most likely to be involved in data production. Importantly, occupations should be considered for the list if the occupation involves tasks which explicitly contribute to adding value to the production of data and the worker undertakes these tasks in a proactive and calculated manner.*

Involvement rates: *NSAs are recommended to apply the same or very similar involvement rates to those listed in this handbook. However, NSAs are encouraged to develop and use involvement rates derived through more systematic objective means.*

Non-labour costs: *It's recommended that non-labour costs are incorporated into the final estimate via a single mark-up applied to labour costs. Such a mark-up represents the costs of inputs, consumption of fixed capital used in production, as well as a return on capital (operating surplus). A single mark-up, based on the ratio of compensation of employees applied against total gross output from the "Computer programming, consultancy and related activities" (ISIC 62) and "Information service activities" (ISIC 63) – or similar available aggregate - is applied to total labour costs. However, NSAs are encouraged to apply specific non-labour information (including mark-ups) into the estimate separately so that differences in COFC and operating surplus across occupations and industries can be applied more accurately and transparently.*

Adjustment for short lived data: *No adjustment is made to the nominal estimate of own account data production to represent data that is consumed within one year, therefore, all data produced on an own account basis is capitalised accompanied by a short asset life. If possible, NSAs should seek to obtain appropriate information that can provide guidance on the proportion of data that is consumed within one year, to make such an adjustment.*

Market transactions: *Data assets that are purchased with exclusive rights are treated as a purchase of an asset (with an offsetting sale of an asse by the seller). However, assuming that the transaction is not a cross-border or inter-sector one, similar to sales of other second-hand assets, this transaction would net off and not impact the overall level of GFCF. Data that is purchased without exclusive rights is treated as a purchase of a*

copy and contributes to the GFCF of the purchaser if it satisfies the necessary conditions of GFCF, (i.e. use in production for more than one year).

Price index used: *Any price index used to deflate nominal estimates of data must reflect the price change observed in both the labour and non-labour costs involved in data production as well as appropriately accounting for the technological and quality improvements that have been observed in the production of digital products over the past several years. Ideally, chain volume estimates of data output should be compiled using a 'pseudo' output price index. This can be created by aggregating appropriate input price indexes and weighted to reflect the actual input costs included in the sum of cost calculation. An adjustment to reflect quality and productivity improvements made to the final output would be added to transform the input price index into a pseudo-output price index. As a default position, compilation can be undertaken using an output price index based on an alternative but similar product.*

Quality adjustment applied to price index: *The handbook recommends including an aggregate adjustment to reflect quality and productivity improvements made to the final output. Such an adjustment may be calculated as*

- *The difference in growth between the input price index for data and the output price index for a similarly produced product where market prices are available.*
- *The growth between the calculated difference in Input price and output price index for similarly produced products where market prices are available.*
- *The total factor productivity estimates for industries that contain a large amount of the occupation identified as data producers.*

Additional deflation consideration: *The intention of any recommendation in this handbook is not to overrule any existing regulations, rather the handbooks' goal is to assist countries compile the most accurate estimates of data output possible. It is the view of the task team that the introduction of a quality adjustment on top of an input price index is conceptually appropriate and would improve the accuracy of the final estimate. It is accepted that countries will continue to adhere to other frameworks and standards that oversee the compilation of their national accounts.*

Parameters used in PIM (Excluding Average service life): *Countries should apply the same parameters in the compilation of depreciation and net capital stock of data as applied currently to other IPPs. However, countries should aspire to continually collect additional information on different assumptions and parameters to refine and improve the estimates of depreciation and capital stock being compiled.*

Average service life: *In the absence of other information, countries should apply a default average service life of 5 years for data assets, Ideally, countries should aspire to break up the estimate of data investment by industry in order to allow for different service lives to be applied based on the industry producing the data.*

Recommended length of back series: *In the absence of additional information to the contrary, this handbook recommends incorporating a time series at least the period covering 1985 – 1995.*

The AEG is invited to:

- confirm if the recommendations associated with the compilation of estimates of data output and GFCF are clear and sound;

- provide detailed written feedback on the full draft handbook until 15 November 2025.