Developing practical compilation guidance on the production of Data in the National Accounts

John Mitchell - consultant to Task Team
on behalf of the Joint Eurostat-IMF Task Team on
"Measuring Data as an Asset in National Accounts"

Meeting of the Group of Experts on National Accounts
Palais des Nations, Geneva, 23-25 April 2024

Recording of Data in the NA: Fundamental recommendations

- Data is an output of production, when capitalized it is considered a **produced asset** → impact on asset boundary
- Data is defined as "*information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities*"
- Data is **distinct from 'observable phenomena'** (OP), which are *'a fact or situation whose characteristics or attributes may be recorded'*. OP are inputs for data
- Most data are produced on an **own-account** basis and it is recommended to value at **sum of costs**
- Only data used in production and providing economic benefit are included in SNA asset boundary. (Standard asset requirement in SNA).

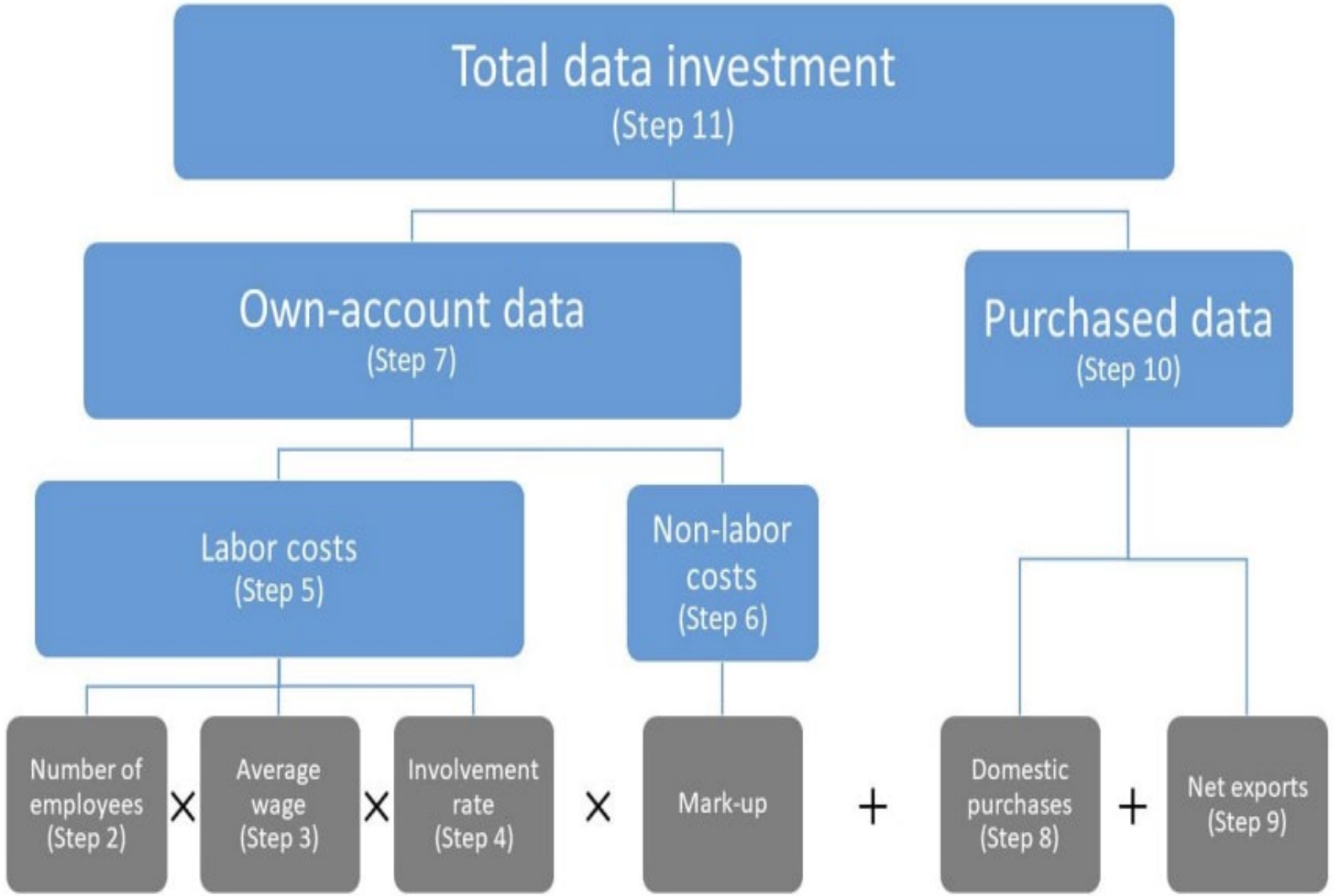
Recording of Data in the NA: Fundamental recommendations (cont.)

- **Expenditure** undertaken to access and record OPs that are **added to an established data asset** is considered **new gross fixed capital formation**
- For simplicity, **all own-account** production of data is considered capital formation.
- Data would be classified to a newly created asset category called "**data and databases**" which would include the current output associated with the production of databases and exclude the current category of computer software
- Like other assets in the National Accounts, **data is subject to economic ownership, valuation**(and re-valuation) and **depreciation** (modelled using PIM).

Simple Methodology Map

Source data or modelling required for grey cells.

Nominal estimate only



Visual courtesy of Destatis

Model for sum-of-costs

$$C_{i,t} = \alpha \sum \tau_{\omega} W_{\omega,i,t} H_{\omega,i,t}$$

Source: BEA, but adopted by several countries (e.g. CAN, JAP, NL, PAK)

Issue: identify suitable parameters

C: cost of investment

$W_{\omega,i,t}$: average annual wage from occupation statistics

$H_{\omega,i,t}$: annual employment from occupation statistics

α: markup for social contributions, capital cost, and intermediate consumption from BEA supply-use tables

Time use: $\tau_{\omega} = \rho_{\omega} s^{*}_{\omega}$ (Blackburn 2021)

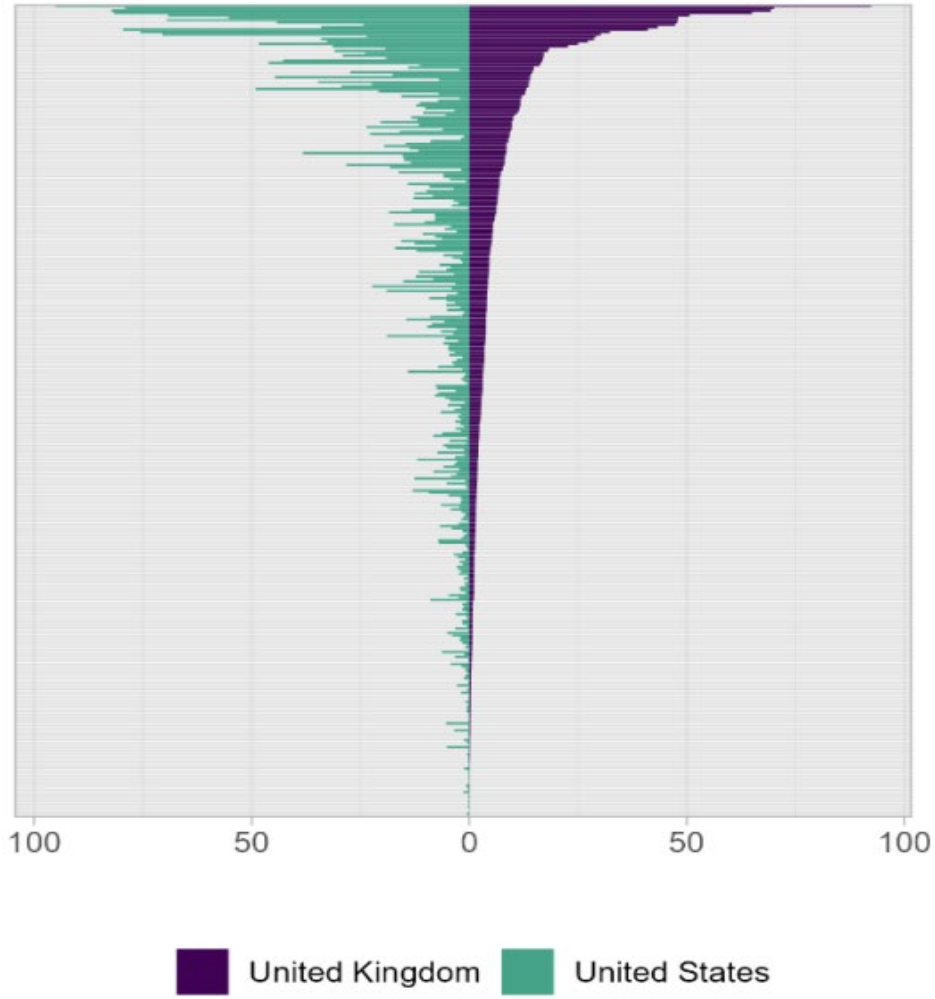
- ρ_{ω}: fraction of jobs engaged in at least 1 data-related activity
- s^{*}_{ω}: share of time allocated to data-related activities

Clarifications being worked through by Task Team

I. Which occupations for labour costs (producing rather than using data / avoid double counting)
II. Involvement rates of chosen occupations
III. Estimating total non labour costs (non labour intermediate costs + return on capital)
IV. Own-account data (how much is GFCF vs Intermediate consumption)
VI. Construction of a suitable price index? (labour/wage index vs non-labour costs)
VII. Asset lives (different lives for different types of data?)
VIII. Back series (also needed for PIM)

Occupations / Involvement rates

- *Task team have surveyed members to derive a more comprehensive list of data producing occupations and their involvement rates.*
- *These lists will be consolidated to identify those occupations consistently identified as well as those considered more country specific.*
- *This result will be contrasted with more systematic approach undertaken by those using online job ads and Machine Learning.*
  - *(OECD, 2023) What is the role of data in jobs in the United Kingdom, Canada, and the United States? A natural language processing approach.*
  - *(United States BEA, 2022) Valuing the U.S. Data Economy Using Machine Learning and Online Job Postings.*



Own-Account Data (GFCF vs Intermediate consumption)

- The Task Team is working towards a consensus recommendation regarding how much output to capitalize.
- Members views on availability of data are being considered along with existing guidelines - *OECD Handbook on Deriving Capital Measures of Intellectual Property Products.*

*'As a general rule, **all expenditures** on intellectual property products, either purchased or produced on own account, **should be recorded as gross fixed capital formation** if they are expected to provide economic benefits for the owner".*

*"This is because the development of IPP products…are **inherently high risk,** and those that undertake them expect that the benefits obtained from the few successes will more than compensate for the cost of the many failures" (not identical circumstances but relevant for data).*

- *Currently the handbook recommends that **all** own account data output is **considered GFCF.***

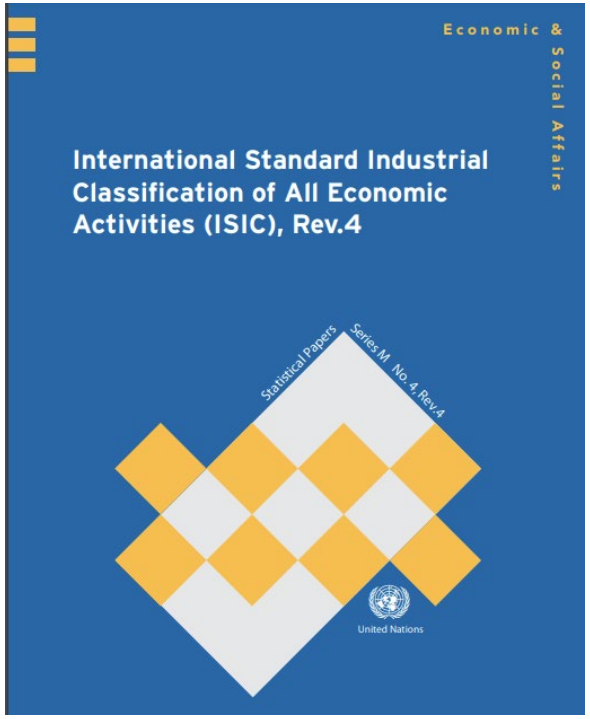
Approach: build on existing work

Country	Year	Value of data asset, % of total GDP	PPT difference in total GDP growth for year	PPT difference in total GCF growth for the year
Australia	2016	2.9%	0.016%	0.57%
Canada	2018	1.9%	-0.037%	-0.09%
Netherlands	2017	3.0%	-0.012%	-0.12%
India	2019	1.0%	0.000%	0.14%
USA	2020	0.8%	0.047%	0.26%

Plus Japan, Germany, Pakistan, OECD, academic work

Data in other classification

- *CPC*
  - New group “Data and data compilation” consisting of two classes – “data” and “compilation services of data” — created in the revised CPC 2.1 (CPC 3.0).
  - Clear category for classification of data production.
  - Will flow into complimentary classifications when revised – including trade (i.e., SITC)
- ISIC
  - Updated explanatory notes take data production specifically into consideration within revised ISIC/NACE categories.



Handbook considerations

- The handbook will be based on **countries' experience** and will include case studies and recommendations sourced from task team members and their organisations.
- A default methodology will be provided to **ensure consistent compilation.** can be undertaken by all countries.
- **Aspirational recommendations** will also be provided allowing for some flexibility to those countries with higher levels of source data and information
- Consideration will be given to **overlaps with other intangible assets**
- Handbook to serve multiple purposes;
  - Recommendations at a technical level for compilers.
  - Conceptual background and explanation of methodology for users.

Envisaged structure of the Handbook

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

Section 2 - Compiling a **nominal estimate** of output and GFCF though the sum of cost method

Section 3 - Compiling **volume estimates**

Section 4 - Compiling **capital stock estimates**

Section 5 - **Overarching measurement and conceptual questions** discussed and explained

Section 6 - Conclusion and condensed list of recommendations

What can countries do now?

- Compilation will benefit from more research and empirical evidence
- Specifically, additional information is required on which occupations are producing data, involvement rates, asset lives etc.
- **Countries can <u>already</u> begin to create estimates based on existing work and guidance notes.**

Contact the task team for assistance or to share experiences!!!

Thank you
for your attention

European Commission