



Study on the monitoring data on ERDF and Cohesion Fund operations, and on the monitoring systems operated in the 2014-2020 period

Contract N° 2019CE16BAT214/2020CE16BAT075

Deliverable 2:
Report on the construction of the Single Database

Written by CSIL
December 2021

EUROPEAN COMMISSION

Directorate-General for Regional and Urban Policy
Directorate B — Policy
Unit B.2 — Evaluation and European Semester

Contact: David Alba and John Walsh

E-mail: regio-eval@ec.europa.eu

*European Commission
B-1049 Brussels*

**Study on the monitoring data on
ERDF and Cohesion Fund
operations, and on the
monitoring systems operated
in the 2014-2020 period**

Contract N° 2019CE16BAT214/2020CE16BAT075

Deliverable D2:

Report on the construction of the Single Database

Manuscript completed in December 2021

[Revised/Corrected/nth] edition

The European Commission is not liable for any consequence stemming from the reuse of this publication.

© European Union, 2021



The reuse policy of European Commission documents is implemented based on Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders.

Table of Contents

Foreword and synthesis	4
1. Background of the study.....	7
2. Strategy for constructing a Single Database	9
2.1. A quick guide for potential users of the Single Database	9
2.2. The overall approach for the data collection, cleaning, harmonisation and enrichment	10
2.2.1. Data collection and preliminary screening	13
2.2.2. Data cleaning, harmonisation and enrichment.....	14
3. Presentation of the Database of Operations and Beneficiaries.....	22
3.1. Coverage of the Database of Operations and Beneficiaries.....	22
3.2. Structure and content of the Database of Operations and the Database of Beneficiaries	23
3.2.1. The structure of the Database of Operations	23
3.2.2. The structure of the Database of Beneficiaries	53
4. The limitations of the Single Database	75
ANNEXES	77
Annex I: Previous data collection projects (as of end 2021).....	78
Annex II: EU regulatory requirements regarding monitoring data	80
Annex III: Availability of public lists of operations and related data coverage ...	82
Annex IV: Data request to the MAs	86
Annex V: Description of the additional .csv files, including operations and variables excluded from the Single Database	89
Annex V.1: The database on non-funded operations.....	89
Annex V.2: The database on beneficiaries of non-funded operations.....	90
Annex V.3: The database on excluded variables at operation level.....	91
Annex V.4 The database on excluded variables at beneficiary level	92
Annex VI: The results of the matching of the database of beneficiaries with the ORBIS database	94

List of Figures

Figure 1: Database structure and interlinkages.....	9
Figure 2: Approach to construct a comprehensive and coherent database	12
Figure 3: Work flow to standardise file formats	16
Figure 4: Example of the variable naming procedure for a French OP	17
Figure 5: Combination of data cleaning, harmonisation and enrichment procedures across different groups of variables.....	20
Figure 6: Total number of operations by Member State	24

List of Tables

Table 1 – Previous data collection initiatives at a comparison.....	8
Table 2 – Data requirements at operation and beneficiary level from EU Regulations	80
Table 3 – Degree of data coverage of variables of the operations database	83
Table 4 – Degree of data coverage of variables of the beneficiaries database	84
Table 5 – Coverage of the database on non-funded operations.....	89
Table 6 – Coverage of the database on beneficiaries of non-funded operations	90
Table 7 – Coverage of the database excluded variables at operation level..	91
Table 8 – Coverage of the database excluded variables at beneficiary level	92
Table 9 – Results of the matching of the database of beneficiaries with the ORBIS database	94

LIST OF ABBREVIATIONS

AT	Austria
BE	Belgium
BG	Bulgaria
BvD	Bureau van Dijk
CF	Cohesion Fund
CP(s)	Cooperation Programme(s)
CPR	Common Provisions Regulation
CY	Cyprus
CZ	Czech Republic
DE	Germany
DB	Database
DG CNECT	Directorate-General for Communications Networks, Content
DG REGIO	Directorate-General Regional and Urban Policy
DK	Denmark
EC	European Commission
ECB	European Central Bank
EE	Estonia
EL	Greece
EN	English
ENI	European Neighbourhood Instrument
ERDF	European Regional Development Fund
ESF	European Social Fund
ESIF	European Structural and Investment Funds
EU	European Union
EUR	Euro (currency)
FI	Finland
FoI	Field of Intervention

FR	France
H2020	Horizon 2020
HR	Croatia
JRC	Joint Research Centre
ICT	Information and Communication Technology
ID	Identification/Identity/Identifier
IE	Ireland
IP	Investment Priority
IPA	Instrument for Pre-Accession Assistance
ISO	International Organisation for Standardisation
IT	Italy
JAP	Joint Action Plan
KETs	Key Enabling Technologies
LT	Lithuania
LU	Luxembourg
LV	Latvia
MA(s)	Managing Authority(ies)
MS	Member State(s)
MT	Malta
N.A.	Not available/Not applicable
NACE	Nomenclature statistique des activités économiques dans la Communauté européenne / Statistical classification of economic activities in the European Community
NGO(s)	Non-Governmental Organisation
NL	(The) Netherlands
NUTS	Nomenclature des Unités Territoriales Statistiques
OP(s)	Operational Programme(s)
PL	Poland
PT	Portugal
R&D	Research and Development

R&I	Research and Innovation
RIS3	Research and Innovation Strategies for Smart Specialisations
RO	Romania
RTD	Research and Technological Development
SFC	System for Fund Management
SGC	Societal Grand Challenges
SI	Slovenia
SMEi	Small and Medium Enterprise Initiative
SMEs	Small and Medium Enterprises
SK	Slovakia
TC	Territorial Cooperation
TO	Thematic Objective
UK	United Kingdom
VIES	VAT Information Exchange System
VAT	Value-Added Tax

Foreword and synthesis

The European Commission awarded CSIL – Centre for Industrial Studies, in partnership with Prognos and PPMI, a contract to carry out the “Study on the monitoring data on ERDF and Cohesion Fund operations, and on the monitoring systems operated in the 2014-2020 period”.

The project aims at providing **reliable and robust monitoring data on expenditure and achievement indicators** that will feed into the Commission’s ex-post evaluation of Cohesion Policy programmes in the 2014-2020 programming period. The work carried out under this contract provides input to the ex-post evaluation in terms of (1) creating a Single Database of the funded operations and classifying them according to their scope of intervention, form of finance and type of beneficiary and (2) gathering, classifying and quality assessing the output indicator data collected by Managing Authorities (MAs).

The first result of the study is the creation of **three interlinked databases** (i.e., Single Database) on the operations funded through the European Regional Development Fund (ERDF) and the Cohesion Fund (CF) during the 2014-2020 programming period under all Thematic Objectives (TOs) and categories of expenditure, on the related beneficiaries and common and programme-specific output indicators at the operation level. These databases synthesise in a harmonised way the vast amount of monitoring data available in the national/regional monitoring systems and in public lists of operations¹ so that it can be processed further and used for later analysis and evaluation work. The data included has four main strong points:

1. **High representativeness:** the Database of Operations (DB_Operations) and Beneficiaries (DB_Beneficiaries) cover 215 national and regional programmes out of 217 and 73 Cooperation Programmes out of 76 (for a total of 288 programmes out of 293, i.e., 98%). The Database of Operations includes descriptive, categorical, and financial information on a total of 584,828 ERDF and CF operations funded during the 2014-2020 programming before a certain cut-off date (generally the end of the year 2020 or the first months of the year 2021). The Database of Beneficiaries includes descriptive and financial information (when available) on different types of beneficiaries. There are 722,787 direct beneficiaries of EU funds in charge of implementing a project, which may be further distinguished in sole beneficiaries (in case of individual projects) and, upon data availability, in lead beneficiaries and related partners (in case of collaborative projects). In addition, there are 437,083 final recipients of operations delivered through an intermediary organisation (e.g., fund manager, financial intermediary, etc.)². Finally, it is also possible to distinguish 8,841 Interreg-specific beneficiaries, namely institutions supporting direct beneficiaries in implementing Interreg programmes. The coverage and representativeness of the Database of Indicators (DB_Indicators) is reported in Deliverable D7.
2. **Interlinkability:** the three databases are interlinked through a unique ad-hoc identifier of the operation, allowing linking the data at beneficiaries and output indicators level to the related operation and vice versa. Moreover, each operation is linked to the Priority Axis, Investment Priority, Thematic Objective

¹ Data has been collected from public list of operations, when available, to fill the gaps of data extractions from monitoring systems.

² The numbers provided corresponds to the total number of beneficiaries for which data was available in the data provided by the Managing Authority or collected from public list of operations. If the same beneficiary has benefitted from more than one operation, it is counted as many times as many operations it benefitted from.

and programme it contributes to so that the causality chain from strategic objectives to funded projects, beneficiaries and delivered output can be established.

3. **Comparability:** Data has been cleaned and harmonised to allow aggregation at MS and EU levels as well as insightful comparative analysis. Moreover, beyond data already available in monitoring systems, the databases include additional information clarifying the nature of operations, beneficiaries and indicators obtained by matching with an external database or own analysis (in particular, a cluster analysis). Operations can be distinguished based on their status (i.e., whether they are completed or ongoing), common typology by TO, financial scale (major and non-major project) and other attributes (e.g. State aid or non-state aid). Beneficiaries are classified based on their NACE sector, ownership, size (for enterprises) and typology (based on their mission). Indicators can be differentiated between common and programme-specific indicators.
4. **Scalability:** Data can be matched with other external databases (e.g., CORDIS data on H2020 projects, Orbis company database, etc.) to retrieve additional information, for instance, on the number of projects carried out under the H2020 framework or other micro-level data, such as data on employment or the turnover. The match can be easily performed by using the CORDIS and Orbis unique identifiers already included in the database of beneficiaries.

This report (**Deliverable D2**) illustrates the data collection process regarding data on operations (DB_Operations) and beneficiaries (DB_Beneficiaries). It discusses the gaps, inconsistencies, and data errors encountered during the integration of data, the solution to overcome these, and the activities undertaken for data cleaning, harmonisation, and enrichment. It also provides detailed information on the structure and key variables included in both databases. The data collection process and the related data cleaning, harmonisation and enrichment procedures, as well as the structure of the database on common and programme-specific output indicators at the operation level (DB_Indicators), are described in Deliverable D7.

The report is organised as follows:

- **Chapter 1** briefly presents the background of the study and the lessons learnt from past attempts to systematically collect detailed expenditure data on Cohesion Policy programmes, which have been taken into account in the design of the strategy to clean, harmonise and enrich the data;
- **Chapter 2** includes a quick guide for the potential users of the Single Database, explaining in general terms how the Single Database could be exploited and presents the process for collecting the data and constructing and harmonising the Database of Operations and Beneficiaries, pointing out the solutions used to overcome and mitigate the main data gaps and inconsistencies;
- **Chapter 3** describes the structure of the Database of Operations (DB_Operations) and the Database of Beneficiaries (DB_Beneficiaries) as well as the content of each variable. If specific to a given variable, the implemented data cleaning, harmonisation and enrichment procedures are described in this section.
- **Chapter 4** lists the main limitations of the Single Database, especially related to the Database of Operations and the Database of Beneficiaries, and the additional refinements which may be implemented while performing the ex-post thematic evaluations.

The report is complemented by six Annexes that expand on different aspects of the methodology:

- **Annex I** provides more details on the previous data collection studies carried out with the aim to collect data on ERDF/CF operations and beneficiaries;
- **Annex II** lays down the main EU regulatory requirements in terms of monitoring data at the level of individual operations and beneficiaries;
- **Annex III** reports the results of the preliminary screening of the public list of operations illustrating their degree of availability;
- **Annex IV** includes the data request submitted to all the MAs;
- **Annex V** describes the additional datasets provided to the European Commission, including operations or variables not retained in the Single Database;
- **Annex VI** describes the results of the matching of the database of beneficiaries with the Orbis database.

The report is accompanied by two datasets on operations (DB_Operations) and beneficiaries (DB_Beneficiaries) in .csv format (**Deliverable D4**), respectively. A third database on output indicators at the operation level (DB_Indicators) is included in Deliverable D7. It is also complemented by an additional report including i) an assessment of the categorisation applied by the national authorities to the operations; ii) per each Thematic Objective, the description of the typologies of operations funded and their economic relevance (i.e., the total, EU, public, and eligible private expenditure allocated to the operations belonging to each operation typology); iii) the description of the beneficiary types (**Deliverable D3** and **Deliverable D5**).

1. Background of the study

The construction of the Single Database responds to the **need to overcome the fragmentation and inconsistencies of data at the operation level** available on the websites as well as in the monitoring systems of the different national and regional Managing Authorities (MAs). Despite the efforts of the European Commission to introduce stricter regulatory requirements regarding the monitoring of operations funded by Cohesion Policy, the degree of data availability, as well as the interpretation of the regulatory framework, still differs across the Member States (and the UK) and even within the same Member State. Therefore, it is critical and highly strategic to collect the data available in the monitoring systems, investigate their quality and scope for aggregation, comparability and use for evaluation purposes at the EU level, and devise a strategy to ensure data cleaning, harmonisation and enrichment.

Drawing from previous experiences of data collection projects launched by the European Commission over the last years to prepare the basis for ex-post evaluations (see Table 1 and Annex I for more details on these projects), this study goes beyond what has already been done. In particular:

- It has a **larger scope and scale**, as it covers all 217 OPs co-financed by ERDF and/or CF³ and 76 Cooperation Programmes (CPs) co-financed during the 2014-2020 period, across the 12 Thematic Objectives;
- It collects a **larger volume of data** from the MAs: in addition to the publicly available variables required by Article 115(2) of the Regulation (EU) No 1303/2013, the study aimed at collecting all the information that according to Art. 24 of the Regulation (EU) No 480/2014 should be recorded and stored in a computerised form (see Annex II for more details on the EU regulatory requirements concerning monitoring data);
- It distinguishes as far as possible among **the different definitions of operations** adopted by national and regional MAs in order to allow future evaluators to better grasp the nature of the unit of analysis;
- It attempted to collect the **full list of beneficiaries** of operations, including the final recipients of those operations being financial instruments or groups of projects.
- It not only focuses on descriptive and financial information on operations and beneficiaries but also on the related **output indicators**;
- It puts significantly higher emphasis on the need to ensure **proper data cleaning, harmonisation and enrichment**, in order to cluster operations and beneficiaries into homogenous typologies and ensure subsequent reuse of data in the ex-post evaluation studies;
- It created a repository of all data gathered from the different monitoring systems that future evaluators could analyse to answer specific research questions.

³ In principle, 220 Operational Programmes were planned, but three were then merged into other OPs, namely the OP SME Initiative in Romania (2015RO16RFSM001) which is now under the Integrated Regional Programme (2014RO16RFOP002), the Smart Growth OP in Spain (2014ES16RFOP001) which is now part of the Multi-regional OP (2014ES16RFOP002) and the Slovakian Research and Innovation OP (2014SK16RFOP001) which has been merged into the Integrated Infrastructure OP (2014SK16M1OP001).

Table 1 – Previous data collection initiatives at a comparison (as of end 2021)

	DG REGIO & CNCT pilot project “Kohesio” ⁴	DG REGIO data scraping project by Balazcs Krich5	JRC dataset of projects co-funded by the ERDF during the multi-annual financial framework 2014-2020 ⁶	Ex-post evaluation RTD programmes ⁷
Coverage	<ul style="list-style-type: none"> 6 Member States (CZ, DK, IE, FR, IT, PL) ERDF, ESF, CF OPs 2014-2020 period 	<ul style="list-style-type: none"> 22 Member States plus 37 Interreg programmes 2014-2020 period 	<ul style="list-style-type: none"> 28 Member States 58 Interreg programmes 	<ul style="list-style-type: none"> 53 ERDF OPs in 18 Member States 2007-2013 period Focus on RTD projects (fields of intervention 01-02)
Data collection	<ul style="list-style-type: none"> Manual collection of public lists of operations and beneficiaries on national/regional websites 	<ul style="list-style-type: none"> Semi-automated collection of public lists of operations on national/regional websites 	<ul style="list-style-type: none"> Manual collection of public lists of operations on national/regional websites Ad hoc requests to some MAs 	<ul style="list-style-type: none"> Lists of operations and beneficiaries provided by the MAs following ad hoc data request
Data processing	<ul style="list-style-type: none"> Partial data cleaning (manual) Partial quality control Translation 	<ul style="list-style-type: none"> Partial data cleaning (automated) No quality control No translation 	<ul style="list-style-type: none"> Manual data cleaning Partial data standardisation Translation Partial quality control on financial figures 	<ul style="list-style-type: none"> Manual data cleaning and standardisation Translation Quality control of financial figures and operations and beneficiaries’ attributes
Data enrichment	No	No	Yes: <ul style="list-style-type: none"> Link with H2020 projects and Amadeus database Attributed KETs and Societal Challenges Goals to operations 	Yes: <ul style="list-style-type: none"> Added attributes to operations and beneficiaries (through text analysis, ad hoc search, link with Amadeus database)

Source: Authors

⁴ <https://kohesio.eu/>

⁵ https://ec.europa.eu/regional_policy/en/information/publications/studies/2020/a-pilot-project-to-collect-clean-and-assess-the-list-of-operations-produced-by-cohesion-policy-programmes-at-national-regional-level

⁶ <https://publications.jrc.ec.europa.eu/repository/handle/JRC120637>

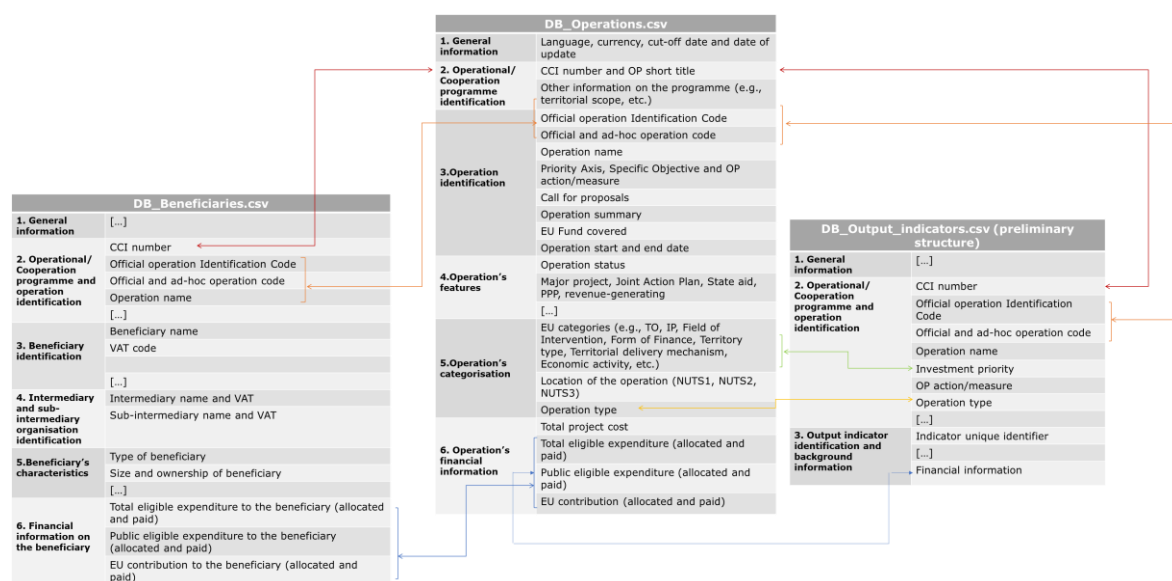
⁷ <https://cohesiondata.ec.europa.eu/stories/s/RTD-EVALUATION-2007-2013-beneficiary-and-project-d/kkj2-8bik/>

2. Strategy for constructing a Single Database

2.1. A quick guide for potential users of the Single Database

The Single Database accompanying this report covers 215 OPs co-financed by ERDF and/or CF and 73 Cooperation Programmes (CPs) co-financed during the 2014-2020 period, across the 12 Thematic Objectives (i.e., 96% of the total number of programmes). It combines three datasets, including data for **584,828 operations** (DB_Operations), **1,168,711 beneficiaries** (DB_Beneficiaries) and the full list of common and programme-specific output indicators which could be collected at the operation level (See Deliverable 7 for more details on the DB_Indicators). The three sets of data are interlinked through a unique ad-hoc identifier of the operation, which allows linking the list of beneficiaries and output indicators to the related operation. In turn, each operation is linked to its respective Priority Axis, Investment Priority, Thematic Objective and OP. In this way, the chain from the OP to the beneficiaries (and related expenditures), as well as output indicators, can be established through the operations and data included in a specific database can be integrated into the others.

Figure 1: Database structure and interlinkages



Source: Authors

This rich set of data makes available to evaluators, policy makers, practitioners and the wider public a unique data source on ERDF and CF programmes funded during 2014-2020. Both in terms of granularity and coverage, it expands much beyond what is available in public lists of operation, public open data platforms at national level as well as in Annual Implementation Reports.

Thanks to this Single Database, it is possible in particular to:

- Filter, count and sum **operations** by country, region and NUTS (until level 3), and key features (e.g., Thematic Objective, Investment Priority, Field of Interventions, including a harmonised typology describing their scope and strategic objective);
- Filter, count and sum **beneficiaries** of operations, including direct beneficiaries, the final recipients of financial instruments or groups of projects, as well as other types of beneficiaries (e.g., partners in collaborating projects) according to their location, typology, size and NACE sector if they are enterprises and other key features.

- Filter, count and sum the **outputs** reported by country, programme and operations as well as the target and implemented values of outputs to measure operations achievements;
- Filter and sum **financial resources** allocated and paid, distinguishing by operation, beneficiary, indicator, programme and territory.

Thanks to the **harmonisation at EU level**, data from this dataset allows comparing data across regions and territories (until NUTS3) of the EU, facilitating benchmarking exercises and cross-regions and cross-countries learnings.

Thanks to the **enrichment with external database**, linking beneficiaries with Orbis and CORDIS databases, it is possible to know:

- If and which of the identified beneficiaries have also been beneficiaries of H2020 project(s) and which ones;
- Whether it is possible to find the beneficiaries (almost exclusively enterprises) in the Orbis database, where key financial and accounting information can be found and downloaded.

There are also **limitations**, which require interpreting data with care to draw meaningful conclusions. In particular:

- The picture provided by this data **cannot be considered complete** and data coverage is varied depending on countries and types of variables. This holds true in particular for financial data, data on beneficiaries (especially those of intermediated instruments and those not being lead partners in collaborating projects), some of the standard EU categories, the list of indicators and the variables allowing the link with external databases.
- While an operation corresponds to a project in most cases, this does not always hold true. Depending on the approach adopted by the MA, an operation can also be an intermediated instrument, a group of projects, a complex investment project or a self-standing component of a larger investment project. Not in all cases it was possible to flag this distinction.
- To ensure compliance with the General Data Protection Regulation (GDPR), **the name of the beneficiaries has been anonymised** whenever it was provided for natural persons.
- The **cut-off date is different**, ranging from the end of December 2020 to July 2021.

The following sections provide a detailed description of the strategy for constructing the single database, highlighting the observed data limitations and the strategies put in place to overcome them.

2.2. The overall approach for the data collection, cleaning, harmonisation and enrichment

Thanks to the lessons learned during the previous data collection projects, the Core Team was fully aware of the challenges of constructing a reliable and comprehensive Single Database. In particular, the data collection strategy was informed by the following considerations:

- Public lists of operations are not enough to allow the categorisation of operations and beneficiaries, and the involvement of the MAs is unavoidable. Public lists of operations generally include insufficient information to categorise the data for ex-post evaluation purposes. The experience with the ex-post evaluation of ERDF

programmes in RTD⁸ highlighted that a much larger amount of data is available in the national and regional monitoring systems to better grasp the scope, nature and typology of implemented projects. For this reason, the data collection targeted the data that the MAs are required to store in a computerised form, even if not made public, following Regulation (EU) No 480/2014.

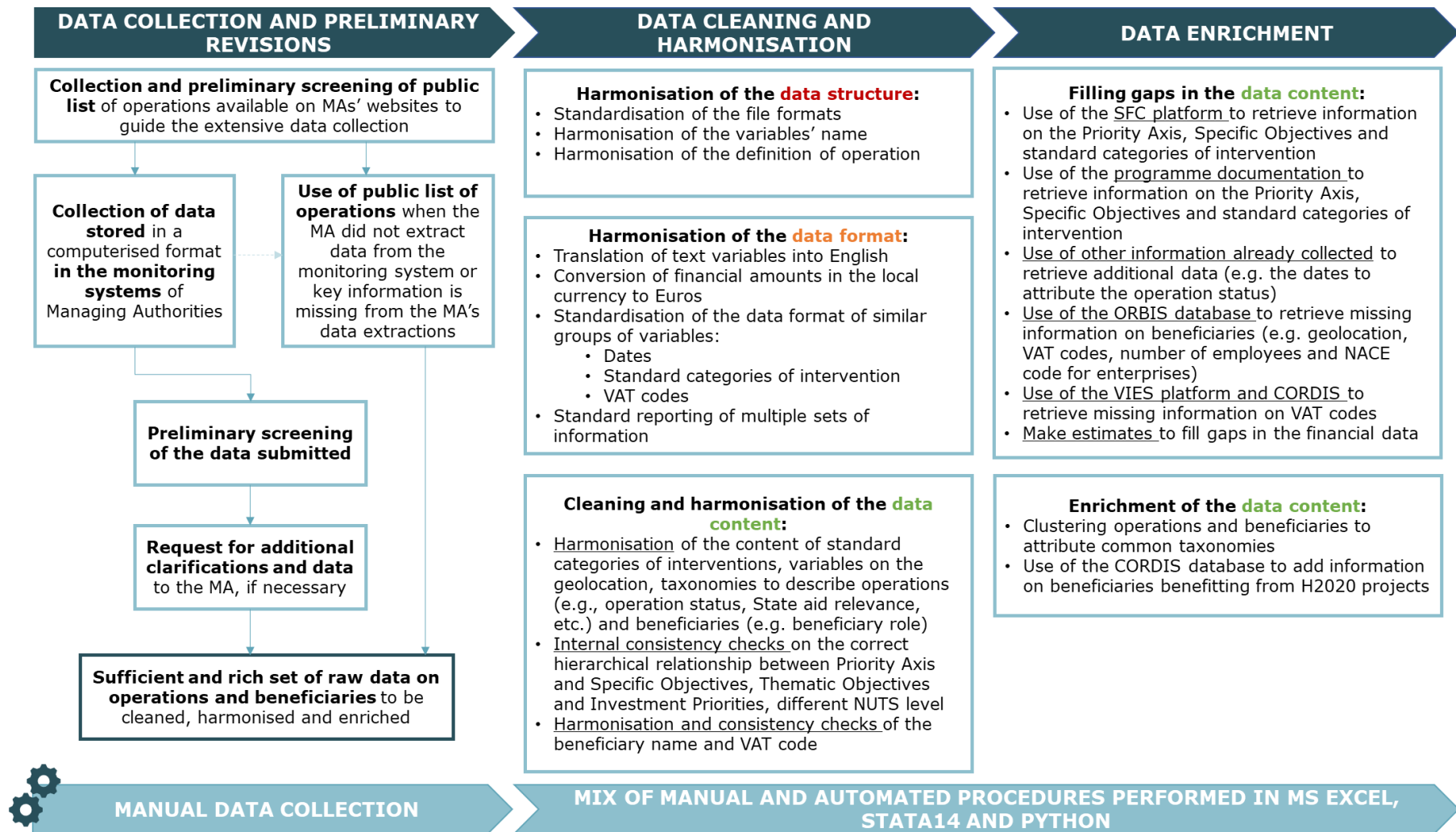
- There is a need to combine automated and manual procedures to collect and harmonise the data. The data scraping project by Balazs Krich confirmed that fully automated processes are not a viable solution for collecting or analysing operation-level data of the 2014-2020 period. Automated procedures and human inspection, along with a good understanding of the functioning and structure of the national and regional monitoring systems and of the meaning of the data therein included, were therefore combined.
- There is a need to integrate information from external databases and through other enrichment procedures. Although the ex-post evaluation of ERDF programmes in RTD already highlighted that a much larger amount of data is available in the national and regional monitoring systems, this study also showed that there are additional attributes of operations and beneficiaries which are key for ex-post evaluations and which are not directly included in monitoring systems, such as common typologies of operations and beneficiaries. Therefore, enrichment activities through clustering or linking data with external databases were deemed necessary.

In a nutshell, Figure 2: presents our approach to constructing the Single Database. In the following sections, the main activities of data cleaning, harmonisation and enrichment procedures implemented to construct the Database of Operations and the Database of Beneficiaries are described more in detail, while the approach followed for the Database of Indicators is described in Deliverable 7. Overall, it should be noted that the design and population of these two databases have been the result of:

- A **dynamic process following an incremental logic**: it has been adjusted and expanded depending on the data availability and the number and nature of the manipulation that has been needed to enrich it.
- The **work centrally managed by the Core Team**: the latter has relied on the network of Country Experts only for the extensive data collection and in residual cases for collecting a specific piece of missing information, while the entire set of data cleaning, harmonisation and enrichment activities have been implemented at a central level.

⁸ European Commission (2021), Evaluation of investments in Research and Technological Development (RTD) infrastructures and activities supported by the European Regional Development Funds (ERDF) in the period 2007-2013 See also <https://cohesiondata.ec.europa.eu/stories/s/RTD-EVALUATION-2007-2013-beneficiary-and-project-d/kkj2-8bik/>.

Figure 2: Approach to construct a comprehensive and coherent database



Source: Authors

2.2.1. Data collection and preliminary screening

The adopted data collection strategy envisaged the collection of data from two main data sources:

- **Data collected and stored in the national/regional monitoring systems**, but not necessarily publicly available: some of the data fields included in Annex III of the Regulation (EU) No 480/2014 (for which computerised storage is required) and any other kinds of data which was considered to be useful for the categorisation of operations and beneficiaries were requested to MAs.
- **Public lists of operations** of OPs and CPs made publicly available by MAs on their websites according to Art. 125(2d) of the Commission Regulation (EU) No 1303/2013 (Common Provisions Regulation - CPR): these lists have been used whenever MAs could not provide data extractions from the monitoring system as well as to enrich the data provided upon ad-hoc request by the MAs, if necessary and possible.

As a preliminary step to guide the data collection from MAs, the lists of operations available on national or regional websites and containing the minimum set of information requested by Annex II of the CPR were searched, collected and screened (see Annex III for more details on the degree of availability of public lists of operations across MS and on the data coverage). The detailed screening of the public lists confirmed that, in the large majority of OPs/CPs, the variable coverage was limited to the minimum set of information required by Regulation (EU) No 1303/2013 Article 115(2). Therefore, an ad-hoc data request was deemed necessary to collect more extensive data at the operation level.

In February 2021, a network of 38 country experts submitted a request for data on operations, beneficiaries and output indicators to all concerned MAs across the 28 Member States, following a standard guideline (see Annex IV). In Italy, Portugal, Poland, Slovakia and the Czech Republic, a central level monitoring system is in place, which provides a good amount of coherent data from both the national and regional MAs. In these cases, the data extraction was therefore requested to the central system. If additional relevant data was retained both at the central level and by regional MAs (i.e., in Portugal, Poland), the latter were contacted and asked to extract only that data not already available in the central monitoring system.

Since the data request was intentionally rather broad to limit as far as possible the data manipulation by the information technology/monitoring departments, several datasets were generally submitted by MAs as a result of the data extractions from monitoring systems. On average, 4-5 datasets, including data on operations and beneficiaries, were provided for each OP/CP. Generally, no metadata nor accompanying notes were provided, and, in most cases, **several rounds of interactions with the MA were deemed necessary** either to clarify the nature of some datasets or variables or to request additional key missing information. This additional request was prepared **based on the results of a preliminary screening of the data submitted**. This preliminary screening envisaged the following steps:

- Verification of the relevance of the data submitted in order to retain only information on approved ERDF and CF operations and to exclude other irrelevant data and information (e.g., data related to operations funded by the European Social Fund (ESF), data at IP and/or measure level, etc.) or duplicates;
- Verification of the existence of a unique operation identifier between the different data provided on operations and beneficiaries in order to identify an effective strategy to merge the different datasets provided;
- Assessment of the variable coverage after a pre-assignment of standard variable names to each variable and exclusion of data and information considered irrelevant;

- Exploration of the possibility to retrieve missing information from other sources in order to reduce the burden of the request of additional clarifications to the MAs.

The data collection process was officially closed on 11th June 2021 for data on beneficiaries and operations⁹ and at the end of July 2021 for output indicators at the operation level. Data on operations and beneficiaries were collected for 288 programmes out of 293. Instead, data on indicators were collected for 270 programmes out of 293.

2.2.2. Data cleaning, harmonisation and enrichment

In spite of the efforts made by the European Commission to ensure consistent expenditure and achievement data at the EU level through the 2014-2020 regulatory requirements and the indisputable progress achieved, the degree of data availability still differs across the Member States and, sometimes, even within the same Member State. Moreover, both the literature (EPRC, 2015¹⁰; Polverari, 2016¹¹; T33, 2017¹²; Nigohosyan and Vutsova, 2017¹³) and Core Team's direct experience highlight that national and regional authorities interpret the regulatory framework in different ways, which leads to a certain level of data fragmentation and incoherence.

Several data cleaning, harmonisation and enrichment activities were hence necessary to cope with the main limitations described above, and they are related to:

- **The data structure.** Data provided by MAs were generally organised in different files, sometimes of different formats, and included variables with different names and levels of granularity and operations having different definitions. Their harmonisation was, therefore, necessary to ensure comparability;
- **The data format.** The same variable was generally provided using different formats, different languages (i.e., text variables), or currencies (i.e., financial variables). In these cases, the data format harmonisation was key to allow the integration of data provided by the different MAs into the Single Database and to ensure their comparability;
- **The data content.** Different activities were necessary to cope with the gaps, inconsistencies and errors in the way data was reported by different MAs.

The different activities performed consisted of a mix of manual and (semi-)automated procedures performed in MS Excel, STATA and Python, depending on the complexity of the procedure implemented. Some procedures were programme-specific and performed separately for the given OP/CP in light of their specificities. Others were applied directly to the Single Database.

On top of these more general procedures, additional variable-specific data cleaning, harmonisation and enrichment procedures were performed. While the former is presented in what follows, the variables-specific activities are described more in detail in the section illustrating the content of the Single Database (see Section 3.2). As part of the data cleaning

⁹ Additional data for some programmes were also sent in July (i.e., East Wales (2014UK16RFOP006), West Wales and The Valleys (2014UK16RFOP005) and Interreg V-A - United Kingdom-Ireland (Ireland-Wales) (2014TC16RFCB048)): they have been nevertheless integrated in the single database.

¹⁰ EPRC (2015), The monitoring and evaluation of the 2014-20 EU Cohesion Policy programmes.

¹¹ Polverari, L. (2016), '2014-2020 EU Cohesion Policy: Results-orientation through better Monitoring', *Eur. Struct. Invest. Funds J.* 4, 26–34

¹² T33 (2017), Development of a system of common indicators for European Regional Development Fund and Cohesion Fund interventions after 2020.

¹³ Nigohosyan, D. and Vutsova, A. (2018), 'The 2014–2020 European Regional Development Fund Indicators: The Incomplete Evolution', *Social Indicators Research*, vol. 137.

activities, **data anonymisation** has also been performed whenever data had not been already anonymised by the Managing Authority. This process aimed to remove personally identifiable information subject to privacy protection to fully comply with the General Data Protection Regulation (GDPR)¹⁴. The variable “beneficiary’s name” in the dataset of beneficiaries has been anonymised when:

- Explicitly requested by the Managing Authority;
- It was possible to identify that the beneficiary was a natural person, not involved in business activity, based on the classification provided by the Managing Authority (see Section 3 of Deliverable D3-D5).

2.2.2.1. Procedures to harmonise the data structure

Overall, the data collected on operations and beneficiaries are characterised by:

- A high number of files, not always in table format;
- A high number of variables with very different names and levels of granularity;
- A different definition of operations within and across programmes;

The following sections briefly describe the procedures put in place to cope with these limitations.

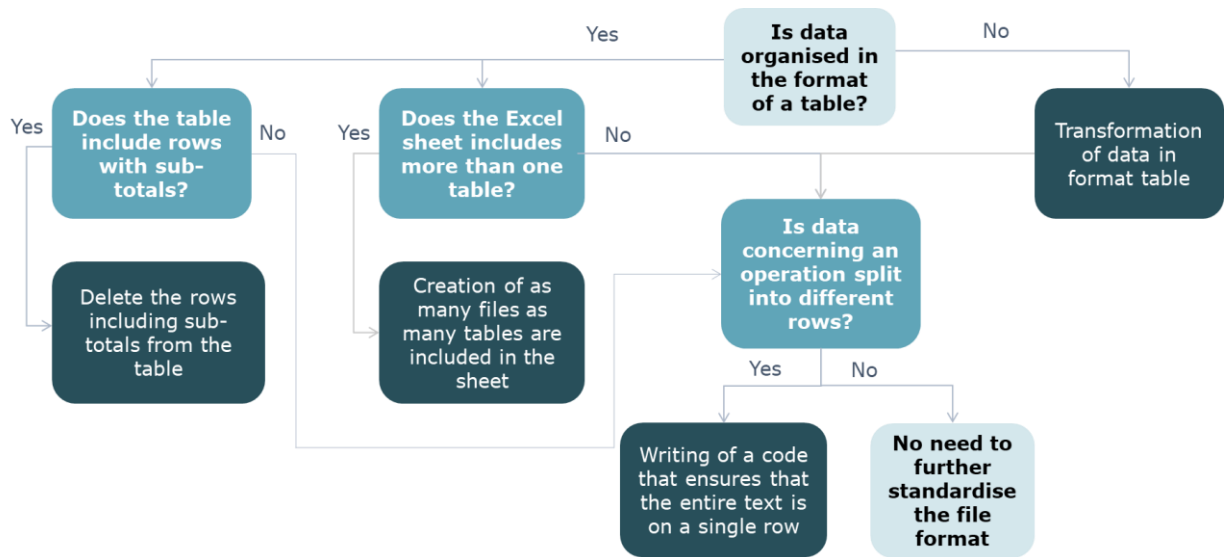
Standardisation of the files format

Although the MAs were asked to provide data extractions preferably in .csv, .xlsx or .xls format, there have been cases in which data was provided in other formats, such as in PDF, MS Word. Moreover, there have also been cases where even if the files were provided in .csv, .xlsx or .xls format:

- They were not directly usable as they were not in the format of a table;
- The information related to one operation was split into more than one row;
- They included multiple tables in the same sheet;
- They included rows indicating sub-totals.
- The following flow chart shows the procedures implemented to solve these specific issues.

¹⁴ Regulation (EU) 2016/679

Figure 3: Work flow to standardise file formats



Source: Authors

Furthermore, in order to cope with the high number of files received, the different datasets provided by each MA were merged or split to eventually obtain two subsets of data and variables, respectively, at the operation and beneficiary level.

Harmonisation of the variable names

A manual harmonisation of the original variable names has been performed to allow the integration of similar information coming from different monitoring systems. To this end, a list of standardised variables' names has been created according to the STATA requirements¹⁵, and each relevant variable included in the original datasets provided by the MA has been renamed accordingly. This activity has been carried out in MS Excel to keep track of all manipulations made to the data. Those variables falling outside the scope of interest were discarded. An example of how this procedure was carried out is presented in the Figure below.

¹⁵ For instance, the variable "Operation name" has been encoded as "prj_name" and a similar approach will be applied for all the other variables of the single database. See the correspondence between the two types of variable name in the READ ME file of the single database.

Figure 4: Example of the variable naming procedure for a French OP

DB n° (each OP)	File name original	Ops	Variable_original	Variable_final
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	Bénéficiaire	benef_name_In
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	Bénéficiaire Privé / Public	benef_ownership_local_In
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	TVA récupérable	benef_vat_regime_In
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	Cf. Onglet Bénéficiaires	/
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	No de l'opération	prj_nr
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	Libellé de l'opération	prj_name_In
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	Description succincte de l'opération	prj_descr_In
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	Statut de l'opération	/
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	Date de la demande de subvention	/
1	France_Bretagne_FEDER_Enquête CE_v2_2021-03 / CE_Étude FEDER 14-20	Bretagne OP - 2014FR16M2OP003	Date de début d'éligibilité	prj_start_date

Source: Authors

Harmonisation of the unit of analysis

In line with the broad definition of operations adopted by Regulation (EU) No 1303/2013, data at the operation level provided by MAs may refer to a project, a project component, a group of projects (i.e., Action Plans, investment strategies, voucher schemes, State aid schemes, etc.) or a financial instrument. Moreover, for some OPs/CPs, financial data was disaggregated by invoice. While in the latter case, the data were aggregated at the operation level exploiting the operation official identifier, the aggregation of different self-standing project components, as well as the disaggregation of groups of projects to provide a database entirely at the project level, was not feasible neither advisable for two main reasons:

1. The harmonisation procedures would have prevented the linkage between financial data and output indicators since the indicators database is at operation level according to the definition adopted by each MA;
2. Except for the specific case of two German OPs¹⁶ for which the MA clearly indicated those operations being components of a project¹⁷, the MAs did not provide precise indications to detect operations having the nature of a project component¹⁸. This identification, therefore, would not be fully automatic and would require manual checks by the Core Team.

Nevertheless, an attempt of harmonisation of the unit of analysis was made by:

- **Collecting data on the final recipients** of groups of projects and of financial instruments, whenever possible.

¹⁶ 2014DE16RFOP009 and 2014DE16RFOP015.

¹⁷ In these cases a higher-level code was added allowing the aggregation of these operations

¹⁸ Some MAs provided a higher-level code beyond the operation unique identifier (i.e., for the Italian programmes, the Wallonia OP), but the meaning was not univocal. For example in the OP Wallonia, the higher-level title in some cases allows the aggregation of components of the same project, but in some others it only aggregates different projects with a common strategic objective. Some other MAs provided more general indications on the criteria that can be used to identify operations being components of larger project (e.g., common operation title, same implementation dates, etc.). This is the case of most of the French MAs that confirmed that operations may be projects as well as components of complex/collaborative projects and that they may be aggregated on the basis on the operation name, description and dates. This also applies to some German OPs, for which the MAs confirmed that the operation name and dates could be used as a basis to identify operations being components of the same projects.

- **Including a flag** at the operation level that specifies whether the operation under analysis is a project, a project component, a group of projects or a financial instrument. The attribution of the flag relied on:
 - The indications provided by the MA;
 - Variables included in the original raw data providing both a higher-level unique identifier and operation unique identifiers, which signalled the presence of operations to be aggregated;
 - Categorical variables, e.g., the form of finance as well as local taxonomies, allowing the identification of financial instruments;
 - The support of the country experts who were asked to identify the nature of operations, if not straightforward.

This flag has always been attributed, but it combines more than one definition in all the cases when precise information on the nature of the operation was not available.

2.2.2.2. Procedures to harmonise the data format

To allow the integration of similar information, their comparability and also further data cleaning activities, additional harmonisation procedures have been implemented to standardise the format of some variables:

- **Translation of text variables into English:** this activity has been key to allowing the comparability of the information provided in the Single Database¹⁹. It also eased subsequent data cleaning and harmonisation procedures, such as the identification and use of common taxonomies for variables such as the operation status and the beneficiary role, as well as the attribution of the category of intervention as per Annex I of the Regulation (EU) 215/2014. An ad-hoc algorithm was developed in Python to automatically translate the dataset on operations and the one on beneficiaries on Google Translate. In case of monolingual OPs/CPs, this algorithm exploited the information included in the variable “Language” (see Section 3.2.1) to identify the source language. In case of multilingual OPs/CPs (i.e., operation name in English and operation description in French) instead, an additional script code was written to first detect the language and then translate the text in English, even though this second approach may have some limitations²⁰. Although the translations are not always perfect due to a number of reasons, e.g., bugs in the algorithm (see Chapter 4), it should be noted that in many cases, through this step, typos included the original text string were fixed.
- **Conversion of financial amounts in the local currency to Euros:** Financial amounts in the final Single Database were entered in EUR. If financial amounts were originally reported in another currency, the average exchange rate in the 2014-2020 period has been computed using the time series of the average yearly exchange rates between the national currency and the Euro (source: the ECB website <http://sdw.ecb.europa.eu/browse.do?node=9691296>). The latter has then been used to convert the amounts from the national currency into Euro terms.
- **Standardisation of the data format of similar variables,** such as dates, standard categories of interventions or the VAT code. For dates, the same format has been used, namely DD/MM/YYYY, while standard categories of interventions have been

¹⁹ Among the variables translated there are: operation name, operation summary, Priority Axis title, Specific Objective title, OP measure/action title, common and specific output indicators names.

²⁰ For instance, the same word can be present in the vocabulary of more than one language and therefore have more than a univocal meaning.

encoded as they are listed in Annex I of the Regulation (EU) 215/2014. The VAT codes associated with each beneficiary has been harmonised as far as possible using the standard format used for EU VAT (e.g., no spaces, the country code is at the beginning, etc.).

- **Standard reporting of multiple sets of information:** operations and beneficiaries may be located in more than one region, or beneficiaries may have been involved in more than one H2020 project, and so on.²¹ Likewise, operations may have been assigned to more than one investment priority, more than one field of intervention and so on. In these cases, such information on operations and/or beneficiaries has been reported in one single row with multiple entries in the corresponding cell, separated by “|”.

2.2.2.3. Procedures to clean, harmonise and enrich the data content

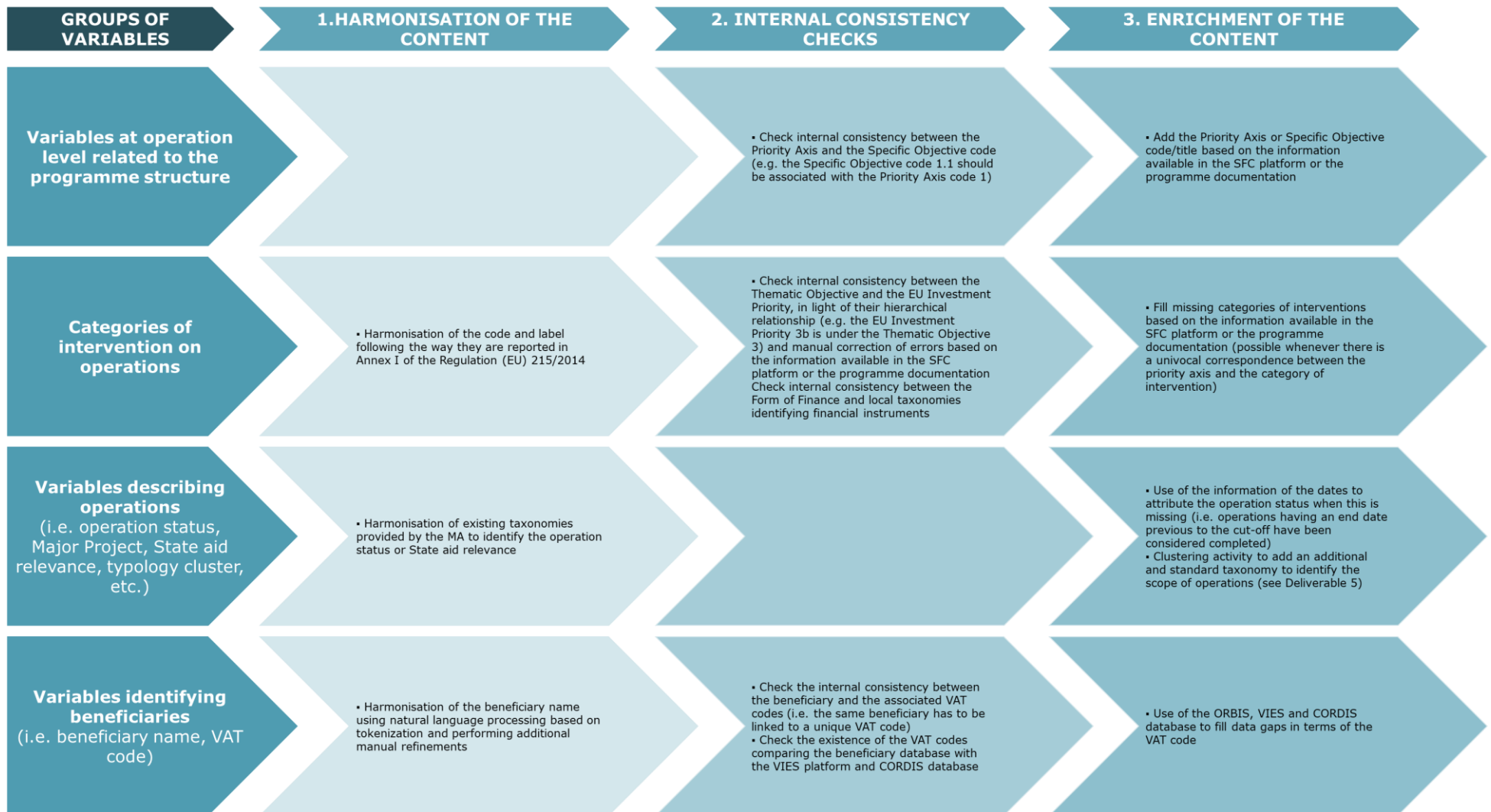
A combination of several data cleaning, harmonisation and enrichment activities has been performed to ensure that each variable was as reliable and complete as possible:

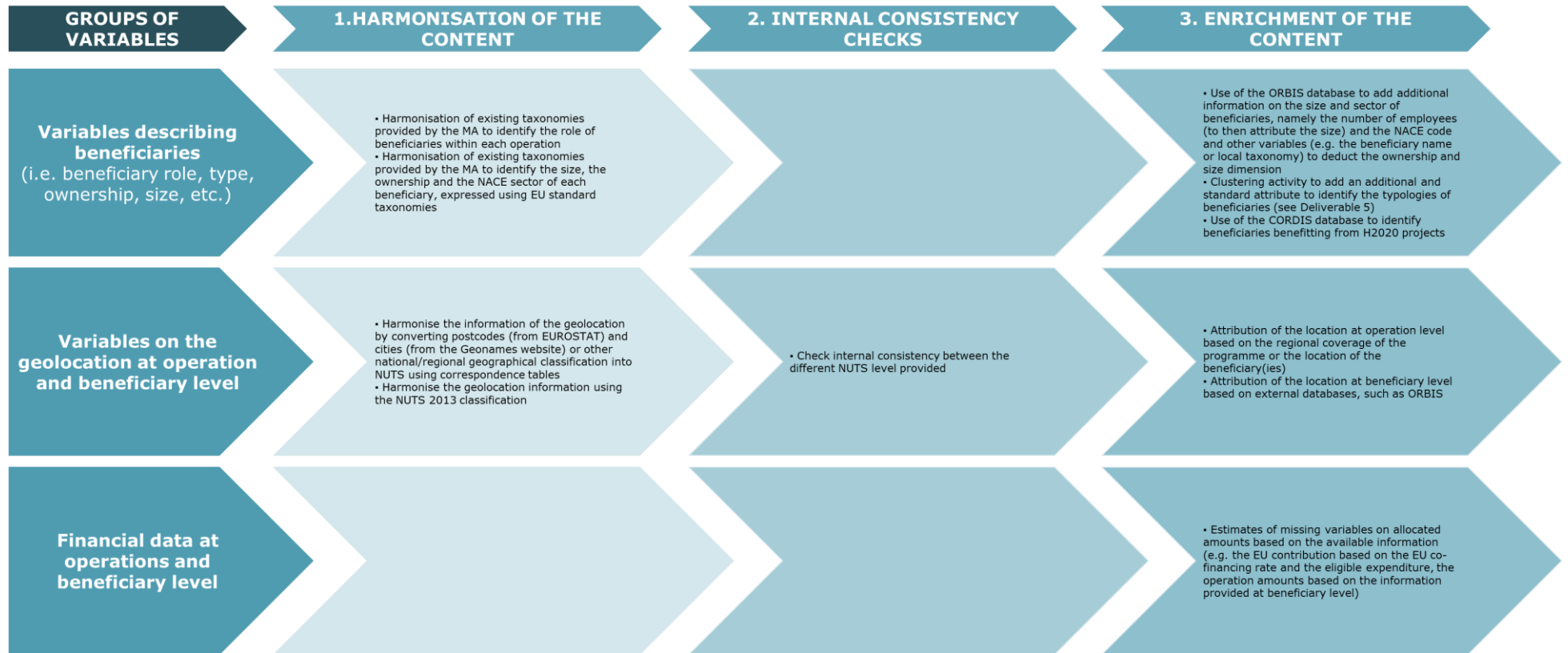
- **Harmonisation of the content:** this activity was necessary for those variables including standard categories of interventions or providing common taxonomies to describe operations and beneficiaries in order to ensure the comparability of the same variables across programmes;
- **Internal consistency checks:** they were performed aiming to detect and correct errors and potential inconsistencies between different sets of information;
- **Enrichment of the data content:** this activity was necessary to fill data gaps of key variables and to enrich the Single Database with additional variables describing operations and beneficiaries by exploiting the information available in external databases or documents and through the clustering exercises.

The combination of such activities varied across groups of variables and is briefly presented in the table below. Additional information regarding the activities put in place to clean, harmonise and fill the gaps of specific variables are instead provided in the next Chapter.

²¹ The H2020 projects in which a beneficiary participated is expected to be included in the single database. This information will be derived from the CORDIS database, as explained in the strategy for database enrichment described in Section 2.2.3.

Figure 5: Combination of data cleaning, harmonisation and enrichment procedures across different groups of variables





Source: Authors

3. Presentation of the Database of Operations and Beneficiaries

3.1. Coverage of the Database of Operations and Beneficiaries

This report is accompanied by two .csv files: the Database of Operations and the Database of Beneficiaries. A third part, including data on output indicators at the operation level, is included in a separate deliverable (Deliverable 7).

The databases on operations and beneficiaries cover 215 OPs out of 217 and 73 CPs out of 76 (for a total of **288 programmes out of 293**, i.e., 98%). Overall, the main data source is the data extracted from the monitoring systems. Only for a small number of programmes (11 OPs and 5 CPs), the data source was the public list of operations either because the MA could not provide a data extraction from the monitoring system²² or because the data provided could not be integrated into the Single Database²³. A total of 5 programmes (3 CPs and 2 OPs) could not be included at this stage in the Single Database because of:

- **The list of operations is not publicly available on the MA website.** This issue concerned the Interreg V-A - Saint Martin-Sint Marten, for which there is no public list of operations on the website of the MA.
- **The impossibility to integrate the data with the Single Database.** This was due to peculiarities of the programme, as in the case of the INTERACT programme, or to data protection issues, as in the case of the programmes managed by the Welsh European Funding Office²⁴. Regarding the INTERACT programme, this is a peculiar CP managed by different MAs. It implements training, seminars, tools and provides support to other programmes (e.g., it allowed the creation of the Keep.eu Portal). As it does not implement projects within the usual ETC sense, this programme has been excluded from the Single Database. Regarding the Welsh data, the privacy policy adopted by the Managing Authority did not allow public disclosure of data. For this reason, although data has been cleaned, harmonised and enriched, they have been finally excluded from the Single Database.

The list of variables included in the Database of Operations and the Database of Beneficiaries has been identified on the basis of the following criteria:

- **Extensive coverage of the variable** in terms of both number operations and number of programmes;
- **Relevance of the information for the clustering of operations and beneficiaries into typologies**, as well as for the future thematic ex-post evaluations;
- **Possibility to harmonise the data across the Member States** and reconcile the information under harmonised labels;

²² The following programmes are concerned: the Austrian national OP, the French Technical Assistance OP, the England and Scotland OP and the seven Hungarian OPs.

²³ This is the case of the ESPON cooperation programme for which the Managing Authority provided data at OP level as well as of the Interreg V-B - Adriatic-Ionian for which data were provided in SQL format and could not be extracted. In both cases, the public list of operations provided in the Keep.eu portal was used.

²⁴ 2014UK16RFOP006, 2014UK16RFOP005, 2014TC16RFCB048.

- **Feasibility of collecting data** from other sources, in case the attribute was not included in the datasets provided by the MAs or collected from the public list of operations.

Additional variables currently excluded from the two databases can be, in any case, retrieved from the raw data provided by the MAs. Indeed, all data provided by the MAs will be made available to the Commission and so to future evaluators. Moreover, **additional .csv files are provided together with the Single Database**, including the data excluded during data cleaning and harmonisation (see Annex V for more details), such as:

- Variables that were initially retained after the preliminary screening of the datasets sent by MAs (see Section 2.2.1) but which have then been excluded because of the low coverage in terms of programmes, operations and beneficiaries;
- Financial variables that have been excluded because of the impossibility to harmonise and reconcile them in a coherent and consistent way²⁵.
- Data on non-funded operations, when available: such operations were not relevant in the context of the Single Database, but they might be useful for future evaluators.

In what follows, the structure of the two databases is presented. Specifically, a brief description of the different parts composing each database is provided, along with a more detailed illustration of the specific data cleaning, harmonisation and enrichment procedures that have been implemented for each variable.

3.2. Structure and content of the Database of Operations and the Database of Beneficiaries

3.2.1. The structure of the Database of Operations

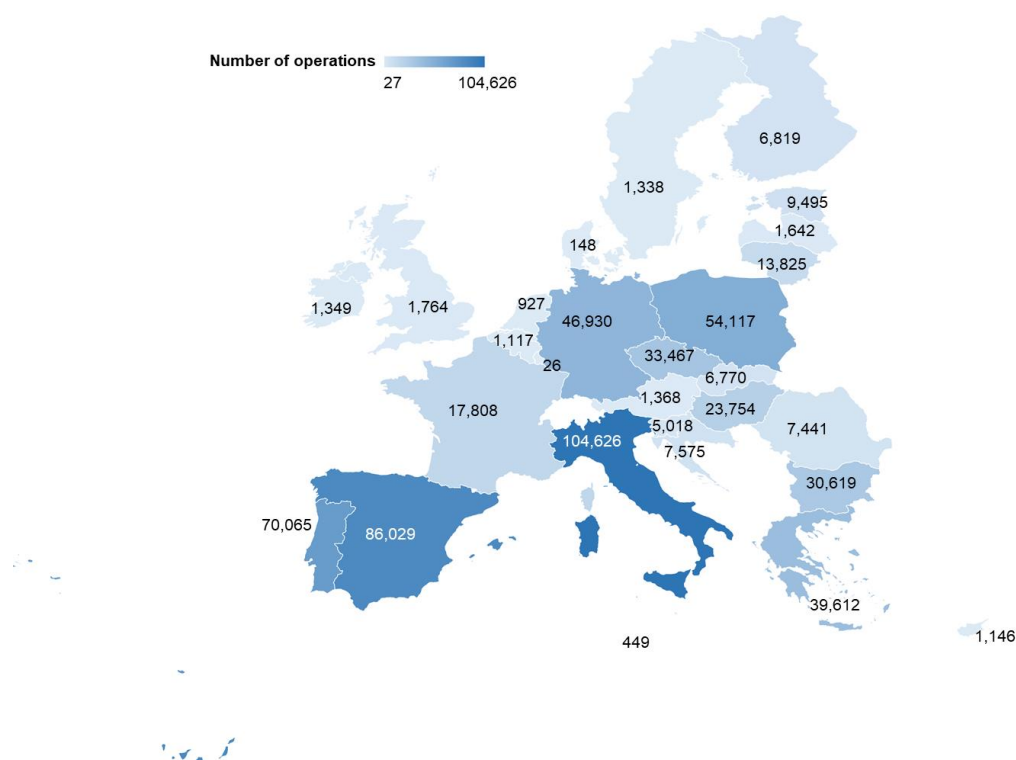
This database (DB_Operations) includes descriptive, categorical, and financial information on the ERDF and CF operations. Each row corresponds to one operation, according to the definition adopted by the MA²⁶.

Although there is high variability in the way MAs define operations, and the latter may therefore have a different nature even within the same programme, an analysis of the distribution of operations across Member States (and the UK) and programmes is provided in what follows. In total, **the database includes data for 584,828 operations**, mostly concentrated in Italy, Spain, Portugal and Poland.

²⁵ As mentioned in Chapter 4, such harmonisation could be attempted at a later stage.

²⁶ In some programmes, data have been provided disaggregated at sub-operation level (i.e., Greece, Malta) or into tranches of payments. In these cases, the Core Team directly re-aggregated the data at operation level.

Figure 6: Total number of operations by Member State



Source: Authors

Note: the Single Database of operations also includes 73 CPs, covering a total of 9,584 operations

In the database, the five OPs with the largest number of operations are:

- The Spanish Multiregional OP with 49,313 operations;
- The Portuguese Competitiveness and Internationalisation OP with 38,510 operations;
- The Bulgarian Innovations and Competitiveness OP with 29,468 operations;
- The Italian Education OP with 25,756 operations;
- The Greek Competitiveness, entrepreneurship and innovation OP with 25,009 operations.

SME Initiative OPs and Cooperation Programmes generally have the lowest number of operations, such as the Interreg V-B – Amazonia (18 operations) and the Interreg V-A - Mayotte-Comores-Madagascar (13 operations).

On average, each programme funded about 2,030 operations. In Eastern and Central Eastern countries such as Latvia, Estonia and Croatia, the average number of operations per OP is the highest (respectively, 14,125; 9,495; 7,575).

The database of operations is structured into six parts:

1. General information
2. Operational/Cooperation programme identification
3. Operation identification
4. Operation's features
5. Operation's categorisation

6. Operation's financial information

In what follows, the different variables are defined and presented. For each variable, the following information are provided:

- Information on the data format;
- Information on the data content resulting from the data cleaning, harmonisation and enrichment procedures and the related data sources;
- Information on whether the variable still has some limitations, despite the data cleaning, harmonisation and enrichment activities carried out; in this respect, a four-point scale was used to make the assessment:

There are no data issues concerning the variable	There are some data issues, but their relevance is low	There are some data issues, but their relevance is medium	There are some data issues and their relevance is high
--	--	---	--

- Information on the coverage in terms of the number of programmes (see the column “OP”) and operations (see the column “Operation”) is provided²⁷. A five-point scale was used to make the assessment:

Very low degree of coverage (less than 20% of the programmes and operations)	Low degree of coverage (between 20 and 40% of the programmes and operations)	Medium degree of coverage (between 40% and 60% of the programmes and operations)	High degree of coverage (between 60% and 80% of the programmes and operations)	Very high degree of coverage (more than 80% of the programmes and operations)
--	--	--	--	---

²⁷ This information will be updated with the final release of the database, at the end of the study.

Part 1. General information

This first part of the database includes some general information on the data collected for each operation. It provides information on the original language in which variables have been provided and the currency used for reporting financial variables, as well as the cut-off date of the data provided and the date of the data extraction.

Data	Data format	Data content and source(s)	Data issue	Data coverage %	
				Operation	OP
Language (data_language)	ISO 639-1	This variable includes information on the language used in the raw data provided by the MA or in the public list of operations collected by the Core Team. It has been used as a basis for the translation exercise in the case of programmes reporting information in only one language. In the case of multilingual programmes (i.e., operation name in English and operation description in French), all the relevant languages have been reported, separated by “ ”, but the variable has not been used as a basis for the translation activity. The Core Team detected and directly attributed the language for each programme after the preliminary revision of the datasets provided by the MA or collected based on the public list of operations.	No	100%	100%
Currency (data_currency)	ISO 4217	This variable includes information on the currency used in the raw data sent by the MA or in the public list of operations collected by the Core Team. It has been used to convert financial variables provided in local currencies into Euros (see Section 2.2.2 for more details on the conversion procedure). The currency was provided by the MA or attributed by the Core Team based on the indications provided by the country expert or the MA.	No	100%	100%
Cut-off date (date_update)	Date	This is the date of update of the data provided by the MA or downloaded from the public list of operations. The variable is the result of the manual attribution of the Core Team based on the information provided by the MA. When the MA did not specify the date, the attribution was made by the Core Team based on the date of the data delivery (it is, therefore, equal to the date of the data extraction) or on the date of update reported in the public list (when the latter was used).	Yes ²⁸	99.65%	97.22%
Date of data extraction (date_extraction)	Date	This is the date of extraction of the data provided by the MA. In the case the public list of operations was used as the main data source, it corresponds to their cut-off date as reported on the MA’s website. The variable is the result of the manual attribution of the Core Team based on the information provided by the MA. When the MA did not specify the date, the attribution was based on the date of the data delivery (it is, therefore, equal to the cut-off date).	Yes ²⁹	99.91%	98.26%

²⁸ Despite the initial ambition to collect data with a cut-off date as of 31/12/2020, the database includes operations updated at different dates. This reflects the different features of monitoring systems across the EU: while in some regions/countries it is possible to filter and extract only operations selected up to a certain date, in other monitoring systems there is not this possibility. Moreover, when this information was not indicated by the MA, the data included in the single database may not correspond to the actual cut-off date.

²⁹ When the information was not provided or clarified by the MA, the date included in the database may not correspond to the actual date of data extraction.

Part 2. Operational/Cooperation programme identification

This second part of the database identifies the OP/CP under which the operation is funded and provides relevant information in terms of its coverage and territorial scope, as well as on the country where the MA responsible for its implementation is located and the EU funds it covers.

Data	Data format	Data content and source(s)	Data issue	Data coverage %	
				Operation	OP
CCI number (op)	Alphanumeric	This is the Operational/Cooperation Programme code. The CCI number was either already included in the data collected or manually attributed by the Core Team based on the scattered information provided by the MA or other variables included in the raw data.	No	100%	100%
OP short title (English) (op_short_title)	Text	This is the short title of the Operational/Cooperation Programme in English. The Core Team automatically attributed the short title of the OP based on the CCI number.	No	100%	100%
OP territorial scope (op_terr_scope)	Text	This variable provides information on the territorial scope of the OP or CP under which the operation is funded. For OPs, the territorial scope may be 'National' or 'Regional', while CPs can be 'Crossborder', 'Transnational' or 'Interregional'. The information was retrieved from the SFC platform based on the CCI number.	No	100%	100%
OP country(ies) (op_country)	ISO 639-1	This variable provides information on the country covered by the programme under which the operation is funded. In the case of Cooperation Programmes, the variable includes the acronym "TC", which stands for Territorial Cooperation. The information was retrieved from the Regio ATLAS platform ³⁰ based on the CCI number.	No	100%	100%

³⁰ https://ec.europa.eu/regional_policy/en/atlas/programmes

OP Member State(s) (op_EU28_MS_covered)	ISO 639-1	This is the list of EU Member State(s) covered by the programme under which the operation is funded. In the case of Cooperation Programmes, the complete list of MS covered has been reported separated by " ". The information was retrieved from the Regio ATLAS platform ³¹ and the programme documentation based on the CCI number.	No	100%	100%
OP MA (op_ma)	ISO 639-1	This variable reports the country code of the MA responsible for the design and implementation of the OP or CP under which the operation is funded. The information was retrieved from the Regio ATLAS platform based on the CCI number.	No	100%	100%
OP EU fund (op_fund)	Text	This is the EU fund(s) under which the OP or CP related to the operation is funded. In the case of multi-fund programmes, the complete list of EU funds covered has been reported, separated by " ". The information was retrieved from the SFC platform based on the CCI number.	No	100%	100%

Part 3. Operation identification

This third part allows the identification of the operation on the basis of the unique identifier, name, description, related call for proposals, EU fund, Priority Axis, Specific Objective and OP measure under which it is funded and the date of start and end of its implementation.

Data	Data format	Data content	Data issue	Data coverage %	
				Operation	OP

³¹ https://ec.europa.eu/regional_policy/en/atlas/programmes

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Official operation identifier (prj_nr)	Text, numeric, alphanumeric	In general, this is an (alpha)numeric string allowing the distinction between different operations in the national and regional monitoring systems, thus taking different formats depending on the Member State and/or region. In combination with the CCI code, this code allowed the link of the database of operations with the output indicators database. This code is available when already included in the raw data provided by MAs or in the public list of operations. There are cases in which the official operation identification code is missing, and the Core Team has assigned only an ad-hoc operation identification code, i.e., prj_ID and prj_row_ID (e.g., some of the programmes for which only the public list of operations could be collected ³²).	Yes ³³	99.42%	97.57%
Operation name (original language) (prj_name_nl)	Text	This is the name of the operation as mentioned in the national language of the Member State or in the language used to record information in the monitoring system. The operation name is available when already included in the raw data provided by MAs or in the public list of operations. For some operations and programmes, this information is missing. In some of these cases, at least an operation acronym was available.	Yes ³⁴	99.50%	96.88%
Operation name (English) (prj_name)	Text	This is the name of the operation translated into English. The operation name in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2.2).	Yes ³⁵	99.50%	96.88%
Operation acronym (prj_acronym)	Text	This is a short name identifying operations. It is available only for some operations, especially those funded by CPs.	No	0.78%	14.93%

³² 2014AT16RFOP001, 2014FR16M0OP011, 2014FR16M2TA001, 2014UK16RFOP001, 2014UK16RFOP004.

³³ The same code may be used across programmes and, therefore, it may not identify univocal operations. Therefore, it cannot be used for univocally identifying an operation across the single database.

³⁴ In some cases, it corresponds to the project acronym, in others it corresponds to the name of the beneficiary, in other cases the information is missing.

³⁵ In some cases, the translation package has not worked properly and the variable in English may still contain the name of the operation in the original language.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Ad-hoc operation identifier (prj_ID)	Alphanumeric	<p>This is an alphanumeric code attributed by the Core Team by numbering the total number of operations of each Operational/Cooperation Programme in ascending order. It allows the link with the beneficiaries database. In general, there is one code for each row. However, there are some exceptions to this rule:</p> <ul style="list-style-type: none"> • <u>The Spanish SMEi OP (2014ES16RFSM001)</u>: in the data extracted by the MA, there were six different operations under the OP (each with a different official operation identifier), but the list of final recipients was provided at the OP level. This is why a unique ad-hoc operation identifier has been assigned to all operations; • <u>Some Slovak OPs</u>: in some cases, the MA provided a unique list of the final recipients covering different operations. This is why, also, in this case, a unique ad-hoc operation identifier has been assigned to all operations. 	Yes ³⁶	100%	100%
Ad-hoc row identifier (prj_row_ID)	Alphanumeric	<p>This is an alphanumeric code attributed by the Core Team by numbering the total number of rows of each operation and Operational/Cooperation Programme in ascending order. It has been included to identify rows in a univocal way (which is not possible based on the ad-hoc operation identifier) to therefore allow the link with the indicators database (which will be delivered at a later stage).</p>	No	100%	100%
Operation definition (prj_definition)	Text	<p>This is the label indicating the definition of operation adopted by each MA. In general, an operation may correspond to a:</p> <ul style="list-style-type: none"> • Project component: in this case, a higher-level code or title (see the variables “<i>portfolio_nr</i>” and “<i>portfolio_name</i>”), if available, allows the aggregation of the different components at project level; • Project: in this case, no additional information is necessary; • Group of projects: they may be Action Plans, investment strategies, vouchers, or other State aid schemes whose direct beneficiary in the data extracted by the MA or available in the public list of operations is an intermediate body transferring EU funds to other entities; in these cases, the information at project level is available in the database of beneficiaries only for those operations for which the MA provided the list of final recipients; • Financial instrument: these are peculiar operations whose mode of delivery is the support through financial instruments, such as venture and equity capital, loan, guarantee, interest rate subsidy, guarantee fee subsidy, technical support or equivalent; • Project of financial instrument: these are the projects funded through financial instruments. <p>As described in Section 2.2.2.1, the Core Team manually attributed the label based on the information provided by the MA/country expert. It should be noted that financial instruments were identified, in general, by using the available information on the form of finance and comparing it with local taxonomies of operations, when available. All those operations falling under the form of finance ‘03 - Support through financial instruments: venture and equity capital or equivalent’, ‘04 - Support</p>	Yes ³⁸	100%	100%

³⁶ It is not univocal for each row and it cannot be used to count the number of operations.

³⁸ It has not always been possible to clearly attribute a unique label to each operation. In fact, in most cases, Managing Authorities provided only general indications on the definition of operation adopted, simply indicating that the list may include either project components, projects or groups of projects. This is why, in some cases, all the possible labels, separated by “|”, have been reported under some OPs/CPs. Further refinements could be made in the context of future ex-post thematic evaluations, as suggested in Chapter 4.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

		through financial instruments: loan or equivalent', '05 - Support through financial instruments: guarantee or equivalent' or '06 - Support through financial instruments: interest rate subsidy, guarantee fee subsidy, technical support or equivalent' were identified as 'financial instruments' in all programmes, except for the Italian and Portuguese OPs. For these programmes, another variable providing a local taxonomy of operations was used to identify these specific operations, in light of the limitations of the variable on the form of finance (see below). In other specific cases, it has also been possible to distinguish between financial instruments and projects funded by financial instruments. This is the case of some German OPs, for which the Core Team noticed that there was a mismatch between the form of finance and the local definition of operation. By comparing the available information with the data on financial instruments available on the Cohesion Data Platform ³⁷ , it has been possible to identify such operations as projects funded by financial instruments directly managed by the MA.			
Official portfolio identifier (portfolio_nr)	Text, numeric, alphanumeric	This is an (alpha)numeric or text string allowing the distinction between different 'groups' of operations in the national and regional monitoring systems, thus taking different formats depending on the Member State and/or region. Such code may allow, if available, the aggregation of operations being project components under the same identification code. This code is available when it was already included in the raw data provided by MAs or in the public list of operations. In particular, it is available for all Italian programmes, where all operations are assigned with a higher-level code identifying the project or groups of projects under which the operation is funded, as well as in the Lithuanian and Cypriot OPs and in some German OPs.	Yes ³⁹	22.18%	17.71%
Portfolio name (original language) (portfolio_name_nl)	Text	This is the name given to groups of operations sharing similar features and objectives (the so-called "portfolio") as the MA has provided it. This is available only for those operations grouping more than one project or being project components of the same project. In particular, it is available for the Italian programmes, the national Cypriot OP and the OP Wallonia.	Yes ⁴⁰	0.48%	5.90%

³⁷ <https://cohesiondata.ec.europa.eu/2014-2020-Finances/ESIF-2014-2020-Financial-Instruments-Implementatio/dcsc-7x87>

³⁹ There is heterogeneity across and within MS. For the Italian programmes, this code cannot be used to identify projects covering several operations since, in some cases which could not be identified, the code simply aggregates operations which are related or that are groups of projects. Instead, in the other programmes for which this code is available (e.g., 2014CY16M1OP001, 2014DE16RFOP009, 2014DE16RFOP015, 2014LT16MAOP001), the code could be used for aggregating project components at project level.

⁴⁰ It cannot always identify operations that are components of the same projects; in some cases, it links different projects being related or having similar objectives, such as in the Italian programmes and the Wallonia OP.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Portfolio name (English) (portfolio_name)	Text	This is the name given to groups of operations sharing similar features and objectives translated into English. The portfolio name in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2).	Yes ⁴¹	0.48%	6.25%
Portfolio definition (portfolio_definition)	Text	This is the label indicating whether the official portfolio code or name identifies a: <ul style="list-style-type: none"> • Project: in this case, the code or name can be used to aggregate operations being components of a single project; • Group of projects: in this case, the code or name is irrelevant. The label is the result of the manual attribution of the Core Team based on the information provided by the MA/country expert, as described in Section 2.2.2.1 ⁴² .	Yes ⁴³	20.65%	12.15%
Call for proposals (original language) (prj_call_code_title_n)	Text	This is the title and/or code of the call for proposals under which the operation has been funded. This information is available when it was already included in the raw data provided by MAs. ⁴⁴	No	48.86%	43.40%

⁴¹ In some cases, the translation package has not worked properly and the variable in English may still contain the name of the portfolio in the original language.

⁴² It should be noted that for 90 operations of the Cypriot OP (2014CY16M1OP001), the operation definition corresponds to the portfolio definition. This is due to the fact that an official portfolio identifier was available for all operations, also for those that can be defined as projects.

⁴³ It has not always been possible to clearly attribute a unique label to each group of operations. In fact, in some cases, such as in the case of Italy or of the Wallonia OP, Managing Authorities provided only general indications on the definition of operation adopted, simply indicating that the higher-level code and/or name may aggregate either project components or projects which share a common objective. This is why, in these cases, all the possible labels, separated by “|”, have been reported under some OPs/CPs. Further refinements could be in the context of future ex-post thematic evaluations, as suggested in Chapter 4.

⁴⁴ Notice that in the case of the Italian programmes, this information is missing for some operations, although it was available in the raw data provided by the MA. In fact, there were some operations linked to more than one call for proposals. In light of the impossibility to check with the competent MA the reason behind it and whether the information was correct or not, the latter has been deleted in accordance with the central monitoring system implementation body.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Call for proposals (English) (prj_call_code_title)	Text	This variable includes the title of the call for proposals translated into English. The call for proposals in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2.2.	No	68.57%	71.53%
Operation summary (original language) (prj_descr_nl)	Text	This variable includes the description of the scope and objectives of the operation, as mentioned in the original list of operations. This information is available when it was already included in the raw data provided by MAs.	Yes ⁴⁵	80.93%	80.90%
Operation summary (English) (prj_descr)	Text	This variable includes the description of the scope and objectives of the operation translated into English. The operation summary in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2.2.	Yes ⁴⁶	82.97%	89.58%
Operation EU fund (prj_fund_type)	Text	This variable includes the information on the type of co-financing fund under which the operation is funded, which can be: <ul style="list-style-type: none"> • The ERDF; • The Cohesion Fund; or • A combination of the two. This information is available when already included in the raw data provided by MAs. When the information was missing, it was manually attributed by the Core Team only in the case of mono-fund programmes (e.g., CPs or OPs funded only by the ERDF or the CF).	Yes ⁴⁷	100.00%	100.00%

⁴⁵ In some cases, it corresponds to the project name, in others it corresponds to the name of the beneficiary, in other cases the information is missing.

⁴⁶ In some cases, the translation package has not worked properly and the variable in English may still contain the name of the operation in the original language.

⁴⁷ For the French OP Technical Assistance, it has not been possible to identify the fund. In the public list of operations used, each operation seems to be under both the ERDF-ESF fund.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Operation other EU fund (prj_other_fund_type)	Text	<p>This variable includes information on other EU funds covering the operation, other than the ERDF or the Cohesion Fund. In fact, there are also cases in which other EU funds were mobilised in combination with the ERDF or CF such as the IPA funding or the ESF.</p> <p>This information is available only when already included in the raw data provided by MAs.</p>	No	0.05%	1.74%
Priority Axis code (prj_priority_axis_code)	(Alpha)numeric	<p>This is the code of the Priority Axis, i.e., a major priority of the OP strategy, under which the operation is funded, as mentioned in the programme. In the case of multiple Priority Axes, the codes have been reported separated “ ”.</p> <p>The information on the Priority Axis code has generally been retrieved from the raw data provided by the MA or the public list of operations. In case both the code and the title were missing but there was the information on the Specific Objective, the latter was used to extract the Priority Axis code.</p> <p>Whenever only the title was available, the code was filled by matching the Single Database with the information available on the SFC platform or relying on the review of the programme documentation.</p>	No	99.73%	99.31%
Priority Axis name (original language) (prj_priority_axis_title_ni)	Text	<p>This is the title of the Priority Axis, i.e., a major priority of the OP strategy, under which the operation is funded, as mentioned in the programme. In the case of multiple Priority Axes, the titles have been reported separated “ ”.</p> <p>The information on the Priority Axis title has generally been retrieved from the raw data provided by the MA or the public list of operations. In case both the code and the title were missing but there was the information on the Specific Objective, the latter was used to extract the Priority Axis code and to then assign the title, following the procedure mentioned in the previous row. Whenever only the code was available, the title was filled by matching the Single Database with the information available on the SFC platform or relying on the review of the programme documentation.</p>	No	99.73%	99.31%
Priority Axis name (English) (prj_priority_axis_title)	Text	<p>This is the title of the Priority Axis translated into English.</p> <p>The Priority Axis title in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2.2.</p>	Yes ⁴⁸	99.73%	99.31%

⁴⁸ In some cases, the translation package has not worked properly and the variable in English may still contain the Priority Axis title in the original language.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Priority Axis share (prj_paxis_pro_rata)	Numeric	This is the share of expenditure attributed under the Priority Axis(es) of each operation. In the case of multiple Priority Axes, the different shares are reported separated by " ". The shares have been attributed by the Core Team by calculating the share of the eligible expenditure attributed to each Priority Axis over the total eligible expenditure allocated to the operation. In the case of a single Priority Axis, the share was attributed by the Core Team equal to 100%.	No	24.86%	19.79%
Specific Objective code (prj_paxis_objective_cod_e)	(Alpha)numeric	This is the code of the Specific Objective under which the operation is funded, as mentioned in the programme. Under each Priority Axis, the MAs can define one or more Specific Objectives. The information on the Specific Objective code has generally been retrieved from the raw data provided by the MA or the public list of operations. In the this case, both the code and the title were missing, but there was the information on the OP measure; the latter was used to extract the Priority Axis code. In other cases, a national code has been used to extract the code. Whenever only the title was available, the code was filled by matching the Single Database with the information available on the SFC platform or relying on the review of the programme documentation. This has not been systematically done for all programmes, and that further enrichment could be performed at a later stage (see Chapter 4).	Yes ⁴⁹	69.70%	72.57%
Specific Objective name (original language) (prj_paxis_objective_title_nl)	Text	This is the title of the Specific Objective under which the operation is funded, as mentioned in the programme. The information on the Specific Objective title generally has been retrieved from the raw data provided by the MA or the public list of operations. In case both the code and the title were missing, but there was the information on the OP measure, the latter was used to extract the Priority Axis code and to then assign the title, following the procedure mentioned in the previous row. In other cases, a national code has been used to extract the code. Whenever only the code was available, the title was filled in by matching the Single Database with the information available on the SFC platform or relying on the review of the programme documentation. This has not been systematically done for all programmes, and that further enrichment could be performed at a later stage (see Chapter 4).	No	65.77%	65.28%
Specific Objective name (English title) (prj_paxis_objective_title_e)	Text	This is the title of the Specific Objective translated into English. The Specific Objective title in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2.2).	No	59.67%	68.40%

⁴⁹ For some specific programmes, the extraction from the national code has not always worked. For instance, despite the indications provided by the country expert, for some operations under the Slovak OPs, the extraction of the Specific Objective code from the local code which was available has not always worked. As a result, it has not always been possible to assign the Specific Objective code and title to all Slovak operations, although the information was in principle available.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

OP measure code (prj_op_measure_code)	(Alpha)numeric	This is the code of the OP measure under which the operation is funded, as mentioned in the programme. Under each Specific Objective, the MAs can define one or more OP measures. The information on the OP measure code and the title have generally been retrieved from the raw data provided by the MA or the public list of operations. Whenever only the code was available, the title was filled by relying on the review of the programme documentation. However, in this respect, it should be noted that this could not have been systematically done for all programmes and that further enrichment could be performed later (see Chapter 4).	Yes ⁵⁰	37.82%	28.47%
OP measure name (original language) (prj_op_measure_title_e_n)	Text	This is the title of the OP measure under which the operation is funded, as mentioned in the programme. Under each Specific Objective, the MAs can define one or more OP measures. In most cases, the information on the OP measure code and the title have been retrieved from the raw data provided by the MA or the public list of operations. Whenever only the title was available, the code was filled by relying on the review of the programme documentation. This has not been systematically done for all programmes, and that further enrichment could be performed later (see Chapter 4).	No	38.03%	30.21%
OP measure name (English) (prj_op_measure_title)	Text	This is the title of the OP measure translated into English. The OP measure title in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2.2).	No	38.03%	30.21%
OP action code (prj_op_action_code)	(Alpha)numeric	This is the code of the OP action under which the operation is funded, as mentioned in the programme. Under each OP measure, the MAs can define one or more OP actions. The information on the OP action code has been retrieved only from the raw data provided by the MA or the public list of operations.	No	2.26%	2.08%

⁵⁰ For some specific programmes, the extraction from the national code has not always worked. For instance, despite the indications provided by the country expert, for some operations under the Slovak OPs, the extraction of the Specific Objective code from the local code which was available has not always worked. As a result, it has not always been possible to assign the OP measure code and title to all Slovak operations, although the information was in principle available.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

OP action name (original language) (prj_op_action_title_n)	Text	This is the title of the OP action under which the operation is funded, as mentioned in the programme. Under each OP measure, the MAs can define one or more OP actions. The information on the OP action title has been retrieved only from the raw data provided by the MA or the public list of operations.	No	0.50%	1.04%
OP action name (English) (prj_op_action_title)	Text	This is the title of the OP action translated into English. The OP action title in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2.2).	No	0.60%	1.39%
Operation start (prj_start_date)	Date	This variable includes information on the date of the start of the operation's implementation. This information has been retrieved from the raw data provided by the MA or the public list of operations.	Yes ⁵¹	97.83%	97.22%
Operation end (prj_end_date)	Date	This variable includes information on: <ul style="list-style-type: none"> • The actual end of operations whose implementation has already been completed; • The expected end of operations which are still ongoing as of the cut-off date. This information has been retrieved from the raw data provided by the MA or the public list of operations.	Yes ⁵²	98.13%	97.92%

⁵¹ In principle, it should correspond to the actual start of the operation's implementation, but there may be cases in which it corresponds to the administrative start of the operation (e.g., the sign of the contract) according to the different interpretation that the MAs may have given to the data field 8 of Annex III of the Regulation (EU) No 480/2014.

⁵² The information available on the end date varies across monitoring systems. Some distinguish between the expected and the actual end date of operations for both ongoing and completed operations. Others provide a variable mixing both the expected end (for ongoing operations) and the actual end date (for completed operations) and two different variables. Still others provide a general end date without specifying if this was the end date expected ex-ante or the actual/updated end date of operations. Depending on data availability, the Core Team tried to distinguish the actual end of completed operations from the expected end of ongoing operations. However, especially for those programmes for which a general end date was available, it is not certain whether the date reported is indeed the actual or the expected end for completed operations.

Part 4. Operation’s features

This part provides information on the individual operation itself. In particular, it informs on whether the operation is completed or ongoing, it is a major project and/or constitute State aid, is a public-private partnership or a Joint Action Plan (JAP). It also includes information on the number of beneficiaries each operation covers for those operations for which the complete list is available.

Data	Data format	Data content	Data issue	Data coverage %	
				Operation	OP
Operation status (prj_status)	Text	<p>This is the status of the operation at the moment of the cut-off date. In particular, it distinguishes between:</p> <ul style="list-style-type: none"> • Completed operations: those which, as of the cut-off date, have been already implemented and closed; • Ongoing operations: those which, as of the cut-off date, are still under implementation; • To be started operations: those which, as of the cut-off date, have not yet started. <p>In most of the programmes, a variable indicating the operation status was available. Since different labels were used across the Member States and even programmes within the same MS, the label was manually harmonised after its translation into English. This was done through the identification of terms that are synonyms of “ongoing” and “completed”. If the information was not directly available in the monitoring system or its harmonisation was not possible because of the impossibility to reconcile the different labels, the Core Team attributed it on the basis of the variable “Operation end date”. More specifically, it assigned the label “closed” if the end date is before the cut-off date of the data provided. In other cases, the MA indicated that all projects are still ongoing and that no operation has been closed as of the cut-off date. In such cases, the label “ongoing” or “completed” is directly attributed to the operation.</p>	Yes ⁵³	95.64%	94.44%

⁵³ In the case the “Operation status” has been assigned based on the “Operation end date” but the “Operation cut-off date” has the limitation presented under Part 1, the Core Team cannot be sure that the information on the “Operation status” has been correctly attributed.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Major project (prj_mp)	Text	This is a dichotomous variable (“Yes” or “No”) identifying whether the operation is a Major Project. An operation is considered as a “Major Project” during the 2014- 2020 programming period if it matches the criteria defined in Article 100 of the Regulation (EU) No 1303/2013. The information has been retrieved directly from the data extracted from the MA or collected from the public list of operations when it was available (in some cases there was a dichotomous variable, in others the Major Project CCI code and in others a local taxonomy of the operations including the label “Major Project”).	No	73.68%	81.25%
Major project CCI (prj_mp_code)	Alphanumeric	This is the official CCI associated with each Major project. It is available only for those MPs for which the information was included in the raw data provided by the MA.	No	41.03% ⁵⁴	27.08% ⁵⁵
Major Project name (original language) (prj_mp_name_nl)	Text	This is the name of the Major project as mentioned in the national language of the Member State or in the language used to record information in the monitoring system. The Major project name is available when already included in the raw data provided by MAs or in the public list of operations. For some operations and programmes, this information is missing.	No	26.14% ⁵⁶	22.92% ⁵⁷
Major Project name (English) (prj_mp_name)	Text	This is the name of the Major Project translated into English. The Major Project name in English is the result of the translation exercise carried out by the Core Team, as described in Section 2.2.2.2)	Yes ⁵⁸	26.14% ⁵⁹	22.92% ⁶⁰

⁵⁴ For this variable, the percentage is calculated not over the total number of operations but over the total number of Major projects.

⁵⁵ *Ibid.*

⁵⁶ *Ibid.*

⁵⁷ *Ibid.*

⁵⁸ In some cases, the translation package has not worked properly and the variable in English may still contain the name of the Major Project in the original language.

⁵⁹ *Ibid.*

⁶⁰ *Ibid.*

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

State aid (prj_state_aid)	Text	<p>This variable includes information on whether the public support for the operation constitutes State aid in the form of a dichotomous variable (“Yes” or “No”).</p> <p>This information has been retrieved from the data extracted from the MA or collected from the public list of operations in different ways, depending on how the original information on State aid relevance was reported:</p> <ul style="list-style-type: none"> • <u>Case of a dichotomous variable at operation level</u>: the Core Team followed the same procedure illustrate for the Major Project variable; • <u>Case of variables about the State Aid schemes/regulation</u>: when necessary, translations were made while pre-processing the data and correspondence between each label and “yes” or “no” or “n.a.” has been used to attribute the correct label; • <u>Case of variables at the beneficiary level</u>: the Core Team extracted the information (following one of the procedures mentioned above) and attributed it at project level: <ul style="list-style-type: none"> ○ In the case of a single beneficiary: the same label used at beneficiary level was also attributed at operation level; ○ In the case of multiple beneficiaries: <ul style="list-style-type: none"> ▪ If all beneficiaries received State aid, the same label used at beneficiary level was also attributed at operation level; ▪ In case only some beneficiaries received State aid, the label “yes” was attributed at operation level; ▪ If no beneficiary received State aid, the label “no” was attributed at operation level. ▪ If the information was missing for all beneficiaries, the label “n.a.” was attributed at operation level. ▪ In some specific cases, the MA provided indications on whether operations fall under the State aid regulation, and the variable was created by the Core Team⁶¹. 	No	75.20%	75.35%
Public-private partnership (prj_public_private)	Text	<p>This variable includes information on whether the operation is a public-private partnership, in the form of a dichotomous variable (“Yes” or “No”) according to the definition or Article 2(24) of the Regulation (EU) No 1303/2013. This information is available only for those programmes for which the raw data provided by the MA included the data field 17 required by Annex III of Regulation (EU) No 480/2014 (“Information whether the operation is implemented under a public-private-partnership structure”).</p>	No	34.54%	28.13%
Joint Action Plan (prj_jap)	Text	<p>This variable includes information on whether the operation is a Joint Action Plan (JAP) according to the definition laid down in Article 104(1) of the Regulation (EU) No 1303/2013, in the form of a dichotomous variable (“Yes” or “No”).</p> <p>This information is available only for those programmes for which the raw data provided by the MA included the data field 14 required by Annex III of Regulation (EU) No 480/2014 (“Information whether the operation is a joint action plan and CCI”).</p>	No	9.13%	11.11%

⁶¹ This occurred for instance for the Gibraltar OP (the MA indicated that all operations fall under the De Minimis Regulation) or for the Romanian Integrated Regional Programme (the MA indicated that only operations under specific measures were subject to State aid).

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Revenue-generating operation (prj_revenues)	Text	<p>This variable provides information on whether the operation is revenue-generating according to the definition laid down in Art.61(1) of the Regulation (EU) No 1303/2013 of 'net revenue', in the form of a dichotomous variable ("Yes" or "No").</p> <p>The information was retrieved from the raw data provided by the MA. In particular, this was possible when:</p> <ul style="list-style-type: none"> There was a dichotomous variable indicating whether the operation generates net revenues; There was a variable indicating the amount of the net revenue generated either at operation or beneficiary level. 	No	21.16%	19.44%
Number of beneficiaries (prj_n_benef)	Numeric	<p>This variable reports the number of beneficiaries taking part in the operation. It is empty for those operations for which the full list of beneficiaries was not available.</p> <p>This information was extracted from different data sources:</p> <ul style="list-style-type: none"> <u>It was retrieved from the DB Beneficiaries.csv</u> for those operations for which the full list of beneficiaries is available in this dataset; <u>It was attributed based on the aggregated information on the number of beneficiaries provided by the MA:</u> this is the case, in particular, of some financial instruments for which the MA did not provide the full list of final recipients but provided only some aggregated figures on the number of final recipients benefitting from the financial instrument. 	No	34.54% ⁶²	85.42%

⁶² This percentage has been calculated considering the total number of operations, including those for which the variable has been left empty because the full list of beneficiaries was not available.

Part 5. Operation’s categorisation

This part provides information on the categorisation of individual operations, based either on standard categories used at EU level or ad-hoc taxonomies adopted by the Core Team.

Data	Data format	Data content	Data issue	Data coverage %	
				Operation	OP
Thematic Objective(s) (prj_to_code & prj_to_title)	Text	<p>This variable includes the code(s) and label(s) of the Thematic Objective(s), under which the operation is funded, separated by “ ” in case of multiple Thematic Objectives.</p> <p>The main data source of information on the Thematic Objective is the data extracted from the MAs or collected from the public list of operations, cleaned and harmonised following the procedures mentioned in Section 2.2.2.3. In particular, following the internal consistency checks performed between the Thematic Objective and the EU Investment Priority, the information originally available in the raw data has been in some cases corrected or deleted by the Core Team whenever some inconsistencies were found⁶³. More specifically, it was corrected whenever the triangulation of the information available in the SFC platform and the programme documentation related to each Priority Axis confirmed that another Thematic Objective is associated with the Priority Axis. Instead, the information was deleted whenever it was not possible to determine which was the correct Thematic Objective associated with the Priority Axis.</p> <p>If missing, the Core Team attributed the information by merging the Single Database with the information on the SFC platform. More specifically, this was possible whenever the information on the Priority Axis was available, and there was a univocal correspondence between the Priority Axis and the Thematic Objective.</p>	No	99.70%	99.65%
Thematic Objective(s) share (prj_to_pro_rata)	Numeric	<p>This is the share of expenditure attributed under the Thematic Objective(s) under which each operation is funded.</p> <p>In the case of multiple Thematic Objective(s), the different shares are reported separated by “ ”. The shares have been attributed by the Core Team by calculating the share of the eligible expenditure attributed to each Thematic Objective over the total eligible expenditure allocated to the operation. In the case of a single Thematic Objective, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more TOs, but it was not possible to estimate the share of each one.</p>	No	99.68%	99.65%

⁶³ This concerned, for instance, some operations under the following programme: 2014SI16MAOP001, 2014FR16M0OP005, 2014FR16M0OP012, 2014FR16M2OP001 2014FR16M2TA001, 2014FR16M2OP006, 2014FI16M2OP001, 2014LU16RFOP001, most Italian OPs, 2014TC16RFCB001, 2014TC16RFCB015, 2014TC16RFCB002.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

EU Investment Priority(ies) (prj_ip_code & prj_ip_title)	Text	<p>This variable includes the code(s) and label(s) of the EU Investment Priority(ies) under which the operation is funded. The possible EU Investment Priority(ies) are listed in Art.5 of the Regulation (EU) No 1301/2013 for operations funded under the ERDF OPs, in Art.7 of Regulation (EU) NO 1299/2013 for operations funded under ERDF CPs and in Art.4 of the Regulation (EU) No 1300/2013 for operations funded under Cohesion Fund OPs.</p> <p>The main data source of information on the EU Investment Priority is the data extracted from the MAs or collected from a public list of operations, which have been cleaned and harmonised following the procedures mentioned in Section 2.2.2.3. In particular, following the internal consistency checks performed between the Thematic Objective and the EU Investment Priority, the information originally available in the raw data has been in some cases corrected or deleted by the Core Team whenever some inconsistencies were found⁶⁴. More specifically, it was corrected whenever the triangulation of the information available in the SFC platform and the programme documentation related to each Priority Axis confirmed that another Thematic Objective is associated with the Priority Axis. Instead, the information was deleted whenever it was not possible to determine which was the correct Thematic Objective associated with the Priority Axis⁶⁵. When missing, the information was attributed by the Core Team by merging the Single Database with the information on the SFC platform or the programme documentation. More specifically, this was possible whenever the information on the Specific Objective was available.</p>	No	91.80%	98.26%
EU Investment Priority(ies) share (prj_ip_pro_rata)	Numeric	<p>This is the share of expenditure attributed under the EU Investment Priority(ies) of each operation. In case of multiple EU Investment Priorities, the different shares are reported separated by " ". The shares have been attributed by the Core Team by calculating the share of the eligible expenditure attributed to each EU Investment Priority over the total eligible expenditure allocated to the operation. In the case of a single EU Investment Priority, the share was attributed by the Core Team equal to 100%.</p>	No	91.80%	98.26%
Category of region (prj_region_category)	Text	<p>This variable includes the type of the category of region(s) covered by the operation, according to the categories listed in Art. 90(2) of Regulation (EU) No 1303/2013. In case the operation covers different categories of regions, these are reported separated by " ".</p> <p>The main data source of the category of region is the data extracted from the MAs or collected from a public list of operations. When missing, the Core Team attributed the information by merging the Single Database with the information on the SFC platform, using the OP Priority Axis as a basis for the merge.</p>	No	97.56%	84.72%

⁶⁴ This concerned, for instance, some operations under the following programmes: 2014FR16M0OP005, 2014FR16M0OP004, 2014FR16M0OP012, 2014FR16M0OP009, 2014FR16M2TA001, 2014TC16RFCB038, 2014TC16RFIR003, some Spanish and Italian programmes, 2014TC16RFTN003, 2014TC16RFCB046, 2014TC16RFCB001.

⁶⁵ For instance, this happened for some Slovak operations funded under the ERDF and/or Cohesion Fund for which the IPs reported in the monitoring system were those related to the ESF.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Category of region share (prj_reg_cat_pro_rata)	Numeric	<p>This is the share of expenditure attributed under the category of region covered by each operation. In the case of multiple categories of regions, the different shares are reported separated by “ ”. The shares have been attributed by the Core Team by calculating the share of the eligible expenditure attributed to each category of region Priority over the total eligible expenditure allocated to the operation. In the case of a single category of region, the share was attributed by the Core Team equal to 100%.</p>	No	96.90%	84.72%
Field(s) of Intervention (prj_foi_code & prj_foi_title)	Text	<p>This is the code and label of the Field(s) of intervention assigned to the operation as per categories listed in Annex I of Regulation (EU) No 215/2014. In case the operation is under multiple fields, this is reported separated by “ ”⁶⁶. The main data source of the Field(s) of Intervention is the data extracted from the MAs or collected from the public list of operations. When missing, the Core Team attributed the information by merging the Single Database with the information on the SFC platform or the programme documentation. More specifically, this was possible whenever the information on the Priority Axis was available, and there was a univocal correspondence between the Priority Axis and the Field of Intervention.</p>	No	97.2%	97.22%
Field(s) of Intervention share (prj_foi_pro_rata)	Numeric	<p>This is the share of expenditure attributed under the Field(s) of Intervention under which each operation is funded. In the case of multiple Field(s) of Intervention, the different shares are reported separated by “ ”. The shares have been either retrieved from the percentage shares included in the data extracted by the MA or attributed by the Core Team by calculating the share of the eligible expenditure attributed to each Field of Intervention over the total eligible expenditure allocated to the operation. In the case of a single Field of Intervention, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more Field(s) of Intervention, but it was not possible to estimate the share of each one.</p>	No	95.96%	96.53%

⁶⁶ It should be noted that in the data extraction of the Polish programmes, there was the information on the secondary Fields of Intervention. However, this information was disregarded following the indications of the country expert and only one Field of Intervention has been reported for each operation. Moreover, for some operations of the Portuguese OPs, the raw data extracted by the MA did not include the Field of Intervention, but the macro-categories under which the Fields of Intervention are distributed (e.g., “1. Productive investment” under which the fields 001, 002, 003 and 004 fall). In this case, all the fields falling under the macro-categories were included under the variable ‘Field(s) of Intervention’ concatenated and separated by “|”.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Form of finance share (prj_form_of_finance_code & prj_form_of_finance_title)	Text	<p>This is the code and label of the information on the Form of Finance assigned to the operation, which corresponds to the modality of support to the realisation of the operation as listed in Annex I of Regulation (EU) No 215/2014. In case the operation falls under multiple Forms of Finance, these are reported separated by “ ”.</p> <p>The main data source of information on the Form of Finance is the data extracted from the MAs or collected from the public list of operations. When missing, the information was attributed by the Core Team by merging the Single Database with the information on the SFC platform or the programme documentation. More specifically, this was possible whenever the information on the Priority Axis was available, and there was a univocal correspondence between the Priority Axis and the Form of Finance.</p> <p>It should be noted that, in some cases, internal consistency checks between the Form of Finance and local taxonomies identifying financial instruments brought about the existence of some inconsistencies. This concerned some operations in the Italian and Portuguese OPs, which, according to the local taxonomy, were financial instruments but had Form of Finance different from 03, 04, 05 and 06 or vice versa. In these cases, following the indications of the MA/country expert, the source of information that was considered most reliable was the local taxonomy and, when inconsistencies emerged, the information Form of Finance was deleted.</p>	No	97.82%	99.65%
Form of finance share (prj_form_of_finance_pro_rata)	Numeric	<p>This is the share of expenditure attributed under the Form(s) of Finance under which each operation is funded. In the case of multiple Form(s) of Finance, the different shares are reported separated by “ ”. The shares have been either retrieved from the percentage shares included in the data extracted by the MA or attributed by the Core Team by calculating the share of the eligible expenditure attributed to each Form of Finance over the total eligible expenditure allocated to the operation. In the case of a single Form of Finance, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more Form(s) of Finance, but it was not possible to estimate the share of each one.</p>	No	97.66%	99.65%
Territory type (prj_territory_type_code & prj_territory_type_title)	Text	<p>This is the code and label of the information on the Territory Type assigned to the operation as per categories listed in Annex I of Regulation (EU) No 215/2014. In case the operation falls under multiple Territory Types, these are reported separated by “ ”.</p> <p>The main data source of information on the Territory Type is the data extracted from the MAs or collected from the public list of operations. When missing, the information was attributed by the Core Team by merging the Single Database with the information on the SFC platform or the programme documentation. More specifically, this was possible whenever the information on the Priority Axis was available, and there was a univocal correspondence between the Priority Axis and the Territory Type.</p>	No	95.39%	98.26%
Territory type share (prj_territory_type_pro_rata)	Numeric	<p>This is the share of expenditure attributed under the Territory Type(s) under which each operation is funded. In the case of multiple Territory Types, the different shares are reported separated by “ ”. The shares have been either retrieved from the percentage shares included in the data extracted by the MA or attributed by the Core Team by calculating the share of the eligible expenditure attributed to each Territory Type over the total eligible expenditure allocated to the operation. In the case of a single Territory Type, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more Territory Type(s), but it was not possible to estimate the share of each one.</p>	No	94.83%	98.26%

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Territorial delivery mechanisms (prj_terr_deliv_mech_code & terr_deliv_mech_title)	Text	<p>This is the code and label of the information on the type(s) of Territorial Delivery Mechanism of the operation as listed in Annex I of Regulation (EU) No 215/2014. In case the operation falling under multiple Territorial Delivery Mechanisms, these are reported separated by “ ”.</p> <p>The main data source of information on the Territorial Delivery Mechanism(s) is the data extracted from the MAs or collected from the public list of operations. When missing, the information was attributed by the Core Team by merging the Single Database with the information on the SFC platform or the programme documentation. More specifically, this was possible whenever the information on the Priority Axis was available, and there was a univocal correspondence between the Priority Axis and the Territorial Delivery Mechanism⁶⁷.</p>	No	98.91%	99.65%
Territorial delivery mechanisms share (prj_terr_deliv_mech_pr_o_rata)	Numeric	<p>This is the share of expenditure attributed under the Territorial Delivery Mechanism (s) under which each operation is funded.</p> <p>In the case of multiple Territorial Delivery Mechanisms, the different shares are reported separated by “ ”. The shares have been either retrieved from the percentage shares included in the data extracted by the MA or attributed by the Core Team by calculating the share of the eligible expenditure attributed to each Territorial Delivery Mechanism over the total eligible expenditure allocated to the operation. In the case of a single Territorial Delivery Mechanism, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more Territorial Delivery Mechanism(s), but it was not possible to estimate the share of each one.</p>	No	98.86%	99.65%
Economic activity (prj_econ_activity_cod e & econ_activity_title)	Text	<p>This is the code and label of the information on the Economic Activity of each operation as listed in Annex I of Regulation (EU) No 215/2014.</p> <p>The main data source of information on the Economic Activity(ies) is the data extracted from the MAs or collected from the public list of operations. When missing, the information was attributed by the Core Team by merging the Single Database with the information on the SFC platform or the programme documentation. More specifically, this was possible whenever the information on the Priority Axis was available, and there was a univocal correspondence between the Priority Axis and the Economic Activity.</p>	No	82.50%	96.18%
Economic activity share (prj_econ_activity_pro _rata)	Numeric	<p>This is the share of expenditure attributed under the Economic Activity(ies) under which each operation is funded.</p> <p>In the case of multiple Economic Activities, the different shares are reported separated by “ ”. The shares have been either retrieved from the percentage shares included in the data extracted by the MA or attributed by the Core Team by calculating the share of the eligible expenditure attributed to each Economic Activity over the total eligible expenditure allocated to the operation. In the case of a single Economic Activity, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more Economic Activity(ies), but it was not possible to estimate the share of each one.</p>	No	80.55%	95.83%

⁶⁷ It should be noted that in the data extracted for the Polish programmes, there were only two broad categories instead of the six listed in Annex I of Regulation (EU) No 215/2014, namely “Integrated Territorial Investments” and “Community-led local development initiatives”. As a result:

- All operations falling under “Community-led local development initiatives” were classified under the territorial delivery mechanism 06;
- Those classified as “Integrated Territorial Investments” were classified under the territorial delivery mechanism 01;
- Those which for which were neither “Integrated Territorial Investments” nor “Community-led local development initiatives” were classified as “07 – Not applicable”..

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Operation country (prj_country_code)	ISO 639-1	<p>This is the country(ies) in which the operation is carried out, reported using the country code. In the case of multiple countries, the different countries are reported separated by “ ”.</p> <p>The information was retrieved as follows:</p> <ul style="list-style-type: none"> From the data extracted by the MA or collected from the public list of operations, in the case this was available; Based on the information on the NUTS1, NUTS2 or NUTS3 region, if this was available. 	No	99.20%	95.14%
NUTS1 region (prj_nuts1_code)	Text	<p>This is the code of the NUTS1 region in which the operation is carried out, reported using the NUTS 2013 classification. In the case of multiple NUTS1 regions covered, the different regions are reported separated by “ ”.</p> <p>The information on the NUTS1 region covered by the operation was retrieved as follows:</p> <ul style="list-style-type: none"> From the data extracted by the MA or collected from the public list of operations, in case this was available, harmonised across programmes using the NUTS 2013 classification; Based on the information on the NUTS2 region, if this was available; Based on the location of the beneficiary, if this was available and univocal across beneficiaries and fell under the territorial scope of the programme under which the operation is funded and after checking, on a sampling basis, that the potential bias is minimised for different types of investments (e.g., business support vs large infrastructural projects). 	No	99.31%	93.40%
NUTS1 share (prj_nuts1_pro_rata)	Numeric	<p>This is the share of expenditure attributed to the NUTS1 region where each operation is implemented. In the case of multiple NUTS1 regions, the different shares are reported separated by “ ”. The shares have been either retrieved from the percentage shares included in the data extracted by the MA or attributed by the Core Team by calculating the share of the eligible expenditure attributed to each NUTS1 region over the total eligible expenditure allocated to the operation. In the case of a single NUTS1 region, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more NUTS1 regions, but it was not possible to estimate the share of each one.</p>	No	96.91%	84.03%
NUTS2 region (prj_nuts2_code)	Text	<p>This is the code of the NUTS2 region in which the operation is carried out, reported using the NUTS 2013 classification. In the case of multiple NUTS2 regions covered, the different regions are reported separated by “ ”.</p> <p>The information on the NUTS2 region covered by the operation was retrieved as follows:</p> <ul style="list-style-type: none"> From the data extracted by the MA or collected from the public list of operations, in case this was available, harmonised across programmes using the NUTS 2013 classification; Based on the information on the NUTS3 region, if this was available; Based on existing correspondence tables between the information on the location of the operation expressed using national/regional statistical codes and the NUTS2 classification; Based on the location of the beneficiary, if this was available and univocal across beneficiaries and fell under the territorial scope of the programme under which the operation is funded and after checking, on a sampling basis, that the potential bias is minimised for different types of investments (e.g., business support vs large infrastructural projects). 	No	99.23%	93.40%

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

NUTS2 share (prj_nuts2_pro_rat a)	Numeric	<p>This is the share of expenditure attributed to the NUTS2 region where each operation is implemented. In the case of multiple NUTS2 regions, the different shares are reported separated by “ ”. The shares have been either retrieved from the percentage shares included in the data extracted by the MA or attributed by the Core Team by calculating the share of the eligible expenditure attributed to each NUTS2 region over the total eligible expenditure allocated to the operation. In the case of a single NUTS1 region, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more NUTS2 regions, but it was not possible to estimate the share of each one.</p>	No	96.72%	84.03%
NUTS3 region (prj_nuts3_code)	Text	<p>This is the code of the NUTS3 region in which the operation is carried out, reported using the NUTS 2013 classification. In the case of multiple NUTS3 regions covered, the different regions are reported separated by “ ”.</p> <p>The information on the NUTS3 region covered by the operation was retrieved as follows:</p> <ul style="list-style-type: none"> • From the data extracted by the MA or collected from the public list of operations, in case this was available, harmonised across programmes using the NUTS 2013 classification; • Based on the information on the postal code using correspondence tables provided by EUROSTAT for the NUTS 2013 classification (https://gisco-services.ec.europa.eu/tercet/flat-files); • Based on the information on the city using correspondence tables provided by Geonames (https://download.geonames.org/export/zip/); • Based on existing correspondence tables between the information on the location of the operation expressed using national/regional statistical codes and the NUTS3 classification; • Based on the location of the beneficiary, if this was available and univocal across beneficiaries and fell under the territorial scope of the programme under which the operation is funded and after checking, on a sampling basis, that the potential bias is minimised for different types of investments (e.g., business support vs large infrastructural projects). 	No	92.00%	87.50%
NUTS3 share (prj_nuts3_pro_ra ta)	Numeric	<p>This is the share of expenditure attributed to the NUTS3 region where each operation is implemented. In the case of multiple NUTS3 regions, the different shares are reported separated by “ ”. The shares have been either retrieved from the percentage shares included in the data extracted by the MA or attributed by the Core Team by calculating the share of the eligible expenditure attributed to each NUTS3 region over the total eligible expenditure allocated to the operation. In the case of a single NUTS1 region, the share was attributed by the Core Team equal to 100%. Instead, the information is missing whenever the operation is associated with more NUTS3 regions, but it was not possible to estimate the share of each one.</p>	No	86.69%	72.57%
Typology cluster (prj_typology_id & prj_typology_label)	Text	<p>This is the typology cluster identified by the k-means algorithm implemented under each Thematic Objective and grouping operations according to their scope based on the EU Investment Priority(ies) and Field(s) of Intervention and enabling the selection of homogeneous clusters of operations relevant for thematic evaluations.</p> <p>The typology cluster was attributed based on the clustering technique illustrated in Deliverable D5, where the resulting list of typology clusters is also listed.</p>	Yes ⁶⁸	90.16%	94.79%

⁶⁸ The typology was assigned running a k-means algorithm and performing additional manual refinements.

Part 6. Operation’s financial information

This part allows the collection of information on the financial resources at the level of each operation (i.e., each individual project, as far as possible). This data includes both allocated and paid contributions.

Data	Data format	Data content	Data issue	Data coverage %	
				Operation	OP
Total operation cost (EUR) (prj_tot_cost)	Numeric	This amount corresponds to the total cost of the operation, irrespective of the sources of funding and of the eligibility of expenses, in EUR (or converted into EUR if provided in another currency in the raw data). The information has been directly retrieved from the data provided by the MA or collected from the public list of operations or estimated as the sum of the total eligible expenditure and the total expenditure made of non-eligible costs ⁶⁹ when the latter was available.	No	66.98%	51.39%
Calculation method for the EU contribution (prj_elig_calc_method)	Text	This variable provides information on the method used to calculate the EU contribution, which according to Art.120(2) of the Regulation (EU) No 1303/2013 is decided for each Priority Axis by a Commission decision and can be: <ul style="list-style-type: none"> Total, if the EU co-financing rate is applied on the total eligible expenditure, including public and private expenditure; Public, if the EU co-financing rate is applied on the eligible public expenditure. The Core Team retrieved this information from the SFC portal for each Priority Axis of all the programmes included in the Single Database. It should be noted that, upon indication of the MA, the calculation method indicated in the SFC portal has been corrected for some Italian programmes, namely the SME Initiative OP (2015IT16RFSM001), the Legality OP (2014IT16M2OP003) and all the Cooperation programmes, except for the Interreg V-A - Italy-Malta (2014TC16RFCB037) ⁷⁰ .	Yes ⁷¹	95.60%	96.53%

⁶⁹ For some programmes, expenditure covered by foreign sources and/or EU funds other than the ERDF of the Cohesion Fund, such as the European Social Fund (ESF), the Instrument for Pre-Accession Assistance (IPA) or the European Neighbourhood Instrument (ENI) fund has been considered as a proxy of non-eligible expenditure under the ERDF and CF.

⁷⁰ For the SME Initiative OP (2015IT16RFSM001) and the Legality OP (2014IT16M2OP003) the calculation method reported in the SFC portal was “Total”, while the MA indicated that it is “Public”. The opposite was true for the Cooperation Programmes, except the Interreg V-A - Italy-Malta (2014TC16RFCB037).

⁷¹ It cannot be excluded that also for other programmes there may be errors in the reporting of the information of the calculation method for the EU contribution on the SFC platform.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

EU co-financing rate at Priority Axis level <i>(paxis_eu_cofinan_level)</i>	Numeric	This is the maximum share of the EU contribution, which could be allocated and paid to the operations under a specific Priority Axis. This information was retrieved from the data provided by the MA or collected from public lists of operations as well as from the SFC portal when this information was missing, and there was a univocal correspondence between the Priority Axis and the EU co-financing rate at the Priority Axis level.	No	99.70%	98.96%
EU co-financing rate at operation level <i>(pri_eu_cofinancing_rate)</i>	Numeric	This is the actual share of the EU contribution which has been allocated or paid to the operation. This information was retrieved from the data provided by the MA or collected from public lists of operations. When missing, it was estimated by the Core Team as: <ul style="list-style-type: none"> • The ratio between the EU contribution and Total eligible expenditure allocated to the operation if the calculation method for the EU contribution was “Total”; • The ratio between the EU contribution and Public eligible expenditure allocated to the operation if the calculation method for the EU contribution was “Public”. 	No	77.03%	85.07%
Total eligible expenditure allocated (EUR) <i>(pri_tot_exp_alloc)</i>	Numeric	This variable includes the amount of the total eligible expenditure of the operation approved in the document setting out the conditions for support (data field 41 of the Regulation (EU) No 480/2014). In general, this information is available only for those programmes for which the EU contribution is calculated on the basis of the Total eligible expenditure (i.e., the calculation method for the EU contribution is “Total”). This information was retrieved from the data provided by the MA or collected from public lists of operations. When the information was missing, when possible, it was estimated by the Core Team as: <ul style="list-style-type: none"> • The sum of the eligible public expenditure and the private contribution allocated to the operation⁷²; or • The ratio between the EU contribution allocated to the operation and the EU co-financing rate at operation level; or • The sum of the total eligible expenditure allocated to the beneficiary(ies) of the operation. 	No	99.93%	98.96%

⁷² It should also be noted that, in some cases, a variable including the beneficiary contribution was also available; however, depending on the ownership of the beneficiary, this source could be either public or private. Considering the impossibility to distinguish between public and private sources, this variable has been considered as a private source of funding and included in the estimate of the total eligible expenditure.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Public eligible expenditure allocated (EUR) (<i>prj_tot_pub_alloc</i>)	Numeric	<p>This variable includes the amount of the total eligible expenditure constituting public expenditure as defined in Article 2(15) of Regulation (EU) No 1303/2013 (data field 42 of the Regulation (EU) No 480/2014).</p> <p>This information was retrieved from the data provided by the MA or collected from public lists of operations. When the information was missing, and it was possible based on the available data, it was estimated by the Core Team as:</p> <ul style="list-style-type: none"> • The sum of the EU and Member State contribution and/or other public contribution⁷³ allocated to the operation; or • The sum of the EU and national, regional and local contribution and/or other public contribution⁷⁴ allocated to the operation; or • The difference between the total eligible expenditure and the private contribution allocated to the operation; or • The sum of the eligible public expenditure allocated to the beneficiary(ies) of the operation. 	Yes ⁷⁵	93.07%	80.21%
EU contribution allocated (EUR) (<i>prj_tot_eu_alloc</i>)	Numeric	<p>This variable includes the amount corresponding to the EU (including both ERDF and Cohesion Fund) contribution allocated to the operation under consideration.</p> <p>This information was retrieved from the data provided by the MA or collected from public lists of operations. When the information was missing, it was estimated by:</p> <ul style="list-style-type: none"> • Applying the EU (ERDF and/or CF) co-financing rate at operation level on the total or public eligible expenditure allocated to each operation, depending on the calculation methods used; or • Summing the ERDF and/or CF contribution allocated to the beneficiary(ies) of the operation. 	No	77.14%	84.72%
Private contribution allocated (EUR) (<i>prj_private_alloc</i>)	Numeric	<p>This variable includes the amount corresponding to the private contribution allocated to the operation under consideration.</p> <p>This information was retrieved from the data provided by the MA or collected from public lists of operations. When the information was missing, and it was possible based on the available data, it was estimated as the difference between the total eligible expenditure and the eligible public expenditure allocated to the operation. As an alternative, it was estimated as the sum of the private contribution allocated to the beneficiary(ies) of the operation when available.</p>	Yes ⁷⁶	27.19%	40.63%

⁷³ These variables are currently not included in the single database of operations, in light of the heterogeneity in the availability of such information across programmes and the impossibility to reconcile the information in a harmonised way in a short period of time. In fact, in some cases it was disaggregated by funding sources, in others it also included non-eligible costs. However, they are included in the other .csv files which accompany the single database of operations and which are further described in Annex V.

⁷⁴ *Ibid.*

⁷⁵ Considering that at this stage it was not possible to harmonise and reconcile the data on the national, regional, local and other public contribution allocated to operations, some estimates may not be so precise and include also non-eligible costs.

⁷⁶ *Ibid.*

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Total eligible expenditure paid (EUR) (prj_tot_exp_paid)	Numeric	<p>This variable includes the amount of the total eligible expenditure paid to the operation as of the cut-off date. In general, this information is available only for those programmes for which the EU contribution is calculated based on the Total eligible expenditure (i.e., the calculation method for the EU contribution is "Total").</p> <p>This information was only retrieved from the data provided by the MA. In some cases, the data provided by the MA also included data on the recoveries subtracted or to be subtracted from the total eligible expenditure. Only when specifically indicated by the MA they have been subtracted from the amount of the total eligible expenditure paid.</p>	No	70.24%	85.42%
Total public eligible expenditure paid (EUR) (prj_tot_pub_paid)	Numeric	<p>This variable includes the amount of the eligible public expenditure paid to the operation as of the cut-off date.</p> <p>This information was only retrieved from the data provided by the MA. In some cases, the data provided by the MA also included data on the recoveries subtracted or to be subtracted from the eligible public expenditure. Only when specifically indicated by the MA they have been subtracted from the amount of the public eligible expenditure paid.</p>	No	42.13%	50.00%
EU contribution paid (EUR) (prj_tot_eu_paid)	Numeric	<p>This variable includes the amount corresponding to the EU (including both ERDF and Cohesion Fund) contribution paid to the operation as of the cut-off date.</p> <p>This information was retrieved from the data provided by the MA. When the information was missing, it was calculated as the sum of the ERDF contribution paid and of the Cohesion Fund paid contribution.</p>	No	33.38%	53.13%
Private contribution paid (EUR) (prj_private_paid)	Numeric	<p>This variable includes the amount of the private contribution paid to the operation as of the cut-off date.</p> <p>This information was only retrieved from the data provided by the MA.</p>	No	8.59%	12.85%
Conversion rate used (conversion_rate)	Numeric	<p>This is the conversion rate used for conversion from national currency to Euros. It corresponds to the amount of national currency for 1 Euro.</p> <p>It has been calculated and attributed by the Core Team following the approach presented under Section 2.2.2.2.</p>	No	100%	100%
Multiple variables identifying estimates	Text	<p>These are different variables that have been created for each financial variable whenever the latter was estimated by the Core Team.</p> <p>These variables have been manually filled by the Core Team whenever an estimate of a financial variable was done.</p>	No	n.a.	n.a.

3.2.2. The structure of the Database of Beneficiaries

This database (DB_Beneficiaries) includes information on the individual beneficiaries of each operation listed in the database of operations. In the case of financial instruments and groups of projects, it also includes information on intermediary organisations transferring EU funds to final recipients. In this context, beneficiaries fall under a two-fold definition:

1. The **public or private body** (or bodies, in case of collaborative operations) **directly benefitting from the EU funds** (direct beneficiary) and in charge of implementing a project to achieve its own goals;
2. The **final recipients** of operations delivered through an **intermediary organisation** (e.g., fund manager, financial intermediary, etc.). The latter is the direct beneficiary of the operation, which then transfers the funding to the final recipients. This typically applies to groups of projects or financial instruments.

Raw data provided by the MAs did not distinguish between intermediary organisations and beneficiaries directly benefitting from the EU funds, but they generally indicate 'beneficiaries' for both categories. In fact, as for Art. 2(10) of the Regulation (EU) No 1303/2013⁷⁷, a beneficiary is a public or private body (or bodies, in case of collaborative projects), which initiates and implements an operation to achieve its own goals, and such operation may be a project as well as a financial instrument or an investment strategy implemented by an intermediary organisation. Moreover, data on the final recipients of EU funds are not generally stored in the monitoring systems, except for some specific cases (e.g., financial instruments directly managed by the MA). As a result of these considerations:

- Whenever it was possible to distinguish the nature of an operation (see Section 2.2.2.1 and Part 3 of the DB_Operations), the Core Team could detect intermediary organisations (this was the case in particular for financial instruments and groups of projects). In the other cases, beneficiaries' data provided by the MAs may refer to both direct beneficiaries and intermediaries;
- Once identified intermediary organisations, the Core Team has made an attempt to collect data on the final recipients of operations either from the MA or from the intermediary organisations (i.e., in the case of the SME Initiative OPs, the European Investment Fund provided such data)⁷⁸.

The key figures of the database of beneficiaries are the following:

- In total, the database includes 1,169,162 rows and data on 1,168,711 beneficiaries⁷⁹ and 471,419 distinct beneficiary institutions⁸⁰;

⁷⁷ Art . 2(10) of Regulation (EU) No 1303/2013 defines the beneficiary of an operation as "a public or private body [...] responsible for initiating or both initiating and implementing operations; and in the context of State aid schemes, as defined in point 13 of this Article, the body which receives the aid; and in the context of financial instruments under Title IV of Part Two of this Regulation, it means the body that implements the financial instrument or the fund of funds as appropriate".

⁷⁸ For some financial instruments, the information on sub-intermediary organisations (e.g., financial instruments managed by funds of funds) has also been included under the variable "Sub-intermediary" if provided by the MA. In some cases, it has also been possible to collect information on the sub-beneficiaries of operations, those implementing specific activities for and/or with the direct beneficiary of the operation.

⁷⁹ The number of rows does not coincide with the number of beneficiaries because for some financial instrument the information collected regards only the intermediated bodies and not the final recipients (see below for further details). It should be noted that institutions beneficiaries of more than one operation are counted as many times as many operations the benefitted from.

⁸⁰ This figure counts the number of distinct beneficiaries that can be found in the Single Database, excluding "Anonymised beneficiaries" and those for which no name is available.

- Out of 1,168,711 beneficiaries, **437,083 are final recipients** of either financial instruments or group of projects and 328,542 are entities for which data has been anonymised;
- Only for 3 operations out of 584,828, the information on the beneficiaries, also in an anonymised form, is missing;
- **The full list of beneficiaries**, including sole beneficiaries in case of individual projects, leaders and partners⁸¹ in case of collaborative projects, and final recipients in case of financial instruments or group of projects, **is available for 245 programmes** out of 281 (i.e., 87% of the total);
- **When considering only active leader/sole beneficiaries and final recipients**, and thus excluding partners, withdrawn beneficiaries and Interreg-specific beneficiaries (e.g., advisors, sub-partners, etc.⁸²), **the total number of beneficiaries is equal to 1,096,700**;
- Out of 568 financial instruments, **the list of final recipients is available for 275 financial instruments** (i.e., 48% of the total). For 292 financial instruments (i.e., 51%), only the information on the intermediary organisation was available, while for only one financial instrument, there is no information on the final recipients nor on the intermediary organisation.

The beneficiary database is structured into six parts:

1. General information
2. Operational/Cooperation programme and operation identification
3. Beneficiary identification
4. Intermediary and sub-intermediary organisation identification
5. Beneficiary's characteristics
6. Financial information on beneficiaries

In what follows, the different variables are defined and presented. Cleaning and harmonisation procedures related to specific variables are also described. For each variable, the following information are provided:

- Information on the data format;
- Information on the data content resulting from the data cleaning, harmonisation and enrichment procedures and the related data sources;
- Information on whether the variable still has some limitations, despite the data cleaning, harmonisation and enrichment activities carried out; in this respect, a four-point scale was used to make the assessment:

There are no data issues concerning the variable	There are some data issues, but their relevance is low	There are some data issues, but their relevance is medium	There are some data issues and their relevance is high
--	--	---	--

- Information on the coverage in terms of the number of programmes (see the column "OP") and beneficiaries (see the column "Beneficiaries") is provided. A five-point scale was used to make the assessment:

⁸¹ See the description of the variable "Beneficiary role" in the following sections.

⁸² See the description of the variable "Beneficiary role" in the following sections.

Very low degree of coverage (less than 20% of the programmes and beneficiaries)	Low degree of coverage (between 20 and 40% of the programmes and beneficiaries)	Medium degree of coverage (between 40% and 60% of the programmes and beneficiaries)	High degree of coverage (between 60% and 80% of the programmes and beneficiaries)	Very high degree of coverage (more than 80% of the programmes and beneficiaries)
---	---	---	---	--

Part 1. General information

This first part of the database includes the same type of information reported in the first part of the database of operations. In fact, it includes some general information on the data collected for the list of beneficiaries available for each operation. It provides information on the original language in which variables have been provided and the currency used for reporting financial variables, as well as the cut-off date of the data provided and the date of the data extraction.

Part 2. Operational/Cooperation programme and operation identification

This second part includes some of the variables included and already described in Part 2 and 3 of the DB Operations that will allow the link among the two databases, such as:

- CCI number (*op*);
- OP short title (English) (*op_short_title*);
- OP country(ies) (*op_country*);
- OP Managing Authority (*op_ma*);
- Official operation identifier (*prj_nr*);
- Operation acronym (*prj_acronym*);
- Ad-hoc operation identifier (*prj_ID*).

Part 3. Beneficiary identification

This part allows the identification of the beneficiary of each operation included in the Database of operations. Each operation can have one or more beneficiaries. As mentioned before, for 3 operations, the information on the beneficiary(ies) is missing. When available, the information on the sub-beneficiaries was included in the database.

Data	Data format	Data content and source(s)	Data issue	Data coverage %	
				Beneficiaries	OP

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Beneficiary name (in the original language) (benef_name)	Text	This is the name of the beneficiary identified for each operation based on the definition of 'beneficiary' adopted in the context of this study (see above). It reports the name in the original language used for reporting the data in the monitoring system or in the public list of operations, depending on the source of the data on operations and beneficiaries, except for natural persons for which the name has been anonymised. A harmonisation of the beneficiaries' names has been carried out, as mentioned in Section 2.2.2.3, using natural language processing ⁸³ and performing additional manual refinements, also with the support of the Country Experts.	Yes ⁸⁴	99.96%	100.00%
Beneficiary VAT code (benef_vat)	Alphanumeric	<p>This the VAT number of the beneficiary, harmonised in the EU VAT format for those beneficiaries for which it has been possible to validate it (see below). The information was either already available in the data extracted by the MA or collected from the public list of operations or it has been retrieved from other variables/external data sources, such as:</p> <ul style="list-style-type: none"> • Other variables included in the data provided or collected containing information on the local fiscal number of beneficiaries, such as the SIRET/SIREN in France: in these cases, the VAT code has been extracted from these local codes and harmonised following the structure of the EU VAT code⁸⁵; • The Orbis and CORDIS database: each beneficiary included in the Single Database has been searched in the Orbis and CORDIS database, which has been used to retrieve missing information on the EU VAT code to fill data gaps as well as to validate the correctness of the VAT code associated with beneficiaries for which the information was available. 	Yes ⁸⁶	62.34%	98.61%

⁸³ The harmonisation exercise relied on tokenisation: each beneficiary's name has been converted into a sequence of tokens (strings with an assigned and thus identified meaning) and the frequency of these tokens across each pair of names has been analysed with the aim to identify beneficiaries for which the name was written in different ways but it represented the same institution. As a result of this analysis, the quality of the match was assessed by a score from 0 to 1, where scores close to 1 indicates a perfect match.

⁸⁴ Only the pair of names whose quality of the match was higher than 0.97 were harmonised. However, it cannot be excluded that also other pairs of names under this threshold could include two different names but identifying the same institution. As mentioned in Chapter 4, further refinements to the harmonisation procedure carried out so far will be implemented in the context of future ex-post thematic evaluations.

⁸⁵ For instance, for the transformation of the SIRET/SIREN into an EU VAT code, the Core Team followed the indications provided on this website: <https://marosavat.com/manual/vat/france/siren-siret/>

⁸⁶ The Core Team could not ensure that all the VAT codes included under this column are valid codes.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Validated VAT (benef_vat_validated)	Text	<p>This variable includes information on whether the VAT code associated with each beneficiary of the Single Database has been validated based on external sources (“True” or “False”), in particular:</p> <ul style="list-style-type: none"> • The VIES platform, an electronic mean to validate VAT-identification numbers of economic operators registered in the European Union for cross border transactions on goods or services: each VAT code included in the database of beneficiaries has been automatically searched on the VIES platform in order to get information on its validity; • The Orbis database, the EU VAT code retrieved from the Orbis database has been extracted and compared to the one already included in the Single Database to check its validity; • The CORDIS database, each beneficiary included in the Single Database has been searched in the CORDIS database and the related EU VAT code compared to the one already included in the Single Database to check its validity. 	No	100.00%	100.00%
Beneficiary VAT regime (benef_vat_regime)	Text	<p>This variable includes information on whether the beneficiary of the operation is able to recover the VAT on expenditure incurred (“Yes”) or not (“No”).</p> <p>This information is available only for those programmes for which the data extracted from the monitoring system included data field 3 ‘Information whether VAT on expenditure incurred by the beneficiary is non-recoverable under national VAT legislation’ of Annex III of the Regulation (EU) No 480/2014.</p>	No	23.50%	93.06%
Orbis BvD (benef_bvd_id)	Alphanumeric	<p>This is an ad-hoc identifier used in the Orbis database. The BvD associated with the beneficiaries was extracted and included in the Single Database to allow the link between the Single Database and the Orbis database. See 0 for more information on the results of the match with the Orbis database.</p>	Yes ⁸⁷	39.70%	91.67%
Beneficiary country (benef_country_code)	ISO 639-1	<p>This is the country(ies) in which the beneficiary is located, reported using the country code. In the case of multiple countries, they are reported separated by “ ”.</p> <p>The information on the country of the beneficiary was retrieved as follows:</p> <ul style="list-style-type: none"> • From the data extracted by the MA or collected from the public list of operations, in the case this information was available; • Based on the information on the NUTS1, NUTS2 or NUTS3 region of the beneficiary, if this was available; • Based on the information retrieved from the Orbis database. 	No	60.40%	98.96%

⁸⁷ The match between the single database of beneficiaries and the ORBIS database has not always worked: if all the name of the beneficiaries was manually searched, a higher number of beneficiaries would find a match in the ORBIS database. See Annex VI for more details on the extent of the match and Chapter 4 for more details on the potential refinements that could be performed in the context of future ex-post thematic evaluations.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Beneficiary NUTS1 (benef_nuts1_code)	NUTS classification	<p>This is the code of the NUTS1 region in which the beneficiary is located, reported using the NUTS 2013 classification. In the case of multiple NUTS1 regions covered, the different regions are reported separated by “ ”.</p> <p>The information on the NUTS1 region of the beneficiary was retrieved as follows:</p> <ul style="list-style-type: none"> • From the data extracted by the MA or collected from the public list of operations, in case this was available, harmonised across programmes using the NUTS 2013 classification; • Based on the information on the NUTS2 region, if this was available; • Based on the information retrieved from the Orbis database. 	No	60.40%	98.96%
Beneficiary NUTS2 (benef_nuts2_code)	NUTS classification	<p>This is the code of the NUTS2 region in which the beneficiary is located, reported using the NUTS 2013 classification. In the case of multiple NUTS2 regions covered, the different regions are reported separated by “ ”.</p> <p>The information on the NUTS2 region of the beneficiary was retrieved as follows:</p> <ul style="list-style-type: none"> • From the data extracted by the MA or collected from the public list of operations, in case this was available, harmonised across programmes using the NUTS 2013 classification; • Based on the information on the NUTS3 region, if this was available; • Based on existing correspondence tables between the information on the location of the beneficiary expressed using national/regional statistical codes and the NUTS2 classification; • Based on the information retrieved from the Orbis database. 	No	60.37%	98.96%
Beneficiary NUTS3 (benef_nuts3_code)	NUTS classification	<p>This is the code of the NUTS3 region in which the beneficiary is located, reported using the NUTS 2013 classification. In the case of multiple NUTS3 regions covered, the different regions are reported separated by “ ”.</p> <p>The information on the NUTS3 region of the beneficiary was retrieved as follows:</p> <ul style="list-style-type: none"> • From the data extracted by the MA or collected from the public list of operations, in case this was available, harmonised across programmes using the NUTS 2013 classification; • Based on the information on the postal code using correspondence tables provided by EUROSTAT for the NUTS 2013 classification (https://gisco-services.ec.europa.eu/tercet/flat-files); • Based on the information on the city using correspondence tables provided by Geonames (https://download.geonames.org/export/zip/); • Based on existing correspondence tables between the information on the location of the beneficiary expressed using national/regional statistical codes and the NUTS3 classification; • Based on the information retrieved from the Orbis database. 	No	59.19%	98.96%

<p>Beneficiary macro-role (benef_macro_role)</p>	<p>Text</p>	<p>This variable provides information on the macro-role of the beneficiary in operation. List of labels included:</p> <ul style="list-style-type: none"> • Direct beneficiary: this is the public or private body (or bodies, in case of collaborative operations) directly benefitting from the EU funds (direct beneficiary) and in charge of implementing a project to achieve its own goals. For operations that are not financial instruments nor groups of projects, these beneficiaries fall under the definition of beneficiary provided by Art. 2(10) of the Regulation (EU) No 1303/2013⁸⁸. • Final recipients: these are the ultimate recipients of EU support delivered through an intermediary organisation (e.g., fund manager, financial intermediary, etc.), namely financial instruments and groups of projects. • INTERREG specific: this category includes specific types of beneficiaries of INTERREG programmes, which generally have a particular governance structure: direct beneficiaries are, in some cases, supported by observers/advisors or sub-partners. 	<p>No</p>	<p>99.96%</p>	<p>100.00%</p>
---	-------------	--	-----------	---------------	----------------

⁸⁸ Art. 2(10) of Regulation (EU) No 1303/2013 defines the beneficiary of an operation as “a public or private body [...] responsible for initiating or both initiating and implementing operations; and in the context of State aid schemes, as defined in point 13 of this Article, the body which receives the aid; and in the context of financial instruments under Title IV of Part Two of this Regulation, it means the body that implements the financial instrument or the fund of funds as appropriate”.

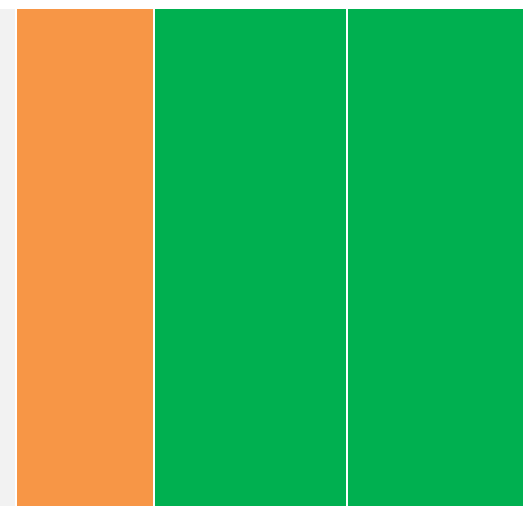
<p>Role of the beneficiary in the operation (benef_role)</p>	<p>This variable provides information on the specific role of the beneficiary in the operation that can be found under each macro-category. List of labels included:</p> <p><u>Under the direct beneficiaries:</u></p> <ul style="list-style-type: none"> • Sole beneficiary: this is the unique beneficiary of an operation carried out individually; • Leader/Sole beneficiary: this label has been attributed whenever an operation had only one beneficiary, but it was not possible to assess whether it is the unique beneficiary or it implements the operation in collaboration with other institutions; • Leader: this is the lead beneficiary of an operation carried out in collaboration with other institutions. The leader is generally the institution signing the contract, and that is the main responsible actor for the implementation of the operation⁸⁹; • Leader (withdrawn): this was the lead beneficiary of an ongoing or completed operation which is no longer responsible for its implementation following a change in the structure of the partnership but who has received some EU funds for the activities carried out before the termination of the contract; • Leader/Partner: this label is specific to the Italian OPs and CPs, where for some operations, more than one beneficiary was reported, but it was not possible to distinguish between leader and partner beneficiaries; • Partner: this is a beneficiary collaborating with a lead beneficiary in the implementation of the operation. Depending on the governance of the operation and/or programme, partners may be direct beneficiaries of EU funds and receive them directly from the MA or may receive their financial part as a result of a transfer of EU funds from the lead beneficiary; • Partner (withdrawn): this was a partner in the context of an ongoing or completed operation that is no longer involved in its implementation following a change in the structure of the partnership but who has received some EU funds for the activities carried out before the termination of the contract. <p><u>Under the final recipients:</u></p> <ul style="list-style-type: none"> • Final recipient (financial instrument): this is the final recipient of a financial instrument; • Final recipient (group of projects): this is the final recipient of a group of projects. • Under the INTERREG specific: • Partner extra-EU: this is a beneficiary collaborating with a lead beneficiary in the implementation of the operations which, however, operates in extra-EU countries and is therefore not beneficiary of EU funds; • Observer/advisor: this is a peculiar partner that does not contribute financially to the project, but which is interested in its results and therefore acts as an observer and advisor in the implementation of the project (in the case there were more than one observer, and it was possible to distinguish the one leading the activity of advisory, the beneficiary has been labelled as "Lead observer/advisor"); • Leader and partner of local partnership: this is a specific role relevant in the context of the Interreg V-B - North Sea Region (2014TC16RFTN005), where, beyond the information on the lead beneficiary and related partners, there are also two additional types of actors: those who are responsible for managing administrative activities for a number of partners of local partnership and acts as the contact to the lead beneficiary ("Leader of local partnership") and those collaborating in local partnerships ("Partner of local partnership")⁹⁰; 	<p>Yes⁹¹</p>	<p>99.96%</p>	<p>100.00%</p>
--	---	-------------------------	---------------	----------------

- **Sub-beneficiary:** in the context of three CPs, the information on the sub-beneficiaries of the operations was also available. In general, they are small institutions with specific expertise and which work closely with the lead beneficiary or a partner of the operation. Their involvement in the project is often limited in terms of time and contribution.

The information was generally available in the data provided by the MA. It should be highlighted that there are great differences in the types and level of granularity of the information recorded on beneficiaries across monitoring systems. Moreover, some programmes deliver operations based on peculiar governance structures. As a result, the harmonisation of this variable has been complex and could not always be performed. When missing, the Core Team proceeded as follows:

- By requesting additional information to the MA;
- By attributing a label based on available information on the related operation (e.g., in case only one beneficiary is reported for an operation, the latter could either be the leader or the unique (sole) beneficiary).

The information is still missing for the beneficiaries of those three operations for which no data was provided (not even in an anonymised form), as well as for those 268 financial instruments for which the information on the intermediary organisation was (not) available, but data on final recipients could not be collected/provided or were provided in an aggregated form.



⁸⁹ Only one beneficiary is generally responsible for the implementation on a collaborative operation. However, an exception is the national Slovenian OP (2014SI16MAOP001). Under this OP, different contracts may be attached to an operation and they may be signed with a unique or different beneficiaries. In the case the contract is signed with different beneficiaries, it is therefore not possible to distinguish which is the one leading the operation and all beneficiaries are on an equal footing.

⁹⁰ See https://northsearegion.eu/media/1154/all-fact-sheets_joint-file.pdf for more details on the governance structure of this programme.

⁹¹ Considering the great differences in the governance structures set up for the implementation of operations across monitoring systems, there is an issue of comparability of the information available on the beneficiary role.

Part 4. Intermediary and sub-intermediary organisation identification

This part allows the identification of the intermediary organisation that, for those operations consisting of financial or other intermediated instruments, coordinates the implementation of the operation and transfer EU funds to the final recipients.

Data	Data format	Data content and source(s)	Data issue	Data coverage %	
				Beneficiaries	OP
Intermediary name (original language) (intermediary_name)	Text	In the case of identified financial or other intermediated instruments, it corresponds to the name of the intermediary body in charge of delivering the EU funds to the final recipients. In the data provided by the MA or collected from the public list of operations, these entities are generally identified as 'direct beneficiaries', in line with the definition adopted by the European Commission (Art. 2(10) of the Regulation (EU) No 1303/2013). However, they have been relabelled as Intermediaries in our database.	No	88.41% ⁹²	98.59% ⁹³
Intermediary VAT code (intermediary_vat)	Text	This is the VAT number of the intermediary, as provided in the data extracted by the MA or collected from the public list of operations.	No	81.85% ⁹⁴	70.42% ⁹⁵

⁹² The percentage was computed out of the total number of financial instruments and group of projects rather than out of the total database.

⁹³ *Ibid.*

⁹⁴ *Ibid.*

⁹⁵ *Ibid.*

Sub-intermediary name (original language) (intermediary_sub_name)	Text	For some specific financial or intermediated instruments, also information on the sub-intermediary organisation is available. In particular, this is the case of instruments managed by an intermediary that is a fund of funds that then select other intermediary organisations to deliver the designed instruments.	No	9.13% ⁹⁶	19.01% ⁹⁷
Sub-intermediary VAT code (intermediary_sub_vat)	Text	This the VAT number of the sub-intermediary organisation, as provided in the data extracted by the MA or collected from the public list of operations.	No	0.12% ⁹⁸	2.82% ⁹⁹

Part 5. Beneficiary's characteristics

This part gives more detailed information on the nature of the beneficiary receiving EU support.

Data	Data format	Data content and source(s)	Data issue	Data coverage %	
				Beneficiaries	OP

⁹⁶ *Ibid.*

⁹⁷ *Ibid.*

⁹⁸ *Ibid.*

⁹⁹ *Ibid.*

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Beneficiary sector (benef_nace_code)	Alphanumeric	<p>This variable includes information on the economic sector of the beneficiary of the operations, which are enterprises by legal status, as defined by the NACE rev.2 classification of economic activities established by the European Commission.</p> <p>The attribution was based on:</p> <ul style="list-style-type: none"> • Variables included in the raw data provided by the MA providing information on the NACE classification, cleaned and harmonised. • The Orbis database, which was matched with the Single Database of beneficiaries. The information on the NACE code has therefore been retrieved for those beneficiaries, which could be found in Orbis. 	Yes ¹⁰⁰	46.40%	92.36%
Beneficiary macro-type (benef_macro_type)	Text	<p>This variable provides information on the macro-category of the type of beneficiary of the operation. List of labels included:</p> <ul style="list-style-type: none"> • Enterprise; • Natural person; • Research and technology transfer organisation; • Higher education institution; • Other education and training institution; • Public administration authority; • Public agency; • Other institution of public interest; • Chamber of commerce and business association; • Business support organisation; • Financial institution; • Trade union; • NGOs and civil society association. <p>See Deliverable D5 for more information on the clustering exercise performed to classify beneficiaries.</p>	Yes ¹⁰¹	100.00% ¹⁰²	100.00% ¹⁰³

¹⁰⁰ It should be noted that for a very small sample of beneficiary institutions, there are some inconsistencies in the NACE sector. This may be due to inconsistencies in the raw data (e.g., the same beneficiary, under different programmes, may have been classified under different NACE sectors) or inconsistencies between the NACE declared in the raw data and the NACE retrieved from the ORBIS database.

¹⁰¹ It cannot be excluded that some misclassifications occurred.

¹⁰² This coverage share includes also those cases where the variable takes value "Unclassified"/"Unclassifiable" because either the information was not applicable or not available. If excluded, the coverage share is 92.40%

¹⁰³ This coverage share includes also those cases where the variable takes value "Unclassified"/"Unclassifiable" because either the information was not applicable or not available. If excluded, the coverage share is 99.30%

Beneficiary type (benef_type)	Text	<p>This variable provides information on the specific type of beneficiary of the operation. It provides a further specification compared to the macro-type, whenever this specification could be attributed. List of labels included:</p> <ul style="list-style-type: none"> • Enterprise; • Natural person - Business; • Natural person – No business; • Research and technology transfer organisation – Research organisation; • Research and technology transfer organisation – Science and Technology Park; • Research and technology transfer organisation – Technology transfer organisation; • Research and technology transfer organisation – Incubator centre; • Research and technology transfer organisation – Competence centre; • Higher education institution; • Other education and training institution; • Public administration authority – national level; • Public administration authority – regional level; • Public administration authority – local level; • Public agency; • Other institution of public interest - Culture, tourism, sport, environment; • Other institution of public interest – Healthcare; • Other institution of public interest – Providers of public services; • Other institution of public interest – Other social services; • Other institution of public interest – Public security services; • Chamber of commerce and business association; • Business support organisation – Cluster organisation; • Business support organisation - Other; • Financial institution; • Trade union; • NGOs and civil society association. <p>See Deliverable D5 for more information on the clustering exercise performed to classify beneficiaries.</p>	Yes ¹⁰⁴	100.00% ¹⁰⁵	100.00% ¹⁰⁶
----------------------------------	------	---	--------------------	------------------------	------------------------

¹⁰⁴ It cannot be excluded that some misclassifications occurred.

¹⁰⁵ This coverage share includes also those cases where the variable takes value “Unclassified”/“Unclassifiable” because either the information was not applicable or not available. If excluded, the coverage share is 92.40%

¹⁰⁶ This coverage share includes also those cases where the variable takes value “Unclassified”/“Unclassifiable” because either the information was not applicable or not available. If excluded, the coverage share is 99.30%

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Legal status of enterprise (benef_enterprise_legal)	Text	This variable indicates those beneficiaries who, by legal status, are limited liability companies ('Yes'). This information has been added in light of the fact that the 'Beneficiary type' does not classify beneficiaries based on their legal status but based on their core mission. This information has been retrieved from local classifications of the types of beneficiaries provided by the MA in the data extraction, after manual checks of the Core Team, as well as on the basis of keywords included in the beneficiary's name, namely acronyms identifying enterprises across the different Member States (e.g., 'GmbH' in Austria and Germany, 'Srl' or 'Spa' in Italy, 'Ltd' in the UK and other countries, etc.)	No	69.42%	97.92%
Beneficiary ownership (benef_ownership)	Text	This variable provides information on the ownership structure of the beneficiary of the operation, namely whether it is a public or private institution. As for the ownership dimension, the taxonomy used is the same adopted by Regulation (EU) 480/2014. The data field 2 of Annex III of this regulation requires that MAs store the information on whether the beneficiary is public law body or a private law body in a computerised format. When this information was missing from the raw data provided by the MAs (i.e., in 45% of the total number of beneficiaries), the attribution was made by the Core Team whenever there were keywords in the beneficiary's name or other variable providing a local taxonomy which could suggest the ownership structure of beneficiaries (e.g., 'Ministry' or 'Public association' for the label 'Public' and 'National sarl' or 'Private company' for the label 'Private').	Yes ¹⁰⁷	64.81%	98.96%
Beneficiary size (benef_size)	Text	This variable provides information on the size of the beneficiary of the operation. It is mainly available for those having the legal status of a limited liability company, but not exclusively. The categories established in the Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises have been used. These labels have been attributed by combining three different sources: <ul style="list-style-type: none"> • The local taxonomy regarding the size already provided by the MAs, if any, after its harmonisation; • Other local taxonomies provided by the MAs, for instance regarding the type, when they included keywords related to the size, e.g., small enterprises, SMEs, large company, etc.; • The variable provided information on the number of employees; this was either already available in the monitoring system or retrieved from the Orbis database¹⁰⁸. 	Yes ¹⁰⁹	38.16%	89.24%
Number of employees (benef_n_empl_oyee)	Numeric	This variable includes information on the number of employees if the beneficiary is a limited liability company. The information has been retrieved either from the data provided by the MA, when available or from the Orbis database. In the case it was extracted from Orbis, the information reported is that of the average number of employees between 2014-2020. Instead, when it was retrieved only from the data provided by the MA, it was not possible to identify the reference period.	No	33.50%	87.50%

¹⁰⁷ It should be noted that for a very small sample of beneficiary institutions, there are some inconsistencies in the ownership. This may be due to inconsistencies in the raw data (e.g., the same beneficiary, under different programmes, may have been classified under different types of ownership).

¹⁰⁸ In the case it was available in the Orbis database, the average number of employees between 2014-2020 was considered.

¹⁰⁹ It should be noted that for a very small sample of beneficiary institutions, there are some inconsistencies in the size. This may be due to inconsistencies in the raw data (e.g., the same beneficiary, under different programmes, may have been classified under different sizes) or inconsistencies between the size declared in the raw data and the size retrieved from the ORBIS database based on the number of employees.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Beneficiary identifier Beneficiary H2020 (benef_h2020)	Text	This is a dichotomous variable (“Yes” or “No”) indicating whether the beneficiary of the operation also benefits from H2020 support. The information has been retrieved by linking the Single Database with the CORDIS database. Only for some specific final recipients of the SMEi OP, the information was available on the raw data provided by the European Investment Fund.	No	99.97%	100.00%
Beneficiary identifier in CORDIS (benef_cordis_id)	Numeric	This is the identifier of the beneficiary used in the CORDIS database and can be used to match the information in the Single Database with the data available in CORDIS. The information has been retrieved by linking the Single Database with the CORDIS database.	No	5.36%	94.10%
H2020 operation ID (benef_cordis_projectID)	Numeric	This variable includes the list of the ID of the H2020 projects granted to the beneficiary, stored in a unique cell. In case of multiple IDs, all IDs have been reported concatenated in one cell, separated by “ ”. The information has been retrieved by linking the Single Database with the CORDIS database.	No	5.36%	94.10%
H2020 operation rcn (benef_cordis_pr_ojectRcn)	Numeric	This variable includes a code identifying all H2020 projects granted to the beneficiary, stored in a unique cell. In case of multiple ‘Rcn’ codes, all the codes have been reported concatenated in one cell, separated by “ ”. The information has been retrieved by linking the Single Database with the CORDIS database.	No	5.36%	94.10%

Part 6. Financial information on the beneficiary

This part allows the collection of information on the financial resources linked to the beneficiary in the framework of a specific operation. This data includes both allocated and paid contribution to each beneficiary.

Data	Data format	Data content	Data issue	Data coverage %	
				Beneficiaries	OP
EU co-financing rate at beneficiary level (benef_eu_cofinancing_rate)	Numeric	<p>This is the actual share of the EU contribution which has been allocated or paid to the beneficiary of the operation.</p> <p>This information was retrieved from the data provided by the MA or collected from public lists of operations. When missing, it was estimated by the Core Team as:</p> <p>The ratio between the EU contribution and Total eligible expenditure allocated to the beneficiary if the calculation method for the EU contribution was "Total";</p> <p>The ratio between the EU contribution and Public eligible expenditure allocated to the beneficiary if the calculation method for the EU contribution was "Public".</p>	No	13.07%	32.64%
Total eligible expenditure allocated (EUR) (benef_tot_exp_alloc)	Numeric	<p>This variable includes the amount of the total eligible expenditure allocated to the beneficiary (i.e., sole beneficiary, leader and or partner) of the operation. In general, this information is available only for those programmes for which the EU contribution is calculated on the basis of the Total eligible expenditure (i.e., the calculation method for the EU contribution is "Total").</p> <p>This information was retrieved from the data provided by the MA or collected from public lists of operations. When the information was missing, and it was possible based on the available data, it was estimated by the Core Team as:</p> <p>The sum of the eligible public expenditure and the private contribution allocated to the beneficiary; or</p> <p>The ratio between the EU contribution allocated to the beneficiary and the EU co-financing rate at beneficiary (or operation if unique) level; or</p> <p>Equal to the amount of total eligible expenditure allocated to the operation, in case of operations having a sole beneficiary.</p>	No	16.78%	50.00%

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Total public eligible expenditure allocated (EUR) (benef_tot_pub_alloc)	Numeric	<p>This variable includes the amount of the total eligible expenditure constituting public expenditure as defined in Article 2(15) of Regulation (EU) No 1303/2013 allocated to the beneficiary (i.e., sole beneficiary, leader and or partner) of the operations.</p> <p>This information was retrieved from the data provided by the MA or collected from public lists of operations. When the information was missing, and it was possible based on the available data, it was estimated by the Core Team as:</p> <p>The sum of the EU and Member State contribution and/or other public contribution¹¹⁰ allocated to the beneficiary; or</p> <p>The sum of the EU and national, regional and local contribution and/or other public contribution¹¹¹ allocated to the beneficiary; or</p> <p>The difference between the total eligible expenditure and the private contribution allocated to the beneficiary; or</p> <p>Equal to the amount of the eligible public expenditure allocated to the operation, in case of operations having a sole beneficiary.</p>	Yes ¹¹²	14.78%	40.63%
EU contribution allocated (EUR) (benef_tot_eu_alloc)	Numeric	<p>This variable includes the amount corresponding to the EU (including both ERDF and Cohesion Fund) contribution allocated to the beneficiary (i.e., sole beneficiary, leader and or partner) of the operation under consideration.</p> <p>This information was included when provided by the MA or collected from public lists of operations. When the information was missing, it was estimated by:</p> <p>Applying the EU (ERDF and/or CF) co-financing rate at beneficiary (or operation if unique) level on the total or public eligible expenditure allocated to each operation, depending on the calculation methods used; or</p> <p>Attributing the ERDF and/or CF contribution allocated to the operation, in case of operations having only a sole beneficiary.</p>	No	16.55%	48.61%
Private contribution allocated (EUR) (benef_private_alloc)	Numeric	<p>This variable includes the amount corresponding to the private contribution allocated to the beneficiary (i.e., sole beneficiary, leader and or partner) of the operation under consideration.</p> <p>This information was provided by the MA or collected from public lists of operations. When the information was missing, and it was possible based on the available data, it was estimated as the difference between the total eligible expenditure and the eligible public expenditure allocated to the beneficiary operation. As an alternative, when this was available, and operations had only a sole beneficiary, it was set equal to the private contribution allocated to the operation.</p>	Yes ¹¹³	13.34%	32.64%

¹¹⁰ This variable is currently not included in the single database of beneficiaries, in light of the heterogeneity in the availability of such information across programmes and the impossibility to reconcile the information in a harmonised way in a short period of time. In fact, in some cases it was disaggregated by funding sources, in others it also included non-eligible costs. However, it is included in the other .csv files which accompany the single database and which are further described in Annex V.

¹¹¹ *Ibid.*

¹¹² Considering that at this stage it was not possible to harmonise and reconcile the data on the national, regional, local and other public contribution allocated to operations, some estimates may not be so precise and include also non-eligible costs.

¹¹³ *Ibid.*

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Total eligible expenditure paid (EUR) (benef_tot_exp_paid)	Numeric	This variable includes the amount of the total eligible expenditure paid to the beneficiary (i.e., sole beneficiary, leader and or partner) of the operation as of the cut-off date. This information is available only for those programmes for which the EU contribution is calculated on the basis of the Total eligible expenditure (i.e., the calculation method for the EU contribution is "Total"). This information was provided by the MA. In some cases, the data provided by the MA also included data on the recoveries subtracted or to be subtracted from the total eligible expenditure paid. Only when specifically indicated by the MA they have been subtracted from the amount of total eligible expenditure paid.	No	14.73%	47.92%
Total public eligible expenditure paid (EUR) (benef_tot_pub_p)	Numeric	This variable includes the amount of the eligible public expenditure paid to the beneficiary of the operation as of the cut-off date. This information was provided by the MA. In some cases, the data provided by the MA also included data on the recoveries subtracted or to be subtracted from the public eligible expenditure paid. Only when specifically indicated by the MA they have been subtracted from the amount of eligible public expenditure paid.	No	7.62%	20.49%
Private contribution paid EU contribution paid (EUR) (benef_eu_paid)	Numeric	This variable includes the amount corresponding to the EU (including both ERDF and Cohesion Fund) contribution paid to the beneficiary (i.e., sole beneficiary, leader and or partner) of the operation as of the cut-off date. This information was provided by the MA. When the information was missing, it was calculated as the sum of the ERDF contribution allocated and of the Cohesion Fund allocated contribution.	No	9.80%	35.42%
Private contribution paid (EUR) (benef_private_paid)	Numeric	This variable includes the amount of the private contribution paid to the beneficiary (i.e., sole beneficiary, leader and or partner) of the operation as of the cut-off date. This information was included only when provided by the MA.	No	7.02%	15.97%
Total amount of the contract allocated to the final recipient (benef_tot_contract_alloc)	Numeric	This variable includes the total allocated amount established in the contract signed with the final recipient of financial or other intermediated instruments. This information was provided by the MA.	No	5.06%	12.50%

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Total amount of the contract covered by EU funds allocated to the final recipient (benef_eu_contract_alloc)	Numeric	This variable includes the total allocated amount covered by EU funds as established in the contract signed with the final recipient of financial or other intermediated instruments. This information was provided by the MA.	No	2.96%	9.72%
Total amount of the contract covered by public funds allocated to the final recipient (benef_pub_contract_alloc)	Numeric	This variable includes the total allocated amount covered by Member State funds as established in the contract signed with the final recipient of financial or other intermediated instruments. This information was provided by the MA.	No	0.43%	2.08%
Total amount of the contract covered by private sources allocated to the final recipient (benef_private_contract_alloc)	Numeric	This variable includes the total allocated amount covered by private sources as established in the contract signed with the final recipient of financial or other intermediated instruments. This information was provided by the MA.	No	0.53%	1.39%

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Total amount of the contract paid to the final recipient (benef_tot_contract_paid)	Numeric	This variable includes the total amount disbursed to the final recipient of financial or other intermediated instruments. This information was provided by the MA.	No	2.88%	8.68%
Total amount of the contract covered by EU funds paid to the final recipient (benef_eu_contract_paid)	Numeric	This variable includes the amount constituting EU expenditure disbursed to the final recipient of financial or other intermediated instruments. This information was provided by the MA.	No	2.50%	7.64%
Total amount of the contract covered by public funds paid to the final recipient (benef_pub_contract_paid)	Numeric	This variable includes the amount constituting Member State expenditure disbursed to the final recipient of financial or other intermediated instruments. This information was provided by the MA.	No	0.41%	1.74%
Total amount of the contract covered by private sources paid to the final recipient (benef_private_contract_paid)	Numeric	This variable includes the amount constituting private expenditure disbursed to the final recipient of financial or other intermediated instruments. This information was provided by the MA.	No	0.51%	1.04%

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Total amount of the loan granted (benef_loan_amount)	Numeric	This is the total amount of the loan granted to final recipients of financial instruments delivered through loans. This information was provided by the MA.	No	5.86%	5.56%
Total amount of the guarantee granted (benef_guarantee_amount)	Numeric	This is the total amount of the guarantee granted to final recipients of financial instruments delivered through guarantees. This information was provided by the MA.	No	0.54%	1.74%
Conversion rate used (conversion_rate)	Numeric	This is the conversion rate used for conversion from national currency to Euros. It corresponds to the amount of national currency for 1 Euro. It has been calculated and attributed by the Core Team following the approach presented under Section 2.2.2.2.	No	100%	100%
Multiple variables identifying estimates	Text	These are different variables that have been created for each financial variable whenever the latter was estimated by the Core Team. These variables have been manually filled in by the Core Team.	No	n.a.	n.a.

4. The limitations of the Single Database

Despite the data cleaning, harmonisation and enrichment activities performed to provide a Database of Operations and the Database of Beneficiaries as coherent and reliable as possible, the Single Database presents some limitations. The latter should be taken into duly consideration while analysing the data included in the Single Database. Some mitigation actions or solutions can be adopted to cope with some of these limitations in the context of future ex-post thematic evaluations.

The limitations for which further data collection, cleaning, harmonisation and enrichment activities could be envisaged while carrying out a detailed analysis in the context of ex-post evaluations are the following:

- **The translation exercise has not always been effective:** As mentioned in the Sections describing the Single Database, and in particular the Database of Operations and Database of Beneficiaries, the algorithm used to translate the database of operations and beneficiaries has not always properly worked, and some variables still contain the information in the original language (this occurred for instance in the CZ programmes and for some EL and CY operations). The translation of some specific variables may be manually revised by future evaluators if necessary for allowing more precise comparative analysis.
- **Not fully harmonised list of beneficiaries:** The harmonisation of the beneficiaries' names has been carried out following the procedures described in Section 3.2.2.3 and Section 3.2.2 (Part 3). However, there may also be other beneficiaries whose names should be harmonised and that were not captured by the automatic and manual harmonisation checks performed. Further manual refinements may be performed by the selected evaluators if necessary.
- **Missing information on financial variables at operation and beneficiary level:** so far, the database includes data on the allocated and paid amount of the i) total eligible expenditure, ii) eligible public expenditure, iii) total EU contribution (both ERDF and/or CF) and iv) private contribution of the selected operations. Moreover, some of these variables have data gaps. However, in the raw data made available to the European Commission, one may find additional financial variables that could not be retained in the harmonised database (e.g., the national or regional contribution). Moreover, additional estimates may be performed by future evaluators to fill gaps in the financial data related to payments or at the beneficiary level based on the available information.
- **The match of beneficiaries with the Orbis and CORDIS databases needs further manual search:** the match of the Database of Beneficiaries with the Orbis and CORDIS databases has been automatically performed on the entire database using the beneficiary's name and/or VAT code. However, a more precise match could be ensured by an additional manual search of beneficiaries in both databases. They could be performed on sub-samples of beneficiaries, if necessary, in the context of future ex-post thematic evaluations.
- **Missing information on the title of the Specific Objective and/or OP measure in some programmes:** in some cases, only the code is available, while the title is missing. Future evaluators may add this information based on the programme documentation, if necessary, for the analyses to be performed in the context of the ex-post evaluations.
- **Inconsistent unit of analysis:** As described in Section 2.2.2.1, it was not possible to construct a Database of Operations with a standardised unit of analysis (i.e., at the project level) due to the great differences in terms of i) definition of operations

adopted by the MA; ii) availability of data on final recipients or other data allowing the aggregation of project components. Moreover, keeping the same operation definition as the one adopted by each MA allows the linkage with the indicator database as the latter was provided at the operation level. The solutions adopted to ensure some consistency in the definition of operations, listed in Section 2.2.2.1, presents additional limitations:

- It has not always been possible to attribute a univocal flag to each operation because for some programmes, no information on the nature of operations was provided by the MAs, and therefore, only a hybrid flag combining the potential different definitions of such operations was included;
- It has not been possible to automatically group those operations which may be components of the same project by assigning a higher-level code to each project component;
- There may be other financial instruments that could not be identified based on the strategy described in Section 3.2.1 (e.g., because the form of finance was missing).

While performing ex-post thematic evaluations, future evaluators shall be aware of this limitation and may further analyse the sub-sample of operations and decide to adopt other solutions to harmonise the unit of analysis.

Limitations for which no additional refinements are possible are the following:

- **Inconsistent cut-off date:** As illustrated in Section 3.2.1, the collected datasets have a different cut-off date. This issue also reflects the great differences across monitoring systems: while some can extract information with a certain cut-off date, others can extract data without imposing restrictions or filters. No solution could therefore be envisaged to cope with this specific limitation;
- **Missing key information on the EU standard categories:** As illustrated in Section 3.2.1, there is no full coverage of the variable Thematic Objective, EU Investment Priority and Field of Intervention, nor of the related shares. It would not be possible to restructure the full databases at TO, IP and/or FoI levels. This would be possible only for those operations for which such information is available;
- **Impossibility to attribute a univocal definition to some operations:** As mentioned in Section 3.2.1, not all operations have a univocal definition because no information was provided by the MAs, and so only a hybrid label combining more than one definition could be assigned. Unfortunately, there is no solution to this limitation;
- **Incomplete list of beneficiaries:** As mentioned in Section 3.2.2, the Single Database does not include the complete list of beneficiaries of all programmes: in some cases, only the information on the lead beneficiary is available; in others, the MA did not even clarify whether the list is complete or not, for financial and other intermediated instruments, data on the final recipients are available only in a few cases. There is also a high variability in the type and degree of granularity of the data provided across programmes and within the Member States. However, this reflects the great differences across monitoring systems, and there is no possibility to overcome these issues;
- **Incomplete data on indicators at operations level:** The Database of Indicators does not cover all programmes funded by the ERDF and CF, since in some cases, data was not provided by the MA. However, it is not possible to retrieve such data from other sources than monitoring systems.

ANNEXES

Annex I: Previous data collection projects (as of end 2021)

In the last few years, the European Commission launched several studies aimed at collecting data on operations and beneficiaries funded by the ERDF and Cohesion Fund. They are:

Kohesio (by DG REGIO and DG CNECT) – With the aim of ensuring transparency, accessibility and reuse of publicly available data on beneficiaries and projects funded by the ESIF, DG REGIO and DG CNECT of the European Commission joined forces to establish a new knowledge management platform, Kohesio. The platform aims at collecting, harmonising and visualising data on operations retrieved from the publicly available lists of ERDF, CF and ESF operations. In its pilot phase, the project focused on collecting data for six Member States - Czechia, Denmark, Ireland, Italy, France and Poland. The project was expected to continue with data collection in all remaining Member States in the second half of 2021. This initiative also served to develop a template and a guidance manual to standardise information from different funds, regions, programmes, languages and formats, with the ultimate goal of simplifying and harmonising the flow of data in the 2021-2027 programming period.

Data scraping pilot project (by Balazs Krich on behalf of DG REGIO) – In 2020, DG REGIO launched a pilot project to test the possibility of automated data collection, cleaning and analysis of public lists of operations. The project aimed at automatically searching the web for the list of operations related to national and regional Cohesion Policy programmes and Interreg programmes of the 2014-2020 period. Automated cleaning and harmonisation procedures were also tested. The project highlighted the impossibility of adopting fully automated and standardised procedures to collect the data and create a single coherent dataset due to the following main factors: i) the impossibility to find all the list of operations online (list for only 22 Member States plus Interreg programmes were actually identified); ii) the extreme variety in the formats and standards of data collected. Reproducible software solutions allowing for the automated access, transformation and partial cleaning of data have been developed by the project. However, the project also highlights the need to perform careful human inspection into each list in order to reduce the possibility of errors or incorrect transformations.

JRC Dataset of projects co-funded by the ERDF during the multi-annual financial framework 2014-2020 (by Julia Bachtrögler, Mathieu Doussineau, Peter Reschenhofer for the JRC). Under an initiative funded by the Stairway to Excellence project, the Joint Research Centre (JRC) aimed at designing and producing a structured dataset of projects co-funded by the ERDF during the multi-annual financial framework 2014-2020. This project was carried out between 2019 and 2020 and focused on variables for which a legal publication requirement exists (as from Regulation (EU) No 1303/2013 Article 115(2)). It led to an Excel dataset of operations, plus detailed information on data quality and completeness in each Member State. In addition to using the lists of operations available online, the authors contacted a number of MAs to obtain any missing data. The database was enriched by linking ERDF operations to the JRC Eye@RIS3 and the Horizon 2020 databases. Text mining techniques were used to include additional attributes, such as key enabling technologies (KET) and societal grand challenges (SGC) associated with operations. However, some key harmonisation procedures (e.g., on the definition of operations across different programmes) were not implemented. The project was relaunched in March 2021, with the aim of updating the existing dataset.

Ex-post evaluation of the ERDF programmes supporting RTD investment in the 2007-2013 period (by CSIL, Prognos and Technopolis on behalf of DG REGIO) – The ex-post evaluation of ERDF-funded investment in Research and Technological Development (RTD) infrastructures and activities in the 2007-2013 programming period included the development of a Single Database of projects and beneficiaries of 53 ERDF programmes. While focusing only on projects funded under the 01 and 02 codes of expenditure, it represented the first systematic attempt to collect project-level expenditure data of the

Cohesion Policy. A wide set of cleaning procedures were carried out in order to find homogeneous typologies of interventions and beneficiaries. This project highlighted the advantages of collecting data at the level of individual operations and beneficiaries, as well as the main underlying challenges.

Annex II: EU regulatory requirements regarding monitoring data

At the European Commission level, the Cohesion Policy monitoring is carried out at the operational programme (OP) level. In particular, provisions on the performance framework and financial data aimed to monitor the progress in programme implementation and achievement of objectives require that data is made available at the priority axis level, aggregating data of individual operations. Precise rules are provided in the Regulations for collecting, storing and making available (even publicly) information on operations and beneficiaries. Monitoring requirements for the 2014-2020 period related to data **at the level of individual operations** are laid down in two Commission's Regulations:

- According to **Art. 125(2d)** of the Commission **Regulation (EU) No 1303/2013** (Common Provisions Regulation - CPR), the Member States and MAs are required to **establish a system to record and store in computerised form data on each operation** necessary for monitoring, evaluation, financial management, verification and audit, including **data on individual participants in operations**. In addition, **Article 115(2)** states that to ensure transparency, the Member States or MAs shall make the list of operations accessible through a single website. This list should be structured by type of fund and operational programme. It should be prepared in spreadsheet data format (e.g., CSV or XML), which could be "sorted, searched, extracted, compared and easily published on the internet". Annex II of the Regulation specifies the minimum set of data to be made publicly available at the operation level.
- According to **Article 24 of the Commission Delegated Regulation (EU) N. 480/2014** supplementing the CPR, and its Annex III, the MA are asked to record and store in a computerised form in their monitoring system a long list of information for each operation and individual participants, additional to the data listed in the CPR. This data is supposed to be used when necessary for the purposes of monitoring, evaluation, financial management, verification and audit. They should be recorded in a manner that allows the data to be aggregated and broken down by OP, priority, Fund or category of region.

The information that should be available at the MAs in compliance with the above Regulation is presented in the following Table.

Table 2 – Data requirements at operation and beneficiary level from EU Regulations

	Minimum set of data to be made publicly available	Data to be recorded and stored in a computerised form in the monitoring systems
	... as from Regulation (EU) No 1303/2013 Article 115(2)	... as from Regulation (EU) No 480/2014
Data at beneficiary level	<ul style="list-style-type: none"> • Beneficiary name 	<ul style="list-style-type: none"> • Name or unique identifier of each beneficiary • Information whether the beneficiary is public law body or private law body • Information whether VAT on expenditure incurred by the beneficiary is non-recoverable under national VAT legislation • Data on payment claims from the beneficiary (in the currency applicable to the operation) – data fields 44-52
Data at operation level	<ul style="list-style-type: none"> • Operation name • Operation summary • Operation start date and end date (expected for 	<ul style="list-style-type: none"> • Name or unique identifier of the operation • Short description of the operation • Date of submission of the application for the operation

	<p>physical completion or full implementation)</p> <ul style="list-style-type: none"> • Total eligible expenditure allocated to the operation • Union co-financing rate, as per Priority Axis • Operation postcode, or other appropriate location indicator • Country • Name of category of intervention for the operation • Date of last update of the list of operations 	<ul style="list-style-type: none"> • Starting date as indicated in the document setting out the conditions for support • End date as indicated in the document setting out the conditions for support • Actual date when the operation is physically completed or fully implemented • Body issuing the document setting out the conditions for support • Date of the document setting out the conditions for support • Information whether the operation is a major project and CCI • Information whether the operation is a joint action plan and CCI • Information whether the public support for the operation will constitute State aid • Information whether the operation is implemented under a public-private-partnership structure • Currency of the operation • CCI of the programme(s) under which the operation is supported • Priority or priorities of the programme(s) under which the operation is supported • Fund(s) from which the operation is supported • Category of region concerned • Code(s) for intervention field • Code(s) for form of finance • Code(s) for territory type • Code(s) for territorial delivery systems • Code(s) for thematic objective • Code(s) for ESF secondary theme • Code(s) for economic activity • Code(s) for location • Amount of the total eligible cost of the operation approved in the document setting out the conditions for support • Amount of the total eligible costs constituting public expenditure as defined in Article 2(15) of Regulation (EU) No 1303/2013 • Amount of public support as set out in the document setting out the conditions for support
--	--	--

Source: Authors based on EU Regulations

Annex III: Availability of public lists of operations and related data coverage

In total, **the lists for 213¹¹⁴ OPs out of 217¹¹⁵ and 75 CPs out of 76 can be collected from public websites**, thus covering 288 Programmes out of 293. No public data on beneficiaries and operations can be found and collected for the “small and medium-sized enterprise initiative” (SMEi) OPs in Bulgaria, Finland, Romania and Malta.

In each Member State (MS), the lists of operations of the OPs are accessible either through OP specific web pages (e.g., in Germany, each of the 16 OPs has a different list of operations, published on a different website), or through single national portals providing information on all the OPs in the MS (e.g., in Poland, Czech Republic). In some MS, there are multiple lists both at the national and the regional level (e.g., Italy, Portugal and France), usually not containing the same type of data.

For what the Cooperation Programmes are concerned, the Keep.eu portal is the main reference for the collection of public data, although in most cases, there are public lists of operations available also on the MA’s website. The Keep.eu portal generally covers a higher number of variables than the public lists of operations available on MA’s website (mainly harmonised) and reports the full list of beneficiaries. This dataset covers 71 CPs out of 76. For the four of the remaining CPs, the public lists on the MA’s website are available, while no public list is available for the CP Interreg V-A - Saint Martin-Sint Maarten (2014TC16RFCB043)¹¹⁶.

Some of the lists of operations publicly available not only provide data at the operation level but also on operations’ beneficiaries and output indicators at the level of the individual operation. This is the case, for instance, of the Italian data portal OpenCoesione, where data on beneficiaries and output indicators at the level of operation are provided for all OPs, and of the Keep.eu portal, where data on beneficiaries at the level of operation are provided for a total of 71/76 Cooperation Programmes.

The degree of data coverage for the variables the Core Team aimed to collect varies from one OP to another, depending on which data is actually available in the public lists. The degree of data coverage of public list of operations by variable is summarised in the tables below.

¹¹⁴ However, for the Competitiveness and sustainable development OP in Cyprus, the list of operations which could be collected from the website is not complete since it was not possible to download the full list directly.

¹¹⁵ Please note that this number does not take into account that the Spanish OP Smart Growth was merged with the Multi-regional OP, the Romanian SMEi OP with the Regional OP and the Slovakian OP Research and Innovation with the OP the Integrated Infrastructure.

¹¹⁶ On the website of the Managing Authority (http://europe.com-saint-martin.fr/CTE_beneficiaires-Saint-Martin-Antilles_41.html) the list of beneficiaries is not available.

Table 3 – Degree of data coverage of variables of the operations database

Variable	Share of programmes for which the variable has been filled in (N = 276)	Share of operations for which the variable has been filled in (N = 429,018)
1. General information		
Language	100.00%	100.00%
Currency	99.64%	99.99%
Cut-off date	61.31%	83.67%
2. Operational/Cooperation Programme identification		
CCI number	100.00%	100.00%
CCI short title	100.00%	100.00%
Country code	100.00%	100.00%
Type of OP EU Fund	99.64%	98.44%
OP territorial scope (cross-border, interregional, national, regional, or transnational, multiregional)	99.64%	98.44%
NUTS coverage	30.66%	1.65%
OP category of region	7.66%	17.88%
3. Operation identification		
Official operation Identification Code	78.47%	86.36%
Ad-hoc Operation Identification Code	100.00%	100.00%
Portfolio name	11.68%	18.07%
Operation name	99.64%	99.99%
Operation name (in English)	72.99%	98.60%
State aid relevance	10.95%	17.94%
Operation summary	96.72%	94.97%
Operation summary (in English)	94.89%	94.79%
Call for proposals - code	20.44%	3.21%
Call for proposals - title	0.36%	1.92%
Operation start and Operation start year	98.91%	97.30%
Expected operation end	13.50%	25.75%
Actual operation end and Operation end year	98.91%	96.27%
Operation delay	13.50%	22.40%
Operation duration (months) and Operation duration (years)	98.91%	95.84%
Operation status	60.22%	79.41%
Major Project and Major Project code	10.95%	17.94%
Major Project name/acronym	0.00%	0.00%
Major Project description (in national language and in English)	3.28%	0.02%
Intervention field	64.96%	95.29%
Thematic Objective	54.74%	50.89%
OP Priority Axis - code	41.61%	65.56%
OP Priority Axis (original)	32.85%	55.86%
OP Priority Axis (in English)	32.12%	55.82%
EU co-financing rate at priority axis level	54.01%	91.21%
Investment priority - code	42.34%	29.06%
Priority Axis Specific Objective - code	13.14%	19.42%
Priority Axis Specific Objective (in national language and in English)	11.68%	18.05%
OP action/measure - code	12.41%	21.17%
OP action/measure (in national language and in English)	8.76%	16.14%
4. Operation's features		
Operation type of EU Fund	77.37%	65.25%
Form of finance - code	18.98%	37.94%
NUTS1 and NUTS2	71.53%	93.16%
NUTS3	63.87%	83.43%
Operation category of region	0.73%	0.48%
Territory Type - code	17.88%	32.34%
Territorial Delivery Mechanisms	18.25%	36.13%
Economic Activity	10.58%	19.74%
Taxonomy of operations	0.00%	0.00%

5. Operation's financial information		
Total operation cost (EUR)	42.70%	21.39%
Total eligible expenditure (allocated) (EUR)	68.25%	97.83%
Total eligible expenditure (paid) (EUR)	14.23%	21.09%
EU co-financing rate at operation level	13.14%	15.62%
EU contribution to the operation (allocated) (EUR)	49.64%	23.24%
EU contribution to the operation (paid) (EUR)	0.36%	0.08%
ERDF contribution to the operation (allocated) (EUR)	46.35%	47.65%
ERDF contribution to the operation (paid) (EUR)	4.01%	9.39%
CF contribution to the operation (allocated) (EUR)	2.92%	2.90%
CF contribution to the operation (paid) (EUR)	2.55%	7.81%
National/regional public contribution to the operation (allocated) (EUR)	16.79%	27.47%
National/regional public contribution to the operation (paid) (EUR)	4.01%	9.33%
Private financing of the operation (allocated) (EUR)	13.87%	16.65%
Private financing of the operation (paid) (EUR)	2.55%	7.81%
Conversion rate used (national currency for 1 €)	99.64%	99.99%

Source: Authors

Table 4 – Degree of data coverage of variables of the beneficiaries database

Variable	Share of programmes for which the variable has been filled in	Share of beneficiaries for which the variable has been filled in
1. Beneficiary identification		
Beneficiary name (original language)	99.64%	99.86%
Beneficiary name (in English)	73.36%	92.83%
Beneficiary VAT code	23.72%	48.25%
Beneficiary other unique identifier	0.00%	0.00%
Beneficiary NUTS1	52.92%	33.69%
Beneficiary NUTS2	52.92%	33.56%
Beneficiary NUTS3	52.92%	33.30%
Country of the beneficiary	53.28%	34.35%
Role of the beneficiary in the operation	87.59%	82.24%
Intermediary name (original) and (in English)	0.00%	0.00%
2. Beneficiary's characteristics		
Beneficiary sector 1 dgt and 2 dgt	0.00%	0.00%
Type of beneficiary	0.00%	0.00%
Ownership structure of the beneficiary	25.91%	6.97%
Number of employees and Size of the beneficiary	0.00%	0.00%
CORDIS	0.00%	0.00%
H2020 Operation ID	0.00%	0.00%
ORBIS BvD	0.00%	0.00%
Type of intermediary	0.00%	0.00%
3. Financial information on the beneficiary		
Total eligible cost expenditure to the beneficiary (allocated) (EUR)	21.53%	6.55%
Total eligible cost expenditure to the beneficiary (paid) (EUR)	0.00%	0.00%
Total EU contribution to the beneficiary (allocated) (EUR) and (paid)	0.00%	0.00%
ERDF contribution to the beneficiary (allocated) (EUR)	18.98%	5.46%
ERDF contribution to the beneficiary (paid) (EUR)	0.00%	0.00%
CF contribution to the beneficiary (allocated) (EUR) and (paid)	0.00%	0.00%
National/regional public contribution to the beneficiary (allocated) (EUR) and (paid)	0.00%	0.00%

Private financing of the beneficiary (allocated) (EUR) and (paid)	0.00%	0.00%
---	-------	-------

Source: Authors

Annex IV: Data request to the MAs

The country correspondents provided the MA with a concise note describing the types of data requested. It is set out below.

“Study on the monitoring data on ERDF and Cohesion Fund operations, and on the monitoring systems operated in the 2014-2020 period.”

by the European Commission DG REGIO

Service Contract No 2019CE16BAT214 I 2020CE16BAT075

TECHNICAL NOTE GUIDING THE DATA REQUEST

The present document has been developed by CSIL in order to provide the MAs with more specific information on the data request regarding ERDF and CF funded operations, related beneficiaries and achievement indicators (i.e. common and programme-specific output indicators).

In order to contribute to the success of the study, we kindly ask the MA to provide data on funded operations and related beneficiaries as well as on achievement indicators. To ease the identification of the useful data, we will refer as much as possible to the ones mentioned in **Annex III of Regulation (EU) No 480/2014**. In particular, we would need the following set of data:

- Data on the beneficiary (fields 1-3)
- Data on the operation (fields 5-22)
- Data on categories of intervention (fields 23-30)
- Data on output indicators (fields 31-34): for common output indicators, data on operation level is needed; for programme-specific output indicators, data on the level of the measure/action/instrument would be sufficient if not available at operation level
- Financial data on operations (allocations) (fields 41-43)
- Final expenditure data on operations: drawing from fields 44-105, as a minimum we would need: the total operation cost, the total amount of eligible expenditure declared to the Commission established on the basis of costs actually incurred and paid (see field 53), and the total paid ERDF/CF, national and private contribution to each beneficiary of the operation;

We also kindly ask you to provide any additional data **at operation level, or at beneficiary level, but that can be consolidated at the operation level**, which is included in your monitoring system beyond the ones mentioned above and which could be useful for future evaluation purposes. For instance:

- On beneficiaries (not only the lead organisation/project manager but also any project partner):
 - Any information that would help us identify and group the typology of beneficiaries and, if any, final recipients (for instance, distinguishing by universities, enterprises, municipalities, etc.); please note that we are not collecting any personal and sensitive data;
 - Any information available on their location (e.g., address, postal code city, or at least the NUTS3 or NUTS2 region) and typology of the location (according to any taxonomy used within the monitoring system);
 - The VAT code of the beneficiary, if available;

- Any information on the role of the beneficiary in the operation concerned (e.g., lead beneficiary or partner).
- On funded operations:
 - Any other type of variable describing the typology or content of the operation or groups of operations, including longer descriptions of the operations, the specific objective of the Priority Axis; the related measure/action/instrument of the Priority Axis and, if available, a short description; the investment priority(ies);
 - The status of the operation at the date of update (e.g. completed or ongoing);
 - If possible/applicable, the reference to the call for projects under which the operation was selected;
 - The EU co-financing rate applied at the level of operation.
- On output indicators:
 - Any complementary information on indicators giving details on indicator definitions, calculation methods, measures/actions/instruments covered by each indicator. This information might be available in the form of handbooks, manuals, notes on indicator usage, etc. If these materials are readily available, we would like to receive them now.

The following general remarks should be considered:

- **Unique operation identifier:** We invite MAs to provide the data on the different sets of information requested (i.e. operations, beneficiaries, expenditure and achievement indicators) **at the level of operation**, when available. Each operation should be identifiable and distinguishable from others based on a unique operation identifier (i.e. operation unique number/code). The latter should be included in all the data extractions provided, including those on beneficiaries and indicators.
- **Definition of operation:** Considering that different definitions of ‘operation’ may be adopted across monitoring systems (see the box below), MAs are asked to specify the definition of ‘operation’ used within the monitoring system and provide any additional identifier codes that could allow subsequent aggregations, if necessary (e.g. at the level of projects or groups of projects).



The definition of ‘operation’ according to Art. 2(9) in the Regulation (EU) N°1303/2013

“An operation means a project, contract, action or group of projects selected by the MAs of the programmes concerned, or under their responsibility, that contributes to the objectives of a priority or priorities; in the context of financial instruments, an operation is constituted by the financial contributions from a programme to financial instruments and the subsequent financial support provided by those financial instruments”.

- **Definition of beneficiary:** For the purpose of this data collection, the definition of beneficiary included in the Regulation (EU) N°1303/2013 (see the box below) generally applies. However, in the context of financial instruments or other OP measures managed through intermediate bodies, MAs are also kindly asked to provide data on the final recipients of EU funding, in case they are available, or to indicate which body owns this kind of data.



The definition of ‘beneficiary according to Art. 2(10) in the Regulation (EU) N°1303/2013

“Beneficiary means a public or private body [...] responsible for initiating or both initiating and implementing operations; and in the context of State aid schemes, as defined in point 13 of

this Article, the body which receives the aid; and in the context of financial instruments under Title IV of Part Two of this Regulation, it means the body that implements the financial instrument or the fund of funds as appropriate”.



The definition of ‘final recipient according to Art. 2(12) in the Regulation (EU) N°1303/2013

“Final recipient means a legal or natural person receiving financial support from a financial instrument”.

- **Financial expenditure:** Regulation (EU) N° 480/2014 foresees to collect a large number of financial information. In general, we are interested in any financial data of allocations and expenditure (paid amounts) at the beneficiary level, aggregable at the level of operation, and possibly comparable with the expenditure data aggregated at the Priority Axis level that are regularly reported to the European Commission. When more variables are available, the MA can provide all of them, but we kindly ask to make explicit the specific definition of each variable to allow subsequent cleaning and harmonisation by the Study Team.
- **Cut-off date:** MAs are kindly asked to provide data on all operations, updated at 31/12/2020. The status of each operation at 31/12/2020 should be specified (whether it is completed or still ongoing). If data at 31/12/2020 are not available yet and not expected to become available before the end of March 2021, the most updated data should be provided, specifying the date of the last update.
- **Currency:** MAs operating in those Member States using a currency different from Euro can provide financial data either in their national currency or already converted into Euros or both. It is important to specify the currency used for reporting financial data. If conversion into Euro has been made, the exchange rate applied should be declared.

Annex V: Description of the additional .csv files, including operations and variables excluded from the Single Database

Additional .csv files accompany the Single Database of operations, beneficiaries and indicators. Such files include operations and additional information which were excluded from the Single Database either because they were not relevant, their coverage was low and varied across and within Member States or because of the impossibility to reconcile and harmonise the information provided. They are:

- **DB_operation_not_funded.csv:** this database includes the list of operations for which an application was submitted but which were not selected as eligible for EU funding;
- **DB_beneficiaries_not_funded.csv:** this database includes the list of beneficiaries of the operations for which an application was submitted but which were not selected as eligible for EU funding;
- **DB_operations_additional_variables.csv:** this database includes some variables available at operation level which were not included in the Single Database of operations;
- **DB_beneficiaries_additional_variables.csv:** this database includes some variables available at beneficiary level which were not included in the Single Database of beneficiaries.

At this stage, the last two databases, including additional variables on operations and beneficiaries are only preliminary and do not include all the information which were excluded from the Single Database. In what follows, the current structure of each database is presented, and the complete list of additional variables which could be provided in the final version of these databases is also included.

Annex V.1: The database on non-funded operations

This database includes the list of non-funded operations of 12 programmes, namely 6 French OPs and 1 Romanian OP and 5 CPs. In total, the database includes 2,388 operations that were not declared eligible for EU funding. Its structure is presented in the table below. It should be noted that the variables included have not been cleaned, harmonised or enriched.

Table 5 – Coverage of the database on non-funded operations

Variable
Date of extraction of the data (<i>date_extraction</i>)
CCI code of the programme (<i>op</i>)
Official operation identifier (<i>prj_nr</i>)
Another official operation identifier (<i>prj_nr_local</i>)
Ad-hoc operation identifier (<i>prj_ID</i>)
Operation name (original language) (<i>prj_name_In</i>)
Operation name (English) (<i>prj_name</i>)
Operation acronym (<i>prj_acronym</i>)
Operation summary (original language) (<i>prj_descr_In</i>)
Start date of the operation (<i>prj_start_date</i>)
Expected end date of the operation (<i>prj_expected_end</i>)
End date of the operation (<i>prj_end_date</i>)
Expected duration of the operation (<i>prj_lenght_months_check</i>)
Priority Axis code (<i>prj_priority_axis_code</i>)
Priority Axis title (original language) (<i>prj_priority_axis_title_In</i>)
Priority Axis title (English) (<i>prj_priority_axis_title</i>)
Specific Objective code (<i>prj_paxis_objective_code</i>)
(<i>prj_paxis_objective_title_In</i>)

Specific Objective title (original language) (<i>prj_paxis_objective_title</i>)
Specific Objective title (English) (<i>prj_op_measure_code</i>)
OP measure title (original language) (<i>prj_op_measure_title_ln</i>)
OP action code (<i>prj_op_action_code</i>)
OP action title (original language) (<i>prj_op_action_title_ln</i>)
OP action title (English) (<i>prj_call_code_title_ln</i>)
Call for proposals (English) (<i>prj_call_code_title_en</i>)
EU fund type of the operation (<i>prj_fund_type</i>)
Thematic Objective code (<i>prj_to_code</i>)
Thematic Objective title (<i>prj_to_title_en</i>)
EU Investment Priority code (<i>prj_ip_code</i>)
Field of Intervention code (<i>prj_foi_code</i>)
Field of Intervention title (<i>prj_foi_title</i>)
Field of Intervention share (<i>prj_foi_pro_rata</i>)
Form of finance code (<i>prj_form_of_finance_code</i>)
Territory type code (<i>prj_territory_type_code</i>)
Territorial delivery mechanism code (<i>prj_terr_deliv_mech_code</i>)
Economic activity code (<i>prj_econ_activity_code</i>)
Region category (<i>prj_region_category</i>)
Major project (<i>prj_mp</i>)
State aid (<i>prj_state_aid</i>)
Public-private partnership (<i>prj_public_private</i>)
Joint Action Plan (<i>prj_jap</i>)
Revenue-generating operation (<i>prj_revenues</i>)
NUTS2 region label (<i>prj_nuts2_label</i>)
NUTS2 region code (<i>prj_nuts2_code</i>)
NUTS3 region code (<i>prj_nuts3_code</i>)
NUTS3 region local code (<i>prj_nuts3_code_local</i>)
NUTS3 region share (<i>prj_nuts3_pro_rata</i>)
Postal code (<i>prj_postal_code</i>)
Other location information (<i>prj_location_mix</i>)
EU co-financing rate at Priority Axis level (<i>paxis_eu_cofinancing_rate</i>)
EU co-financing rate at operation level (<i>prj_eu_cofinancing_rate</i>)
Total eligible expenditure allocated (<i>prj_tot_exp_alloc</i>)
ERDF contribution allocated (<i>prj_erdf_alloc</i>)
Foreign contribution allocated (<i>prj_foreign_alloc</i>)
Contribution of the public beneficiary allocated (<i>prj_pub_benef_alloc</i>)
Contribution of the private beneficiary allocated (<i>prj_pri_benef_alloc</i>)

Annex V.2: The database on beneficiaries of non-funded operations

This database includes the list of beneficiaries of the non-funded operations of the 12 programmes mentioned above. In total, the database includes 7,835 beneficiaries. Its structure is presented in the table below. It should be noted that the variables included have not been cleaned, harmonised nor enriched.

Table 6 – Coverage of the database on beneficiaries of non-funded operations

Variable
Date of the data extraction (<i>date_extraction</i>)
CCI number of the programme (<i>op</i>)
Official operation identifier (<i>prj_nr</i>)
Ad-hoc operation identifier (<i>prj_ID</i>)
Operation acronym (<i>prj_acronym</i>)
Beneficiary name (original language) (<i>benef_name</i>)
Beneficiary VAT code and other identifier (<i>benef_vat_mix</i>)
Beneficiary SIRET (<i>benef_siret</i>)
Beneficiary other identifier (<i>benef_ID</i>)
Beneficiary role (<i>benef_role</i>)
Beneficiary type (English) (<i>benef_type</i>)
Beneficiary type (original language) (<i>benef_type_ln</i>)
Beneficiary sub-type (English) (<i>benef_type2</i>)
Beneficiary ownership (<i>benef_ownership</i>)

Beneficiary size (<i>benef_size</i>)
Beneficiary number of employees (<i>benef_n_employee</i>)
Beneficiary VAT regime (<i>benef_vat_regime_ln</i>)
Beneficiary receiving revenues (<i>benef_revenues_generating</i>)
Beneficiary country (<i>benef_country_ln</i>)
Beneficiary NUTS1 code (<i>benef_nuts1_code</i>)
Beneficiary NUTS1 label (<i>benef_nuts1_label</i>)
Beneficiary NUTS2 code (<i>benef_nuts2_code</i>)
Beneficiary NUTS2 label (<i>benef_nuts2_label</i>)
Beneficiary NUTS3 code (<i>benef_nuts3_code</i>)
Beneficiary NUTS3 label (<i>benef_nuts3_label</i>)
Beneficiary city (<i>benef_city</i>)
Beneficiary postal code (<i>benef_postal_code</i>)
EU co-financing rate at beneficiary level (<i>benef_eu_cofinancing_rate</i>)
Total cost allocated to the beneficiary (<i>benef_tot_cost</i>)
Total eligible expenditure allocated to the beneficiary (<i>benef_tot_exp_alloc</i>)
Public eligible expenditure allocated to the beneficiary (<i>benef_public_alloc</i>)
ERDF contribution allocated to the beneficiary (<i>benef_erdf_alloc</i>)
Member State contribution allocated to the beneficiary (<i>benef_ms_alloc</i>)
Private contribution allocated to the beneficiary (<i>benef_private_alloc</i>)
Beneficiary contribution allocated (<i>benef_pub_pri_benef_alloc</i>)
Foreign contribution allocated (<i>benef_foreign_alloc</i>)
Allocated contribution of private beneficiary (<i>benef_pri_benef_alloc</i>)
Allocated contribution of public beneficiary (<i>benef_pub_benef_alloc</i>)
Revenues amount received by the beneficiary (<i>benef_revenues_amount</i>)

Annex V.3: The database on excluded variables at operation level

This database includes the variables at operation level, which were excluded from the Single Database of operations. Its structure, presented in the table below, is preliminary, and it does not include the entire list of variables excluded from the Single Database (which are, however, listed below). It should be noted that the variables included have not been cleaned, harmonised nor enriched.

Table 7 – Coverage of the database excluded variables at operation level

Variable	Reason for exclusion
CCI code of the programme (<i>op</i>)	n.a.
Official operation identifier (<i>prj_nr</i>)	n.a.
Ad-hoc operation identifier (<i>prj_ID</i>)	n.a.
Ad-hoc operation identifier (<i>prj_row_ID</i>)	n.a.
Operation duration (<i>prj_lenght_months</i>)	Low coverage
Typology of financial instrument (<i>prj_fin_instrument_type</i>)	Low coverage
Type of projects (individual or collaborative) (<i>prj_collab_type</i>)	Low coverage
Local taxonomy of the operation (<i>prj_local_taxonomy</i>)	Not relevant
Local sub-taxonomy of the operation (<i>prj_local_taxonomy2</i>)	Not relevant
Information on operations phased in 2007-2013 or 2021-2027 (<i>prj_phased</i>)	Low coverage
City of the operation (<i>prj_city</i>)	Low coverage and no standard taxonomy available
Postal code of the operation (<i>prj_postal_code</i>)	Low coverage and no standard taxonomy available
Local Administrative Unit(s) of the operation (<i>prj_lau</i>)	Low coverage
Non eligible cost of the operation (<i>prj_non_eligible_cost</i>)	Low coverage
Member State allocated contribution, including a mix of undefined public sources (<i>prj_ms_alloc</i>)	Impossibility to reconcile information
Member State paid contribution, including a mix of undefined public sources (<i>prj_ms_paid</i>)	Impossibility to reconcile information
National allocated contribution (<i>prj_national_alloc</i>)	Low coverage and impossibility to reconcile information
National paid contribution (<i>prj_national_paid</i>)	Low coverage and impossibility to reconcile information

Regional allocated contribution (<i>prj_regional_alloc</i>)	Low coverage and impossibility to reconcile information
Regional paid contribution (<i>prj_regional_paid</i>)	Low coverage and impossibility to reconcile information
Department/district allocated contribution (<i>prj_department_alloc</i>)	Low coverage and impossibility to reconcile information
Local allocated contribution (<i>prj_municipality_alloc</i>)	Low coverage and impossibility to reconcile information
Local paid contribution (<i>prj_municipality_paid</i>)	Low coverage and impossibility to reconcile information
Undefined public expenditure allocated contribution (<i>prj_exp_pub_alloc</i>)	Low coverage and impossibility to reconcile information
Other undefined expenditure allocated contribution (<i>prj_other_alloc</i>)	Low coverage and impossibility to reconcile information
Other undefined expenditure paid contribution (<i>prj_other_paid</i>)	Low coverage and impossibility to reconcile information
Mix of national and private allocated contribution (<i>prj_ms_private_alloc</i>)	Low coverage and impossibility to reconcile information
Mix of national and private paid contribution (<i>prj_ms_private_paid</i>)	Low coverage and impossibility to reconcile information
Mix of Member state and EU allocated contribution (<i>prj_erdf_ms_alloc</i>)	Low coverage and impossibility to reconcile information
Mix of Member state and EU paid contribution (<i>prj_erdf_ms_paid</i>)	Low coverage and impossibility to reconcile information
Allocated contribution of the public/private beneficiary (<i>prj_pub_pri_benef_alloc</i>)	Low coverage and impossibility to reconcile information
Paid contribution of the public/private beneficiary (<i>prj_pub_pri_benef_paid</i>)	Low coverage and impossibility to reconcile information
Foreign allocated contribution (<i>prj_foreign_alloc</i>)	Low coverage/Not relevant
Foreign paid contribution (<i>prj_foreign_paid</i>)	Low coverage/Not relevant
ENI allocated contribution (<i>prj_eni_alloc</i>)	Low coverage/Not relevant
IPA allocated contribution (<i>prj_ipa_alloc</i>)	Low coverage/Not relevant
IPA paid contribution (<i>prj_ipa_paid</i>)	Low coverage/Not relevant
Revenues amount (<i>prj_revenues_amount</i>)	Low coverage and impossibility to reconcile information

Annex V.4 The database on excluded variables at beneficiary level

This database includes the variables at beneficiary level, which were excluded from the Single Database of operations. Its structure, presented in the table below, is preliminary, and it does not include the entire list of variables excluded from the Single Database (which are, however, listed below). It should be noted that the variables included have not been cleaned, harmonised nor enriched.

Table 8 – Coverage of the database excluded variables at beneficiary level

Variable	Reason for exclusion
CCI code of the programme (<i>op</i>)	n.a.
Official operation identifier (<i>prj_nr</i>)	n.a.
Ad-hoc operation identifier (<i>prj_ID</i>)	n.a.
Ad-hoc beneficiary identifier (<i>benef_row_ID</i>)	n.a.
City of the beneficiary (<i>benef_city</i>)	Low coverage and no standard taxonomy available
Postal code of the beneficiary (<i>benef_postal_code</i>)	Low coverage and no standard taxonomy available
Local Administrative Unit(s) of the beneficiary (<i>benef_lau</i>)	Low coverage
Member State allocated contribution, including a mix of undefined public sources (<i>benef_ms_alloc</i>)	Impossibility to reconcile information

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

Member State paid contribution, including a mix of undefined public sources (<i>benef_ms_paid</i>)	Impossibility to reconcile information
National paid contribution (<i>benef_national_paid</i>)	Low coverage and impossibility to reconcile information
Local allocated contribution (<i>benef_municipality_alloc</i>)	Low coverage and impossibility to reconcile information
Local paid contribution (<i>benef_municipality_paid</i>)	Low coverage and impossibility to reconcile information
Other undefined public expenditure paid contribution (<i>benef_other_public_paid</i>)	Low coverage and impossibility to reconcile information
Other undefined expenditure allocated contribution (<i>benef_other_alloc</i>)	Low coverage and impossibility to reconcile information
Mix of national and private allocated contribution (<i>benef_ms_private_alloc</i>)	Low coverage and impossibility to reconcile information
Mix of Member state and EU allocated contribution (<i>benef_erdf_ms_alloc</i>)	Low coverage and impossibility to reconcile information
Mix of Member state and EU paid contribution (<i>benef_erdf_ms_paid</i>)	Low coverage and impossibility to reconcile information
Allocated contribution of the public/private beneficiary (<i>benef_pub_pri_benef_alloc</i>)	Low coverage and impossibility to reconcile information
Paid contribution of the public/private beneficiary (<i>benef_pub_pri_benef_paid</i>)	Low coverage and impossibility to reconcile information
Foreign allocated contribution (<i>benef_foreign_alloc</i>)	Low coverage/Not relevant
Foreign paid contribution (<i>benef_foreign_paid</i>)	Low coverage/Not relevant
ENI allocated contribution (<i>benef_eni_alloc</i>)	Low coverage/Not relevant
IPA allocated contribution (<i>benef_ipa_alloc</i>)	Low coverage/Not relevant
Revenues amount (<i>benef_revenues_amount</i>)	Low coverage and impossibility to reconcile information

Annex VI: The results of the matching of the database of beneficiaries with the ORBIS database

The match of the database of beneficiaries with the ORBIS database has been performed following several steps:

1. After the harmonisation of the beneficiary names, beneficiaries including specific keywords clearly identifying beneficiaries other than private entities were excluded from the sample to reduce its size and the time needed for the match. The keywords used were: “University”, “Minist”, “Municipality”, “Region”, “City”, “Authority”, “Polytechnic”, “Chamber”, “Commune”, “District office”, “County”, “Council”, “National School”, “Museum”, “Hospital”;
2. Search in the ORBIS database of all the beneficiaries retained in the sample after Step 1, using several variables such as the name, VAT code, city and country¹¹⁷;
3. Fine-tuning of the search in the ORBIS database for only those countries where the coverage of the BvD ID variable was lower than 10% and of those beneficiaries classified as “enterprise”, for which the VAT code was available.

As a result of the above-mentioned steps, the extent of the match of the database of beneficiaries with the ORBIS database is presented, by country, in the table below.

Table 9 – Results of the matching of the database of beneficiaries with the ORBIS database

Country	Total number of unique beneficiaries classified as enterprises	% of enterprises found in ORBIS
HU	9,278	99.11%
BE	2,542	98.03%
AT	953	96.54%
PT	40,552	93.08%
EE	2,293	91.63%
BG	29,746	89.50%
CY	629	88.08%
CZ	21,919	86.91%
UK	211	86.73%
SI	2,729	85.20%
IE	649	79.97%
LV	2,758	77.96%
ES	62,094	77.53%
FI	4,335	75.85%
PL	19,744	73.43%
HR	7,631	66.95%
FR	12,103	65.62%
DK	28	57.14%
DE	19,438	55.01%
IT	57,321	52.37%
MT	957	51.10%
RO	6,456	38.77%
GR	15,452	28.37%
LT	6,674	18.18%
SK	5,832	15.79%
NL	1,354	1.55%
SE	925	0.43%

¹¹⁷ These variables were still not cleaned nor harmonised at the moment of the match.

REPORT ON THE CONSTRUCTION OF THE SINGLE DATABASE

LU	2	0.00%
----	---	-------

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at:

https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696 or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: <https://op.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

