

Handbook on Data as an asset

Erich Strassner – IMF

Nicola Massarelli – Eurostat

Slides prepared by John Mitchell – consultant

27th meeting of the
Advisory Expert Group on National Accounts
28-30 October 2024, Washington DC, USA

Data TT – General timeline



Progress of task team

- Seven meetings between July - 23 and Oct-24
- Task team has reviewed and provided feedback on handbook chapters
- The handbook reflects the testing that has occurred over the past 12 months, specifically on:
 - Choice of occupation and involvement rates
 - Non-Labor mark up methodology.
- Several countries (e.g. Japan, Germany, Pakistan) have undertaken new research or tested alternative methodologies to improve source data available.

Structure of the Data handbook

Chapter 1 - Defining the *conceptual boundary* of data for inclusion in the SNA

Chapter 2 - Compiling a *nominal estimate* of output and GFCF through SoC method

Chapter 3 - Compiling *volume estimates*

Chapter 4 - Compiling *capital stock estimates*

Chapter 5 - *Overarching measurement and conceptual questions* discussed and explained;
Conclusion and condensed list of *recommendations*

Handbook to serve multiple purposes:

- Recommendations at a technical level for *compilers*
- Conceptual background and explanation of methodology for *users*

Definition of Data

- Definition of data tweaked on a minor basis, removal of “**organizing**” to provide better demarcation between data and databases.

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

- This reflects the change in the 2025 SNA draft.
- Non-digital data and Auxiliary data excluded

Measurement concepts have remained as proposed...and in line with 2025 SNA text 1/2

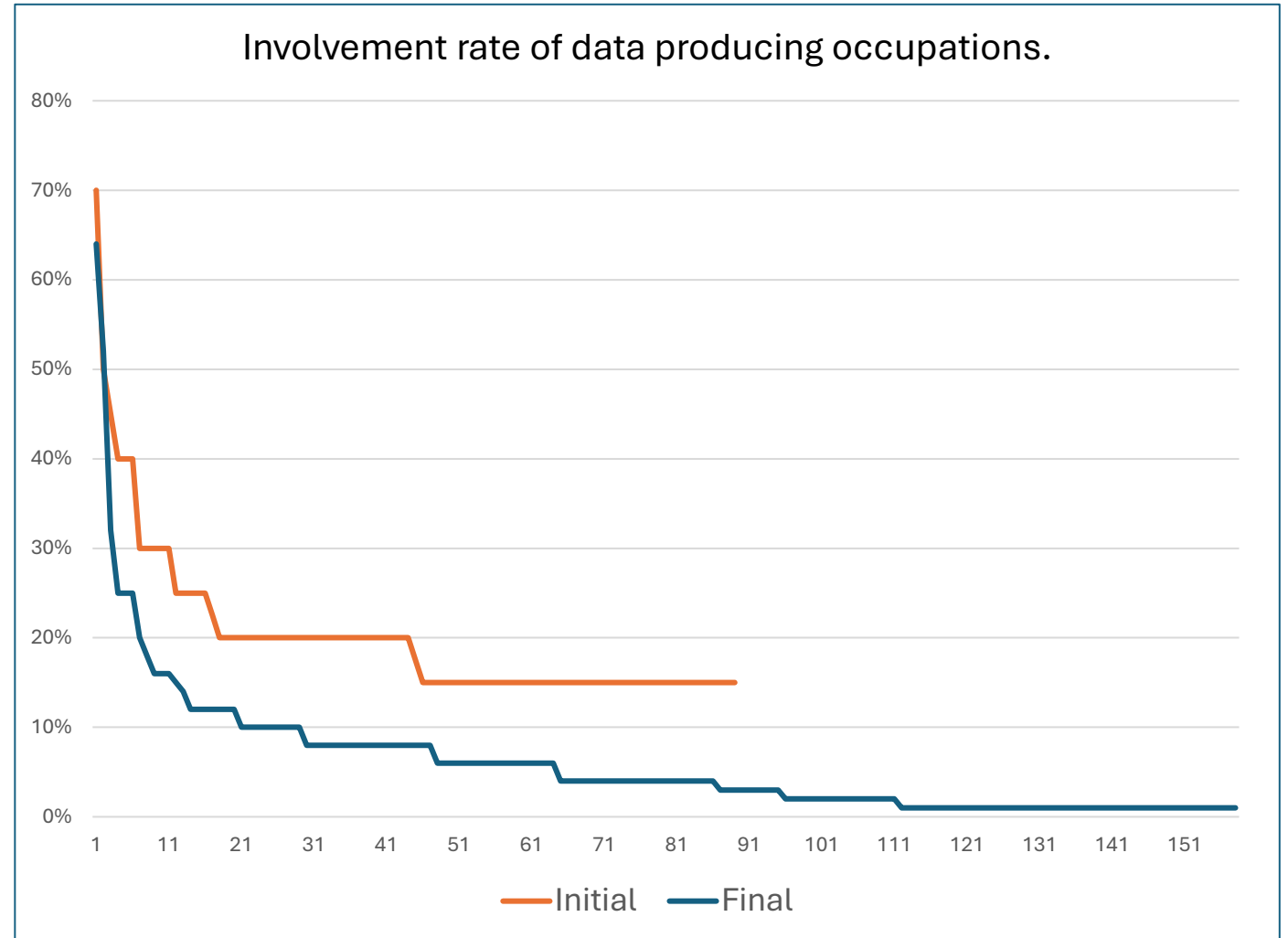
- Within the SNA, output of, and investment (GFCF) in **data should be reported together with databases as a single IP product**, however, this should be **reported separately from computer software**.
- All expenditure on production of data on an own account basis is regarded as a capitalised expense 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**.

Measurement concepts have remained as proposed...and in line with 2025 SNA text 2/2

- Data assets that are purchased *with exclusive* rights are treated as **a purchase of an asset** (with an offsetting sale of an asset by the seller). 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).
- 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**, *however* **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**.

Occupations and Involvement rates

- Significant refinement on default occupation list based on empirical evidence (USA, OECD, GER & JAP) rather than based simply on subjective expert opinion.
- Results in a longer tail (more occupations) but with lower involvement rates.
- Testing has resulted in similar estimates to those obtained independently (i.e. entirely through ML).



Clarity on double counting

Conceptual delineation

“expenditure relating to the creation of a data asset should be categorized as production of data, regardless of the subsequent use of the data asset **with the exception of data integral to the production of Research and development and data assembled in a database created solely as a step in the production of an AI computer program** and that cannot be re-used”.

- Purchases of R & D and Software should reflect the use of data assets that were used in production of these assets.
- 90% of recommended occupations for data have involvement rates of less than 10% so **double counting based on using the same labor cost is considered unlikely.**
- Countries are encouraged **to reviewed own account production of computer software and R & D to ensure that the same expenditure is not recorded twice** and classified to both assets.

Recommendations on deflation and backcasting

Deflation

- 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 technological and quality improvements**.
- It is recommended that volume estimates of data are compiled **using a ‘pseudo’ output price index**. Either **output price index of a similar product** or Input Price index adjustment to reflect quality and productivity improvements.

Backcasting: time series length

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

Recommendations on capital stock

Capital stock, feedback reflected a preference to maintain a connection with existing IPPs rather than change without evidence.

- It is recommended that countries **apply similar parameters** (Age-efficiency decline, Age-price assumption, retirement profile) **as applied currently to other IPPs** in the compilation of depreciation and net capital stock of data
- 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 to allow for different service lives to be applied based on the industry producing the data.

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 [written feedback](#) (including any specific suggestions) on the full draft handbook [until 15 November 2025](#).**