Recording of Data Assets in the National Accounts: 2025 SNA Recommendations

5TH INTERNATIONAL SEMINAR ON BIG DATA FOR OFFICIAL STATISTICS: MEASURING THE DIGITAL ECONOMY
29-31 MAY 2024

Erich Strassner, IMF & Jorrit Zwijnenburg, OECD

Presentation reflects work of the Digitalization Task Team with special thanks to John Mitchell (Consultant)

Agenda

- Motivation: Importance of data
- How is data created?
- Overview of new recording
- How to value data?
- Outcomes of Global Consultation
- Final Recommendations
- Next Steps

Introduction

Introduction

- Since the 2008 SNA was published, production of data is now fundamental to many business models, while most traditional businesses now produce some form of data to either increase returns or lower costs.
- A large amount of this data is used repeatedly in production for more than a year, but no fixed asset category currently exists outside of possible inclusion within databases.
- Because of the current classification and definition within the SNA, there is growing concern that the (production of) data is not being appropriately represented in the accounts, impacting the estimates of GDP, balance sheets and productivity.

Data as an asset within the SNA

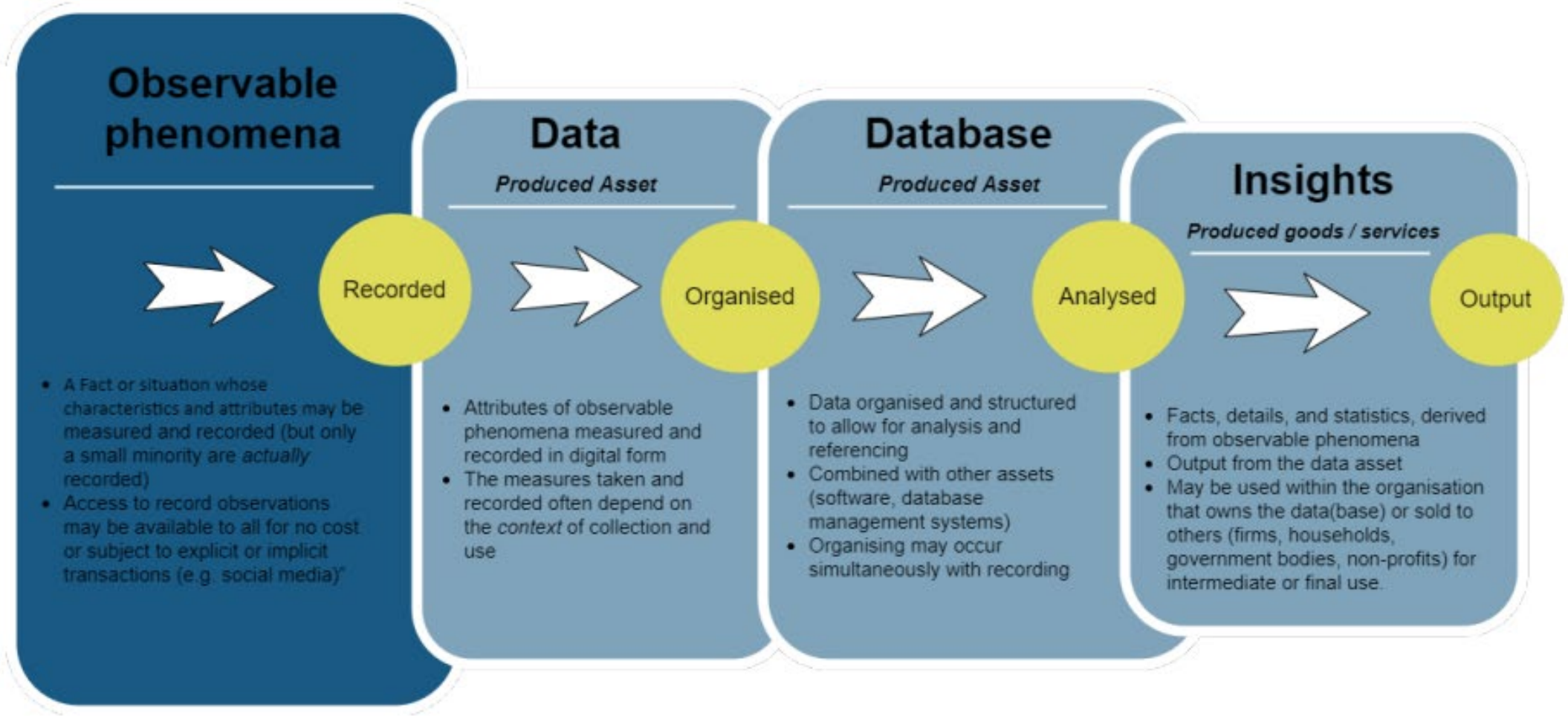
- Data contributes to improved productivity and efficiency in a range of outputs across different industries
  - Just-in-time supply chain control
  - Customer loyalty programs
  - Improved demand forecasts
  - …
- Data has become an important input to do more targeted advertising
- Not just the private sector; public sector data has transformed many aspects of non-market output.
- For all these reasons, it was decided to include data as an asset in the updated SNA

Data as an asset within the SNA

- Due to this we must,
  - Explicitly define what data is, how it is created, and assess how it might be recorded in the national accounts.
  - Refine the current definition of databases in order align appropriately and consistently with any new data classification.
- The current definition of databases in the 2008 SNA (10.112):
  - “Databases consist of files of data organized in such a way as to permit resource-effective access and use of the data”.
  - It also includes the following guidance, “the cost of preparing data in the appropriate format is included in the cost of the database but not the cost of acquiring or producing the data”

How is data created?

Data Information Chain from an SNA Perspective



Source: Mitchell, Ker & Lesher (2021)

What are observable phenomena?

- Data is distinct from the information elements of ‘observable phenomena’ (OPs) which are inputs for data
- Examples of OP are ‘the weather,’ ‘online activities of persons,’ ‘traffic,’ etc.
- Data consists of recorded and organized information elements of the OPs
  - Examples of information elements are ‘the temperature,’ ‘search requests for hotels in Paris,’ ‘number of cars on highways on Monday between 9-10am’
- Once the information elements of the OP is recorded and stored in a digital format, it becomes data
- OPs are regarded as non-produced (they exist, they are just ‘out there’) and in general have no value, expect if purchased.
- The explicit purchases/payments for OPs are considered as either a purchase of a service (output) or payment of rent for providing access to the specific OPs (which implies giving up privacy).
- For pragmatic reasons, the information elements of OPs are excluded from the asset boundary.

What is data?

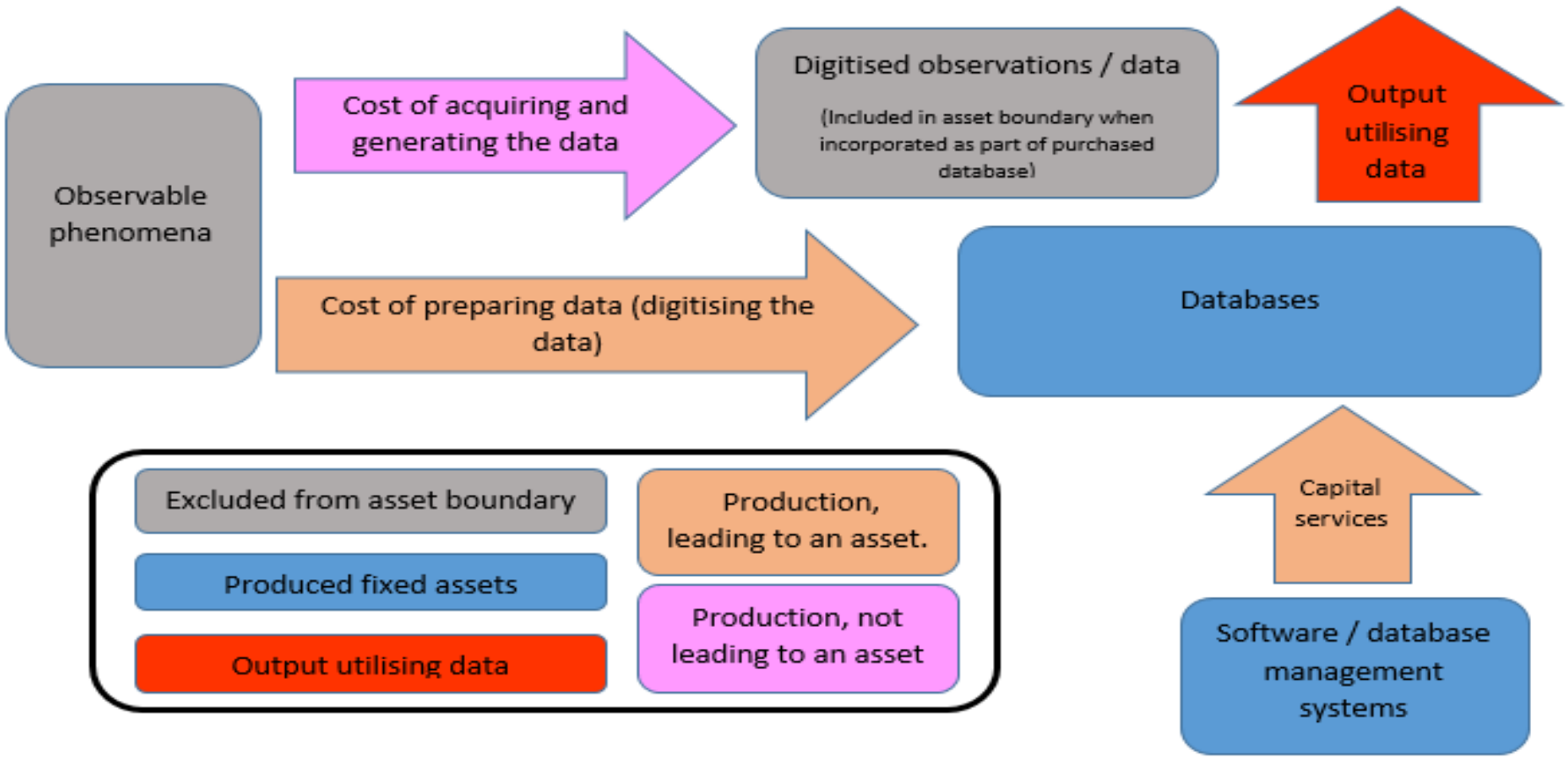
- Data is defined as "*information content that is produced by accessing and observing phenomena; and recording, organizing and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities*"
- Data can still be many different things to many different people … but this is how we define data for statistical purposes in the SNA
- For pragmatic reasons data in the national accounts focusses only on digitized data

Is data fully produced?

- Similarly produced data sets may have different value in view of the difference in content
- Should compilers consider that some of the assets' value is resulting from the intrinsic value of the information embedded in the dataset?
- Two options:
  1. Yes, recommend creation of new hybrid asset category, consisting of produced and non-produced part
  2. No, recommend that data is entirely the result of production
- The recommendation is to consider data assets entirely the result of production
  - The higher value may also stem from other reasons
  - There are no (other) assets in the SNA that are considered produced and non-produced
  - Option 1 will be very difficult to apply in practice and the results difficult to explain to users

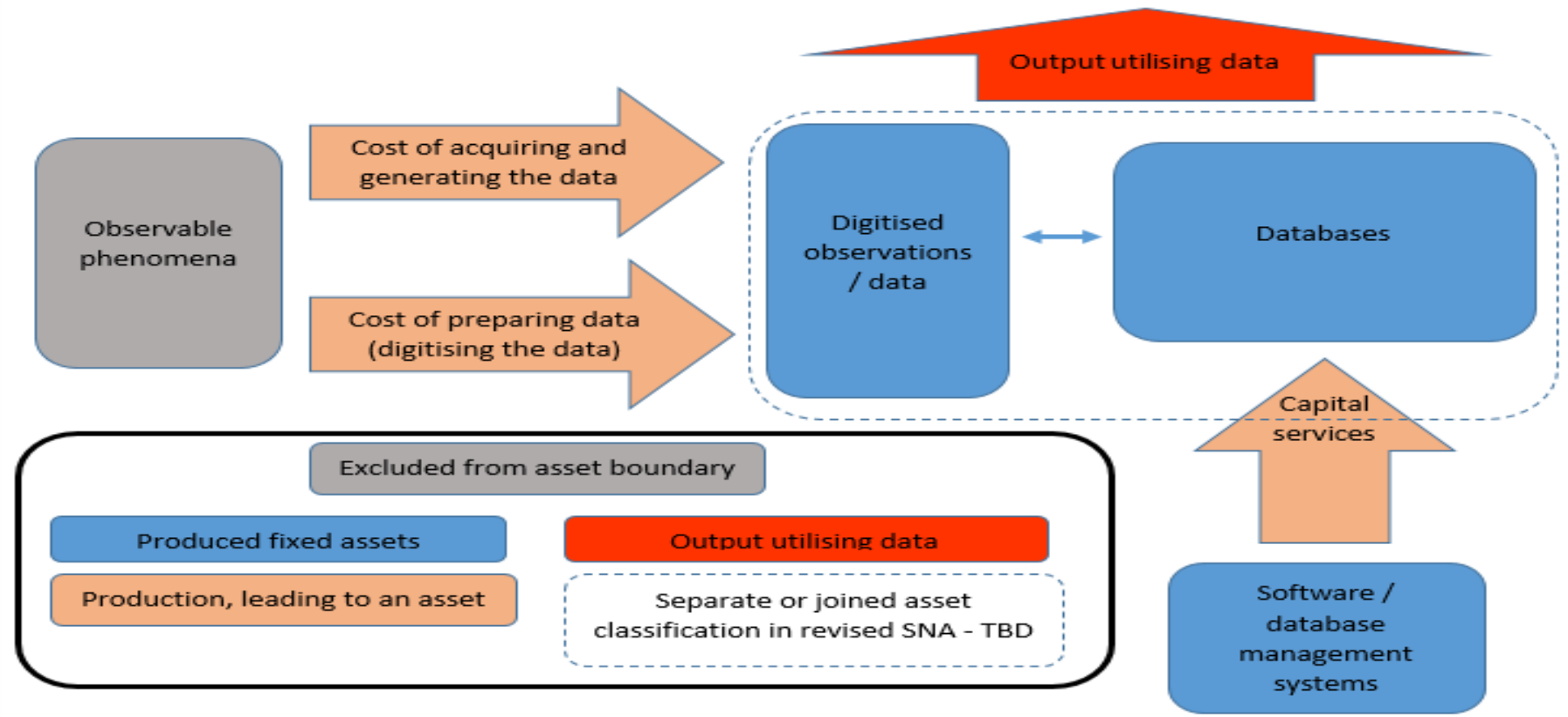
Overview of new recording

Current recording



- Some expenditures explicitly excluded, some included as GFCF
- Data themselves are not capitalized

New recording



- Data and databases are considered produced
- All expenditures associated with the data asset contributes to GDP

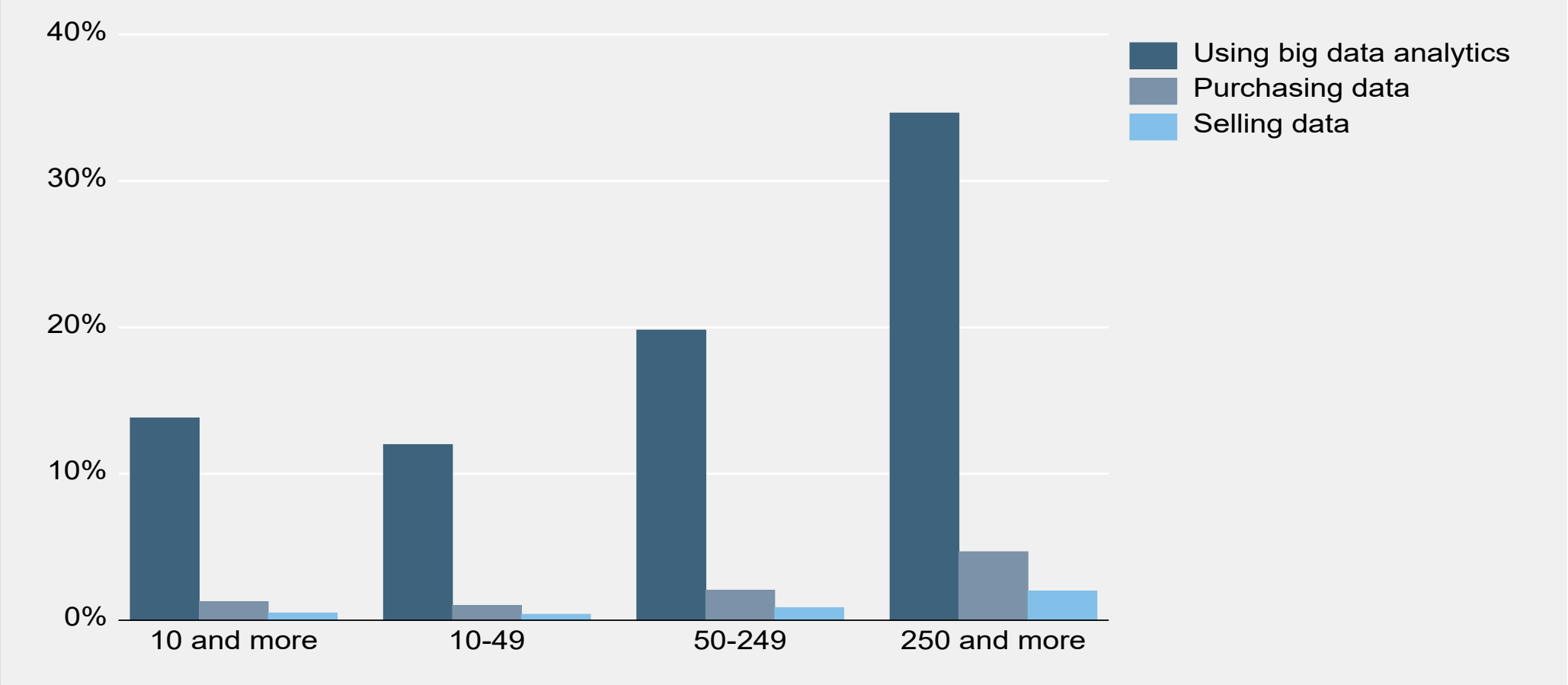
How to value data?

Valuation: Data versus other assets

- There are not many sales/purchases of data
- Most data is produced on an own account basis
- Data doesn’t depreciate in a conventional sense
- Data is extraordinarily heterogeneous
- Quantity is often unrelated to the value

Purchasing or selling big data are rare, even among firms that conduct big data analytics

Average across 23 European countries by firm size (total number of employees), 2020



Valuation of data assets

Three methods to value data:

- Market price:
  - However, most data is not sold at market in large enough quantities
- Net present value:
  - However, issue that not enough of the required information is available
- Sum of cost:
  - Successfully tried already (see Australia, Canada, Netherlands, USA, Pakistan, Germany, academia)
  - A lot of similarities with other intangible assets in the SNA
- Due to the clear practical advantages, the sum-of-cost approach is recommended for own account produced data

What costs to include in the sum-of-cost?

- The value of databases includes "*the costs of preparing data in the appropriate format […] but not the cost of acquiring or producing the data*" (2008 SNA, 10.113)
- With the inclusion of data within the asset boundary, the last part has now become relevant
- However, more specific guidance is needed, reflecting the complexity of producing data
- Furthermore, the costs of acquiring and producing data might incorporate a large range of input costs that may also be involved in the creation of other assets
- For that reason, it is proposed to consider the following costs:
  - Costs of planning, preparing, and developing a data production strategy
  - Costs associated with accessing, recording and storing information embedded in OPs
  - Costs associated with processing, cleaning, and organizing the data to allow for use in productive activities
  - In addition, an estimate for consumption of fixed capital and a mark-up for net operating surplus for market producers should be included

How to record ongoing expenditures?

- Most assets receive maintenance or undergo major improvements
- Maintenance concerns a current expense that is not capitalized, whereas major improvements concern gross fixed capital formation
- While each addition of an individual information element may not seem to constitute a major improvement, when considered in their entirety they certainly do have a material impact on the fixed asset
- The expenditure improves the relevance of the asset which in turn "improves the assets performance, increase[s] their capacity or prolong their expected working lives" (SNA 1.156)
- For that reason, it is recommended that expenditure undertaken to update a data asset with new OPs should be considered as investment (GFCG) rather than repair and maintenance

Data produced by nonmarket producers

- Often data might be made available to users for no cost
- However, it is considered that the data producers are still receiving an economic benefit, even if it's not a direct monetary benefit
- Therefore, the creation of the asset would still be recorded as production and recorded on the balance sheet of the relevant entity
- However, since the data is provided for free, no transaction takes place vis-à-vis any users (so it won't show up as an asset on the balance sheet of the user)

Data consumed within a year

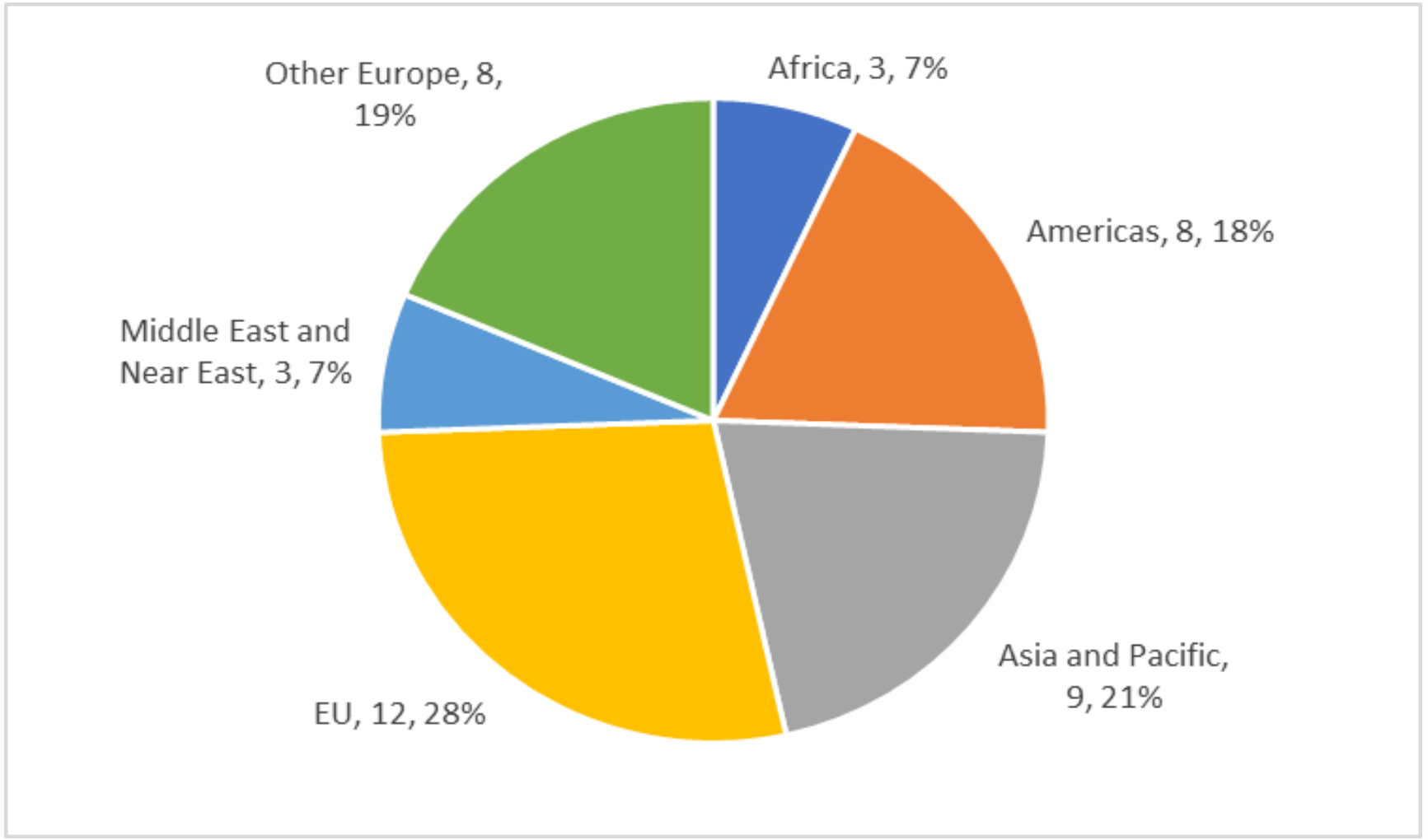
- Data may often be used in production but for a period less than one year
- In that case, it should be recorded as intermediate consumption
- This holds for ancillary data, i.e., data generated as an externality of production supporting the principal or secondary activities (e.g., HR data), but possibly also for other types of data
- However, data is not liable to physical wear and tear, so may be kept for more than a year
- It will be difficult to delineate between data with a life length of more and less than a year
- Two options were considered:
  1. A set percentage of total expenditure could be capitalized and the remaining part treated as intermediate consumption
  2. All expenditure could be capitalized but with applying a very short asset life and highly skewed retirement distribution
- While ideal to chose option 1, for practical reasons, option 2 is recommended

Global consultation

Global consultation

- Global Consultation held as part of SNA research agenda
- 43 Countries / Institutions provided responses
- There was unanimous agreement that conceptually, data was the result of production and should be capitalised in the SNA

Consultation received broad representation and support



Outcomes of global consultation

- Although *practical* challenges were noted, there was *near* unanimous conceptual support for the specific proposals covering:
  - The sum of costs valuation as the preferred method for own-account production of data
  - The creation of a single data and databases asset category, which is separate from software
  - The recording of sold data (non-exclusive basis - production and sale of a copy of an original, exclusive basis - sale of a produced asset)
  - The recommendation that guidance covering service life was not a mandatory feature of this guidance note
  - The exclusion of ancillary data from the capitalized data

Final Recommendations

Final recommendations (1)

- Data is the result of production, i.e., when capitalized in the national accounts, it is classified as a produced asset
- Data is distinct from the information elements of 'observable phenomena' (OP), which are inputs for data
- Expenditure undertaken to access and record OPs that are added to an established data asset is considered new gross fixed capital formation, as it prolongs the life of the data asset
- Only data that provides an economic benefit when used in the productive activities of its owner for more than one year is included in the SNA asset boundary
- Data is considered part of the intellectual property product suite of assets
- Data will be classified to a newly created asset category called "data and databases"

Final recommendations (2)

- Data sold in a market transaction with no exclusive rights to the data is considered a sale of a copy of an original
- If exclusive rights are granted, it is recorded as a sale of a produced asset (in case it meets the one year criterion)
- All own account production of data is considered capital formation
- Own account production of data should be valued at the sum of costs, which would include:
  - costs of planning, preparing and developing a data production strategy,
  - costs associated with accessing, recording and storing information embedded in OPs,
  - costs associated with designing, organizing, testing and analyzing the data in order to draw information and conclusions from it
- Costs include an estimate for consumption of fixed capital and a mark-up for net operating surplus for market producers

Next steps

Development of Practical Measurement Guidance

- The ISWGNA agreed with the development of a compilation handbook to address practical implementation challenges and to produce internationally comparable estimates of data
- A joint Eurostat-IMF Task Team on *Data as an Asset in the National Accounts* has been launched to look into the drafting of the handbook. Several clarifications and decisions underway:
  - Which occupations for labour costs (producing rather than using data / avoid double counting)
  - Involvement rates of chosen occupations
  - Estimating total non labour costs (non labor intermediate costs + return on capital)
  - Own-account data (how much is GFCF vs Intermediate consumption)
  - Construction of a suitable price index? (labor/wage index vs non-labor costs)
  - Asset lives (different lives for different types of data?)
  - Backcasted series (also needed for perpetual inventory method)

Structure of Compilation Guide

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**

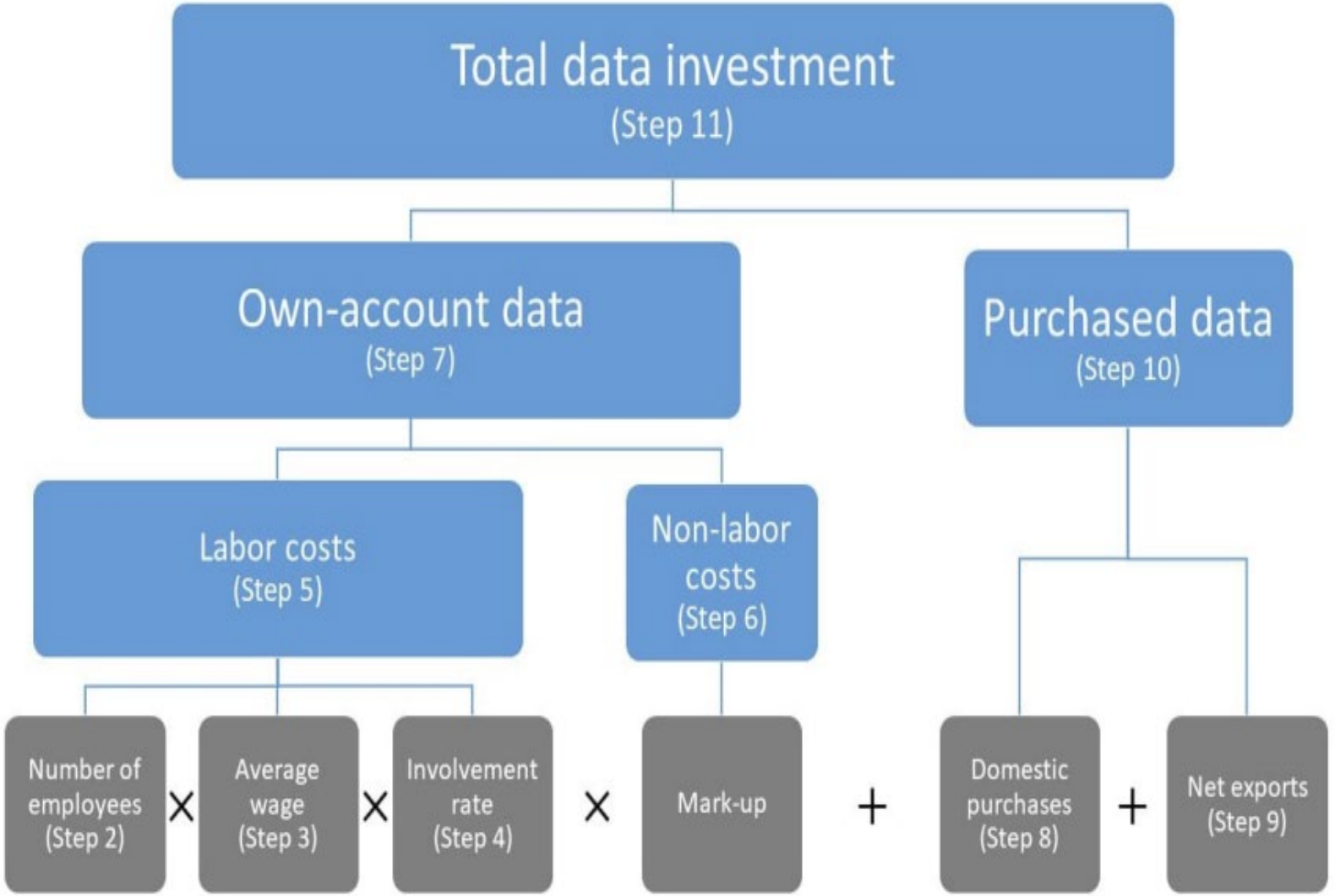
Section 5 – **Overarching measurement and conceptual questions** explained

Section 6 – Conclusion and condensed list of recommendations

Simple Methodology Map

Source data or modelling required for grey cells.

Nominal estimate only



Visual courtesy of Destatis

Model for sum-of-costs

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

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

Primary issue: Identify suitable parameters to calibrate sum of costs

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

Measurement Points: 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 the more systematic approach undertaken by several countries 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.*

Measurement Points: Own Account Data
How Much is Gross Fixed Capital Formation of Intermediate Consumption

- Task Team is working towards a consensus recommendation regarding how much data 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". [also relevant for data].

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

Measurement Points: Build on Existing Efforts!

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

Measurement Points: Data and Updates to Classification Systems

- Central Product Classification (CPC)
  - Clear category for classification of data production
  - Will flow into complimentary classifications when revised – including trade and international standard industrial classification (ISIC)
- ISIC
  - Updated explanatory notes take data production specifically into consideration within revised ISIC/NACE categories.

Additional Compilation Guide Considerations

- Guide will be based on **country experiences** and will include case studies and recommendations sourced from task team members and their organizations
- In addition to ‘best practice’ recommendations; ‘Second best’ methods will be provided to ensure consistent compilation can be undertaken by all countries
- Consideration will be given to **overlaps with other intangible assets**
- Guide to serve multiple purposes:
  - Recommendations at a technical level for compilers
  - Conceptual background and explanation of methodology for users

Thank you

The Digitalization Task Team includes Andrew Baer (IMF), Allison Derrick (BEA), Andreas Dollt (Eurostat), Kevin Fox (UNSW), Ziad Ghanem (Stats Canada), Richard Heys (ONS), Stanimira Kosekova (ECB), Nicola Massarelli (Eurostat), John Mitchell (OECD), Marshall Reinsdorf (consultant), Jennifer Ribarsky (IMF), Sebastián Rébora (Central Bank of Chile), Carol Robbins (NSF), Benson Sim (UN), Michael Smedes (ABS), Erich Strassner (Chair, IMF), Cliodhna Taylor (ONS), Jim Tebrake (IMF), John Verrinder (Eurostat) and Jorrit Zwijnenburg (OECD)

Former members : Sri Soelistyowati (Statistics Indonesia), Teck-Wong Soon (Statistics Department Singapore) and Dylan Rassier (BEA)

Any omissions are accidental