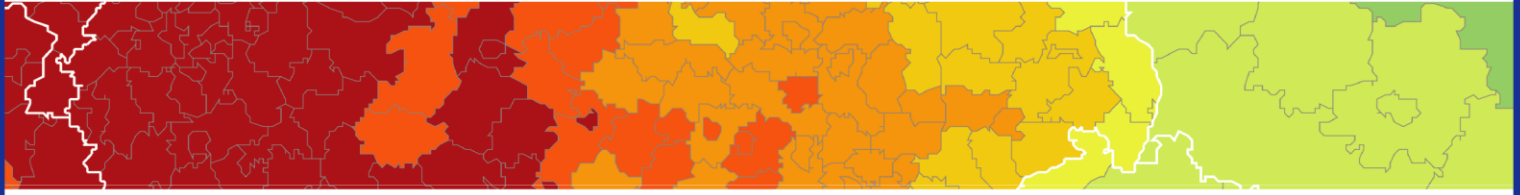**Inspire policy making by territorial evidence**

ESPON 2020 Database Portal

Scientific Platform

Executive Summary

Version 20/05/2020

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
National University of Ireland Maynooth - NUIM (Ireland)
- Martin Charlton

ESPON 2020 Database Portal
## Sharing Essential ESPON Data

1 Objective

The ESPON 2020 Database Portal ensures the availability of harmonised and accurate data on the European territory and neighbouring countries, which enables policy-makers, practitioners, researchers and other stakeholders to address a wide range of territorial issues in today's rapidly-evolving society, economy and environment. The ESPON 2020 Database Portal gathers, manages and disseminates international statistical (as well as geospatial) data, in order to make them available in a user- and machine friendly manner, ready for direct use in analyses and policy work for the ESPON community and the outer World.

2 The ESPON 2020 Database content

2.1 Types of data

The main data in the ESPON 2020 Database are statistical indicators of different kinds (Figure 1).

Figure 1. Indicators in the ESPON 2020 Database
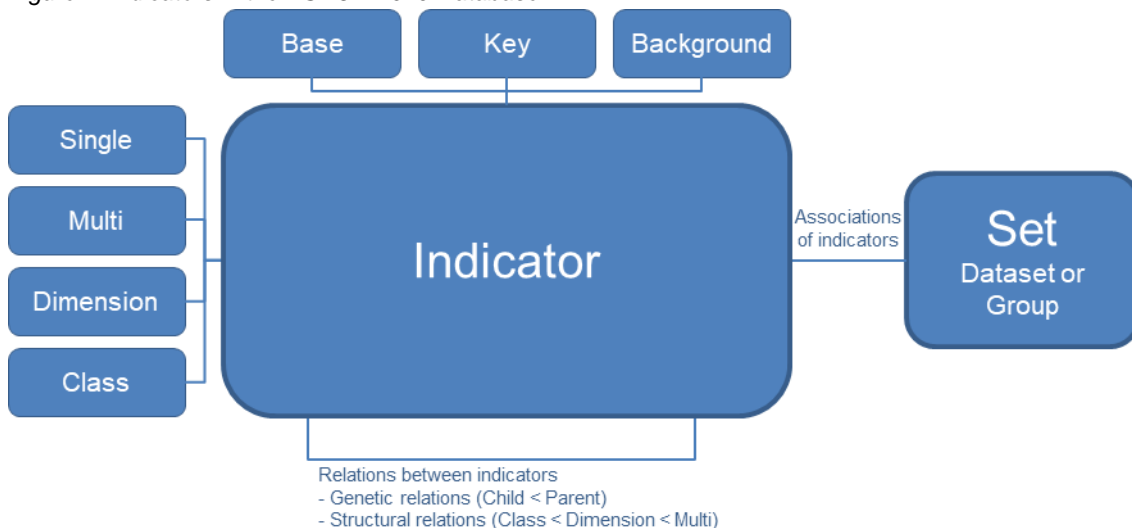
To begin with, the ESPON 2020 Database maintains a set of **Base indicators** of the highest possible quality, regularly updated, useful and usable for the work of the ESPON Programme and projects. The Base indicators are time-series of base data (e.g. population, Gross Domestic Product) originating from official sources, such as Eurostat or National Statistical Offices, and complemented by estimations in case of data gaps.

Beyond these harmonised Base indicators, the richness of the ESPON 2020 Database Database lies in the data (and metadata) collected from ESPON projects, i.e. the **Key indicators**, which are considered by the projects as their main results. Key indicators usually use a standard nomenclature, they may have multiple parent (background) indicators. The genealogy of Key indicators describes the **Background indicators** used in the process of creating a Key indicator (either original compiled data or a significant step in the calculation process). Standard Background indicators use standard nomenclatures, contrary to non-standard background indicators (e.g. grid data, flow data).

The originality of the data in the ESPON 2020 Database Portal is also the extent to which relations between indicators are described, providing a more readable information structure. Indicators may be linked on two bases:

(1) **Structural** relations (M**ulti** indicators / **Dimension** indicators / **Class** indicators). A Multi indicator is an indicator that has one or more Dimensions and/or Classes, e.g. "Total population" (Multi indicator) broken down by "Age groups" (Dimension) and specific age intervals (Classes).

On the contrary, a **Single** indicator is not broken down into Dimensions and Classes.

(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 **Parent(s)** – was used in the process of calculating the other – the **Child**. A methodology that describes the calculation step is associated to each genealogy relation.

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).

In addition to the Base indicators and Key (along with Background) indicators, the ESPON 2020 Database Portal provides other data:

- **Project Archives**, which gathers the whole data content produced by each ESPON Project, including other data of interest, in addition to the above Project indicators.
- **Resources**, i.e. various types of information resources such as reports, mapkits.
- The Database also stores **Spatial data**, which includes the definition of territorial nomenclatures (NUTS, FUA, etc.), spatial extents (EU28, EU 28+4, Territorial Cooperation Areas, etc.) and geometries (GIS data).

Each content is described by **Metadata**, which allows its discovery through search tools.

2.2   Indicators statistics

The ESPON 2020 Database contains 812 statistical indicators (as of 20 March 2020). Table 1 and Figure 2 below show the distribution of the various kinds of indicators:

Table 1. Number of ESPON 2020 Base and Project by type of indicators

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

Figure 2. ESPON 2020 indicators, Base/Key/Background and by types
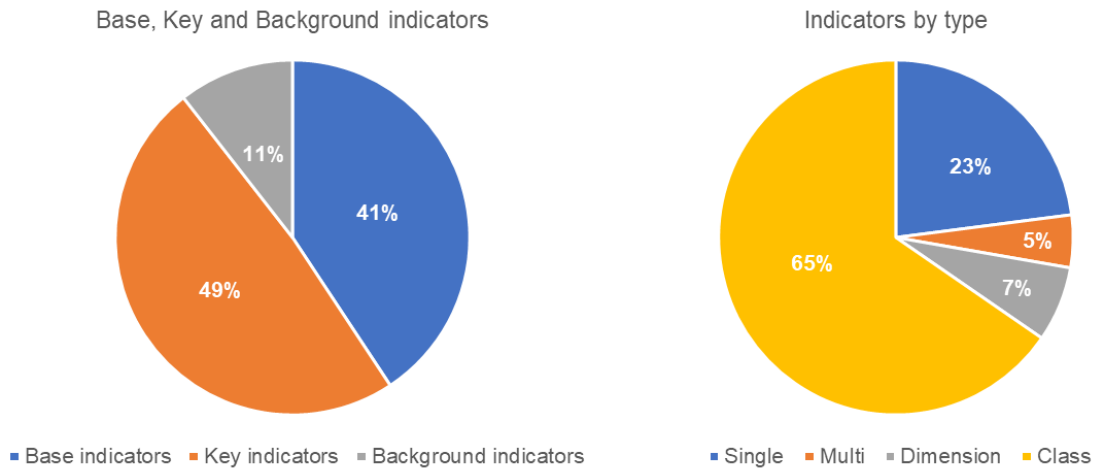
The number of indicators available in the ESPON 2020 Database is growing as projects continue to deliver their data.

3 Functionalities of the ESPON 2020 Database Portal

The ESPON 2020 Database Portal is organised in several components, as illustrated in Figure 3, associated with various functionalities for the collection, management and dissemination of data.

3.1 The Database system, data and management tools

At the core of the ESPON 2020 Database Portal is the Database system with its data and management tools.

The Database model, implemented in PostgreSQL/PostGIS and Django technologies, allows for 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 Data and Metadata Upload interface (https://database.espon.eu/data-upload/) enables the description, quality checking and upload of data from projects. These steps are framed within a robust Data Delivery Process (https://database.espon.eu/doc/espon-2020-data-delivery-process.html): an overview table ensures that the essential data related to the main outputs of projects are delivered, data reviewers have been trained to support current and future projects in their data delivery, by means of the Upload interface. 34 ESPON 2020 projects have been accompanied in their delivery process so far, i.e. 15 Applied research (SO1), 15 Targeted analysis (SO2), 4 Monitoring and tools (SO3).

Two other specific processes allow for the migration of 2013 indicators (https://database.espon.eu/doc/espon-2013-data-migration.html) and for the update of Base indicators (https://database.espon.eu/doc/espon-2020-core-data-updating-strategy.html). 420 indicators from the ESPON 2013 Programme were imported from the previous M4D database into the new 2020 data structure, i.e. 130 indicators from 8 projects (13 datasets) as well as

390 Base indicators, representing 18 Multi indicators and 29 Single indicators and covering the thematic fields of demography, employment, education, economy, energy, society, research&development, information society. The Base indicators have been updated, when possible, up to years 2017 and 2018.

Figure 3: ESPON 2020 Database Portal components
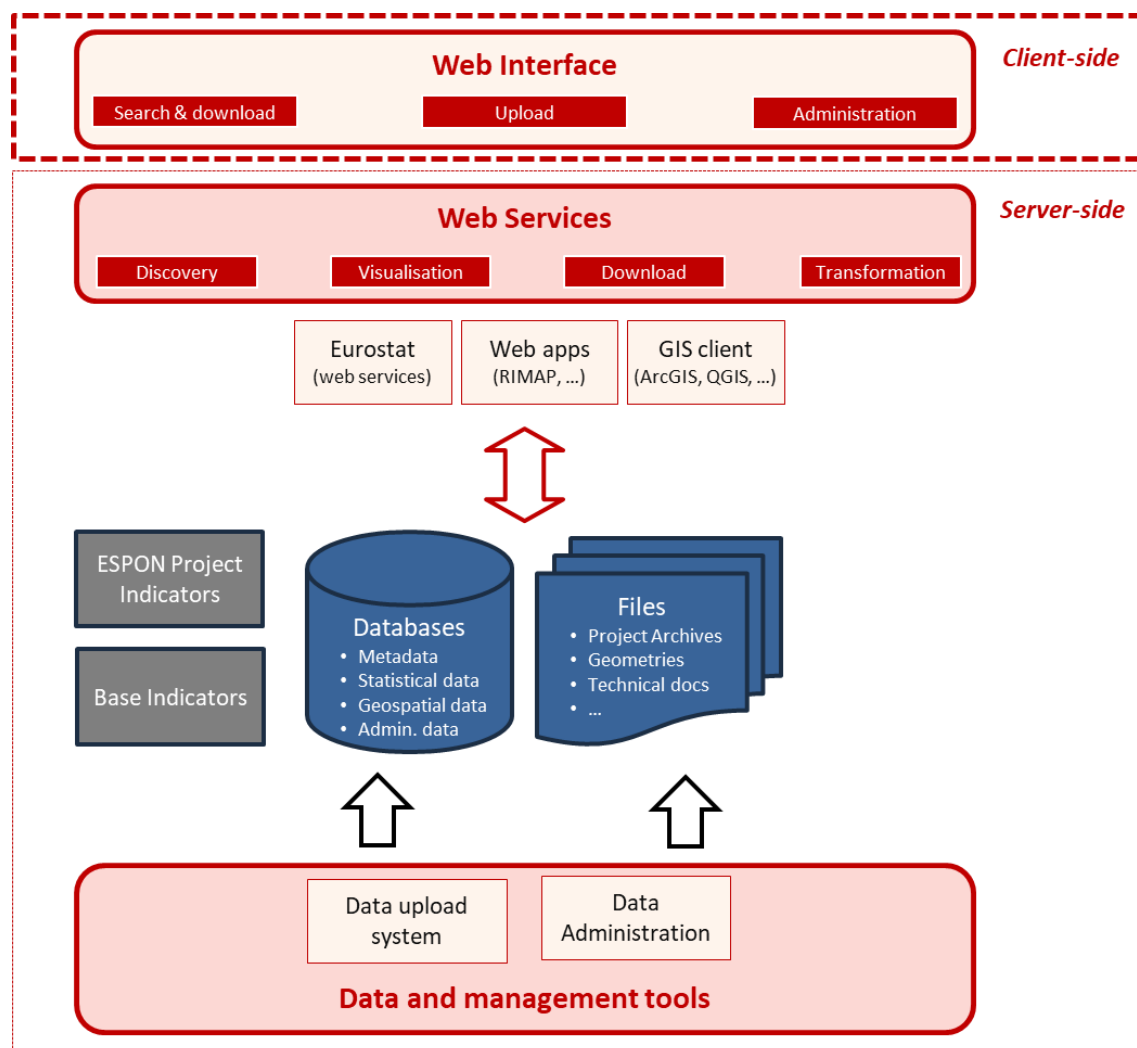
The Administration Module is used for various tasks such as the management of users, data (including nomenclatures and geometries) and other information resources, quality checks, the generation of project archives and indicator packages (disseminated via the User interface). The module is composed of several tools and processes, which are documented at https://database.espon.eu/doc/administrator-module.html.

The entire system (including the dissemination interfaces described in the next section) is packaged in a Docker container, making the ESPON 2020 Database Portal deployable on any operating system.

3.2 Dissemination interfaces

Data are made available to internal and external applications through a series of Web services. The REST Web services (https://database.espon.eu/api) provide ESPON 2020 data in a customized and secured manner, and they can be easily further developed and adjusted in the future. OGC Web services expose metadata and data according to international standards, including WMS, WFS, GML, CSW, INSPIRE (https://database.espon.eu/geoserver, https://database.espon.eu/geonetwork).

The User interface (https://database.espon.eu/) is the web interface providing users with an interactive access to the various data of the ESPON 2020 Database Portal, i.e. main statistical indicators, project archives, other information resources, documentations. The "Main Data" section of the User interface allows for the explorative search and the download of the Base, Key and Background indicators.

3.3 Support activities

The ESPON 2020 Database Portal provides on-demand support for various data processing and mapping activities, such as the preparation of harmonised geometries and map kits, the production of maps for publications, the refinement of quality checks, the update of functionalities in the interfaces and tools. The requests for support, expressed as new needs, are periodically identified, and they are covered based on the financial and human resources available.

4 Lessons learned, future perspectives

The ESPON 2020 Database Portal, a central tool for the ESPON Programme, remains a complex undertaking in all respects, 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 processes (EGTC, the ESPON 2020 Database partners, ESPON projects). The ESPON 2020 data collection is successfully carried out 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 for a direct use by diverse users and systems (e.g. GIS and spreadsheet software, third party web applications).

For the next ESPON Database project, a number of further activities are suggested:
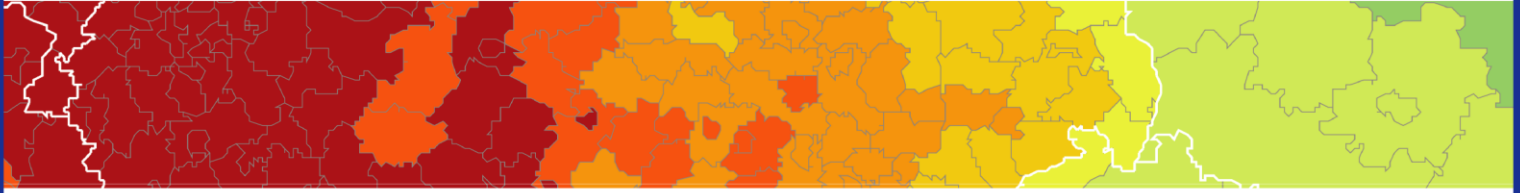
- Modify the system (data model, Search & Download and Upload interfaces) to allow the discovery of more data types in the main search interface : GIS data, non standard indicators (i.e. statistical data not compliant with a territorial nomenclature stored in the Database or data delivered in multiple nomenclatures). This would also

enable the dissemination of all kinds of data as unique packages (this is currently the case only for the Main statistical indicators, all "Other data" are only accessible through the Project Archives).

- Modify the system to allow the insertion 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" types.

- 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 (statistical data and other data).

- Implement more data exploration tools, in particular:

  o add base thematic mapping functionalities 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);

  o 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