



8th International
Conference on
BIG DATA
& Data Science for Official Statistics

BILBAO 2024

Informing Climate Change and
Sustainable Development Policies
with Integrated Data

BILBAO. SPAIN

10-14 JUNE 2024

#UNBigData2024

Roadmap for an MPD project

Esperanza Magpantay

Senior Statistician, ITU

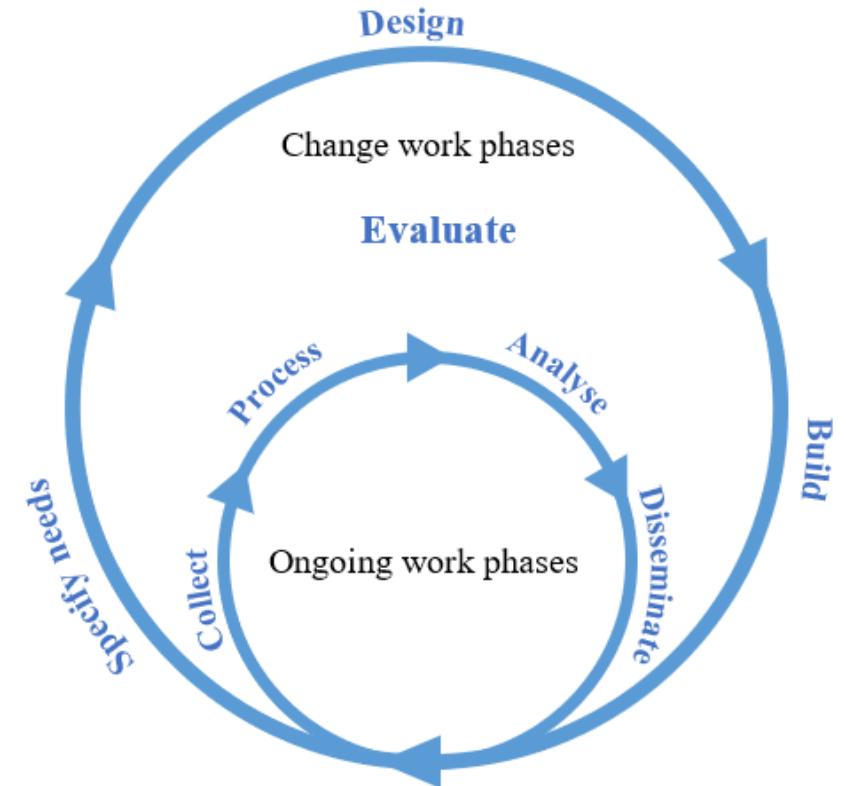
UN-CEBD Mobile Phone Task Team Lead

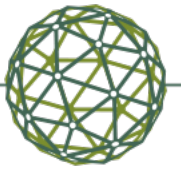




GSBPM Process

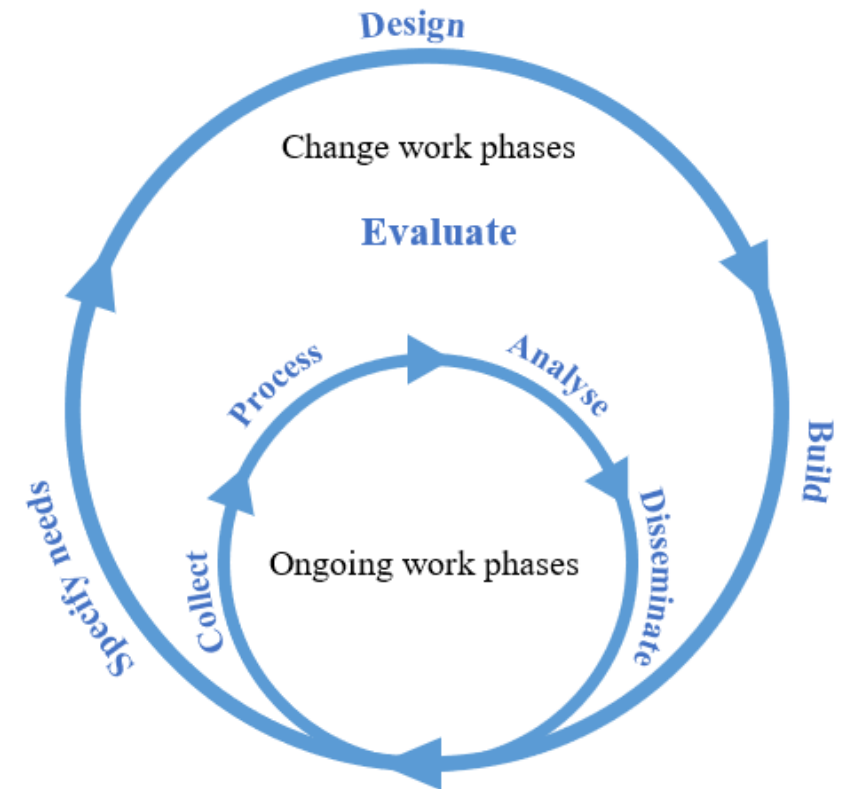
- Generic Statistical Business Process Model (GSBPM) - describes and defines the set of business processes needed to produce official statistics.
- It provides a standard framework and harmonised terminology to help statistical organisations to modernise statistical production processes, as well as to share methods and components.

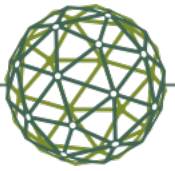




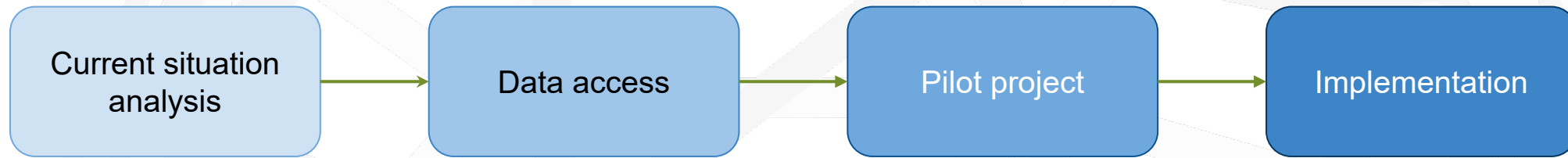
GSBPM can also be used for:

- integrating data and metadata standards
- a template for process documentation and for harmonizing statistical computing infrastructures
- to provide a framework for process quality assessment and improvement





General process flow

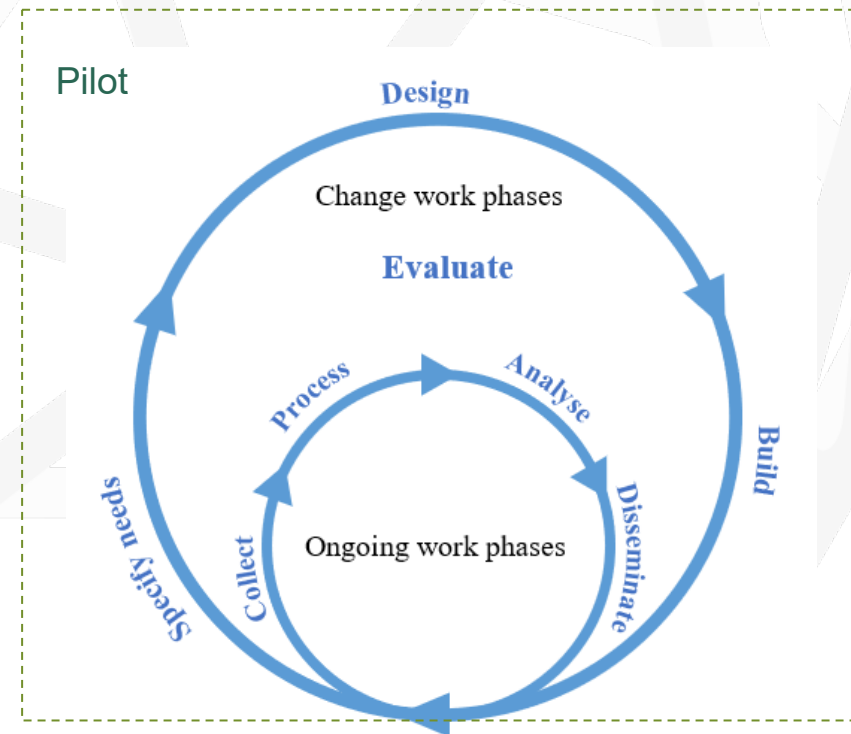


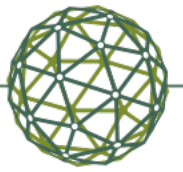
Pre-project activities

- Specifying needs
- Tendering
- Choosing the right partner

Preparations for the Pilot Project

- Legal situation assessment
- Technical situation assessment
- Business model
- Hosting environment agreement and set-up
- Discussion with MNOs
- Preliminary data validation (QA)

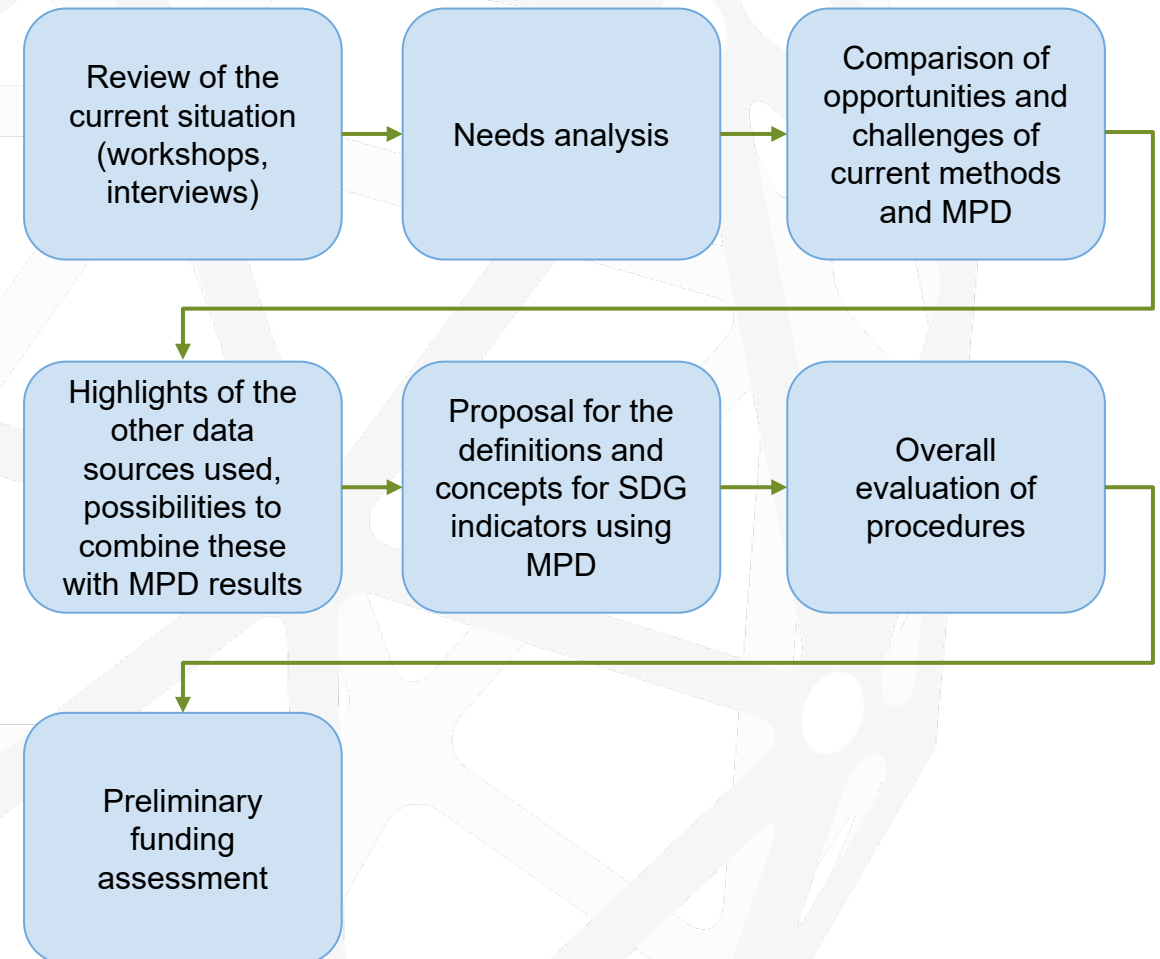


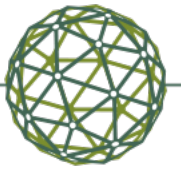


Current situation analysis

The main tasks in the current situation analysis phase:

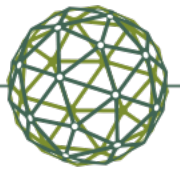
1. Comparison of opportunities and challenges of current methods and MPD
2. Highlights of the other data sources used (outside the surveys), possibilities to combine these with MPD results
3. Proposal for the definitions and concepts for indicators using MPD
4. Overall evaluation of procedures to be implemented for generating statistics from MPD based on the experience in other countries





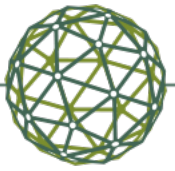
Data access

- The access to MPD is one of the key challenges for government organisations.
- Data is governed by privacy, telecommunication legislation, and MNO internal rules
- Time consuming process. Hence, it is better to start as early as possible.
- Some elements of the data access process are also carried out during the project, together with the situation analysis.



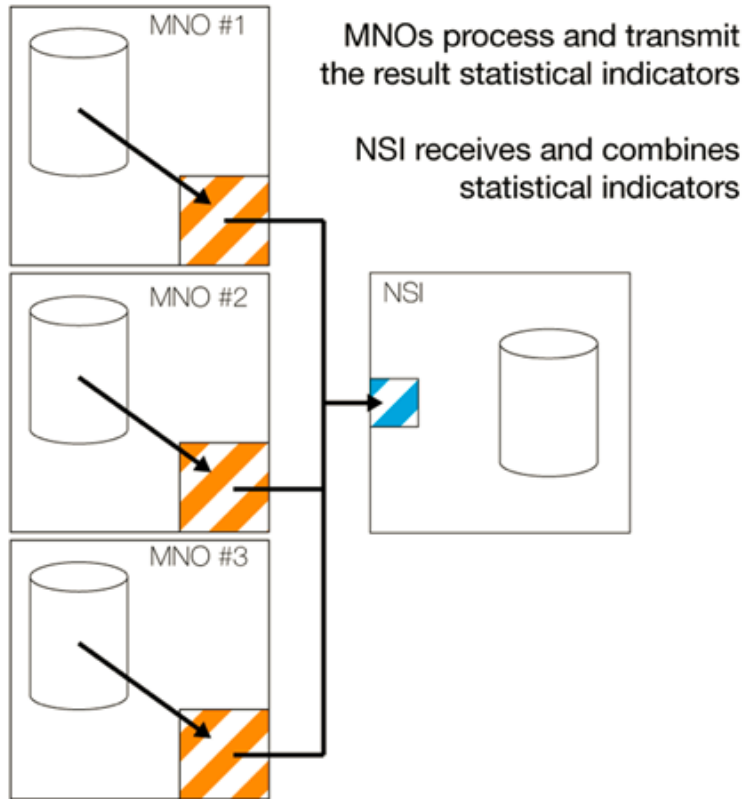
Main tasks in data access

Tasks	Description/Questions fo think through
1. Legal situation assessment	How is it possible to get MPD necessary for generating agreed statistical indicators in agreed domains, including all possible legal activities?
2. Business model assessment	Where will the data be processed? (Centralised/Distributed), Who are the stakeholders?
3. Technical situation assessment	How will the raw MPD extraction from data warehouses/billing systems work? Where will the data be hosted? Which hardware and software are necessary for storing and processing the data (technical specifications)? Who is responsible for providing the hardware and upkeep according to the technical specifications? Who is responsible for software installations, maintenance and upkeep? How is the data access provided (VPN/SSH etc.)?

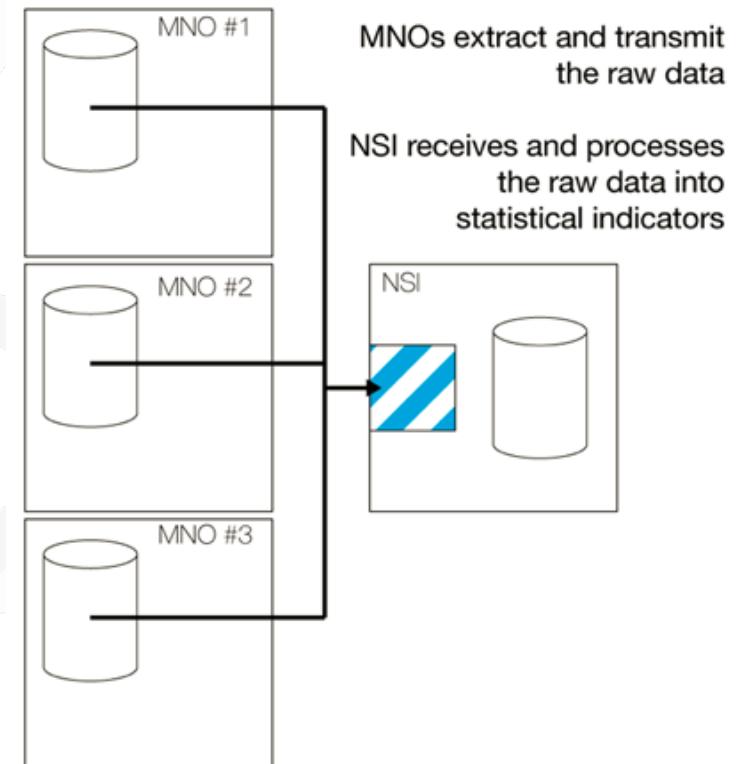


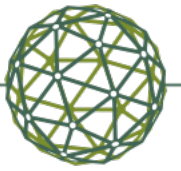
Where is the data processed?

Distributed (Type A)



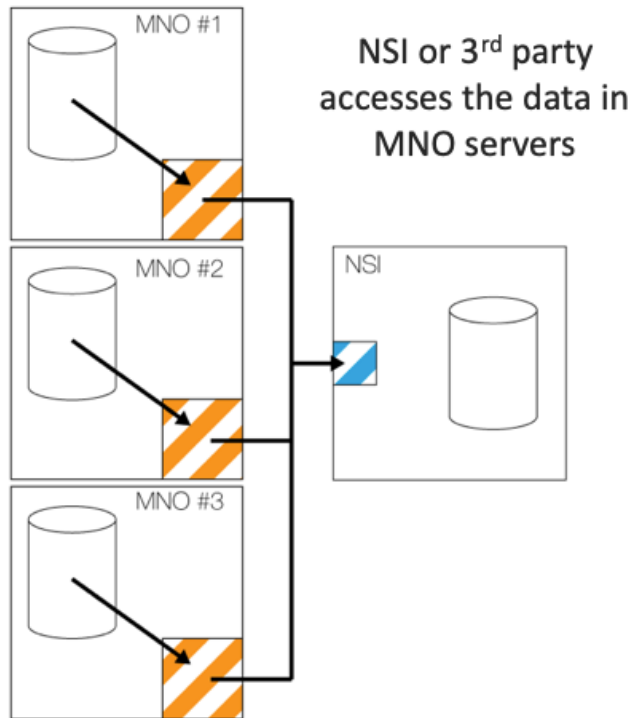
Central (Type B)





Where is the data processed?

Public-Private Partnership (Type C)

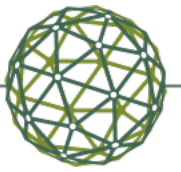


Advantages

- Higher data security / Lower risk of (accidental) disclosure
- Higher confidence for MNO
- Higher granularity of data
- Higher control/confidence for statistical authority (trust, transparency, accuracy)

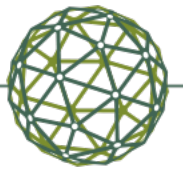
Disadvantages

- More complex environment
- Higher cost of the project



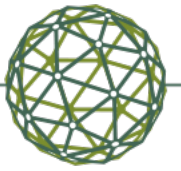
Main tasks in data access

Tasks	Description/Questions fo think through
4. Hosting environment agreements and set-up	Based on the answers received with the previous steps, agreements on the hosting environment and set-up are done.
5. Discussions with MNOs	The main goal for discussions with MNOs in this stage is to get access to a sample of the data that has been extracted according to the data requirements checklist.
6. Preliminary data validation (QA)	During this step, the accordance and logic validation are done. This means, it will be checked if the format of the data set is OK and all minimum fields required are covered. Also, it will be checked if the cell locations provided are inside the country of reference etc.

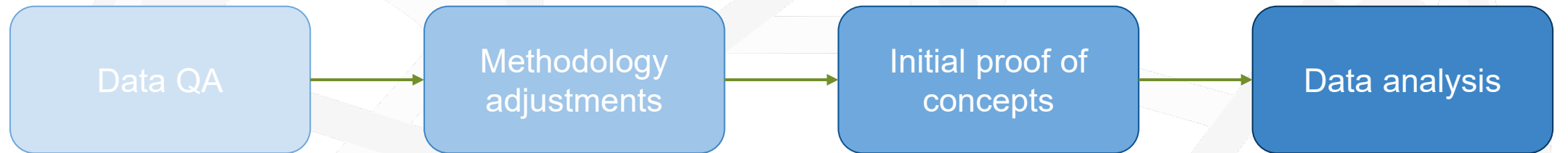


Various scenarios/examples of data access

Countries	Key organisation	Data received from MNOs	Number of MNOs	Base for data collection	MPD is official statistics	Domains	Processing algorithms
Indonesia	Ministry of Tourism	Results	1	Commercial agreement	Yes	Tourism	Third party developer
Estonia	Central Bank	Raw, pseudonymised	1	Statistics act	Yes	Tourism / travel (BoP)	Third party developer
Oman	NSO	Raw, pseudonymised	2 (all)	Statistics act	No	Population, mobility, tourism	Third party developer
Netherlands	NSO	Results	1	Commercial agreement	No	Population	MNO
Finland	Ministry of Transportation	Results	1	Commercial agreement	No	Transportation	MNO
Saudi Arabia	Ministry of Tourism	Results	1	Commercial agreement	No	Tourism	MNO



Pilot Project

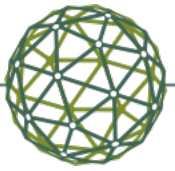


- To ensure compatibility of input data with project outcomes.
- Typically conducted several times.

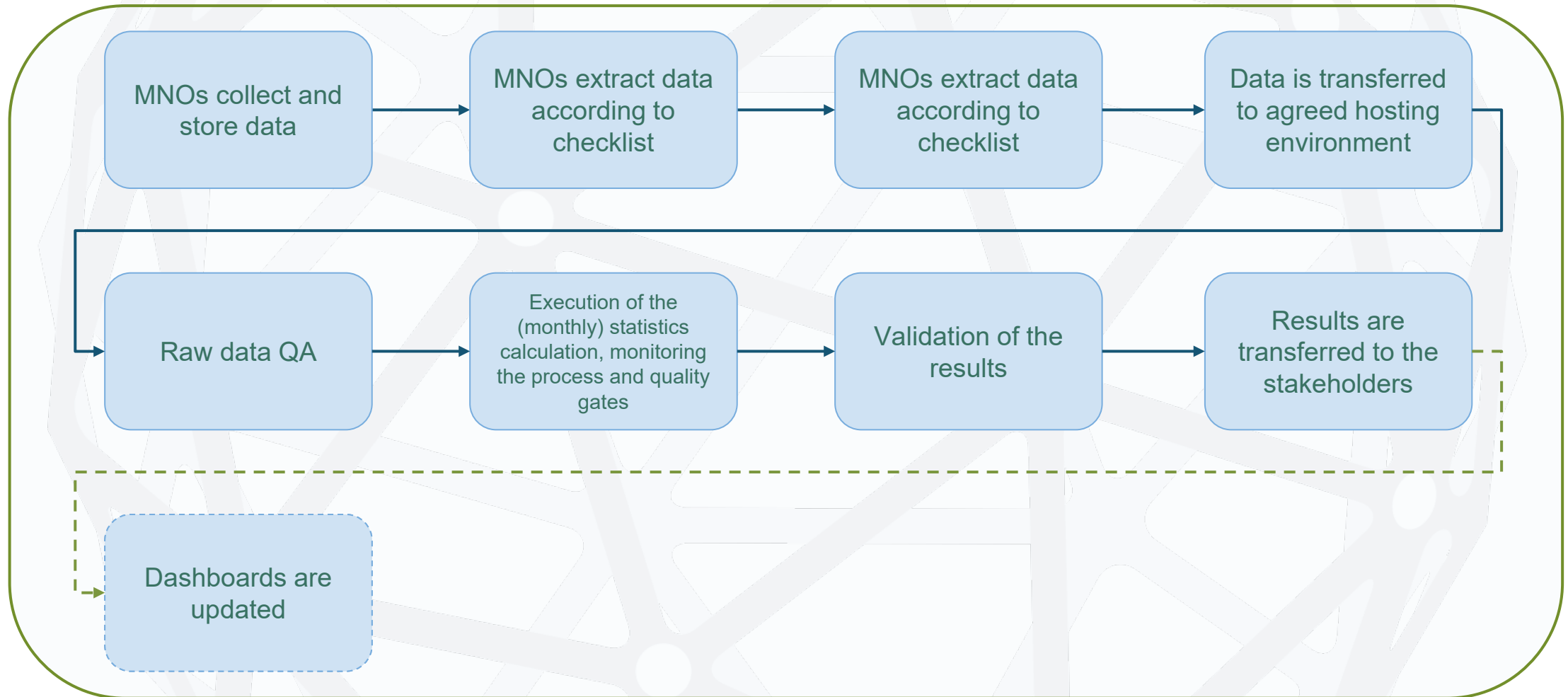
- General methodology is being adjusted according to the use case.
- It involves:
 - in depth analysis of the possible parameters,
 - creating test scenarios,
 - running tourism statistics calculations with several scenarios
 - analysing the results of these together with the relevant stakeholders

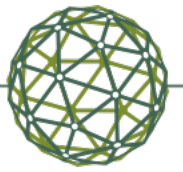
- All agreed statistics in all agreed domains are calculated based on actual data received from the MNOs (usually for 6-12 months data)
- Data visualisations might be used for presenting the output statistics

- The statistics are analysed together with the stakeholders, compared to reference data available etc



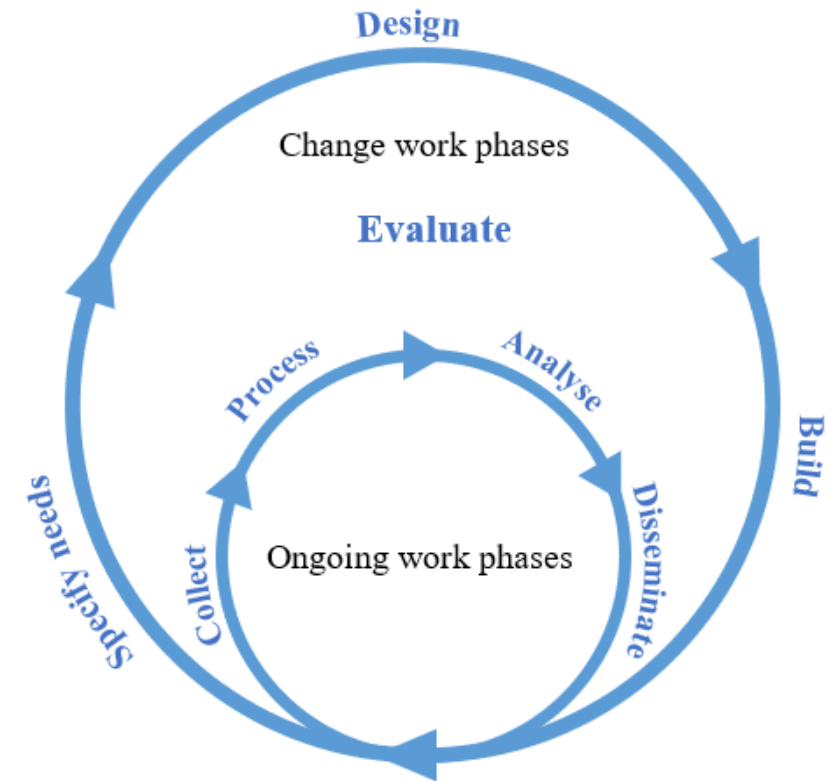
Implementation

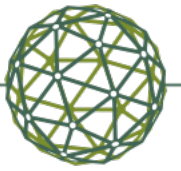




Implementation

- When the pilot project is a success and the method has been validated to work, implementation to regular statistical production commences.
- Production is focusing on providing continuous data updates (e.g. monthly statistics).
- It includes all steps related to the data collection, processing and calculation of the indicators.





Summary

- **Think of use of MPD** as the use of any new data in a statistical business process (GSBPM)
- **Assess the current situation** by specifying needs exactly and making a business case for MPD
- **Prepare for access to data** by selecting appropriate legal, business and technical combinations
- **Run a pilot project** first to properly assess data quality, design the solution and establish feasibility for statistics
- Upon success, **continue to implementation** for regular production where the statistical calculation process is optimized

Thank you very much!
For any question, please contact:

[magpantay\[at\]itu.int](mailto:magpantay[at]itu.int)



#UNBigData2024