Study on monitoring data and monitoring systems

Objective and timing

- Objectives:

  1) to provide reliable and robust monitoring data for the thematic ex-post evaluations (Tasks 1,2,3)

  2) to enhance the Commission's and other stakeholders' understanding of the functioning of the programmes' monitoring systems (Task 4)

- Contractor will work for 14 months, started in January 2021

- Country experts are contacting Managing Authorities to collect data

European Commission

State of play: TASK 1&2 project database

The study will:

- collect the list of operations/projects supported under the programmes

- integrate the datasets obtained from managing authorities into a single database

- populate the database with common attributes for all projects/operations: financial information, project description and content, and any other common datafields (+ data cleaning, standardisation & enriching)

- typology clustering of the projects/operations into interventions (beyond the categorization by intervention fields)

- clustering the beneficiaries by type and size

- See one example of use of data in presentation for RTD eval

Overview of the Tasks

Task 1:

- Collect information and **expenditure data** about the operations and beneficiaries
- Synthesize them into a single searchable database

Task 2:

Cluster the operations and beneficiaries into typology groups according to their characteristics

Task 3a:

- Collect and validate **common monitoring indicators**, as far as possible at the level of operations
- Analysis of plausibility/ reliability

Task 3b:

- Collect information on **programme-specific indicators**
- Aggregation of similar indicators

Task 4:

- Assessment of ERDF/CF **monitoring systems**
- Recommendations for improvements

Task 5: Workshop

Task 6: Final Report

WE ARE HERE

March 7, 2021: Database of expenditure + report

June 7, 2021: Complete database of expenditure, analysis of typologies of operations and beneficiaries

September 7, 2021: Databases of achievement indicators + reports
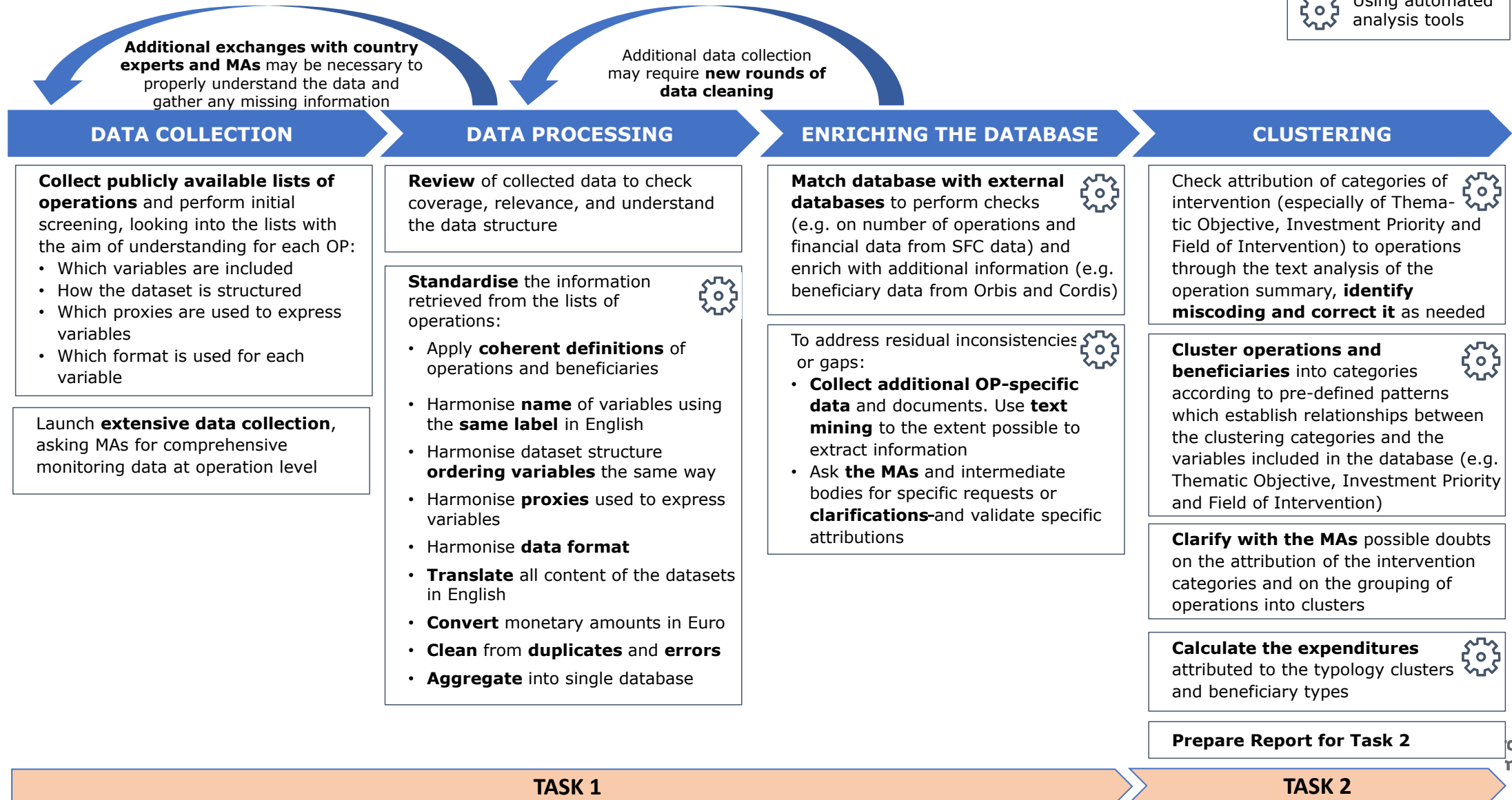
December 7, 2021: Country sheets, + report

January 7, 2022

February 7, 2022

European Commission

4

T1 and T2 - Approach to the database construction

Using automated analysis tools

Additional exchanges with country experts and MAs may be necessary to properly understand the data and gather any missing information

Additional data collection may require **new rounds of data cleaning**

DATA COLLECTION	DATA PROCESSING	ENRICHING THE DATABASE	CLUSTERING

Collect publicly available lists of operations and perform initial screening, looking into the lists with the aim of understanding for each OP:
- Which variables are included
- How the dataset is structured
- Which proxies are used to express variables
- Which format is used for each variable

Launch **extensive data collection**, asking MAs for comprehensive monitoring data at operation level

Review of collected data to check coverage, relevance, and understand the data structure

Standardise the information retrieved from the lists of operations:
- Apply **coherent definitions** of operations and beneficiaries
- Harmonise **name** of variables using the **same label** in English
- Harmonise dataset structure **ordering variables** the same way
- Harmonise **proxies** used to express variables
- Harmonise **data format**
- **Translate** all content of the datasets in English
- **Convert** monetary amounts in Euro
- **Clean** from **duplicates** and **errors**
- **Aggregate** into single database

Match database with external databases to perform checks (e.g. on number of operations and financial data from SFC data) and enrich with additional information (e.g. beneficiary data from Orbis and Cordis)

To address residual inconsistencies or gaps:
- **Collect additional OP-specific data** and documents. Use **text mining** to the extent possible to extract information
- Ask **the MAs** and intermediate bodies for specific requests or **clarifications** and validate specific attributions

Check attribution of categories of intervention (especially of Thematic Objective, Investment Priority and Field of Intervention) to operations through the text analysis of the operation summary, **identify miscoding and correct it** as needed

Cluster operations and beneficiaries into categories according to pre-defined patterns which establish relationships between the clustering categories and the variables included in the database (e.g. Thematic Objective, Investment Priority and Field of Intervention)

Clarify with the MAs possible doubts on the attribution of the intervention categories and on the grouping of operations into clusters

Calculate the expenditures attributed to the typology clusters and beneficiary types

Prepare Report for Task 2

European Commission

TASK 1	TASK 2

Questions raised during data collection by the Contractor

- *What is asked of Managing Authorities approached by the Contractor?*

To share monitoring data which is recorded and stored in computerised form in their monitoring system (as per Annex III of the Delegated Regulation of No. 48 2014 of the CPR 03/03/2014);

- *Does this include personal data?* Data related to natural persons should not be shared and is not needed for this study;

- The Contractor (CSIL Milano, Prognos and PPMI) is bound by a confidentiality clause ensuring the protection of the shared data.