ESPON

EUROPEAN UNION

Co-financed by the European Regional Development Fund

Inspire policy making by territorial evidence

ESPON 2020 Database Portal

Scientific Platform

Delivery 7 – Final Report

Version 20/03/2020

This scientific platform activity is conducted within the framework of the ESPON 2020 Cooperation Programme, partly financed by the European Regional Development Fund.

The ESPON EGTC is the Single Beneficiary of the ESPON 2020 Cooperation Programme. The Single Operation within the programme is implemented by the ESPON EGTC and co-financed by the European Regional Development Fund, the EU Member States and the Partner States, Iceland, Liechtenstein, Norway and Switzerland.

This delivery does not necessarily reflect the opinion of the members of the ESPON 2020 Monitoring Committee.

Authors

University of Geneva (UNIGE), Lead Partner

- Hy Dao, Andrea De Bono, Cédric Gampert, Gregory Giuliani, Jean-Philippe Richard, Denisa Rodila

Autonomous University of Barcelona - UAB (Spain)

- Roger Milego Agrás, Jorge Lopez Perez

MCRIT (Spain)

- Oriol Biosca, Ramon Català, Jaume Jorba, Efrain Larrea, Andreu Ulied, Harold del Castillo

Spatial Foresight (S4S)

- Clément Corbineau, Erik Gloersen, Sebastian Hans

Project Steering Group

- Project Support Team: Radu Necsuliu (Romania), Ádám Móricz (Hungary),
- ESPON EGTC: Martin Gauk (Project Expert), Marjan van Herwijnen (Project Expert) Laurent Frideres (HoU E&O), Ilona Raugze (Director), Piera Petruzzi (Outreach), Johannes Kiersch (Financial Expert).

Acknowledgements

RIATE (France)

- Timothée Giraud, Ronan Ysebaert, Chloé Didelon

National University of Ireland Maynooth - NUIM (Ireland)

- Martin Charlton

Information on ESPON and its projects can be found on www.espon.eu.

The web site provides the possibility to download and examine the most recent documents produced by finalised and ongoing ESPON projects.

This delivery exists only in an electronic version.

ISBN: 978-2-919795-58-1

Contact: info@espon.eu

ESPON 2020 Database Portal

Sharing Essential ESPON Data

Table of contents

List of Figures

List of Tables

Abbreviations

API	Application Programming Interface
CDS	Core Database Strategy
CRUD	Create, Read, Update and Delete
CSV	Comma Separated Values
CSW	Catalog Service for Web (OGC)
DB	Database
DBMS	Database Management System
DBS	Database System
EC	European Commission
EGTC	European Grouping of Territorial Cooperation
ESPON	European Territorial Observatory Network
EU	European Union
FAQ	Frequently Asked Queries
GEOJSON	Geographical JavaScript Object Notation
GIS	Geographical Information System
INSPIRE	Infrastructure for Spatial Information in Europe
JSON	JavaScript Object Notation
LAU	Local Administrative Units
M4D	Multi-Dimensional Data Design and Development
NA	Not Applicable
NSO	National Statistical Office
NUTS	Nomenclature of Territorial Units for Statistics
OGC	Open Geospatial Consortium
PDM	Project Database Manager
REST	Representational State Transfer
SQL	Structured Query Language
UI	User Interface
UMZ	Urban Morphological Zones
URL	Uniform Resource Locator
WFS	Web Feature Service
WMS	Web Map Service
WP	Work Package
XML	Extensible Markup Language

1 Introduction

1.1 About this report

This report is the Delivery 7 of the ESPON 2020 Database Portal project (March 2017-March 2020), Service Contract EE/503/029/2017. In accordance with the terms of reference, this Final Delivery (Delivery 7) should include the following elements:

- Final version of the technical background document
- Updated version of the ESPON Database Portal including new, innovative and improved functionalities
- Report on the work done in relation to the updated version of the database portal
- Final version of administrator module to manage the database systems
- Final version of administrator module to manage the user interface of the Database Portal
- Final version of guidance document for both administrator modules
- Update according to the updating strategy
- Report on the work done in relation to the updating strategy
- Source code of all software developed in relation to the ESPON 2020 Database Portal
- A power point presentation (max. 30 slides), made up of the main components of the ESPON 2020 Database Portal, to be agreed with the ESPON EGTC after Delivery 6.

1.2 ESPON 2020 Database Portal

ESPON 2020 Database Portal project is one of the key projects of the ESPON 2020 Programme in terms of supporting the ESPON EGTC with regard to:

- Provision of core data and data-related support to the ESPON 2020 projects;
- Harmonisation and quality control of the main data outcomes from the ESPON 2020 projects;
- Facilitation of the searching and finding of ESPON 2020 data and data-related products;
- Publication of main ESPON 2020 results in an attractive way;
- Any task related to data, spatial data and data-related products.

The main objectives of the ESPON 2020 Database Portal project could be simplified and summarised in short as:

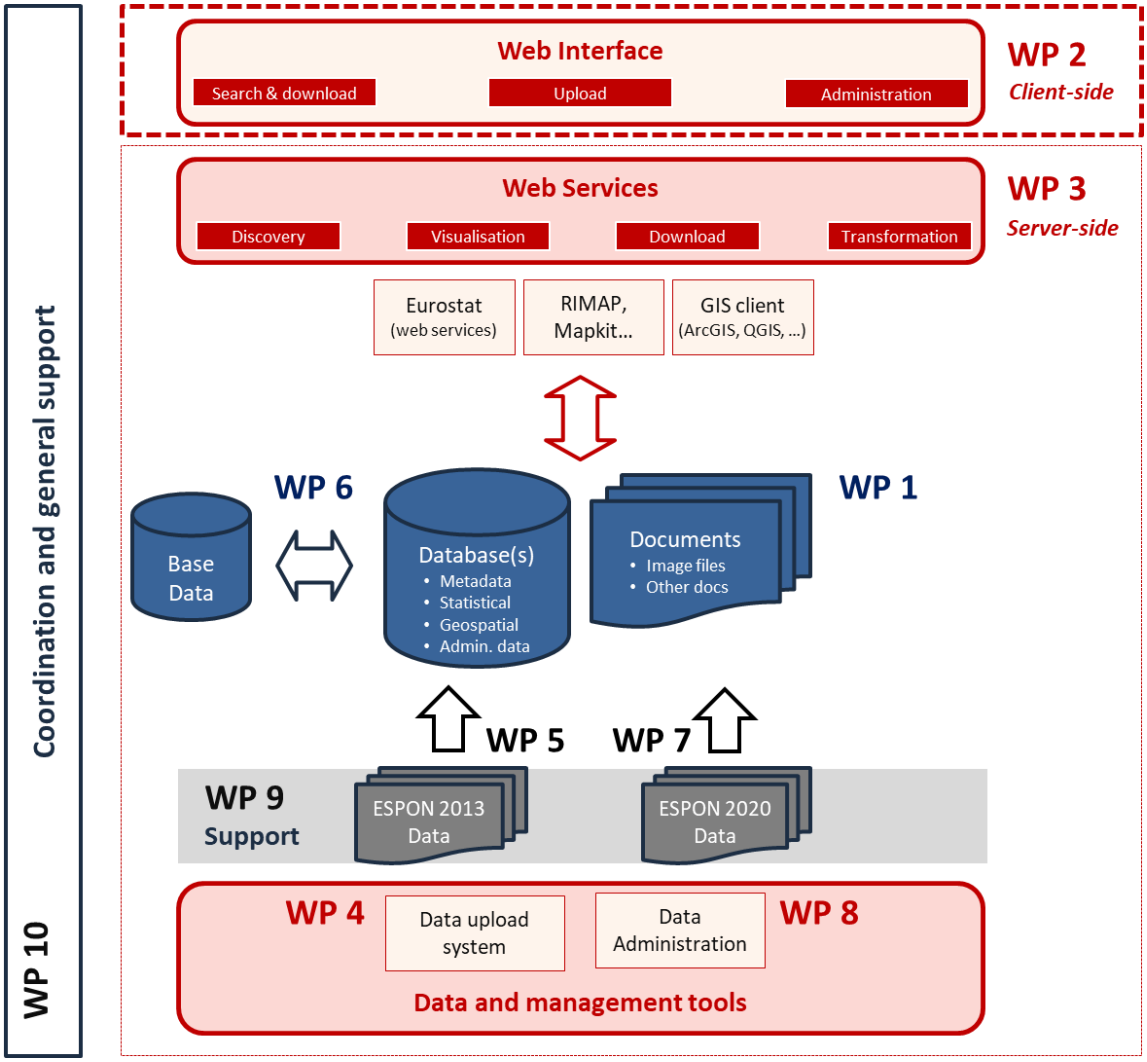
1. Maintain a set of core data of the highest possible quality and regularly updated, useful and usable for the work of the ESPON Programme and projects;

2. Collect the most relevant data (and metadata) from ESPON projects;
3. Disseminate the above data in a user- and machine-friendly manner to the ESPON community and to the outer World.

The ESPON 2020 Database Portal is organised in ten components, namely the work packages (WPs), as illustrated in Figure 1. The work packages are related to three main categories of activities:

- Data (WP 5, 6, 7), i.e. migration of data from the ESPON 2013 Database, Core datasets and new data from ESPON 2020 activities ;
- Systems (WP 1, 2, 3, 4, 8), comprising Database system, data & management tools, web services and web interface;
- General support and coordination (WP 9, 10).

Figure 1: ESPON 2020 Database Portal - Work Packages Organization



2 ESPON 2020 Data

2.1 ESPON 2013 Data Migration (WP5)

420 indicators (290 Base indicators and 130 Project indicators) from the ESPON 2013 Programme were incorporated in the new ESPON database 2020 schema. Table 1 lists the 2013 Projects and datasets imported. When relevant, the new multi-indicator structure was applied to the 2013 indicators (e.g. the Corine land cover), which resulted in a reduced list of multi and single indicators.

Table 1: List of the 2013 project datasets imported in the 2020 database

2013 Project	Dataset
ECR2 - Economic Crisis	Economic Resilience
ESPON CLIMATE - Climate Change and Territorial Effects on Regions and Local Economies in Europe	Exposure to climate change
	Potential impact of climate change
	Vulnerability and response to climate change
ESPON DB - M4D	Land Use Data in NUTS 2006 delineation
GREECO - Territorial Potentials for a Greener Economy	Greeco Emissions
	Regional Green Economic Performance
	Index of Green Economy Theoretical Potentials
	Greeco Potential
SeGI - Indicators and Perspectives for Services of General Interest in Territorial Cohesion and Development	SGI indicators
SEMIGRA - Selective Migration and Unbalanced Sex Ratio in Rural Regions	Population Structure
NSS - North Sea Star	Presence of North Sea Region Programme energy projects
TIPTAP - Territorial Impact Package for Transport and Agricultural Policies	Territorial impact assessment - Common Agricultural Policy

After the migration of the selected 2013 indicators, the exploration of data revealed missing territorial units for some indicators, as compared to those expected from the declaration of spatial extent, nomenclature, level and version. These missing rows in the data tables are considered as “no data”. A script was developed to insert the missing rows in the data table and assign them NULL values, in order to explicitly express the absence of data.

Moreover, some ESPON 2013 data have been detected to be orphaned in the sense that there were declared as Class indicators but assigned no upper Dimension nor Multi indicators, as it would be expected in these cases. The assignation of the adequate Multi/Dimension/Class or Single types was done after a thorough revision of these orphan indicators.

2.2 ESPON 2020 Core Data Updating Strategy (WP6)

Regarding the Updating strategy, the main objectives in the WP5 and WP6 have been achieved.

On the one hand, an application called **Updating tool** has been developed. This application is now totally integrated in the ESPON Data Portal and will allow a user to be guided in the whole process to update an ESPON base indicator, from generating a downloadable local environment with required functionality until the final process to integrate the free-outlier checked and approved changes in the production ESPON Database. The procedure about how the updating tool works is not complex, but it requires a supervision all along the process for each indicator by executing the scripts directly from a dashboard when required.

On the other hand, a new version of all the ESPON base indicators is delivered with this report by incorporating the most recent figures from a trusted source like Eurostat, as well as estimated values where no trusted sources are available, following the guidelines from ESPON M4D ESTI[1] framework and the catalogue of the approximation methods for ESPON time series. This new update with Eurostat values from official datasets, as downloaded in 22/10/2019, have been achieved following the methodology that has been described in the Updating Strategy, and it replaces the old indicators that were imported from Excel files and dated 09/03/2017.

2.2.1 Updating tool

All the source code of the developed tools can be found in the official git repository in a module called Updating[2] and this is not but another separate application in the ESPON Data Portal. That module is fully integrated in the ESPON Database and it takes advantage of other applications that have been developed in another work packages.

To plug this module into the Database Portal, an accessible folder in the server with writing/reading permissions is needed. The files generated by this application will be stored in this folder, to be further retrieved on demand by the application (Table 4).

This module contains five tables under the namespace updating: tdta_metadata, tdta_eurostat, tdta_missing_pattern, tdta_nulls y tdta_report. Table 2 provides more details for these tables. Those three first tables are the most important ones because they store data that is relevant for a correct working of the tool. The last two tables are optional.

This module also contains seven command scripts, namely python scripts, to be launched in the ESPON server via line commands by an expert user. These scripts have been renamed to reflect the order of execution. Table 3 provides more details about these scripts.

Table 2: Table names and a short description of the data that stores in the ESPON database

Table name	Description
tdta_metadata	This table contains information that is relevant to ESPON

[1] ESPON (2011) ESPON M4d - The Core Database Strategy – A new paradigm for data collection at regional level, pp. 70-103.

[2] https://git.unepgrid.ch/espon-egtc/edp/src/branch/update_base_indicator/django-edp/Updating - (for credentials to access the repository please contact the admin of the Database - database@espon.eu)

	indicators, basically metadata and a unique identifier that allows to group indicartors that are related (multi indicators)
tdta_eurostat	This table contains information about Eurostat tables as well as parameters that are needed to extract the counterpart ESPON indicators from those datasets
tdta_missing_pattern	This table contains information related to missing patterns detected for each of the countries within the same indicator
tdta_reports	This table contains a detailed report about the updates for a given indicator, providing the number of rows and affected tables in the ESPON database
tdta_nulls	This table contains detailed information about detected nulls for a given indicator

Table 3: Script names and a short description of the functionality

Script name	Description
1st_insert_metadata.py	This script populates tdta_metadata table
2nd_download_eurostat.py	This script download Eurostat tables from Eurostat services and stored those tsv files in a private folder in the server
3rd_insert_eurostat.py	This script populates tdta_eurostat table
4th_insert_nulls.py	This script populates tdta_nulls table
5th_insert_mp.py	This script populates tdta_missing_pattern table
6th_create_EUS.py	This script creates the Excel Updating Suite
7th_importer2ESPONDB.py	This script updates tdta_data_item, trel_data_item_source, trel_data_item_method and tdta_metadata tables

Table 4: Folder name in the server and a description of the files /folder that contains

Folder name	Description
<root_path>/eurostat/	The location of the folder in which datasets from Eurostat web services are downloaded. Not empty.
<root_path>/macros/	The location of the folder in which macros are generated for creating the Excel Updating Suite. A nested subfolder structure is generated. Initially empty.
<root_path>/eus/	An initially empty folder but EUS suite is generated in a subfolder called **eus**, as well as xlsx files are stored in

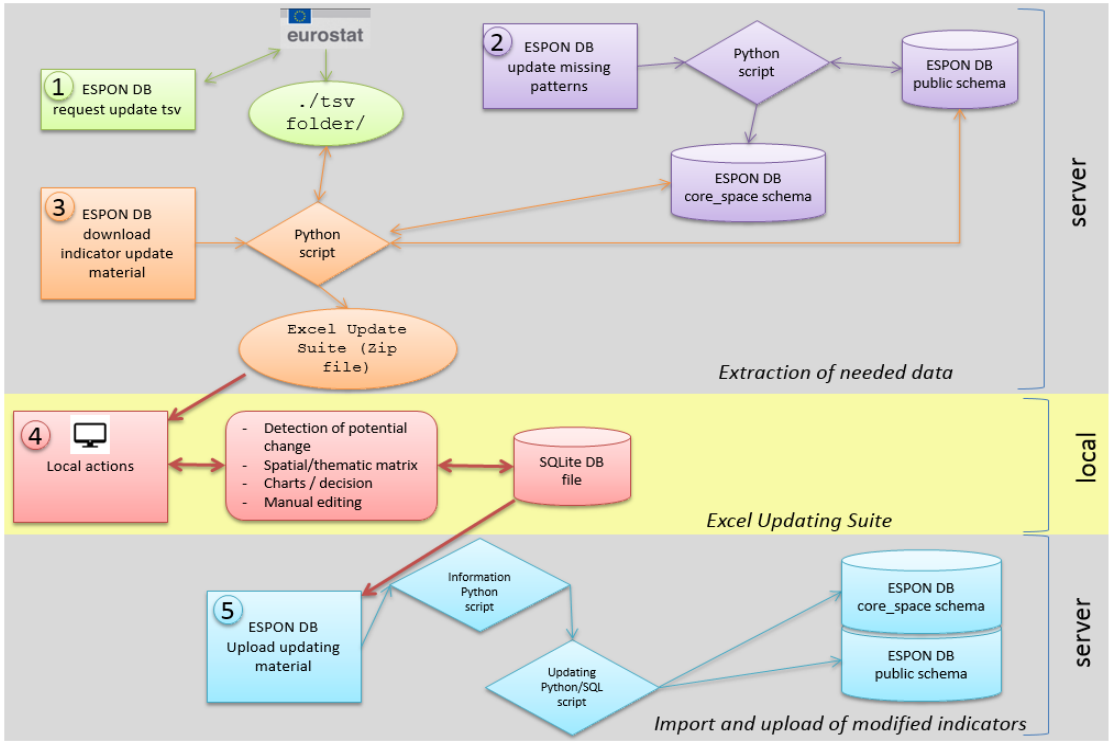
	another subfolder called **xslx** to make the zip automatically. A nested subfolder structure is generated to get zip files per country basis or per all countries.
<root_path>/db_updates/	The location of the folder in which user uploads database dumps in sqlite3 format.
<root_path>/db_reports/	The location of the folder in which a report is generated in text format is generated when user launch updates files against the ESPON database. Initially empty.

2.2.2 Updating procedure

Figure 2 gives a schematic illustration of how the updating process of a given ESPON indicator can be achieved with the Updating too, through two different but connected ways for the user to interact with the developed technology are underlying: local and server machines. There are three sequential stages from the beginning to the end of the updating:

- Building the Excel Updating Suite (ESPON Server)
- Updating Workflow (Local)
- Updating ESPON Database (ESPON Server)

Figure 2: Skecth of the three stages for the Updating an Indicator. First grey block corresponds to the Building the Excel Updating Suite, second yellow block corresponds to the Updating workflow and third grey block is the final Updating ESPON Database. Each stage contains numbers that represents the steps to follow



3.3.2.1 Building the Excel Updating Suite

Step 1, 2 and 3 in Figure 2 correspond to the execution of the first six pythons scripts as described above and they have to be run by an expert user that controls the ESPON server. Better improvements could be achieved by developing a User interface to control the launch of those scripts. Behind the scenes, several events take place:

- Download Eurostat tsv files (see 1 in Figure 2)
- Detecting nulls in ESPON Database
- Comparing between Eurostat figures and ESPON database to generate virtual changes, namely changes that are susceptible to modify the ESPON database but this decision will be taken in the Updating workflow
- Detection of missing patterns (see 2 in Figure 2)
- Parametrize scripts to be adapted for different countries in a given indicator
- Generation of Excel sheets to be downloaded as zip (see 3 in Figure 2)

Table 5: Sequential steps in the Updating workflow to updating a base indicator

#Step	Action	Description
1	Preparation of the data for the update (creation of the Excel Updating Suite)	After selecting the indicator to be updated, a process is run on the server to generate an Excel Updating Suite for each country, containing, in different sheets, the current indicator figures, the virtual changes derived from the comparison with Eurostat data, and a specific sheet with macro buttons to launch the following steps. Those Excel sheets are downloaded from the server within a ZIP file.
2	Detection of potential changes from Eurostat data, which are compliant with the spatial and thematic constraints of the specific indicator, when estimated data are involved.	After unzipping the Excel sheets, the process starts with one of the countries. In order to detect the potential and compliant changes, the corresponding macro should be run from the Excel Updating Suite, i.e. by pressing the *Launch* button
3	Creation of spatial and thematic validation matrices for the whole time-series.	In order to create the matrices, the corresponding buttons (*Spatial matrix* and *Thematic matrix*) in the Excel Updating Suite should be pressed.
4	Creation of supportive charts (line plots) of the indicator time-series.	Optional. In case of need, the expert in charge of the update can generate automatic line plots of the data, to have a graphical overview of the indicator time-series by pressing the *Charts* button.
5	Review of the resulting matrices to eventually decide if potential changes should be integrated or rejected.	The matrices provide the deviation of potential changes with regard to the spatial and thematic rules. It is up to the expert in charge of the update, to decide whether a specific deviation is acceptable or not.
6	Manual editing of potential changes to be rejected or modified.	All the potential changes which are not edited (or removed) in the corresponding Excel sheet will be eventually actual changes in the database. If some changes have been considered unacceptable, should be removed from the corresponding Excel sheet at this stage.

7	Detection of potential outliers in ESPON time series.	In case of need, the expert in charge of the update can launch two algorithms to detect potential outliers in the ESPON time series by pressing the *Outliers* button.
8	Save potential changes to the local DB.	By means of the *Save* button, the potential changes that remain in the Excel sheets are stored in a local database.
9	Introduction of additional figures from other sources and save the changes to the local DB.	There is the possibility to manually add potential changes coming from other sources (including estimated values). This is done in this step by pressing the corresponding buttons (*Start manual session* and *Detect manual changes*). If manual changes have been introduced, the potential changes need to be saved to the local DB again.
10	Repeat processes 1 to 9 for all the countries covered by the specific indicator.	
11	Prepare a single file for the specific indicator including all the countries.	
12	Upload the file to the ESPON DB Server and actually update the indicator.	

The main objective of those steps is the generation of the Excel Updating Suite (EUS) to be downloaded as a zip by the user in charge to update the ESPON indicators. EUS is a bunch of Excel files and folders with macros that are pre-generated in the server but different for each indicator and for each country Figure 3. This is the main requirement to start with the updating workflow.

Figure 3: Extracted files and folders from zip EUS file. Python scripts act as macros thanks to open-source Excel extension named xlwings (http://www.xwlings.org)

An Excel file with extension xlsm contains the data and metadata as extracted from the ESPON database. A detailed description of all the sheets in this Excel file can be found in the *Description of the main worksheets* section Deliverable D5 pag. 8-15. A Dashboard sheet (see Figure 3) controls the execution of the macros.

3.3.2.2 Updating workflow (Local)

Table 5 contains fully-explained sequential steps to follow in the process of updating a base indicator in a local machine by an expert user. As it has been emphasized that, due to the complexity and heterogeneity of the indicator changes and the sources of current indicator figures, the updating procedure cannot be totally automatized and there is a need for manual supervision by a trained expert.

There are no remarkable changes in names neither in the EUS suite nor those steps when compared to work as reported in previous deliverables. However, some improvements in functionality have been developed while the process to update all the ESPON indicators of the eight categories was in progress.

By observing the detected missing patterns, an automation procedure has been developed that does not cover all the possible patterns but the most common ones. Main new improvements in the tool are as follows:

- Generation of the estimated values with gap-filling techniques:
    - interpolation or/and retropolation in a concrete year by using values belonging to a totally complete year
    - completion of values in a thematic dimension for a multi indicator when possible
- Inclusion of the transfer keys for 2016-to-2013 nuts changes in the Updating tool for the affected countries
- Generation of parametrized scripts depending of the nature of indicator: single/multi, stock/ratio
- Control of the derived changes when new figures in Eurostat break the missing pattern
- Control and correction of null values in the database
- Correction of some metadata errors from previous import

In Figure 4 and Figure 5 a visual representation of interpolation and retropolation approaches is displayed for the multi-indicator named “Unemployment population by age and gender groups” in Austria. An image of the virtual changes as saved in local sqlite3 file (step 8) is also included.

Figure 4: Missing pattern for class indicator Unemployment male population with 15-74 age group in Austria. Interpolation for nuts2 in 2018 by using NUTS1-2 proportions in year 2017. Virtual change row as stored in local db (sqlite3)

unempl_males_15-74

	2015			2016			2017			2018		
c	dataitem_id	decimal_value	source	dataitem_id	decimal_value	source	dataitem_id	decimal_value	source	dataitem_id	decimal_value	source
tunit_code												
AT	2633704	142,2	S1	2640542	153,5	S1	4693401	141,6	S1	4693402	121,3	S1
AT1	2633705	85,5	S1	2640543	92,7	S1	4693403	85,9	S1	4693404	78,4	S1
AT11	2633706	3,5	S1	2640544	4,2	S1	4693405	3,6	S1			
AT12	2633707	24,6	S1	2640545	25,6	S1	4693406	24,3	S1	4693407	18,6	S1
AT13	2633708	57,3	S1	2640546	62,8	S1	4693408	57,9	S1	4693409	56,8	S1
AT2	2633709	24,5	S1	2640547	24,4	S1	4693410	23,3	S1	4693411	17,7	S1
AT21	2633710	9	S1	2640548	7,2	S1	4693412	6,9	S1	4693413	5,5	S1
AT22	2633711	15,5	S1	2640549	17,2	S1	4693414	16,4	S1	4693415	12,2	S1
AT3	2633712	32,2	S1	2640550	36,4	S1	4693416	32,4	S1	4693417	25,3	S1
AT31	2633713	17,3	S1	2640551	19,4	S1	4693418	17,2	S1	4693419	12,6	S1
AT32	2633714	5,3	S1	2640552	5,9	S1	4693420	5,1	S1	4693421	4,3	S1
AT33	2633715	6,1	S1	2640553	7,4	S1	4693422	5,7	S1	4693423	5,1	S1
AT34	2633716	3,6	S1	2640554	3,7	S1	4693424	4,3	S1	4693425	3,3	S1

unempl_males_15-74

	2015			2016			2017			2018		
c	dataitem_id	decimal_value	source	dataitem_id	decimal_value	source	dataitem_id	decimal_value	source	dataitem_id	decimal_value	source
tunit_code												
AT	2633704	142,2	S1	2640542	153,5	S1	4693401	141,6	S1	4693402	121,3	S1
AT1	2633705	85,5	S1	2640543	92,7	S1	4693403	85,9	S1	4693404	78,4	S1
AT11	2633706	3,5	S1	2640544	4,2	S1	4693405	3,6	S1		3,3	S2
AT12	2633707	24,6	S1	2640545	25,6	S1	4693406	24,3	S1	4693407	18,6	S1
AT13	2633708	57,3	S1	2640546	62,8	S1	4693408	57,9	S1	4693409	56,8	S1
AT2	2633709	24,5	S1	2640547	24,4	S1	4693410	23,3	S1	4693411	17,7	S1
AT21	2633710	9	S1	2640548	7,2	S1	4693412	6,9	S1	4693413	5,5	S1
AT22	2633711	15,5	S1	2640549	17,2	S1	4693414	16,4	S1	4693415	12,2	S1
AT3	2633712	32,2	S1	2640550	36,4	S1	4693416	32,4	S1	4693417	25,3	S1
AT31	2633713	17,3	S1	2640551	19,4	S1	4693418	17,2	S1	4693419	12,6	S1
AT32	2633714	5,3	S1	2640552	5,9	S1	4693420	5,1	S1	4693421	4,3	S1
AT33	2633715	6,1	S1	2640553	7,4	S1	4693422	5,7	S1	4693423	5,1	S1
AT34	2633716	3,6	S1	2640554	3,7	S1	4693424	4,3	S1	4693425	3,3	S1

Table: app_updating_virtual_changes

	indicator_id	indicator_m	indicator_code	country_code	year	virt_chng_type	tunit_code	old_src	new_src	old_value	new_value	dataitem_id	vch_date	status_chg	updating_type
	234	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	234	226	unempl_females_15-74	AT	1999	R3m	AT11	NULL	S2	NULL	3.1	NULL	2020-02-28...	not_saved	INSERT
2	234	226	unempl_females_15-74	AT	2000	R3m	AT11	NULL	S2	NULL	3.2	NULL	2020-02-28...	not_saved	INSERT
3	234	226	unempl_females_15-74	AT	2001	R3m	AT11	NULL	S2	NULL	2.5	NULL	2020-02-28...	not_saved	INSERT
4	234	226	unempl_females_15-74	AT	2002	R3m	AT11	NULL	S2	NULL	3.1	NULL	2020-02-28...	not_saved	INSERT
5	234	226	unempl_females_15-74	AT	2003	R3m	AT11	NULL	S2	NULL	3.4	NULL	2020-02-28...	not_saved	INSERT
6	234	226	unempl_females_15-74	AT	2004	R3m	AT11	NULL	S2	NULL	4.4	NULL	2020-02-28...	not_saved	INSERT

Table 6 displays the transfer keys to transfer indicators reported with NUTS 2016 version to NUTS 2013 version, as required from initial objectives. At the moment, Eurostat provides NUTS 2013 datasets for the counterpart ESPON demographic indicators, being the rest of indicators reported with the newest NUTS-2016 version. So, this value transfer has been automatized, especially for updating the stock class indicators under the employment categories.

As explained before, one single script cannot cover all the variability in nature of indicators and the heterogeneity of the missing pattern founds for ESPON indicator time-series. By parametrizing the scripts, an expert user could modify one or two varibles in these scripts, especially the variables related with the missing pattern (e.g., first year complete, FYC) when missing pattern was incorrectly estimated, and re-run the script from the Dashboard sheet.

Figure 5: Missing pattern for class indicator Unemployment female population with 15-74 age group in Austria. Retropolation for nuts2 in 1999-2004 range year by using NUTS1-2 proportions in year 2005

unempl_females_15-74

1999			2000			2001			2002			2003			2004			2005		
dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source
2526487	80,6	S1	2533340	78	S1	2540189	70,3	S1	2547038	78,7	S1	2553891	76,4	S1	2560744	102,7	S1	2567597	105,8	S1
2526488	38,9	S1	2533341	40,3	S1	2540190	31,7	S1	2547039	39,8	S1	2553892	42,6	S1	2560745	55,3	S1	2567598	53,1	S1
4516941			4516942			4516943			4516944			4516945			4516946			4516947	4,2	S1
2526489	14,9	S1	2533342	14,3	S1	2540191	10,4	S1	2547040	16,3	S1	2553893	12,6	S1	2560746	17,5	S1	2567599	18,3	S1
2526490	21,1	S1	2533343	23,4	S1	2540192	18,9	S1	2547041	20,9	S1	2553894	26,4	S1	2560747	34,4	S1	2567600	30,6	S1
2526491	14,1	S1	2533344	15,4	S1	2540193	18,9	S1	2547042	18,7	S1	2553895	12,9	S1	2560748	19,1	S1	2567601	20,5	S1
4516959			4516960	4,4	S1	4516961	5,5	S1	4516962	5,6	S1	4516963	4,9	S1	4516964	7,9	S1	4516965	7,9	S1
2526492	9,9	S1	2533345	11	S2	2540194	13,4	S1	2547043	13,1	S1	2553896	8	S2	2560749	11,2	S1	2567602	12,5	S1
2526493	27,6	S1	2533346	22,3	S1	2540195	19,7	S1	2547044	20,2	S1	2553897	21	S1	2560750	28,4	S1	2567603	32,3	S1
2526494	16,1	S1	2533347	13,8	S1	2540196	11,2	S1	2547045	11,8	S1	2553898	11,8	S1	2560751	14,8	S1	2567604	15,7	S1
4516977			4516978			4516979			4516980			4516981			4516982	4,9	S1	4516983	4,6	S1
4516995			4516996			4516997			4516998			4516999			4517000	5	S1	4517001	6,5	S1
4517013			4517014			4517015			4517016			4517017			4517018			4517019	5,5	S1

unempl_females_15-74

	1999			2000			2001			2002			2003			2004			2005		
	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source	dataitem_id	cimal_val	source
tunit_code																					
AT	2526487	80,6	S1	2533340	78	S1	2540189	70,3	S1	2547038	78,7	S1	2553891	76,4	S1	2560744	102,7	S1	2567597	105,8	S1
AT1	2526488	38,9	S1	2533341	40,3	S1	2540190	31,7	S1	2547039	39,8	S1	2553892	42,6	S1	2560745	55,3	S1	2567598	53,1	S1
AT11	4516941	3,1	S2	4516942	3,2	S2	4516943	2,5	S2	4516944	3,1	S2	4516945	3,4	S2	4516946	4,4	S2	4516947	4,2	S1
AT12	2526489	14,9	S1	2533342	14,3	S1	2540191	10,4	S1	2547040	16,3	S1	2553893	12,6	S1	2560746	17,5	S1	2567599	18,3	S1
AT13	2526490	21,1	S1	2533343	23,4	S1	2540192	18,9	S1	2547041	20,9	S1	2553894	26,4	S1	2560747	34,4	S1	2567600	30,6	S1
AT2	2526491	14,1	S1	2533344	15,4	S1	2540193	18,9	S1	2547042	18,7	S1	2553895	12,9	S1	2560748	19,1	S1	2567601	20,5	S1
AT21	4516959	4	S2	4516960	4,4	S1	4516961	5,5	S1	4516962	5,6	S1	4516963	4,9	S1	4516964	7,9	S1	4516965	7,9	S1
AT22	2526492	9,9	S1	2533345	11	S2	2540194	13,4	S1	2547043	13,1	S1	2553896	8	S2	2560749	11,2	S1	2567602	12,5	S1
AT3	2526493	27,6	S1	2533346	22,3	S1	2540195	19,7	S1	2547044	20,2	S1	2553897	21	S1	2560750	28,4	S1	2567603	32,3	S1
AT31	2526494	16,1	S1	2533347	13,8	S1	2540196	11,2	S1	2547045	11,8	S1	2553898	11,8	S1	2560751	14,8	S1	2567604	15,7	S1
AT32	4516977	4,8	S2	4516978	3,8	S2	4516979	3,4	S2	4516980	3,5	S2	4516981	3,6	S2	4516982	4,9	S1	4516983	4,6	S1
AT33	4516995	4,9	S2	4516996	3,9	S2	4516997	3,5	S2	4516998	3,6	S2	4516999	3,7	S2	4517000	5	S1	4517001	6,5	S1
AT34	4517013	4,7	S2	4517014	3,8	S2	4517015	3,4	S2	4517016	3,4	S2	4517017	3,6	S2	4517018	4,8	S2	4517019	5,5	S1

Table: app_updating_virtual_changes

	indicator_id	indicator_m	indicator_code	country_code	year	virt_chng_type	tunit_code	old_src	new_src	old_value	new_value	dataitem_id	vch_date	status_chg	updating_type
	229	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	229	226	unempl_males_15-74	AT	2018	R3m	AT11	NULL	S2	NULL	3.3	NULL	2020-02-28...	not_saved	INSERT
2	229	226	unempl_males_15-74	BE	2001	R1	BE34	S1	S1	0.0	NULL	2537922.0	2020-02-28...	not_saved	UPDATING
3	229	226	unempl_males_15-74	DE	1999	R3m	DE5	NULL	S2	NULL	22.5	NULL	2020-02-28...	not_saved	INSERT

Table 6: Transfer keys for NUTS changes in Ireland and Hungary

Country code	Level NUTS	NUTS 2016	NUTS 2013	Transfer keys
IE	2	IE04	IE01	1
IE	2	IE06	IE01	0.52
IE	2	IE05	IE02	1
IE	2	IE06	IE02	0.48
HU	2	HU11	HU10	0.08
HU	2	HU12	HU10	0.92

During the updating process of the ESPON indicators, the treatment of null values has been a serious issue, by considering null values as either present or absent in the ESPON database. At the end of that process, a local sqlite3 file containing the approved virtual changes is created and the script has to be aware of which kind of missing values is present in a given indicator.

Even the main goal of the Updating strategy is to provide complete time series for ESPON indicators, missing data is unavoidable. Missing data was already present in the original Excel

files derived from ESPON Update of Maps project, by empty cells in value indicators or/and associated sources. Null data is technically the model instance (indicator/source) that no values can be attached to. Technically, a null data could be stored in a database server, with a special null type. Or a null data is not optionally stored. More generally, that instance could be associated with more than a model or table, for instance, an indicator value only has sense if source from which it comes is also reported, so that indicator data can be considered as a tuple, pair of (value, source).

Five kinds of null data have to be dealt with when updating:

- Value that does not exist in ESPON database (no rows in ESPON tables).
- Null value that does exist in ESPON database, but no source is attached to.
- Null value that does exist in ESPON database and source is attached to.
- No null value that does exist in ESPON database and no source is attached to.
- Wrongly-zero-assigned value from previous import.

This casuistic in ESPON database has to be present when approved virtual changes are saved because virtual changes are simply but rows with many fields ready to be integrated in the ESPON production database, filling the relevant tables and we have to be aware beforehand if those changes will be inserting new rows or updating existing ones.

3.3.2.3 Updating ESPON Database

When the updating process for an ESPON indicator has finished, expert user has to upload the sqlite3 file with all changes (step 12) to a folder in the ESPON server and trigger the execution of 7 command scripts (see Table 3). A text report is generated and stored in the database. It contains relevant information about data to be inserted in ESPON database (Figure 6).

Figure 6: Report containing data to be inserted/updated in ESPON database

```
is indicator_m = 226 ? ... True
There are some fails in tunits ? ------------- NO  |
There are some fails in indicator codes ?  --- NO  |
There are new years   ?  --------------------- YES  |  2017, 2018
There are new sources ?  --------------------- NO  |
There are new methods ?  --------------------- YES  |  N/A

There are 11013 insertions and 77144 updates
 Country: AT
  -- Insertions  : 288
  -- Updates     : 2208
 Country: BE
  -- Insertions  : 354
  -- Updates     : 2573
 Country: BG
  -- Insertions  : 222
  -- Updates     : 1647
 Country: CH
  -- Insertions  : 210
  -- Updates     : 1019
 Country: CY
  -- Insertions  : 24
  -- Updates     : 25
 Country: CZ
  -- Insertions  : 144
  -- Updates     : 1033
 Country: DE
  -- Insertions  : 2730
  -- Updates     : 12925
 Country: DK
  -- Insertions  : 108
  -- Updates     : 432
 Country: EE
  -- Insertions  : 48
  -- Updates     : 355
 Country: EL
  -- Insertions  : 420
  -- Updates     : 5058
 Country: ES
  -- Insertions  : 516
  -- Updates     : 6080
 Country: FI
  -- Insertions  : 156
  -- Updates     : 594
```

Basically, in general terms, this new update for the base indicators incorporates the new years for the majority of indicators, for instance 2017 and 2018 when time series ended in 2016. Indicator data have been transfered to NUTS 2013 for these countries: France, Poland, Ireland, Finland, Hungary, Deutschland, Lithuania, United Kingdom, Netherlands, only for the Eurostat data reported in NUTS 2016 version, that is, for all indicators except for Demographics.

2.3 ESPON 2020 Data Delivery Process (WP7)

A delivery procedure was set up for data produced by ESPON projects. This procedure entails multiple interactions between the Project Team (ESPON 2020 Database Portal), Service Providers (ESPON project that produced data) and ESPON EGTC. Data delivery is a contractual obligation for Service Providers at the end of their project.

The first of the following sections illustrates the ESPON 2020 Database Portal approach to project data deliveries and defines the main concepts behind it. The second section provides a state of affairs with regard to the actual delivery of data by ESPON Service Providers.

2.3.1 Conceptual Approach to Data Delivery

The aim of the ESPON Database Portal is to receive, structure and display all data from ESPON Service Providers. To do so, the Project Team developed a dual approach to data, allowing a selection and structuration of the most meaningful results and, at the same time, a

collection of all kinds of data that a project may have produced. The two delivery channels are referred to as the Main Data and the Other Data.

Main Data gather the most original and policy relevant indicators ("key indicators"), indicators used to calculate these indicators ("background indicators"), and their metadata, including mutual inter-indicators relations (various kind of groupings), indicator sources and possible preliminary processing applied to data during compilation. Main Data correspond to the queryable content of the Database Portal. Indicators to be included in Main Data are selected based on a dialogue between the Service Provider, ESPON EGTC (project expert) and the ESPON 2020 Database Portal (Project Team). These data and related metadata are provided by the Service Providers through the ESPON 2020 Data and Metadata Upload System (http://database.espon.eu/data-upload/ - for credentials to access the repository please contact the admin of the Database - database@espon.eu). Main Data are mostly based upon indicators provided in standard nomenclatures (e.g. NUTS, Functional Urban Area).

Other Data section gathers all indicators produced (compiled or calculated) by the ESPON Service Provider in the course of their project, exclusive of data already provided as part of "Main Data". These indicators may be delivered in the form of statistical or geographic data. Collection of statistical indicators is completed through Excel templates adapted from the M4D project. It is gathered by individual file or bulk of individual files, each described briefly in the database system.

Main Data and Other Data gather **indicators**. An indicator designates a measure with a unique definition, which may cover the whole or part of the ESPON Space and which may or may not be associated with a standard nomenclature (e.g. "population density", "population potential", "female employment rate"). Indicators under Main Data are the basic items provided in the list shown in the Search & Download section of the User Interface.

Main Data: concepts and structure

A. Datasets and indicators

Main Data is the major ESPON 2020 Database Portal entry point for the final user. Main Data is based on the following concepts:

- A **dataset** gathers one or more indicators in a thematically or statistically coherent whole. Indicators in a dataset pertains to the same theme and/or are related one between another by genetic relations. Datasets are just a vehicle for data collection. They are not visible as such for the final user of the ESPON database.
- A **key indicator** is considered by a Service Provider as one of the main output of the project. As an output of different types of data processing, it may have multiple parent indicators. All key indicators use a standard nomenclature.
- A **background indicator** is an indicator which was used in the process of creating a key indicator (either original compiled data or a significant step in the calculation

process). A standard background indicator uses a standard nomenclature and is delivered using the same workflow). A non-standard background indicator is not presented in a standard nomenclature and must therefore be delivered through a dedicated sub-channel (e.g. grid data, flow data).

Metadata associated with datasets and indicators are listed and described in the Manual "How to Deliver My Data" (online document).

B. Indicators groupings (indicators groups, indicator genealogy, dimensional indicators)

The originality of the Main Data under the ESPON 2020 Database Portal is the extent to which it identifies relations between indicators. Indicators may be linked on two basis:

(1) **Structural** relations (multi indicators / dimension indicators / class indicators). A multi indicator is an indicator that has one or more dimensions and/or classes. A Dimension is a particular way to break down a multi indicator according to a criteria. A Class is an individual instance of a type under the criteria taken as dimension. For instance, "Total population" (Multi indicator) can by broken down by "Age group" (Dimension) which corresponds to specific age interval (Classes), e.g. 0 to 4, 5 to 9, 10 to 14.

(2) **Genetic** relations (indicator genealogy). A genealogy relation can be established between one (or several) indicator(s) and another, when one (or several) indicator(s) - the **Parents** - were used in the process of calculating the other – the **Child**. A methodology that describes the calculation step is associated to each genealogy relation. Genealogy relations are declared at the level of indicators. For indicator with multiple dimensions and multiple classes, genealogy is defined at the level of the multi-indicator. The genealogy then applies to all dimensions and classes.

These relations are at the core of the cumulative and networking approach to indicators. This approach is cumulative as new indicators may be related to existing ones (e.g. when an ESPON project refer to indicators already available in the ESPON Database Portal). This approach is network-oriented as it allows the final user of the ESPON 2020 Database Portal to navigate between indicators on the basis of structural or genetic relations.

Furthermore, any indicator can be associated to a **Dataset** (sets of indicators defined by Project) or a **Group** (any desired grouping of indicators, e.g. a thematic category).

C. Sources and preliminary processing

Each individual data record (value of a standard indicator for an individual region) is associated with a 'Source' and may be assigned a 'Preliminary Processing'.

A **Source** is a set of information that allows to locate external data used in the process of compiling or calculating a data record. It includes a name, a URL and a description.

A **Preliminary Processing** is a set of information that identifies calculating procedures that were applied to external data before compilation. It may be an aggregation or disaggregation, an estimation, overlay procedure between geographic layers, etc.

'Sources' and 'Preliminary Processing' ensure the systematic reference to external data.

Other Data: concept and structure

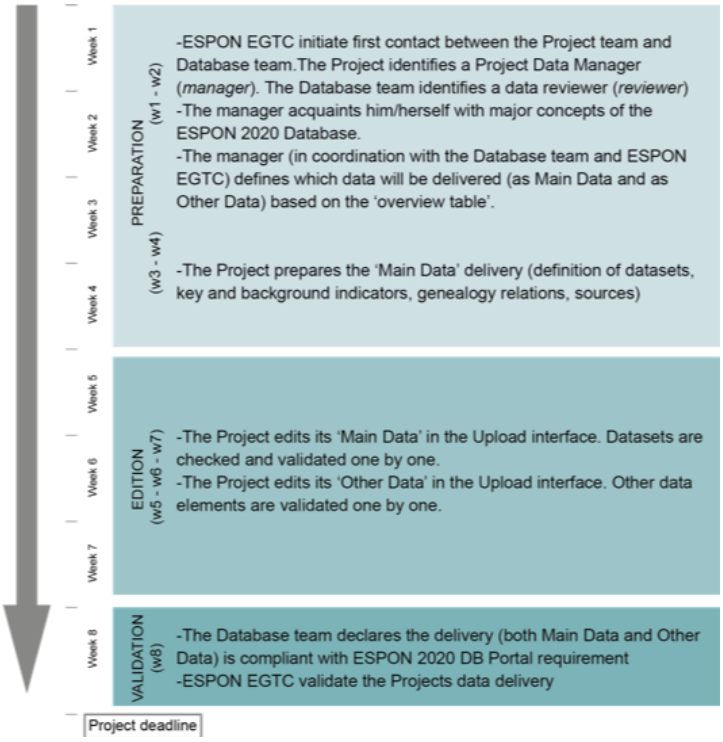
Data to be included as 'Other Data' may be statistical data (which may or may not use standard spatial nomenclatures), geographical data or any other relevant data (e.g. survey results). Metadata associated to each indicators and file delivered as Other Data are less extensive; these are described in the online documentations in the section "How to Deliver My Data". In contrast with Main Data, no relation is identified between indicators provided under Other Data.

Once a Project delivery is completed, the content of Main Data and Other Data is merged in a **Project Archive** that one may download from the ESPON 2020 Database Portal. The Project Archive gathers the whole data content produced by an ESPON Project.

2.3.2 Data delivery process

Data delivery from ESPON Service Providers requires: (1) coordination between several Parties (workflow) (2) delivery systems and templates and (3) instructions sent to ESPON Service Providers. The whole process shall take no longer than eight weeks. It should ideally start at the delivery of the draft final report of the project. Figure 7 provides the general timeline for the delivery.

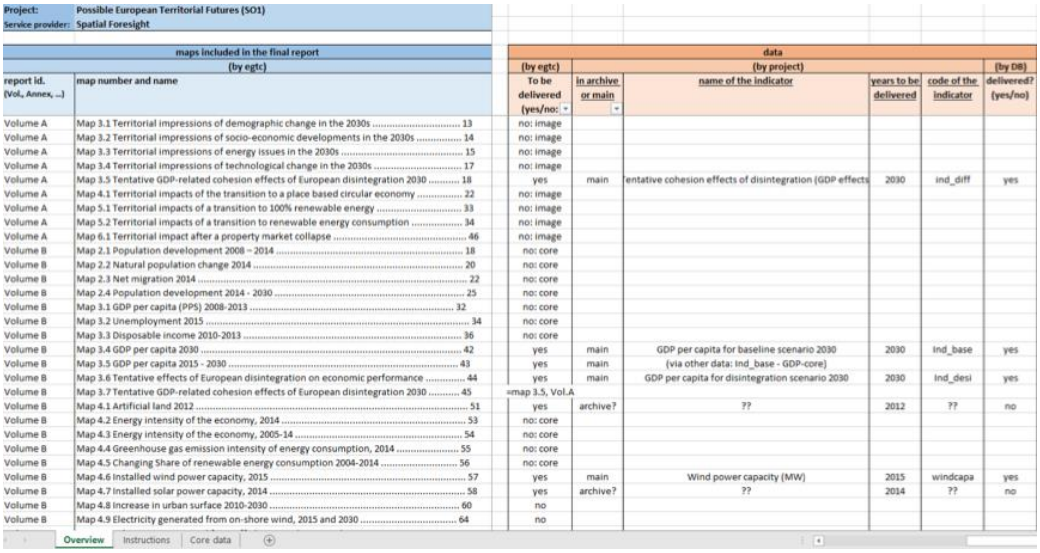
Figure 7: Timeline of project data delivery



The "overview table": ensuring a common understanding of the delivery content

The delivery is performed through the ESPON 2020 Upload interface (see Section 5.2). The entire process is monitored via an "overview table". The overview table relates maps and figures of the draft final report to indicators, which are then assigned a status ('key', background', 'other'). This table is then used as a tracking tool to validate the completion of a projects delivery (see exemple in Figure 8). The interactive process between ESPON EGTC (project experts), the Services provider and the Database team, as well as all steps of the delivery process, are described in detail in the online documentation.

Figure 8: Data Delivery - Overview Table



Project:	Possible European Territorial Futures (SO1)									
Service provider:	Spatial Foresight									
	maps included in the final report (by egtc)		data							
			(by egtc)	(by project)						(by DB)
report id. (Vol., Annex, ...)	map number and name		To be delivered (yes/no:	in archive or main	name of the indicator	years to be delivered	code of the indicator	delivered? (yes/no)		
Volume A	Map 3.1 Territorial impressions of demographic change in the 2030s	13	no: image							
Volume A	Map 3.2 Territorial impressions of socio-economic developments in the 2030s	14	no: image							
Volume A	Map 3.3 Territorial impressions of energy issues in the 2030s	15	no: image							
Volume A	Map 3.4 Territorial impressions of technological change in the 2030s	17	no: image							
Volume A	Map 3.5 Tentative GDP-related cohesion effects of European disintegration 2030	18	yes	main	entative cohesion effects of disintegration (GDP effects	2030	ind_diff	yes		
Volume A	Map 4.1 Territorial impacts of the transition to a place based circular economy	22	no: image							
Volume A	Map 5.1 Territorial impacts of a transition to 100% renewable energy	33	no: image							
Volume A	Map 5.2 Territorial impacts of a transition to renewable energy consumption	34	no: image							
Volume A	Map 6.1 Territorial impact after a property market collapse	46	no: image							
Volume B	Map 2.1 Population development 2008 – 2014	18	no: core							
Volume B	Map 2.2 Natural population change 2014	20	no: core							
Volume B	Map 2.3 Net migration 2014	22	no: core							
Volume B	Map 2.4 Population development 2014 - 2030	25	no: core							
Volume B	Map 3.1 GDP per capita (PPS) 2008-2013	32	no: core							
Volume B	Map 3.2 Unemployment 2015	34	no: core							
Volume B	Map 3.3 Disposable income 2010-2013	36	no: core							
Volume B	Map 3.4 GDP per capita 2030	42	yes	main	GDP per capita for baseline scenario 2030	2030	ind_base	yes		
Volume B	Map 3.5 GDP per capita 2015 - 2030	43	yes	main	(via other data: ind_base - GDP-core)					
Volume B	Map 3.6 Tentative effects of European disintegration on economic performance	44	yes	main	GDP per capita for disintegration scenario 2030	2030	ind_desi	yes		
Volume B	Map 3.7 Tentative GDP-related cohesion effects of European disintegration 2030	45	=map 3.5, Vol.A							
Volume B	Map 4.1 Artificial land 2012	51	yes	archive?	??	2012	??	no		
Volume B	Map 4.2 Energy intensity of the economy, 2014	53	no: core							
Volume B	Map 4.3 Energy intensity of the economy, 2005-14	54	no: core							
Volume B	Map 4.4 Greenhouse gas emission intensity of energy consumption, 2014	55	no: core							
Volume B	Map 4.5 Changing Share of renewable energy consumption 2004-2014	56	no: core							
Volume B	Map 4.6 Installed wind power capacity, 2015	57	yes	main	Wind power capacity (MW)	2015	windcapa	yes		
Volume B	Map 4.7 Installed solar power capacity, 2014	58	yes	archive?	??	2014	??	no		
Volume B	Map 4.8 Increase in urban surface 2010-2030	60	no							
Volume B	Map 4.9 Electricity generated from on-shore wind, 2015 and 2030	64	no							

Overview | Instructions | Core data

Projects whose data delivery was supported

The following projects were supported in their data deliveries (Table 7):

Table 7: Data upload - supported projects

Applied research (SO1)	Targeted analysis (SO2)	Monitoring and tools (SO3)
LOCATE	SPIMA	Big Data and Housing
FUTURES	LinkPAS	TIA UPGRADE
SME	MIGRATUP	FUORE
EMPLOY	Alps2050	EMTMT
FDI	CPS	
PROFECY	HERITAGE	
GRETA	BIG DATA CBC	
COMPASS	URRUC	
BRIDGES	TEVI	
YUTRENDS	MSP-LSI	
Financial Instruments	BT2050	
CIRCTER	ENSURE	
ETRF	SHARING	
MIGRARE	ACPA	
T4	TIA CBC	

Training for data reviewers

In order to improve the capacity of the team to guide and support ESPON projects in their delivery, reviewers need to be properly trained. Therefore a training workshop has been designed and implemented to ensure that reviewers have adequate capacities to guide Service providers in their delivery.

The workshop is structured in three parts and may be implemented in one day (8 hours):

- Introduction to the main concepts associated with data delivery.
- Presentation of the tools that were prepared to ensure collaboration among reviewers: the “reviewers toolbox” and the “monitoring file”.
- Exercises based on actual ESPON data.

The participation to a training session is crucial for the reviewers to understand the main challenges associated with data collection and to ensure the integrity of the data delivery. This is why, the training is subject to an internal certification. All reviewers appointed to support projects shall attend a training session. Training materials are made available as an internal “toolbox” in the online documentation.

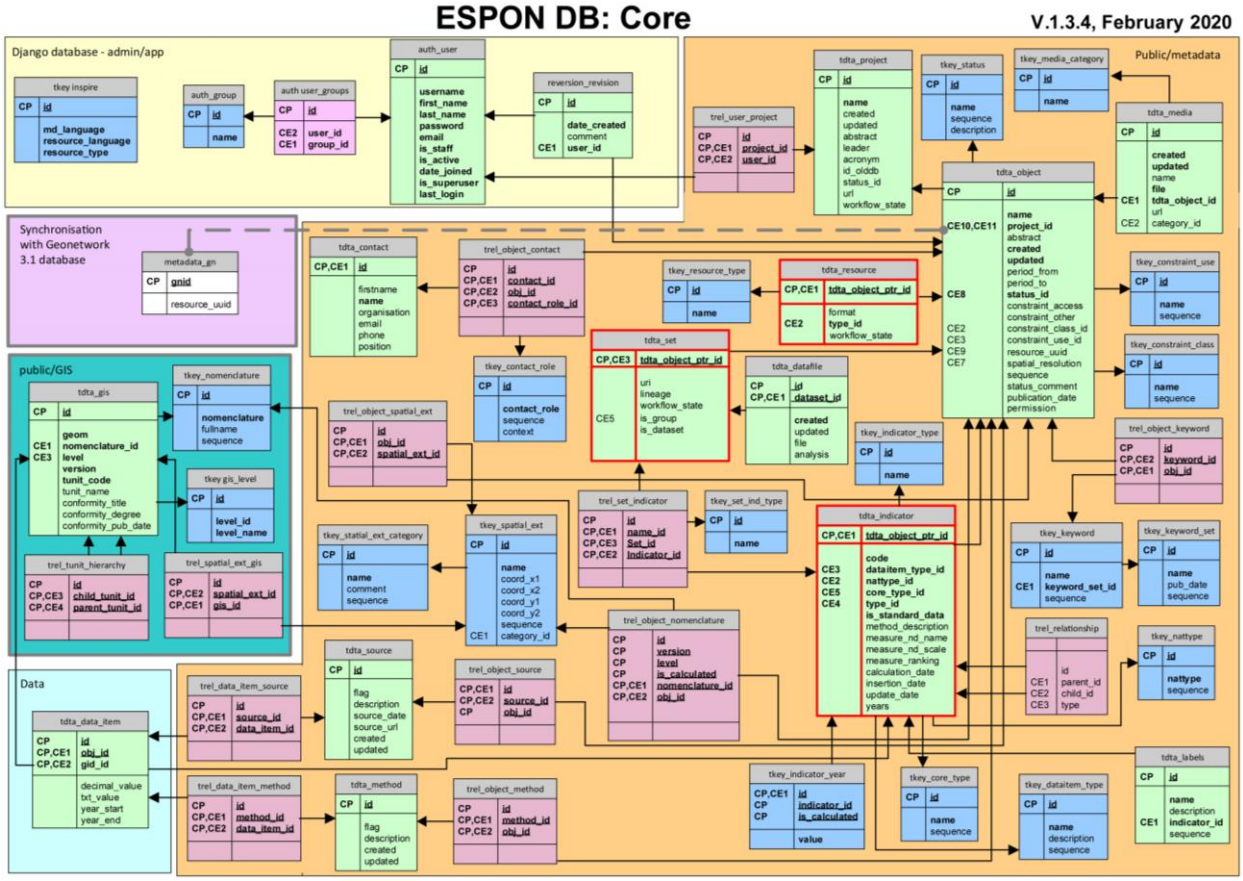
3 Database and Web Services

3.1 Database Systems (WP1)

The ESPON 2020 Database system (DBS) has been built around PostgreSQL/PostGIS. PostgreSQL is a very well-known relational database system, used and maintained by a huge worldwide community of users.

The data model of ESPON 2020 system was developed to be as simple as possible despite the high complexity imposed by the heterogeneity of ESPON data. The data model is implemented in the Django framework (https://www.djangoproject.com/) and is illustrated in Figure 9.

Figure 9: Database model



3.1.1 Data model

Naming of tables follows a prefix rule. A first "edp_core" is created by the Django datamodel process from a python script. In Figure 9, the yellow background includes (1) application tables, which represent the conceptual schema called "admin/app", containing the Django standard tables including the users/roles information and (2) tailored tables allowing to track the various actions undertaken by the users. The green background is used for GIS table(s) including geometry. The light blue background includes the data table containing the actual data values. The orange background includes the whole metadata tables. The purple

background designates the Geonetwork database which is synchronised with ESPON database.

Tables are prefixed and also coloured in this model. Light green is used for 'tdta_' tables, containing different data inputs (tables modelling the main used concepts, such as: project, dataset, indicator, t-indicators, territorial unit, etc.). Purple tables, 'trel_', are relational tables linking records between several tables. Blue 'tkey_' tables contain defined values stored in the database (used for example for drop down list selection in the user interface). Tables with red lines are the specific tables for the three main object type (set, indicator and resource).

The files stored on server folders are linked in dedicated fields of the database. The GIS nomenclatures are stored directly in dedicated PostGIS table(s). Some synchronisation scripts maintain the coherence between the main database and a GeoNetwork associated application PostgreSQL database.

3.1.2 Data and Metadata Workflows

Workflows are defined for all relevant processes involving the validation and quality control of datasets, as well as upload and update procedures. ESPON EGTC approvers are responsible for the final approval.

Mainly three different workflows were implemented according to the user rights and the three main types of input (Standard data; Non standard data; Non data).

3.1.3 Main object types

The three main object types are: sets, indicators and resources:

1. Sets

   Sets include datasets and light groups which allow to group indicators independently of the project or dataset from which they belong.

2. Indicators

   This type includes the comprehensive information of indicator. It includes a genealogy and indicator types hierarchy. Genealogy allows to describe the methodology and history of an indicator. The indicator type hierarchy gives the possibility to describe sub-indicators hierarchy. In this architecture, we define two main types of indicators: single indicators and multi indicators. These are the indicators that will appear in the list of Search and Download interface. In addition, the multi indicators have two possible sub-types, dimension indicators and class indicators:

- A **single indicator** - is the general indicator that we considered so far, with no dimensions and no classes attached. This type of indicator should have genealogy (if exists) and data directly attached to it. Example: GDP, GDP per capita.
- A **multi indicator** - is an indicator that has one or more dimensions and/or classes (hereafter named **dimension indicators** and **class indicators**). Example of a multiple indicator: ***population by age*** which has one dimension (***age***) with several

classes (***pop0-4, pop5-9, pop10-14, etc.***). Note: we can only have one level of dimensions (no sub-dimensions) and one level of classes (no sub-classes). E.g. “Corine land cover, second level” is not defined as a sub-dimension of “Corine land cover, first level”, but as a separate dimension (of the multi indicator “Corine land cover”). Consequently, the classes of “Corine land cover, second level” are not included in the classes of “Corine land cover, first level”.

3. Resources
   Resources include documents such as manuals or technical reports, but also mapkits, OLAP data cubes, non standard data as well as project archives.

3.1.4 Roles

Six roles have been implemented in the system (see for detailed privileges).

General

1. **Anonymous** – web site visitors.
2. **Authenticated User** - restricted access to particular downloads[3].
3. **Administrator** - SuperUser with permissions to manage users, nomenclatures, Upload System, etc.
4. **Upload System** - These roles are used within the *Upload System*, which allows to deliver Data and Metadata.
5. **Project Manager** – can insert metadata and data - acting as main project contact for data provision to the database system.
6. **Project Reviewer** - semantic, genealogy, syntactic, quality control check and validation.
7. **Project Approver (EGTC) -** validates and gives the final approval for data publication in the Portal (core, key indicators and other non standard data).
8. **Admin Superuser** - manage (create, edit, delete) users, manage backups, database system updates, and probably manage nomenclatures.

3.1.5 Data and Metadata Statuses

Within a workflow, data and metadata can receive various statuses according to the steps they have achieved. A certain user, depending on the assigned role, can push a status forward or backward in the workflow.

The following statuses were defined:

1. Open (once metadata is created and until submission).

[3] Notice that potentially registered and privileged user roles could be implemented later on if necessary (when identification is required to download some information such as Mapkit, or copyrighted dataset such as GISCO data).

2. Submitted (metadata is ready for 'Semantic and genealogy check').
3. SG checked (metadata has successfully passed through 'Semantic and genealogy check').
4. SD checked (syntactic data check has been successfully done).
5. QC checked (quality control has been successfully done).

Notice that the statuses described here follow the most complicated workflow (for ‘standard data’) and are simpler for ‘non standard’ data and 'non data' metadata.

Other status exists in the database, for example for project which can be opened, approved or archived, as well as for Resources which can be opened, submitted or accepted.

3.2 Web Services (WP3)

3.2.1 OGC Web Services

The Open Geospatial Consortium (OGC) is an international not for profit organization committed to making quality open standards for the global geospatial community. These standards are made through a consensus process and are freely available for anyone to use to improve sharing of the world's geospatial data.

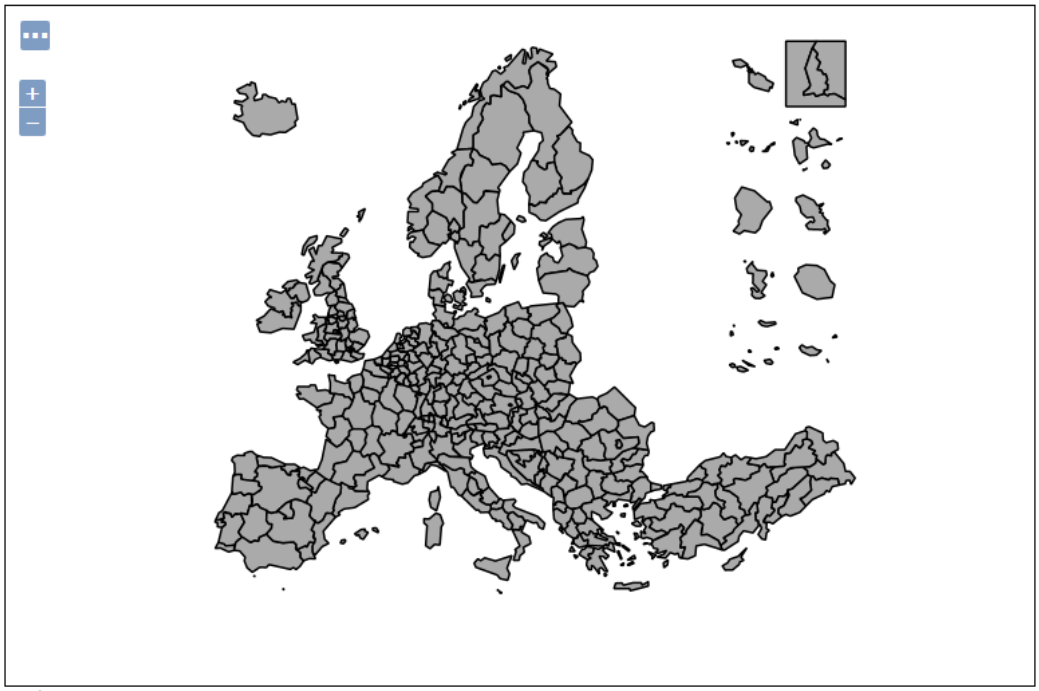
Currently 17 layers are available for both OGC Web Map Service (WMS) and Web Feature Service (WFS) services as summarized in Table 8.

Table 8: Nomenclatures available for WMS and WFS

Geographic unit	***Nomenclature version***
NUTS0	2003, 2006, 2010
NUTS1	2003, 2006, 2010
NUTS2	2003, 2006, 2010
NUTS3	2003, 2006, 2010
LAU2	2013
FUA	2015-2018
UMZ	2000, 2006
Metro Regions	2013

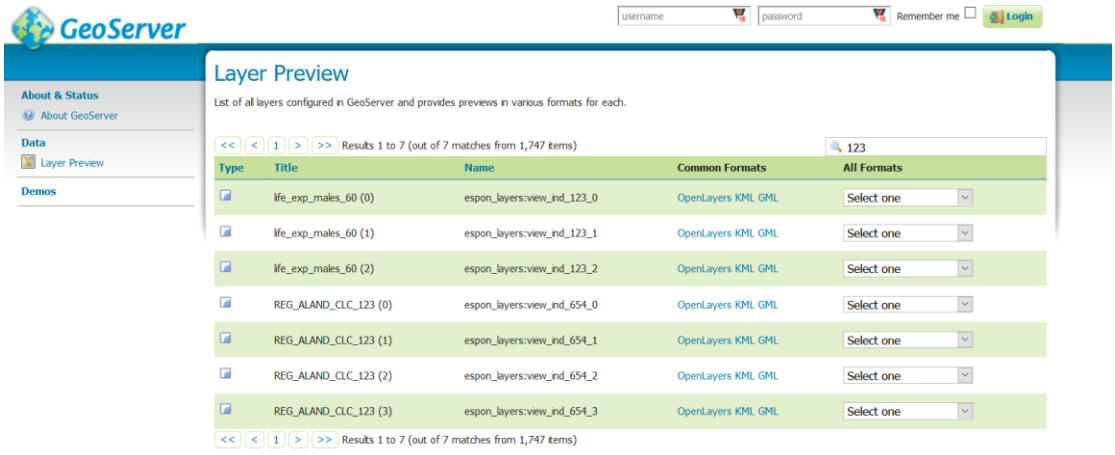
It is possible to visualise the spatial coverage of those spatial extents through a WMS webservice. For example, for NUT2 version 2006, the URL is (Figure 10): http://database.espon.eu/geoserver/espon_layers/wms?service=WMS&version=1.1.0&request=GetMap&layers=espon_layers:gis_nuts2_2006&styles=&bbox=-2823913.87226075,-3076144.55029697,1.00261246655394E7,5415981.70375715&width=768&height=507&srs=EPSG:3035&format=.

Figure 10: Spatial Coverage for NUTS2, 2006



Every published standard indicator has associated WMS/WFS web services in ESPON Geoserver. WMS indicator service is available by indicator and level (Figure 11).

Figure 11: ESPON 2020 Database Geoserver



They can be called by WMS service or a WFS service for example to download a zip file containing geometries in shape file and attribute table. For a example, WMS for indicator life_exp_females_80 NUTS1:
https://database.espon.eu/ghttps://database.espon.eu/geoserver/espon_layers/wms?service=WMS&version=1.1.0&request=GetMap&layers=espon_layers:view_ind_134_1_2013_1&style

s=&bbox=2641758.46955152,1428695.35991733,7313157.28662197,5456938.51107795&width=768&height=662&srs=EPSG:3035&format=application/openlayers.

Figure 12: WMS for indicator life_exp_females_80 NUTS1

tunit_code	tunit_name	processes	sources	y_2002	y_2003	y_2004	y_2005	y_2006	y_2007	y_2008	y_2009	y_2010	y_20
CH0	SCHWEIZ/SUISSE /SVIZZERA		S1- : http://ec.europa.eu /eurostat /data/database	9.6	9.3	9.9	9.9	10.3	10.3	10.4	10.4	10.4	10.6

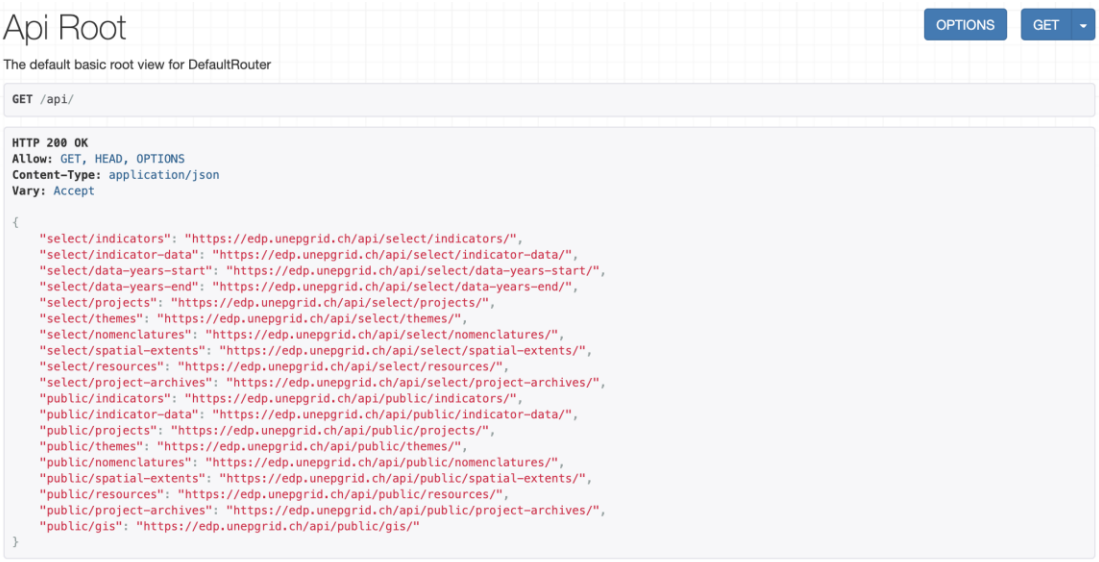
3.2.2 REST Web Services

The REST Web Services used in the ESPON DB were developed using the Django REST Framework (http://www.tomchristie.com/rest-framework-2-docs/), which is a flexible and powerful toolkit for easily building Web APIs. This framework was chosen due to numerous features, such as: authentication policies, serialization support, flexible views and extensive documentation and community support.

The REST Web Services are currently hosted at: http://database.espon.eu/api/ (Figure 13) and they are in a stable version since Delivery D4. After the finaly deployment of the Portal on the ESPON servers, a new address will be provided.

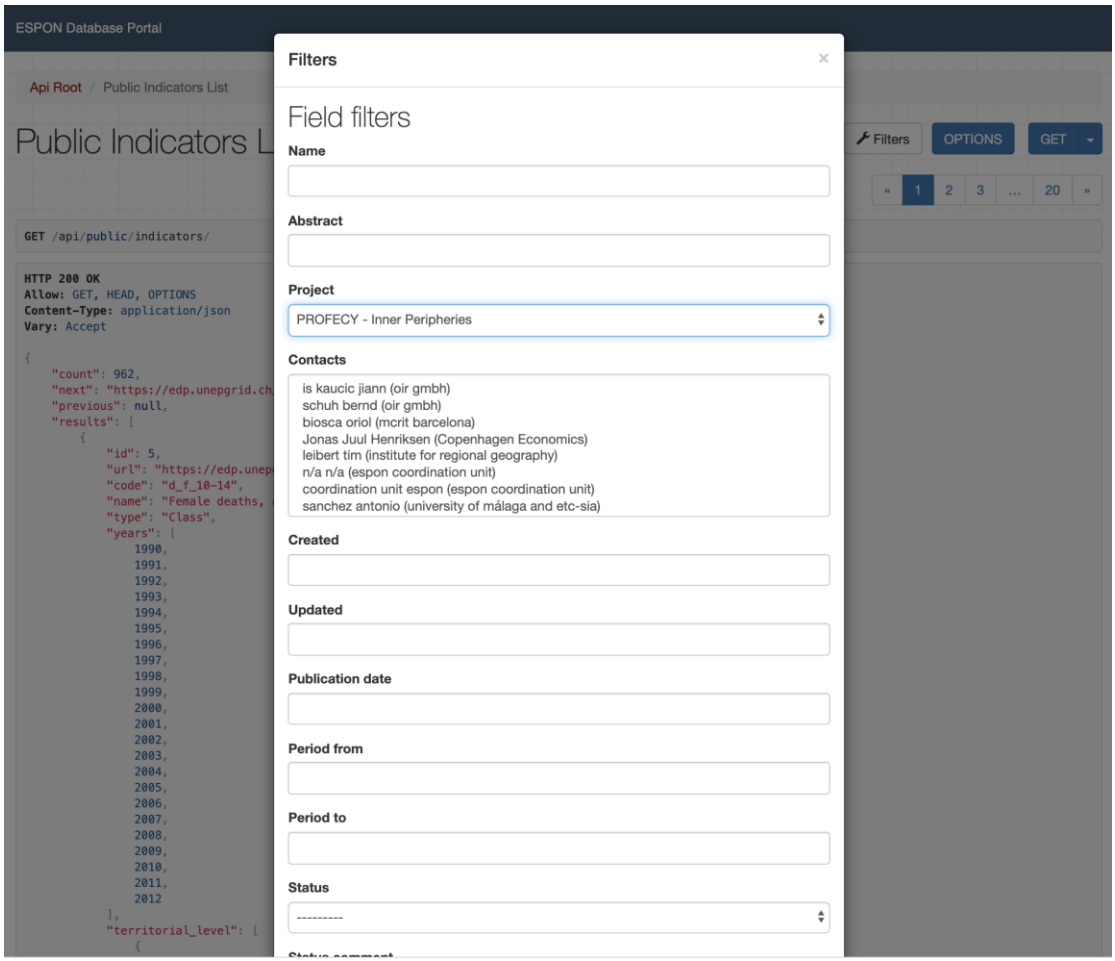
Some of the web services were updated to better deliver the infomartion in the User Interface but also to improve their efficiency as much as possible.

Figure 13: ESPON DB Web Services API



All the REST Web Services offer pagination (considering the increasing number of data that are now available in the database) and filtering options in the API interface. The filtering is available for all the listed attributes available for the selected table and it is provided using the Django REST Framework support for the generic filtering backends that allow to easily construct complex searches and filters. Generic filters can also present themselves as HTML controls in the browsable API and admin API (Figure 14).

Figure 14: ESPON DB REST Web Services - Filtering



The REST Web Services are grouped in 3 main categories, serving:

- The ESPON DB User Interface
- ESPON Tools (TIA Tool, and more recently the ESPON Functional Urban Regions)
- External users

The last two categories are served by the same list of Web Services for the moment (which are public - example: http://database.espon.eu/api/public/indicators/). When there will be some special requests from other ESPON Tools to develop dedicated/personalized Web Services, they will be added in the second category and made private.

The list of Web Services developed for the Database User Interface contains services similar to the public ones but customized to serve the special request necessary in the interface.

Among these services, we enumerate:

- http://database.espon.eu/api/select/indicators/
- http://database.espon.eu/api/select/indicator-data/
- http://database.espon.eu/api/select/projects/

- http://database.espon.eu/api/select/spatial-extents/
- http://database.espon.eu/api/select/nomenclatures/
- http://database.espon.eu/api/select/themes/
- http://database.espon.eu/api/select/resources/
- http://database.espon.eu/api/select/project-archives/

All the information required for display, search, filter and download in the User Interface is retrieved using the above Web Services and their attached filters.

The full online documentation of the Web services can be found at: https://database.espon.eu/doc/how-to-use-the-espon-db-web-services.html.

3.2.3 Services related to metadata exchange: catalogue services

Methodology

In order to ensure the interoperability between the ESPON metadata system and the external world it is essential to be compliant with international recognized standards. The first part of the process was to convert each ESPON metadata record to INSPIRE metadata (in ISO/TS 19139 based XML format).

Subsequently with the purpose of providing "discoverable services", the INSPIRE compliant metadata are integrated in the GeoNetwork catalogue system. GeoNetwork[4] has been developed to connect spatial information communities and their data using a modern architecture, it supports Catalogue Services for the Web (CSW), Geo Portal REST and several more protocols.

The processes of metadata conforming to the INSPIRE profile, together with their integration into the GeoNetwork platform, are summarized in the following operations:

1. Mapping schema from ESPON metadata to INSPIRE
2. Extraction of selected metadata records from ESPON database
3. Generation of ISO 19139 XML template
4. Import in Geonetwork database

Mapping from ESPON metadata to INSPIRE

The INSPIRE Implementing Rules for Metadata requires that a set of metadata elements shall be provided. These are grouped in the following categories:

1. Identification
2. Classification of spatial and data services
3. Keyword

[4] https://geonetwork-opensource.org/

4. Geographic location
5. Temporal reference
6. Quality and validity
7. Conformity
8. Constraint related to access and use
9. Organisations responsible for the establishment, management, maintenance and distribution of the resource
10. Metadata on metadata

A mapping table was generated keeping in account all the mandatory fields required from the INSPIRE and those present and compatible with ESPON metadata.

As one might expect, the correspondence is not perfect: different fields specific to ESPON are not present in the INSPIRE profile. In this case, they have not been incorporated. Let us remember that the main purpose of standardization is to make the metadata discoverable and accessible. For specific/detailed information, the user can always refers to the original metadata.

On the other hand not all the information necessary to satisfy the INSPIRE standards is contained in the ESPON metadata fields. There are two possible different options to overcome this lack depending on the case:

- The fields are automatically generated as for example the UUID.
- The fields are filled by default: this is the case for values that are shared for all the metadata such as the metadata language (en) or the character coding (UTF8).

The following table (Table 9) illustrate an extract of matching for Metadata sections from INSPIRE including identification, Classification and Keywords. The full table is available in the annex.

Table 9: mapping table INSPIRE-> ESPON2020 (full table available in the annex)

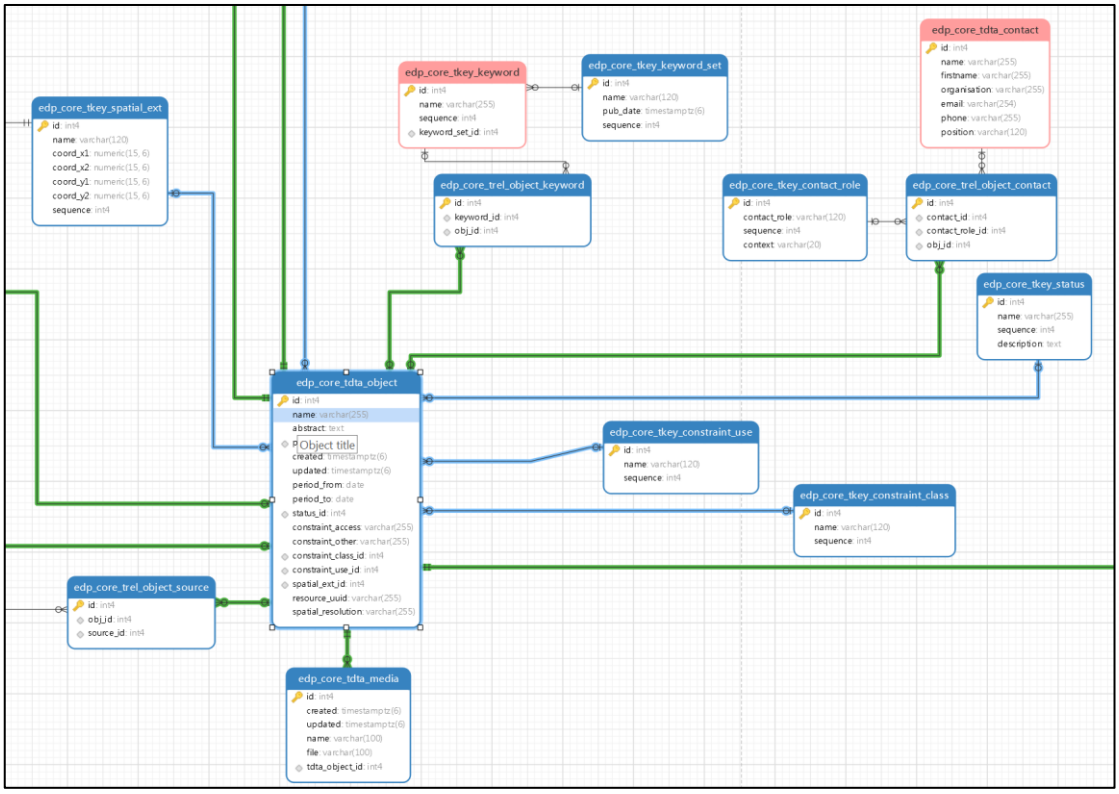
Item Inspire	Item from ESPON metadata	Suggestions (how to match Inspire)	Multiplicity	Obligation	ISO Data type	Example	ISO Definition
INSPIRE-Metadata Identification							
Resource Title	**Name**		1	M	Character String	*Total population change*	known
Resource Abstract	**Abstract**		1	M	Character String	*Difference of population between two time-periods*	Brief narretive summary of the content of the resource
Resource Type	**None**	**Default**	1	M	MD_ScopeCode	***dataset***	Scope to which metadata applies
Resource Locator	**None**	**Default/automatic**	* 0..1	M	URL	http://projects.mcrit.com/espondb/index.php/main-data	Location (Address) for onb-line access using a Uniform Resource Locator address or similar addressing scheme
Unique Resource Identifier	**code**		* 1..1	M	MD_Identifier	*pop_chxx*	Value uniquely identifying an object within a namespace
Resource Language	**None**	**Default**	* 0..1	M	LanguageCode (ISO/TS 19139)	***eng***	Language(s) used within the datasets, **ISO 639-2**
INSPIRE Metadata Classification of spatial data and services							
Topic Category	**Topic category (datasets only)**	*Heritage from dataset*	* 1..1	M	MD_TopicCategory	*Society*	Main theme(s) of the dataset **Code List B.5.27**
INSPIRE Metadata Keyword							
Keyword Value	**keywords**		* 1..1	M	CharacterString	Administrative boundaries, administrative units, airport….	word(s) or phrase(s) used to describe the subject

Extraction of selected metadata records from ESPON database

For each “object” from the table edp_core_tdta_object, all the INSPIRE compatible information contained in the related tables is retrieved and stored in a new table called inspire_main.

Subsequently these records will fill the ISO 19139 template.

Figure 15: edp_core_tdta_object and related tables (zoom in) IN PROGRESS



Generation of ISO 19139 XML template

Once the correspondences between the ESPON and INSPIRE fields have been defined, an (empty) template has been created in xml format, compliant with the ISO 19139 standard and the INSPIRE profile.

This template is composed of the fixed base XML structure, which will be filled later by the different records: for each ESPON-INSPIRE metadata one XML template is filled and stored successively in a dedicate field of the database.

Figure 16 shows an extract of the XML template: in red color the fixed structure; the text in black corresponds to the variable elements, that will be extracted and filled from the ESPON database.

Figure 16: XML template (extract)

```
    '<?xml version="1.0" encoding="UTF-8"?>
<gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gco="http://www.isotc211.org/2005/gco" xmln
xmlns:xlink="http://www.w3.org/1999/xlink" xsi:schemaLocation="http://www.isotc211.org/2005/gmd http://schemas.
  <gmd:fileIdentifier>
      <gco:CharacterString>'||my_uuID||'</gco:CharacterString>
  </gmd:fileIdentifier>
  <gmd:language>
      <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="'||metaLang||'"/>
  </gmd:language>
  <gmd:characterSet>
      <gmd:MD_CharacterSetCode codeSpace="'||metaStandard||'" codeListValue="'||metaChar||'"
codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxC
  </gmd:characterSet>
  <gmd:hierarchyLevel>
      <gmd:MD_ScopeCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/re
  </gmd:hierarchyLevel>
  <gmd:contact>
      <gmd:CI_ResponsibleParty>
         <gmd:organisationName>
            <gco:CharacterString>'||metaOrg||'</gco:CharacterString>
         </gmd:organisationName>
         <gmd:contactInfo>
```

Import in Geonetwork database

The GeoNetwork has an independent PostgreSQL database separated from the main ESPON 2020 database, but physically located on the same server.

GeoNetwork stores every single xml arborescence relative to a metadata record in a specific field of the "metadata" table of its database. This table is therefore generated and filled in a separate section (schema) from the ESPON database. Along with this table, which contains information on INSPIRE metadata, an auxiliary table (“operationallowed”) is also generated to manage the publication and the access to metadata.

These two tables are then automatically transferred via scripting to the Geonetwork database. Through a Web Service the database is re-indexed to take into account the latest changes.

Workflow

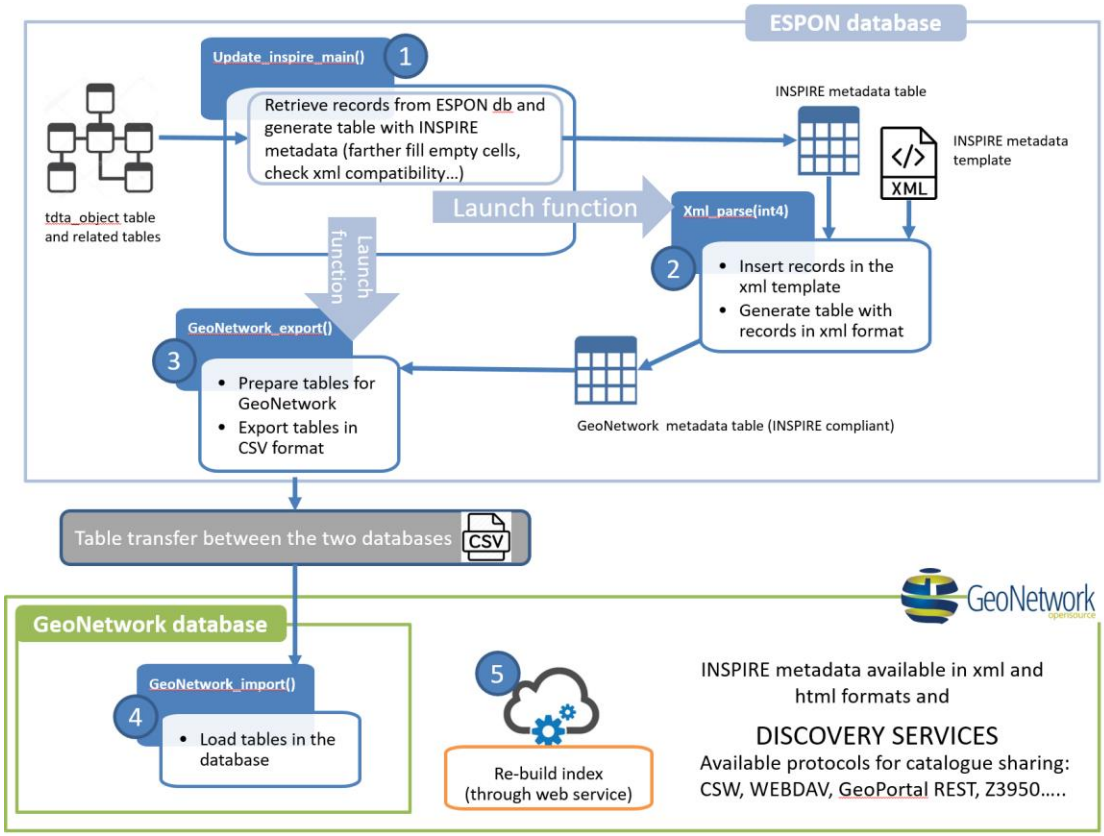
Ideally, for each update or new metadata entry in the *edp_core_tdta_object* table the corresponding INSPIRE metadata must be generated and integrated into GeoNetwork.
The workflow is managed by a Python script that launches a set of internal database functions written in PL/pgSQL language.

1. The function Update_inspire_main() generates the main table with the most INSPIRE metadata elements retrieved from the related tables in the ESPON database. Moreover, It takes care of those characters not compatible with xml encoding (ex.:“&”), and it fills empty text cells using a generic “not set” or “not available” where values from ESPON metadata were not available: this is necessary because null values are not allowed. Then it will launch functions 2 and 3.
2. The functionxml_parse(int) inserts corresponding values for each record in the xml template.
3. The third function, Geonetwork_export(), generates the tables to export in the GeoNetwork database. The tables are converted in csv format and stored in a specific folder of the server.

4. The function geonetwork_import() loads the tables in the GeoNetwork database.

Finally the Index is rebuild to allow GeoNetwork to take into account the changes made to the database. Figure 17 summarises the workflow:

Figure 17: From ESPON to INSPIRE - the metadata workflow



Access

Single metadata

The individual metadata (xml) can be accessed directly through the ESPON platform through http://database.espon.eu or through the GeoNetwork: http://database.espon.eu/geonetwork.

Example for “Predominant relations in the GI network by regions” id 1050 is shown below:

xml

- http://database.espon.eu/indicator/1050/metadata-inspire.xml
- https://database.espon.eu/geonetwork/srv/api/records/d52d8dc6-aa6c-11ea-903f-005056b63a53/formatters/xml

html:

- https://database.espon.eu/geonetwork/srv/eng/catalog.search#/metadata/d52d8dc6-aa6c-11ea-903f-005056b63a53

Catalogue

The whole catalogue can be shared through a harvesting process: it consists in ingesting metadata from remote sources and storing it locally in the catalog for fast searching. It is a scheduled process, so local copy and remote metadata are kept aligned. Several sources can be harvested including:

- OGC-CSW 2.0.2 ISO Profile,
- OAI-PMH,
- Z39.50 protocols,
- Thredds,
- Webdav,
- Web Accessible Folders,
- ESRI GeoPortal
- Other GeoNetwork node.

Example for CSW GetCapabilities endpoint:

http://database.espon.eu/geonetwork/srv/eng/csw?request=GetCapabilities&service=CSW&acceptVersions=2.0.2&acceptFormats=application%2Fxml.

User Interface

3.3 Search and Download (WP2)

3.3.1 Main Data Application

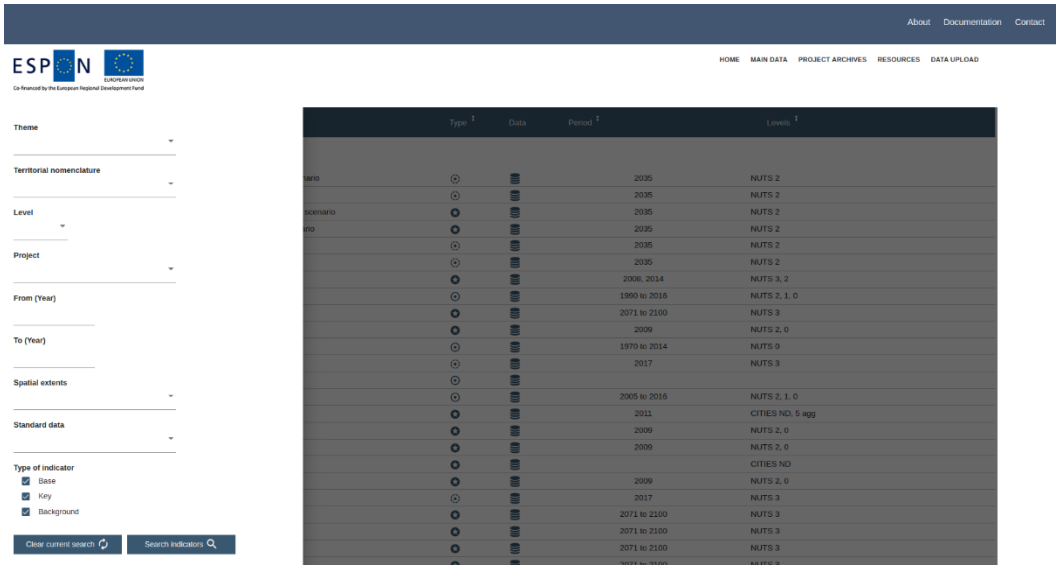
The main data application is the section of the user interface containing the indicators list of the ESPON database as well as the search and download tools.

It contains a guided search allowing to select different elements from closed lists that are automatically generated through web services based on the current database contents.

The available search fields are:

- Theme
- Territorial nomenclature (NUTS, FUAs, LAU…)
- Level (for NUTS)
- Project
- From/To year
- Spatial extents
- Standard data
- Type of indicator (base/key/background) – as checkboxes instead of dropdown

Figure 18: Search options of the main interface

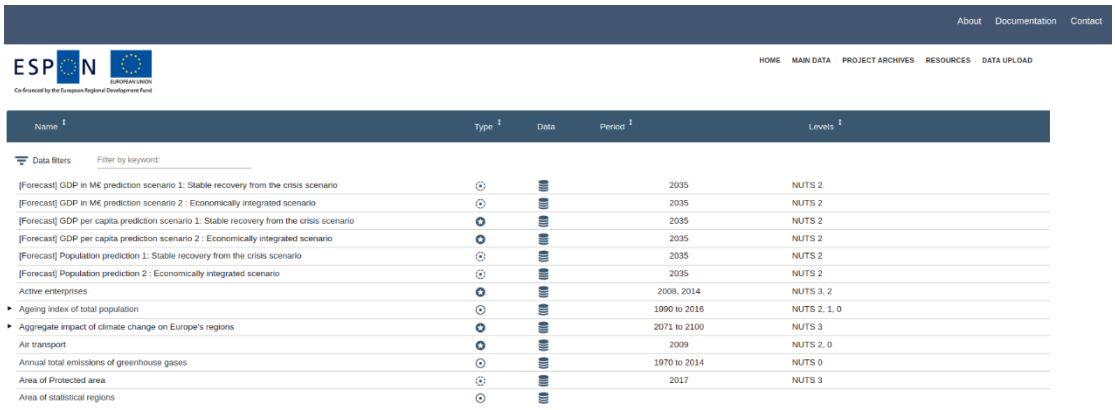


The tree of indicators shown on the interface is read from a data.json file obtained from the server. This file is generated periodically by an automated cron task in the server that gathers all current data on the database to create a new tree structure file. This architecture increases dramatically the performance of the system, as creating the tree on the fly directly from the

database would take too long (about a minute). The guided searches also rely on this cached version of the indicators tree to speed up the searches.

The order of the indicators is predefined in the web services response through a field called “sequence”, but the user can sort the first level of the elements of the tree by clicking the column Name. The levels nested underneath are always displayed according to the sequence field. Sortable fields are marked with a small icon.

Figure 19: Main data section. Indicator tree



By clicking on an indicator, the user gets a new dialog that integrates all the available options in relation to that indicator:

1. Metadata: list of fields with information on the selected indicator. Automatically populated from the database.
2. Explore data: tools to display the indicator data as a graph or a map and to download it.
3. Genealogy: information on the indicator(s) that have been used to compute the current indicator. Available only on indicators for which it has been defined.

Figure 20: Explore data dialog

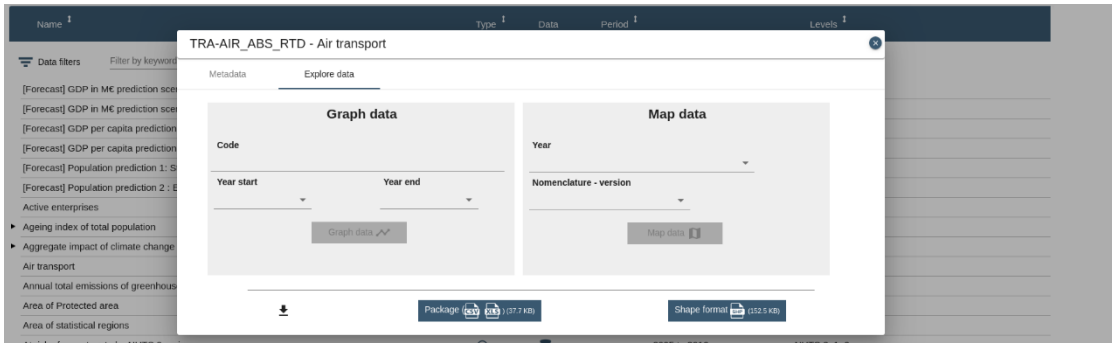


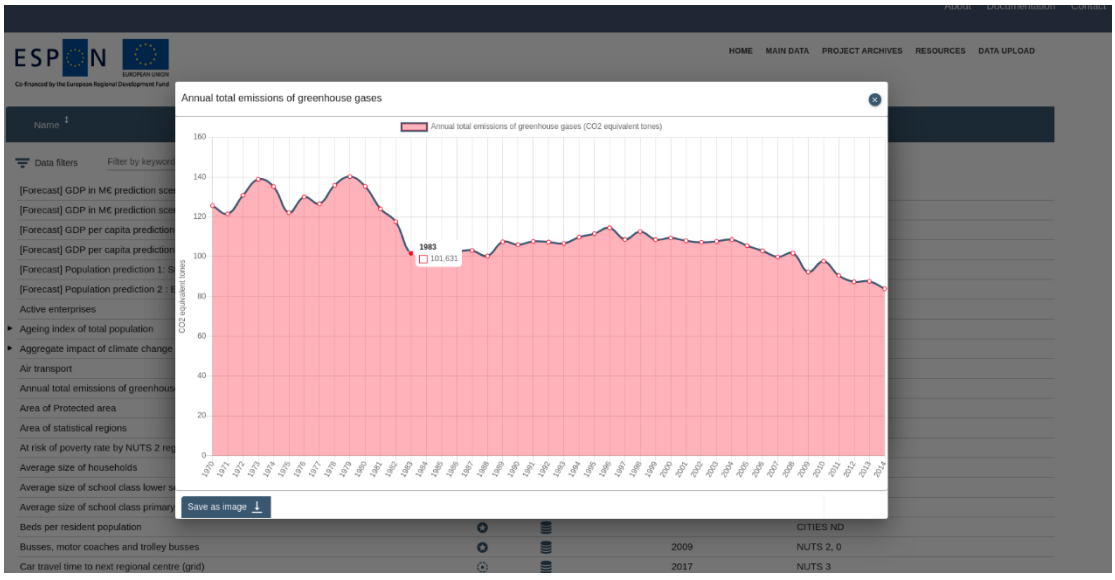
Figure 21: Displaying the genealogy of an indicator

The graphical tool provides filters to narrow down the data to be displayed. There are three possible filters:

- Region: a selector to choose any territorial unit for which data exists for the selected indicator. The choices are automatically populated from the contents of the database.
- Year start: a selector to choose the year to start the graph. The choices are automatically populated from the contents of the database (optional parameter).
- Year end: a selector to choose the year to end the graph. The choices are automatically populated from the contents of the database (optional parameter).

Once the graphic is ready it can be saved as an image. The style of the graphic is preset and cannot be customized by the user.

Figure 22: Example of graphic



The mapping tool shows geographical extent of the availability of data for a given indicator. This is shown as dark grey for regions with data and light grey for regions with no data. Clicking on a region brings up a dialog showing the territorial unit code, name and value for the indicator.

The map follows the ESPON mapping layout with:

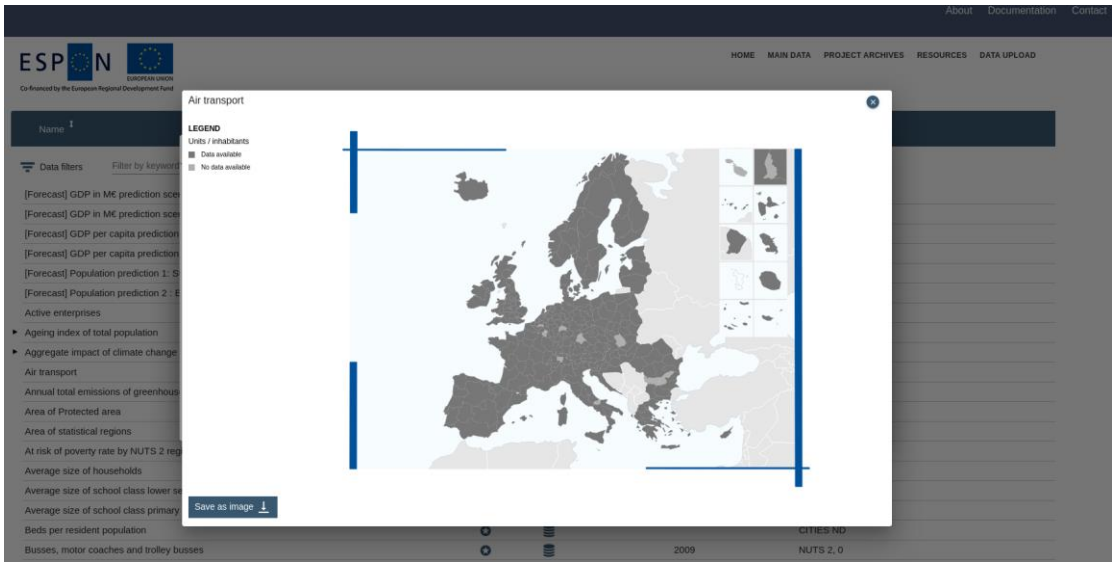
- Ultraperipheral territories as insets in the top right.
- Sea in light blue.
- Non-ESPON countries in light grey.
- Blue lines framing the map.

The map is created by selecting two elements from dropdown lists:

- Year: a selector to choose the year of the data to be mapped. The choices are automatically populated from the contents of the database.
- Nomenclature-version: a selector to choose the geographical division to be mapped (NUTS/LAU… and reference year). Options here could be NUTS2-2006, NUTS3-2013…

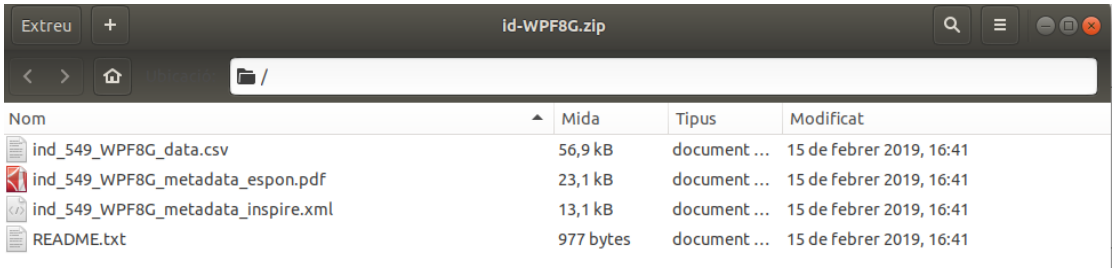
Once the map is ready it can be saved as an image, zoomed in/out and regions clicked to get information. The boundaries shown on the maps are automatically retrieved from ESPON DB through web services.

Figure 23: Mapping an indicator



The Download section allows getting a zipped package with all data and metadata linked to the selected indicator. The link is retrieved from the database through a web service, thus whenever it changes, the interface will automatically show the updated link to the package.

Figure 24: Contents of a data package



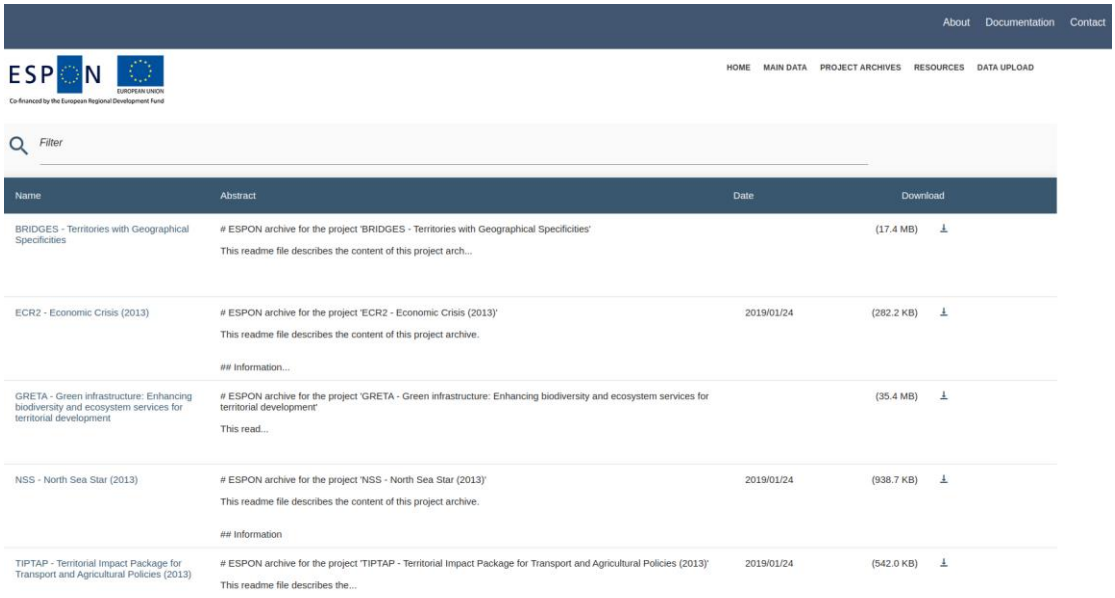
3.3.2 Project Archives Application

This section shows the project archives, a set of information and data packages related to the ESPON projects. Data is shown in table format with a search field at the top that allows looking for any text chain in all the fields of the table. All fields of the table can be sorted in ascending or descending order.

The field “Abstract” contains usually very long texts, therefore it is shown abbreviated. It can be expanded in a dialog box by clicking in the cell.

The “Name” field contains a link to the project website. The “Attachment” field has a link to download the associated information of the project in the form of a zipped file that contains a variety of files.

Figure 25: Project archives section



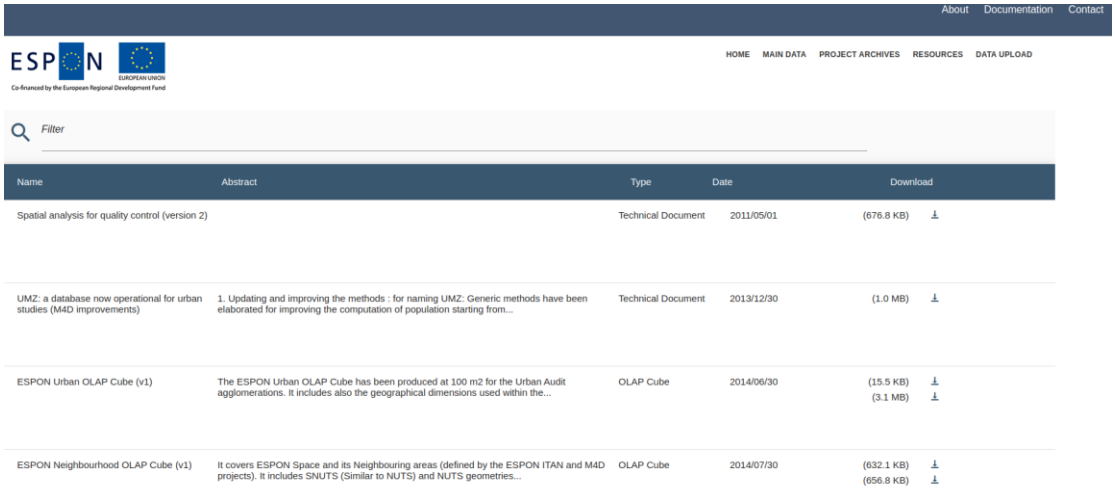
3.3.3 Resources Application

This section shows the resources, a set of varied information such as reports, mapkits or others. Data is shown in table format with a search field at the top that allows looking for any text chain in all the fields of the table. All fields of the table can be sorted in ascending or descending order.

The field “Abstract” contains usually very long texts, it is therefore shown abbreviated. It can be expanded in a dialog box by clicking in the cell.

The “Name” field contains a link to the project website. The “Attachment” field has a link to download the associated information of the project in the form of a zipped file that contains a variety of files.

Figure 26: Resources section



3.4 ESPON 2020 Data and Metadata Upload System

The Upload System allows Project Managers to deliver their Data and Metadata using the following components:

- User interface (deliver, browse and track project Metadata and Data)
- Validation process
- Search tool with facets
- Project archive and Indicator package generation

During the upload process various spatial and data checking are undertaken to analyse the data and compare it to the declared metadata entered previously. It involves the collaboration of multiple stake holders (Project manager, Project reviewer and Project Approver).

3.4.1 Delivery User interface

This Web user interface allows Projects Managers to deliver their Data and Metadata across the **validation Process** (see section bellow) by providing the following elements:

- Various **entry forms** which allow to input Metadata, Data and switch the validation process states.
- A comprehensive **navigation** which allows to browse across the structure of the delivered data and its validation process states.
- A **faceted search** to inspect across different features the delivered data.

The User manual for Project Manager is available here:

https://database.espon.eu/doc/how-to-deliver-my-data.html

3.4.2 Validation process

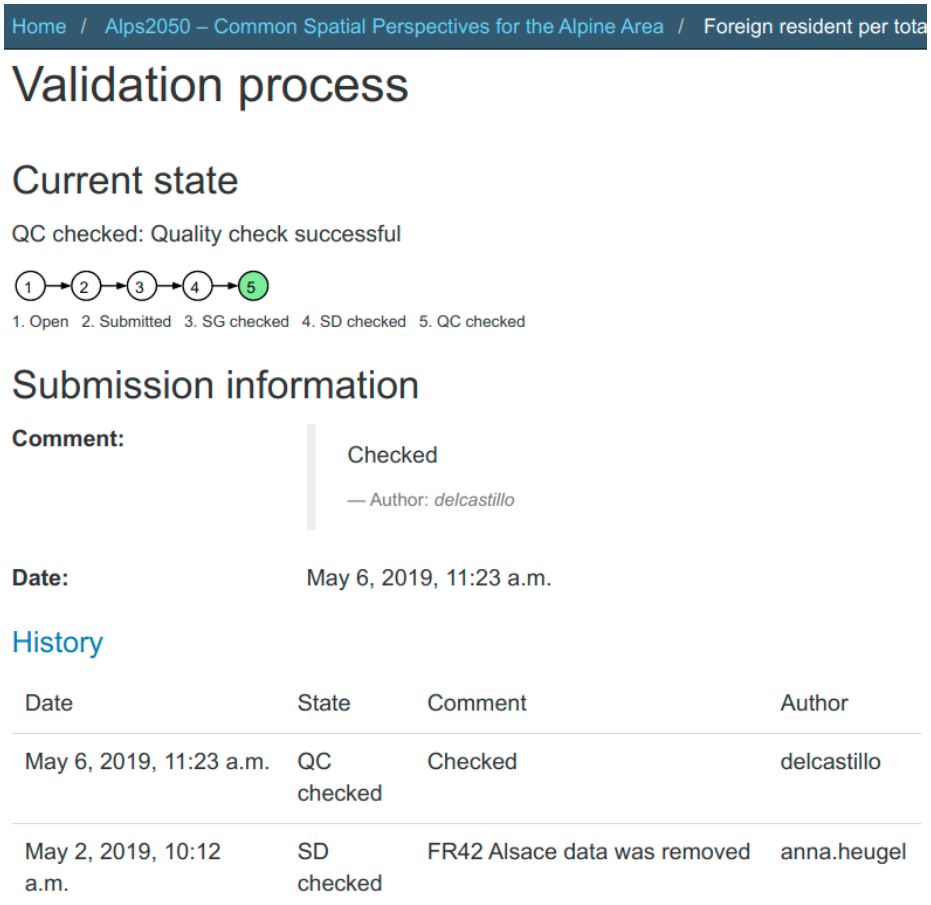
The validation process allows a Project Manager to deliver its data jointly with a Project Reviewer across multiple steps, which ensures that the material delivered reaches a certain level of “normalization” and quality before being published via the ESPON Database Portal *Search and Download* by the Project Approver (ESPON EGTC).

This is done through a workflow which controls Project, Main Data (Data + Metadata) and Other Data delivery.

Summarized workflow:

- Project reviewer creates a Project to which he assign a Project Manager
- Project Manager inserts Main Data and/or Other Data in the Project
- Project Reviewer reviews the delivered Metadata
- Project Manager uploads the related Data for the newly reviewed Metadata
- Uploaded Data is checked against the related Metadata for Main Data, which are described by Indicators
- Project reviewer proceeds to a quality control on the successfully checked Main Data and Metadata pair and accepts Other Data submission(s)
- Once the Main Data quality control is done and Other Data submission(s) accepted the Project Reviewer can set the Project as approved
- Project Approver (ESPON EGTC) proceeds to the final approval of the Project that publishes its items via the User Interface

Figure 27: Other Data validation process screenshot of delivery interface



Further information about the validation process: https://database.espon.eu/doc/project-validation-process.html

Semantic and genealogy check (SG check)

SG check consists in a manual verification of the Metadata content done by the Data Reviewer. The following points should be checked:

- language proofreading of all textboxes (dataset abstract, indicator abstracts, genealogy methodology, etc.);
- genealogy: all background indicators contributes directly or indirectly to the genealogy of (at least) a key indicator.

The *SG check* results either in:

- a validation of the metadata (*SG checked*). In such case, the manager may validate the metadata, which freezes metadata edition and allows the manager to upload data. The manager may alternatively decide to edit metadata.
- a request for further elaboration in the form of a report highlighting changes to be implemented and other points to be checked.

Spatial and Data check

The Spatial and Data check (SD check) is performed automatically by the system upon upload of the standard data file.

The Spatial and Data check results in a report which appears in dataset workflow tab. This report includes:

- **'Errors'** (syntactic inconsistencies between metadata and data). These shall be corrected to go on with the upload procedure.
- **'Warnings'** (information which are of interest for the manager and may induce him/her to revise data and/or metadata).
- **'Information'** (information which could be of interest for the manager and which provides an overview of the content of the dataset).

'**ERRORS'** include **logical input errors detection** resulting from inconsistencies between metadata and data. The procedure checks for:

- Structural consistency of XLS table (existence of expected columns).
- Consistency of Spatial Objects (or unit codes) used in 'data' with the nomenclatures declared in 'metadata'.
- Consistency of Spatial Objects (or unit codes) used in 'data' with the spatial extent declared in 'metadata'.
- Consistency of declared indicator codes and years used in 'data' with the indicator declared in 'metadata'.
- Consistency of the data type in 'data' with the Data Type property declared in 'metadata'.
- Consistency of value labels used as 'data' with the value labels declared in 'metadata' (for NAT Type 'typologies').
- Consistency of the 'sources' used (flags) in 'data' with the sources declared in metadata.
- Consistency of the 'preliminary processing' used in 'data' with the 'preliminary processing' declared in metadata.
- Existence of a source flag to all data cell.

WARNINGS include automatic **completion test and outlier detection**. It provides:

- Completion rate for each indicator/year.
- A list of potential statistical outliers [for stock and ratio indicators]. The system spots unusually high/low values in a data column, by reporting any value outlying from the

median/quartiles (distance to the quartile Q1 and Q3, respectively lower or greater than 1.5 times the interquartile difference) in comparison to similar spatial entities.

- A list of potential trend outliers [for stocks and ratio indicators]. The system spots suspicious movements in time series using median/quartiles method.

'**INFORMATION'** includes:

- Basic descriptive statistics on each indicator/year.
- Typology issues [for typology indicators]. The system provides the proportion of entities under each category at the level of each indicator/year. It may raise attention on over- or under-represented categories of the typology (i.e. categories with a doubtful number of associated spatial entities).

Based on this report the manager has three options:

- If there is no 'error', **submit data for quality control** to the manager.
- **Discard data** and upload a revised version of the file. This option implies that metadata are deemed correct but data file needs to be revised.
- **Discard data and revise metadata**. This option implies that metadata will be checked again for semantic and genealogy consistency.

Quality check (QC check)

Once the manager submits data for quality control, the Database Reviewer performs a last visual check for the consistency of data and metadata based on the warnings which were raised through the SD check procedure.

The DB Reviewer has three options:

1. **Validate the dataset**.
2. **Discard data** and request that the manager upload a revised version of the file, taking into account one or several warnings. This option implies that metadata are deemed correctly but some data files need to be revised.
3. **Discard data and open metadata for revision**. This option implies that metadata need revision based on the information provided through the data file. Any revision of the metadata implies that metadata needs to be checked again for semantic and genealogy consistency.

Validated datasets are not anymore editable by the project manager.

'Main Data' delivery is considered complete when all indicators foreseen in the overview table are included in valid datasets.

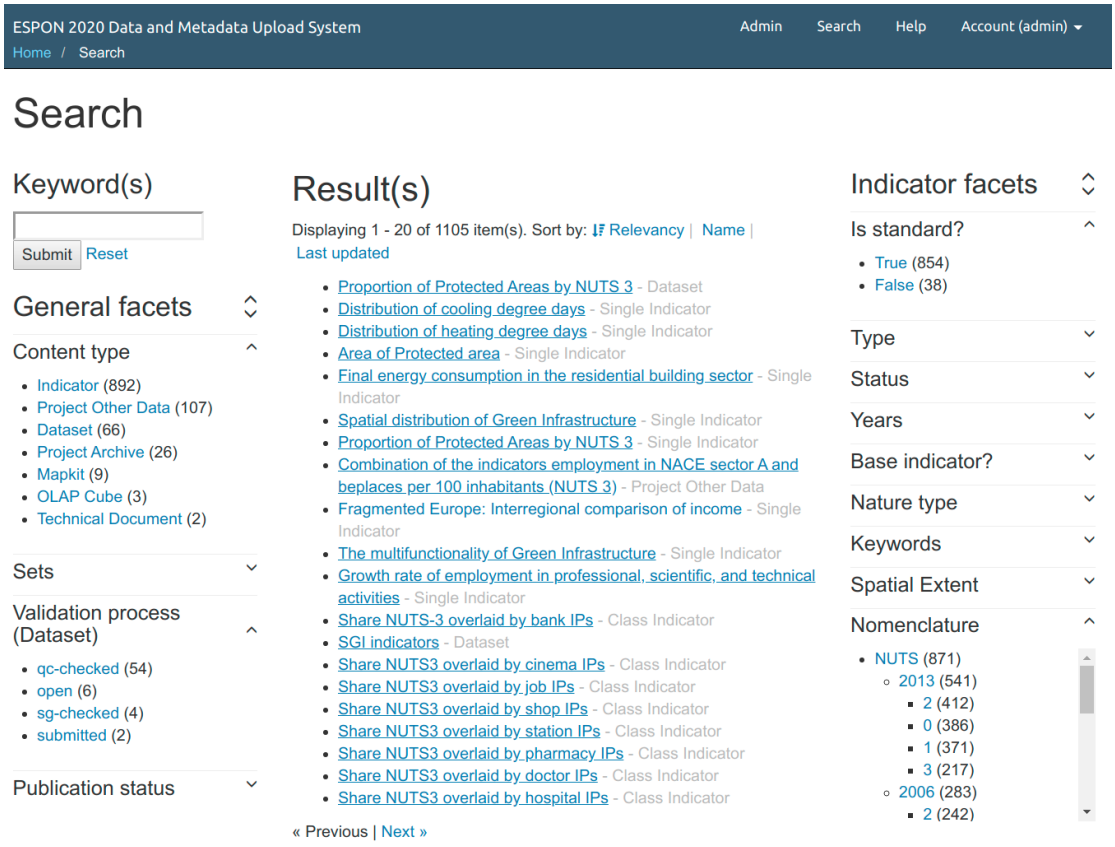
A final version of the "How to Deliver My Data" Manual is found at: https://database.espon.eu/doc/how-to-deliver-my-data.html

3.4.3 Search tool with facets

This tool is available at http://database.espon.eu/search/ for all authenticated user of the Data and Metadata Upload System: *Project Managers*, *Project Reviewers* and *Project Approvers* (ESPON EGTC).

This search tool allows to find elements across the database (Main Data, Other Data, Indicators, Resources, etc, ...) using Facets that expose some chosen "dimensions" of the data set (e.g.: Keywords, Spatial nomenclature, Validation process state, Content type, etc.). This allows to discover and find items by drilling down the result while clicking the Facets items links and observing their count states.

Figure 28: Search tool with facets



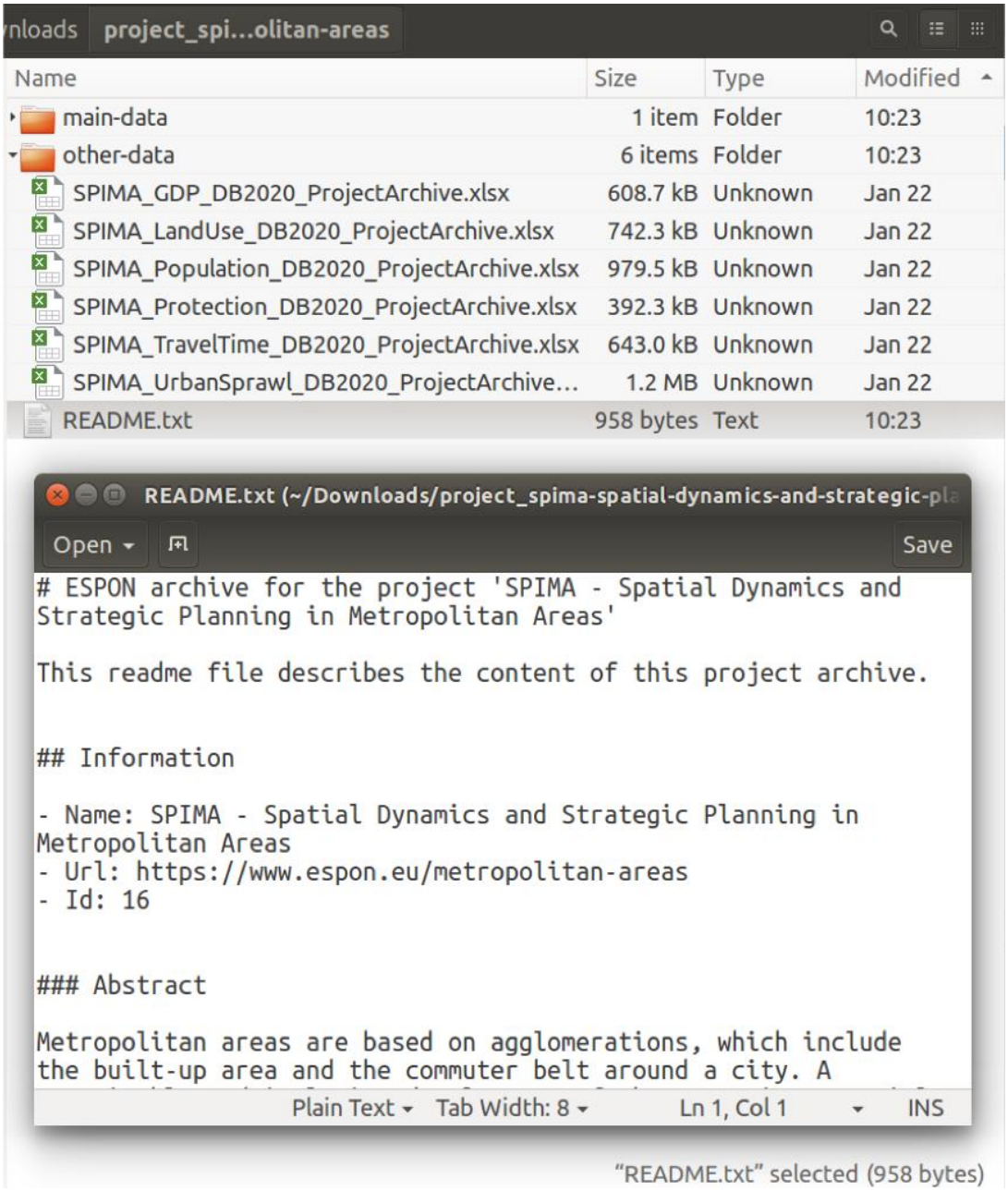
3.4.4 Project archive and Indicator package generation

These elements are built during dedicated transitions of the Validation Process and are made available via the Portal User Interface. See the following page for more information:

Project Archive

A *Project archive* is a zip file that combines each *Other Data* and *Main Data* items within a project.

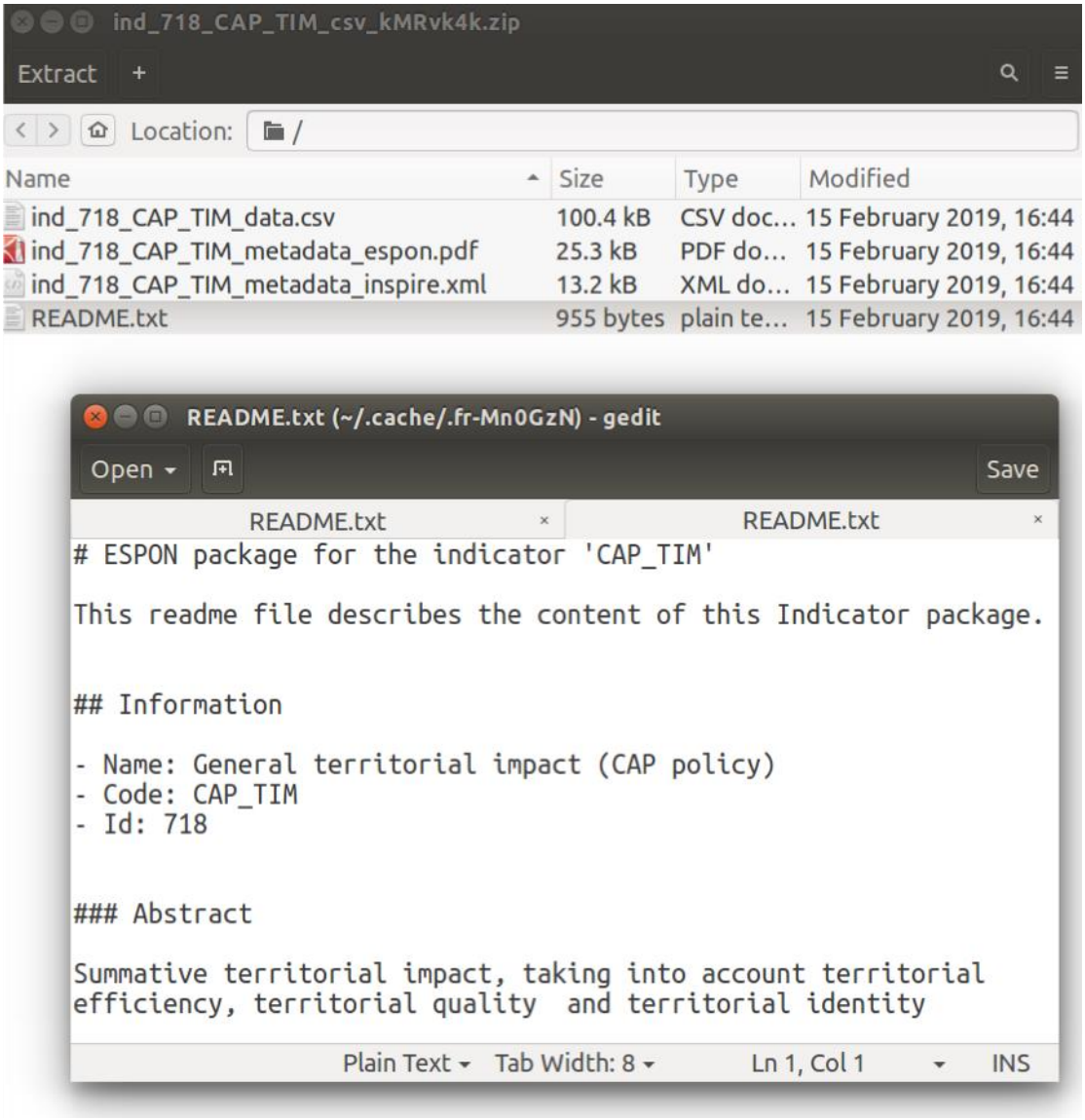
Figure 29: Project Archive



Indicator Package

The *Indicator Package* is a zip file that includes Metadata INSPIRE (XML), Metadata ESPON (PDF) and Data (CSV, SHP).

Figure 30: Indicator Package



Documentation on *Project archive* and *Indicator package* generation is found at: https://database.espon.eu/doc/project-archive-and-indicator-package-generation.html

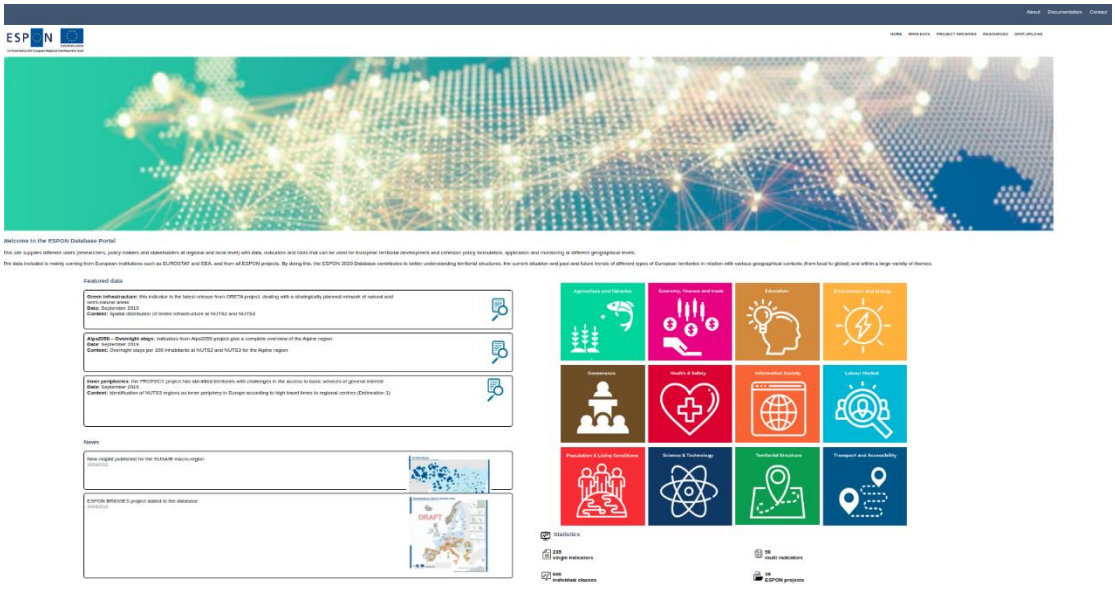
3.5 Administrator Module

3.5.1 User Interface Administration Module

The main data, project archives and resources applications are developed in Angular7, and have been integrated into the Django framework that handles the database, data upload, administration and web services.

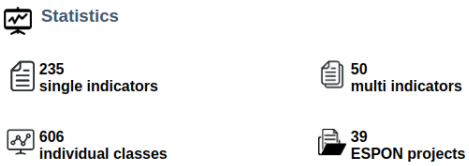
The landing page and the “About” section of the user interface have been developed as static pages within the Django framework.

Figure 31: Landing page



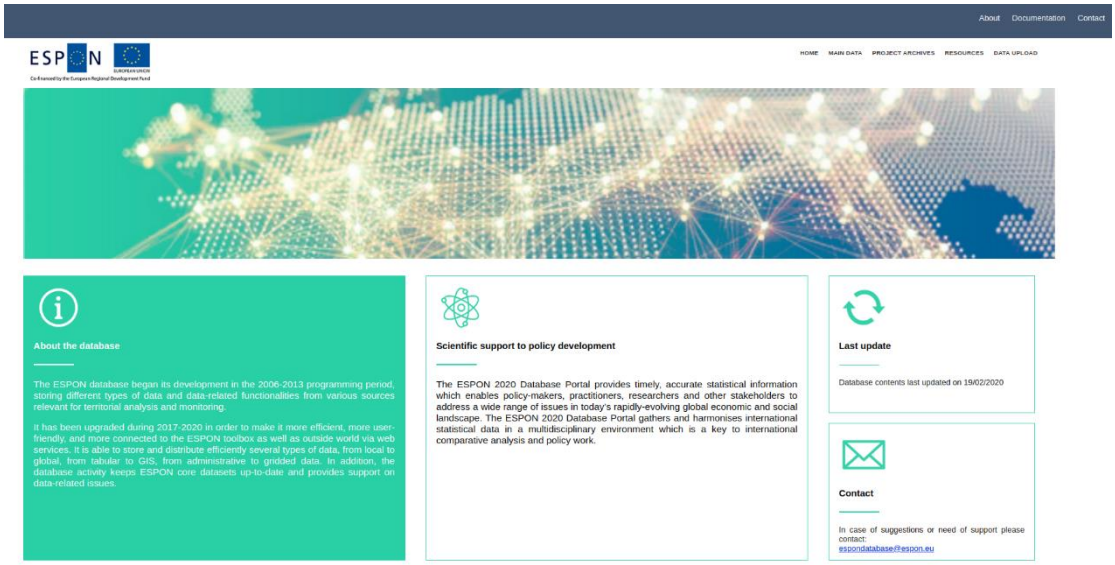
The landing page contains javascript code to make calls to web services that return the necessary data to fill in the non-static parts of the page such as the statistics:

Figure 32: Dynamic statistics



The “about” page contains general information on the ESPONDB portal and is programmed as a static content html page.

Figure 33: About page



3.5.2 Database Systems Administration Module (UNIGE – 2 – 3 pages)

This module allows to manage the following elements:

- Data (for Standard and Non standard Indicators)
- Data tracking (of the Data Upload Validation Process)
- Resources
- Tasks
- New nomenclature insertion

Data management (for standard indicator)

Data management for Standard Indicator must be done through the Data and Metadata Upload interface by re-submitting the existing Standard Data file. This ensures that the Data file will pass through the Data and Metadata checkings.

See Data management (for Standard Indicator) at:
https://database.espon.eu/doc/data-management-for-standard-indicator.html

Management of Data tracking

This tracking allows to manage the Validation Process elements (Projects, Main Data and Other Data). It allows to find Validation Process elements into a particular state and trigger the workflow transitions.

See Management of Data Tracking at:
https://database.espon.eu/doc/management-of-data-tracking.html.

Resources management

Resource management is done using the standard Django Admin interface for the related model (tdta_resource).

See Resource Management at:
https://database.espon.eu/doc/resource-management.html.

Tasks management

Tasks are processes which are triggered along the different transitions of the Validation Process (see https://database.espon.eu/doc/espon-2020-data-delivery-process.html). In certain case it might be necessary to trigger those tasks independently of the workflow and this is possible to do using the Django admin by accessing the related model editing interface.

Here we are to referring to the following tasks:

- Project Archive build
- Project publication (and its elements)
- Dataset Standard Data insertion
- Indicator Package build

See Tasks management documentation at:
https://database.espon.eu/doc/tasks-management.html

User management

Users are managed through the Django Admin system where some common operations can be done on users such as add, edit, turn active/inactive, delete and also Group assignation, which allows to setup users' roles that are relevant for the Delivery Process workflow.

See User management documentation at:
https://database.espon.eu/doc/user-management.html

How to add new nomenclatures (NUTS) to the database

Introduction

This section describes how it is possible for the system administrator to add new geospatial data for NUTS nomenclatures.

It is important to underline that with this operation it is only possible to add new nomenclatures (NUTS) without deleting the existing ones. Deleting preexisting nomenclatures also means eliminating the indicators associated with them.

For these operations it is necessary:

- to have administrator access to the database,
- to use a GIS software for data preparation,
- to use a PostgreSQL client to perform some database operations.

Detailed step by step information is available in the online documentation at:
https://database.espon.eu/doc/how-to-add-new-nomenclatures-nuts-to-the-database.html

1) Prepare the GIS file according to the guidelines

In order to be coherent with the already included nomenclatures, the data should come from Eurostat Reference Data (GISCO database) available at:
https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data

The data must have a well-defined format, including all the geometries and the attribute values in a single file, which includes the four NUTS levels (0 to 3) of territorial units.

The following table gives a guideline for the attributes. It is important to respect the field naming.

Table 10. Required attributes for new standard geometries (nomenclatures)

field name	Type	Lenght	comment
stat_levl_	int4	32	Mandatory: 0 = NUTS0 -> 3 = NUTS3
nuts_id	varchar	5	Mandatory: Official nut ID from GISCO
name_ascii	varchar	255	Mandatory: it is the Official GISCO nomenclature name in ASCII
eu28_4	int4	32	Mandatory: 1 = nomenclature belongs to EU 28 + 4 Partner States
eu28_cc_4	int4	32	Mandatory: 1 = nomenclature belongs to EU 28 + 4 Partner States + Candidate Countries
eu28	int4	32	Mandatory: 1 = nomenclature belongs to EU 28
tc	int4	32	Mandatory: 1 = nomenclature belongs to Territorial Cooperation (TCA)
version_	int4	4	Mandatory: year of reference for nomenclatures in four digits (example: 2016)
gid	int4	32	Mandatory: unique identifier. NB values will be successively automatically re-calculated in order to avoid any duplicates

The spatial reference system of the dataset must be ETRS 1989 **(EPSG 3035)** in Lambert Azimutal projection with center in E52N10, coordinates in meters.

2) Import the GIS file in the (geo)database

This section explains how to import the GIS file into the database. The open source software QGIS[5] is used, but obviously other solutions are possible (e.g. ArcGIS, Manifold).

We strongly suggest saving the GIS file in GeoJson or Geopackage format. The shapefile format could possibly cut the name of some locations if they were particularly long.

3.5.2.1.1Connect to the ESPON database

One can connect to PostgreSQL easily through the QGIS interface, as explained in the online documentation at:
https://database.espon.eu/doc/how-to-add-new-nomenclatures-nuts-to-the-database.html.

3.5.2.1.2Import the GIS file into the "gis" schema of the database

Once the connection is established, it is possible to start importing the GIS file into the database.

[5] https://www.qgis.org

3) Process and check the data through a PostgreSQL function

Once the nomenclatures have been introduced into the database, the administrator simply has to launch a specific PostGreSQL function through a PostgreSQL client such as PgAdmin[6], Navicat[7] or other.

This function allows to:

- Generate unique identifiers for each new geometry.
- Establish hierarchical relationships between the various NUTS levels, which will be written in a specific table.
- Establish the spatial extension of each specific extent (e.g. EU 28, EU 28+4) in terms of NUTS units. This information is written in a dedicated table.
- Finally update the main table that contains all the database nomenclatures.
- Generate a temporary table called *test_nomenclature* in order to check the outputs before to commit to the operational database.

4) Commit to public schema

All these operations are performed in a reserved area of the database, it is therefore possible to verify the outcome of the operations before definitively transferring the new nomenclatures to the operating area of the database.

Once the checks have been carried out, viewing the table created by the function previously, the second and last Postgres functions can be launched. It allows you to permanently transfer the tables to the database's operating scheme.

[6] https://www.pgadmin.org/

[7] https://www.navicat.com/

4 Conclusions and Final Steps

4.1 Achievements

4.1.1 D7 deliveries

The following expected D7 deliveries have been provided:

- The final version of the technical background document has been delivered as an online document https://database.espon.eu/doc/.
- An updated version of the ESPON Database Portal including new, innovative and improved functionalities is running at https://edp.unepgrid.ch/. By June 2020, the Portal will be accessible at ESPON URL http://database.espon.eu/.
- The present report on the work done in relation to the updated version of the database portal.
- A final version of administrator modules to manage the database systems and the user interface of the Database Portal are delivered as a series of tools and procedures. Final version of guidance document for both administrator modules are available at https://database.espon.eu/doc/administrator-module.html and https://database.espon.eu/doc/search-and-download-user-interface.html.
- An update according to the updating strategy has been carried out on the Base indicators, a report on the work done in relation to the updating strategy is included in the present document, a document on the updating process and tools can be found at https://database.espon.eu/doc/espon-2020-core-data-updating-strategy.html
- Source code of all software developed in relation to the ESPON 2020 Database Portal are put on a GIT repository https://git.unepgrid.ch/espon-egtc/ (for credentials to access the repository please contact the admin of the Database - database@espon.eu).
- A power point presentation (max. 30 slides), made up of the main components of the ESPON 2020 Database Portal, has been prepared and delivered with the present report.

4.1.2 Indicators stored in the database (WP 5, 6, 7)

In addition to Base indicators (WP6), Project indicators are the main content of the ESPON Database. Table 11 presents summary statistics on the Base and Project indicators currently published in the database:

Table 11. Number of Base and Projects' indicators, by indicator type (Single, Multi, Dimension, Class)

Indicator Type	Total	Base indicators	Project indicators		
			Total	Key	Background
Single	**211**	29	182	122	60
Multi	**44**	18	26	20	6
Dimension	**63**	27	36	33	3
Class	**601**	300	301	273	28
Total	**919**	374	545	448	97
Single + Multi	**255**	47	208	142	66
Data (Single + Class indicators)	**812**	329	483	395	88

Data of the ESPON 2013 Programme, i.e. 390 Base indicators and 130 Project indicators, were imported from the previous M4D database (WP5) and have been cleaned in order to comply with the current structure of the database (reorganisation into the new 2020 Multi indicator structure when needed, renaming of indicators, coding of null values). Base indicators have been updated, when possible up to years 2017 and 2018 (WP6).

The process for the delivery of ESPON 2020 data (WP7) is robust: an overview table ensures that the essential data related the main outputs of projects are delivered, Data reviewers at UAB and MCRIT have been trained, which enlarges the capacities to support current and future projects in their data delivery. 34 ESPON 2020 projects have been supported in their data delivery (see section "Projects whose data delivery was supported", p. 17):

- 15 Applied research (SO1) projects
- 15 Targeted analysis (SO2) projects
- 4 Monitoring and tools (SO3) projects

4.1.3 Database system, data & management tools, web services and web interface (WP 1, 2, 3, 4, 8)

The database model (WP1) allows the management of ESPON 2020 data types and structures, nomenclatures, spatial extents (e.g. EU28, EU28+4, Transnational Cooperation Areas), geometries in a flexible and extensible way.

The REST Web services http://database.espon.eu/api/ (WP3) provide ESPON 2020 data in a customized and secured manner, for their use in third party applications. OGC Web services are fully functional for exposing metadata and data in international standard formats, including WMS, WFS, GML, CSW, INSPIRE (http://database.espon.eu/geoserver/, http://database.espon.eu/geonetwork).

The Data and Metadata Upload interface (http://database.espon.eu/data-upload/) is fully functional (WP 4) and it has been extensively used for the delivery of ESPON 2020 project data. The User Interface (http://database.espon.eu/), so far developed as a separate Joomla

application, has been migrated to the same Django environment as the Upload Interface for a better integration and an easier deployment on the external Web hosting provider Hypertech (WP2). This deployment is carried out by means of a Docker container making the applications portable on any operating system. The various components of the administror module (WP8) have been developed as tools or processes, their descriptions are available at https://database.espon.eu/doc/administrator-module.html.

4.1.4 General support and coordination (WP 9, 10)

Seven WP9 (Support) Request of Services have been issued:

Table 12. WP9 issued

RoS nr	Task
5	1) New functionalities for handling transnational cooperation areas (TCA), 2) Update indicator fields and related interfaces, 3) Bulk validation of deliveries
13	Provide support to EGTC on map production for publications
14	Review ESPON SoET (State of the European Territory Report) maps
16	Create an evidence pack for Czech country fiche
17	Develop 30 new MapKits for country fiches
18	Additional functionalities (light grouping, smart completeness indicators) and harmonisation of indicator (naming, "no data"" values)
19	Harmonised geometries for SNUTS/NUTS 2016

The RoS mechanics proved to be an efficient mean to add unplanned functionalities during the project. The four last RoS (16 to 19) shall be progressively completed by the closure of all ESPON 2020 Database Portal activities (May 2020).

4.2 Lessons learned, future perspectives

The Database, a central tool for the ESPON Programme, remains a complex undertaking in every respect, whether in terms of definitions, processes, tools and communication, as it was already underlined by the previous ESPON 2013 M4D Database project.

Objectives, concepts, procedures and tools of the ESPON 2020 Database are now well established and shared among the persons and teams involved in the data delivery process (EGTC, the ESPON 2020 Database partners, ESPON projects). The ESPON 2020 data collection was successfully completed, it is done in coordination with the final deliveries of the projects. Data are structured efficiently and flexibly (extensible data model), they are disseminated by various means (Search & Download web interface, web services) that allow a their direct use by multiple systems (e.g. GIS and spreadsheet software, third party web applications).

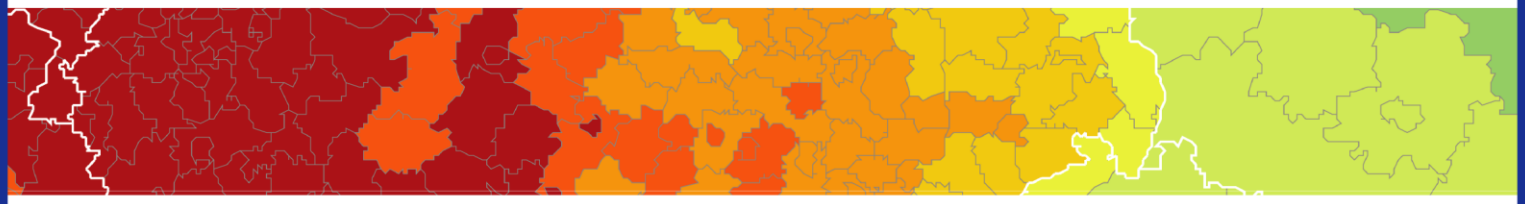
For the next ESPON Database project, a number of further activities are suggested:

- Modify the system (data model, Search & Download and Upload interfaces) to allow the discovery of more data types in the main search interface : GIS data, non

standard indicators (i.e. statistical data not compliant with a territorial nomenclature stored in the Database, or data delivered in multiple nomenclatures (or nomenclature levels)). This would also allow to disseminate all kinds of data as unique packages (this is currently the case only for the Main indicators, all “Other data” are only accessible through the Project Archives).

- Modify the system to allow the insertion into the database of new delineations that are not strictly territorial nomenclatures, e.g. macroregional delineations constructed with units from various nomenclatures/levels (SNUTS, NUTS0, 1, 2, 3). This would imply creating a third type of indicators “Quasi Standard” or “Candidate Standard” in addition to “Standard” and “Non Standard”.
- Update the Base indicators when time series will be available from Statistical Offices (in particular Eurostat) in the 2016 nomenclature.
- Implement a proper search engine (such Apache Solr) with an appropriate tagging of all information resources in order to offer one single entry to the entire Database content (data and non data).
- Implement more data exploration tools, in particular:
    - add base thematic mapping functionnalities showing the data content as quick-looks following established semiological rules based on the nature type of the indicators (e.g. stock data represented as proportional symbols)
    - connect the ESPON 2020 Database to a Business Intelligence (BI) tool (such as Apache Superset) in order to take advantage of its exploration capabilities.

  These exploration tools, without replacing proper web mapping applications or monitoring tools, have a huge potential to reveal unexpected structures in the wealth of data that the ESPON 2020 Database gathered from the ESPON community.



ESPON 2020 – More information

ESPON EGTC
4 rue Erasme, L-1468 Luxembourg - Grand Duchy of Luxembourg
Phone: +352 20 600 280
Email: **info@espon.eu**
www.espon.eu, **Twitter**, **LinkedIn**, **YouTube**

The ESPON EGTC is the Single Beneficiary of the ESPON 2020 Cooperation Programme. The Single Operation within the programme is implemented by the ESPON EGTC and co-financed by the European Regional Development Fund, the EU Member States and the Partner States, Iceland, Liechtenstein, Norway and Switzerland.