# EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

## 2024 ESTORIL - PORTUGAL

Instituto Nacional de Estatística
Statistics Portugal

eurostat

The conference is partly
financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

# A processing pipeline for European Official Statistics: towards standardization of mobile network operator data processing

*Erika Cerasti*

**ISTAT - National Institue of Statistics - Italy**

**erika.cerasti@istat.it**

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly
financed by the European Union

The conference is partly financed by the European Union

# The consortium

Consortium leader

GOPA WORLDWIDE CONSULTANTS

Consortium partners

positium NOMMON

Subcontractors

Methodology and quality in official statistics

Advisory Board for the Use of MNO's Data for Official Statistics

REPUBLIC OF SLOVENIA STATISTICAL OFFICE

Statistics Poland

Individual experts

Testing MNOs

Vodafone Italy

Orange España

Vodafone España

GOPAcom.

Communication and dissemination

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

# Specific objectives:

- Development of an **open end-to-end methodological framework** for the processing of MNO data for official statistics;
- Development of a **reference open-source software pipeline** implementing the proposed methodological framework;
- Practical **demonstration of the processing** pipeline across *5 MNOs in four EU countries* to produce a set of experimental statistical indicators.

***Use-cases domains**:* inbound tourism, outbound tourism, population, spatial statistics (e.g., functional urban areas), local commuting, and international commuting.
***Period:*** *February 2023 - June 2025*

***Development of a complete, open, end-to-end PROCESSING PIPELINE as a PROPOSAL for the production of future official statistics based on multiple MNOs data***

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

# Context

- **ESS Task Force on the use of MNO Data for official statistics:** established in 2021 to address the methodological aspects involved in the reuse of MNO for official statistics. Chaired by Eurostat – 18 EU NSIs

- **Research grant: ESSnet MNO-MINDS "Mobile Network Operator Methods for Integrating New Data Sources". Object**: methodologies for Integration of MNO and non-MNO data sources. Funded by Eurostat – 10 EU NSIs

***Tailored to the European Statistical System (ESS)context and needs the project develops a proposal for a methodological framework that will be reviewed by the ESS TF MNO and may be adopted as ESS standard.***

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

Reference Position paper:
***"Reusing mobile network operator data for official statistics: the case for a common methodological framework for the European Statistical System – 2023 edition"***

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Key considerations to ensure the framework's ***flexibility, adaptability and applicability*** **to** diverse scenarios:

- **Methodological soundness**
- **Integration of previous findings** (*from past work by the ESS and NSIs at EU national level)*
- **Stakeholder consultation**
- **Evolvability**
- **Methodological challenges and recommendations**
- **Modularity**
- **Use case specific methods in the pipeline**
- **Consistency**
- **Quality assurance**
- **Explainability**
- **Adherence to standards**
- **Openness and reproducibility**
- **Multi-MNO orientation**
- **Privacy protection**

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly
financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

# Reference and demonstrator scenario

Set of assumptions about the context in which the proposed pipeline will be used by NSIs.

- **Demonstrator scenario**:
  - ***data processing environment*** - disaggregated data will always remain within the MNO infrastructure.
  - ***SDC methods*** are implemented at the MNO level due to legal, technical, and privacy considerations.

- **Reference scenario**:
  - ***SDC methods*** are implemented at the NSIs level.
  - ***advancements*** in integration and privacy preserving techniques.

This project is not devoted to the investigation of methods for integrating MNO data with other non-MNO data sources or SDC methods

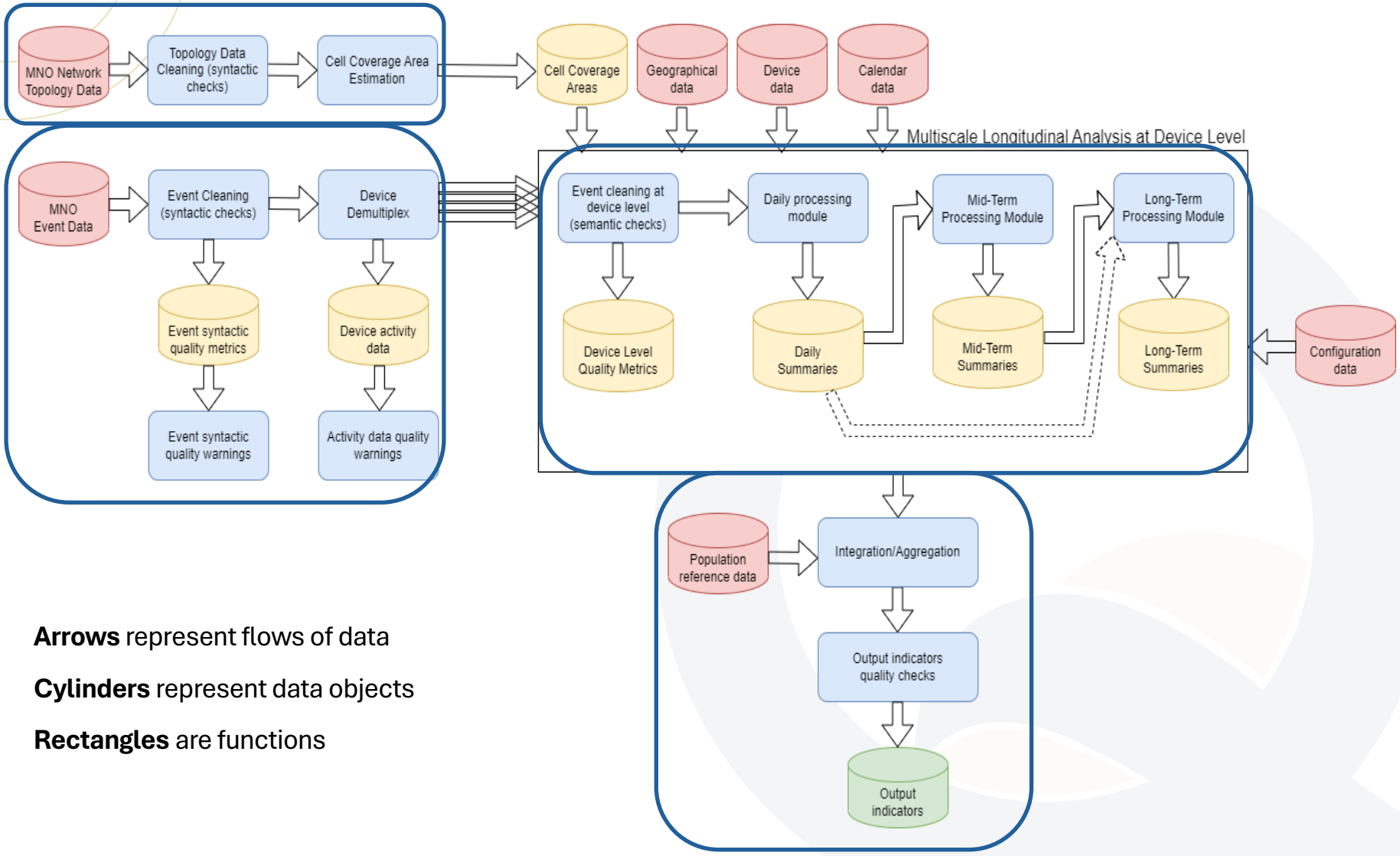
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

# PIPELINE PROCESS FLOW DIAGRAM




**Arrows** represent flows of data

**Cylinders** represent data objects

**Rectangles** are functions

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

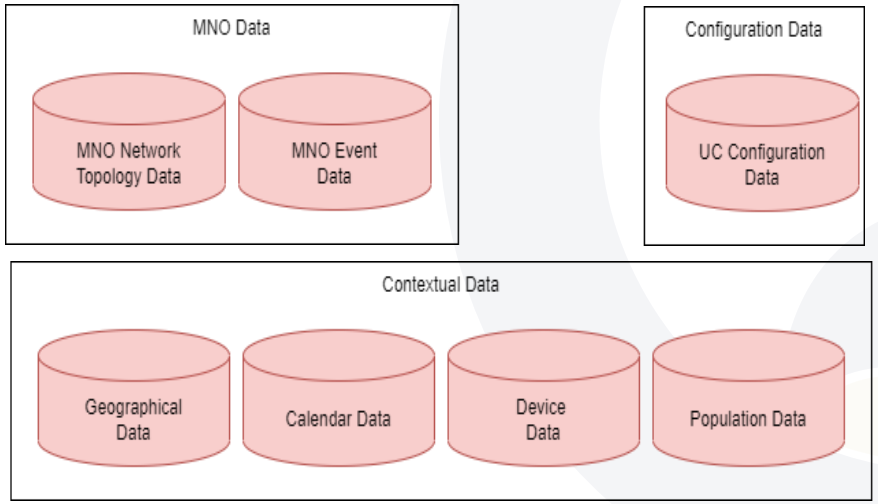
The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

# INPUT DATA OBJECTS

The process take as input three data categories containing different types of data objects:

- **MNO Data:** MNO network topology data, MNO event data
- **Contextual Data:** geographical data, Calendar Data, Device Data, Population Data
- **Configuration Data**: data used to specify the use case (i.e. selected indicators, time resolution, zoning system, use case specific requirements, etc)




INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

# PROCESSING NETWORK TOPOLOGY DATA

## Cell footprint estimation module

The cell footprint determines how much a cell "covers" each tile of a spatial grid.

Module: generate ***a standardised spatial representation*** of the cell footprint for each cell based on the available MNO Network Topology Data.

When latitude and longitude information about the device location are not provided, we use the estimation of the geographic area covered by the cell the device is connected to, as the probable location of the device.

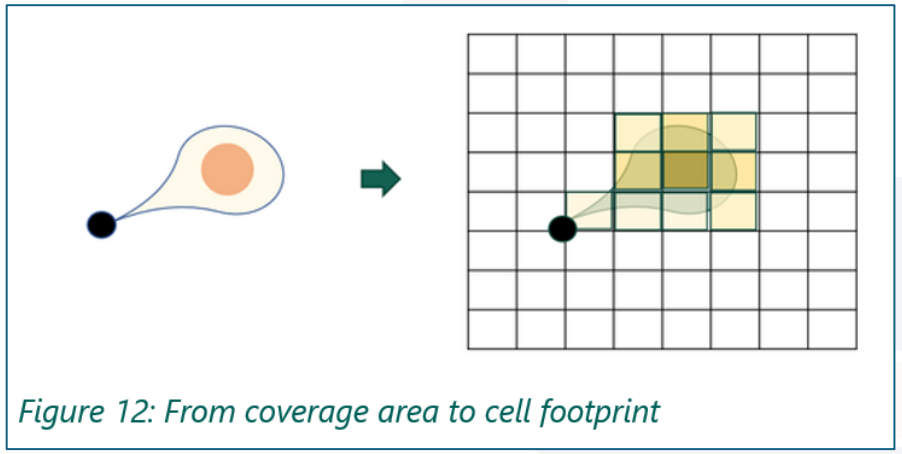


*Figure 12: From coverage area to cell footprint*

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

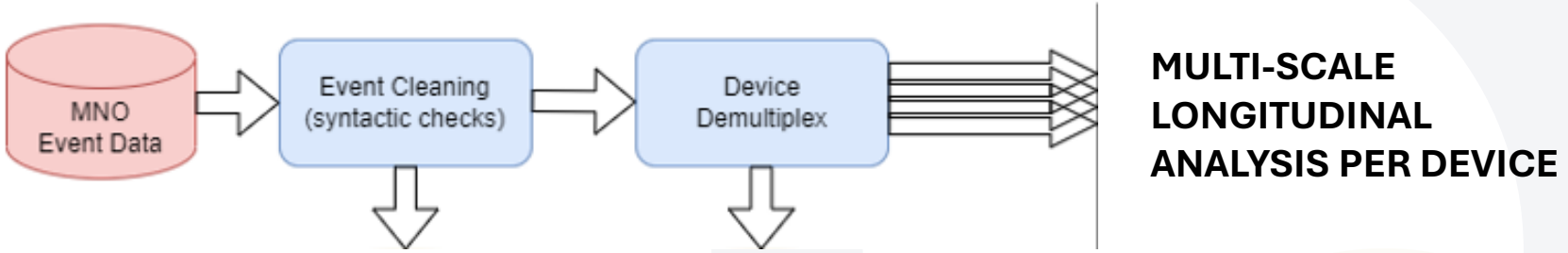
The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

# PROCESSING EVENT DATA

The pipeline receives daily flows of Event Data.
Each flow is processed according to the following chain:

- **Event Cleaning function**
  Filter out malformed or missing data
  A syntactic check is applied
- **Device Demultiplex module**
  Cleaned Data are **grouped per device and per day**
  sub-flow of temporally ordered events for each device




***Quality controls are implemented at different stages the processing flow: input data, intermediate results and final outputs.***

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

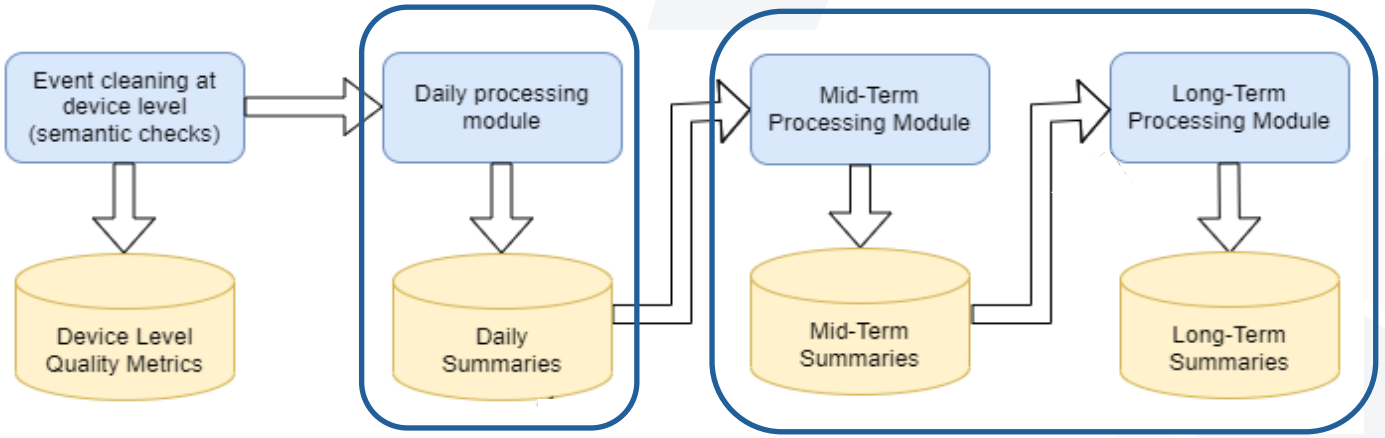
# MULTI-SCALE LONGITUDINAL ANALYSIS PER DEVICE

Events of each device are processed independently in a multi-scale dimension:

- **Event cleaning at device level**
- **Daily processing**
- **Longitudinal functions:** Mid-Term processing and Long-Term Processing modules

Unidirectional data flow, from the smaller to larger scales, without feedback loops.

Information at the individual device level never leaves the secure environment, only aggregated data are send to NSIs.




**Integration of several types of contextual information with summary objects**

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

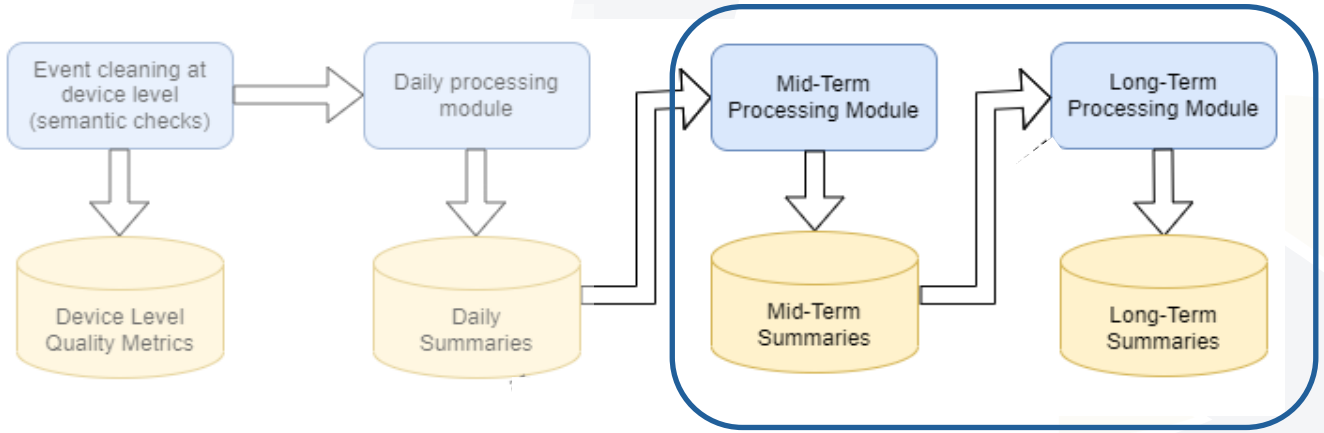
# MULTI-SCALE LONGITUDINAL ANALYSIS PER DEVICE

## Mid-Term processing output example

- the most common overnight place of a user, over a certain month. Some use cases will require the production of outputs related to the mid-term interval, as the tourism statistics use case.

## Long-Term processing output example

- anchor points (e.g. usual environment, home-location, workplace) and other labelled/tagged information that is required by the use cases of the pipeline if new needs arise.




INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

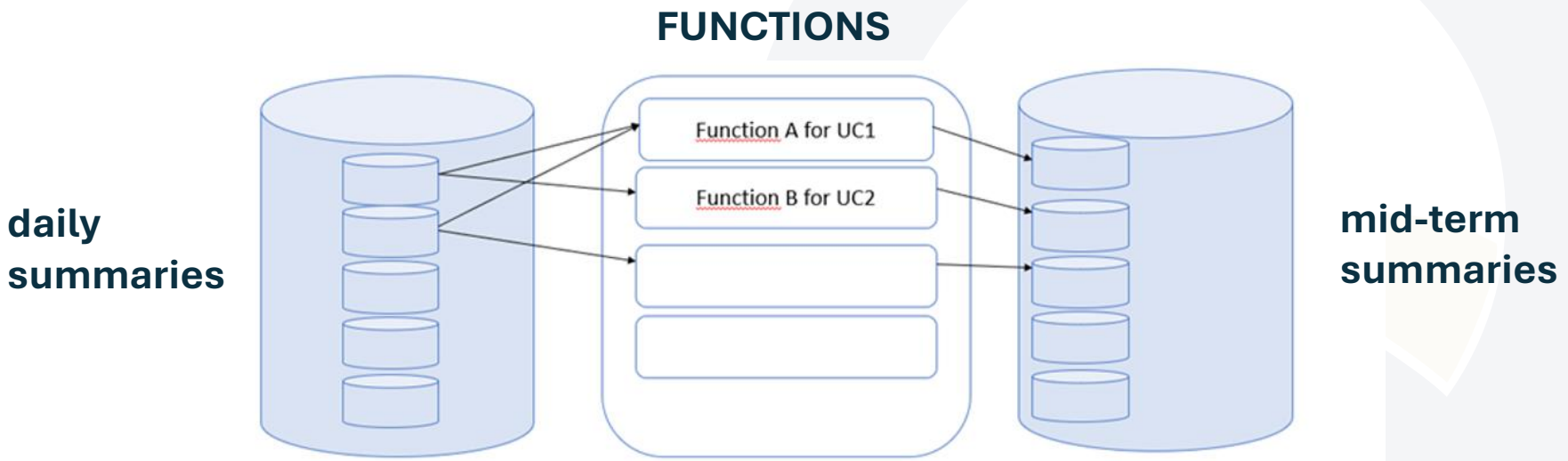
EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

# MODULARITY OF FUNCTIONS AND DATA OBJECTS

**From Daily processing to Mid/Long-Term processing**

Daily summaries serve as input for different functions of the mid/long-term processing module.

The modularity of the pipeline enables ***different use cases*** providing different statistics of interest and facilitates the evolvability of the pipeline if new needs arise.




INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

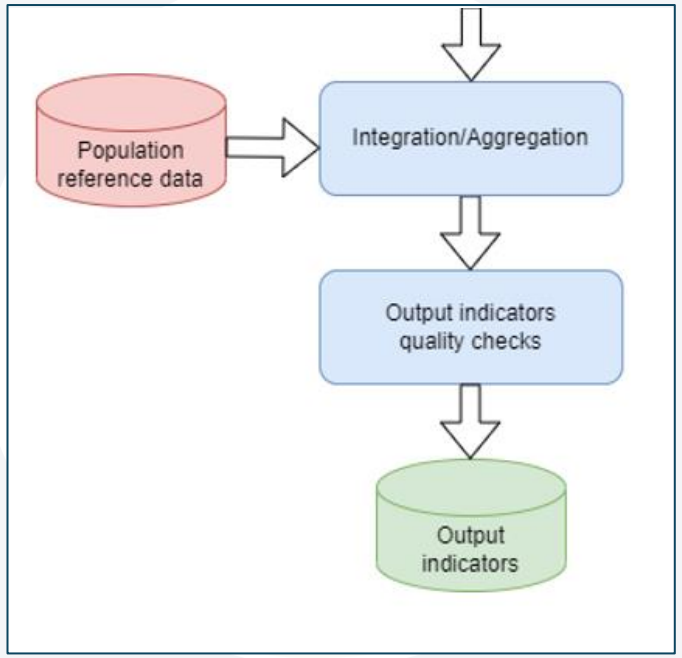
# AGGREGATION AND INTEGRATION MODULES

Methods for aggregation for integration of MNO data with other non-MNO data sources **are not the focus of the project.**

Modules corresponding to these processes are included in the pipeline and simple methods are implemented in order to provide an end-to end workflow.

Advanced methods will be developped in the research project ***ESSnet MNO-MINDS Mobile Network Operator Methods for Integrating New Data Sources***

Eventually included in the pipeline in future developments




INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

eurostat

The conference is partly financed by the European Union

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL
eurostat
The conference is partly financed by the European Union

# For further details or if you wish to be informed on the progress of the work:

**Project website**
https://cros.ec.europa.eu/multi-mno-project

**Contact: Florabela Carausu** *Florabela.Carausu@gopa.de*

*Project Manager, Service Contract Eurostat ref. 2021.0400 TSS Multi-MNO(Trusted Smart Statistics, multi-Mobile Network Operators)*

# Thanks for your attention