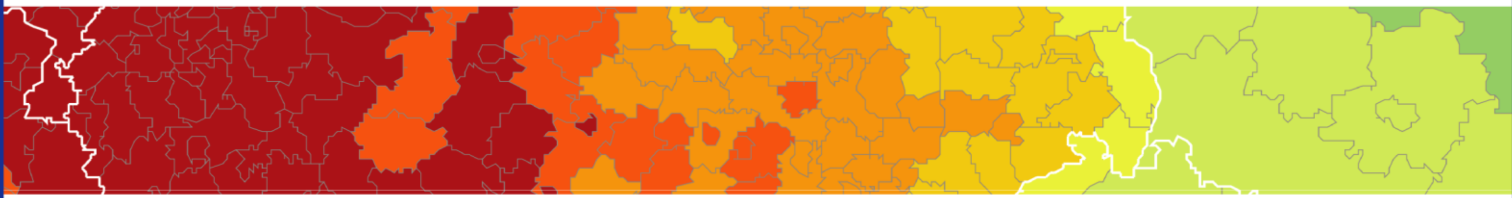
ESPON
EUROPEAN UNION
Co-financed by the European Regional Development Fund

Inspire policy making by territorial evidence



ESPON FUORE - Functional Urban Areas and Regions in Europe

Monitoring and Tools

Final Report

Version 10/07/2020

This monitoring and tools 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.

Authors
Roger Milego (coordinator), Maria José Ramos, Francisco Domingues, Jorge López, UAB (Barcelona)
Jacques Michelet, UNIGE (Geneva)
Juan Arévalo, Beatriz Martín, Gloria Passarello, Ana Barbosa, RANDBEE (Málaga)
Katerina Jupova, GISAT (Prague)
Efrain Larrea, Oriol Biosca, MCRIT (Barcelona)

Advisory Group

Project Support Team:
Natasa Kaspari, Ministry of Interior, Department of Town Planning and Housing, Cyprus
Valerie Lapenne, Commissariat général à l'égalité des territoires (CGET), France
Liviu Băileșteanu, Ministry of Regional Development and Public Administration, Romania

External expert: Valeriya Angelova, Eurostat

ESPON EGTC: Zintis Hermansons (Project Expert), György Alfoldy (Financial Expert)

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.

Contact: info@espon.eu

ESPON FUORE - Functional Urban Areas and Regions in Europe

Table of contents

List of Figures

List of Tables

List of Maps

Abbreviations

CLC	Corine Land Cover
DB	DataBase
EBM	EuroBoundaryMap
EC	European Commission
EGTC	European Grouping of Territorial Cooperation
ERG	European Reference Grid
ESM	European Settlement Map
ESPON	European Territorial Observatory Network
EU	European Union
FR	Final Report
FUA	Functional Urban Areas
FUORE	Functional Urban areas and Other REgions (acronym applied to this project)
GHS-BUILT	GHS-BUILT layer produced in the GHSL (Global Human Settlement Layer) framework
GI	Green Infrastructure
GI PN	Green Infrastructure Potential Network
IR	Interim Report
KO	Kick-Off
LMA	Labour Market Area
MSA	Maritime Service Area
MUA	Morphological Urban Area
NACE	Nomenclature generale des Activites economiques dans les Communautes europeennes
NSI	National Statistical Institute
NUTS	Nomenclature of Territorial Units for Statistics
OLAP	OnLine Analytical Processing
PST	Project Support Team
SMD	Spatial Multidimensional Database
ToR	Terms of Reference
UAB	Universitat Autònoma de Barcelona (Autonomous University of Barcelona)
UNIGE	Université de Genève (University of Geneva)
URL	Uniform Resource Locator
UX	User Experience
WP	Work Package

1 Introduction

The current document is the Final Report (FR) of the “ESPON Functional Urban Areas and Regions in Europe” (ESPON FUORE) project and it corresponds to its fourth and last deliverable (D4). This report has been elaborated by the awarded service provider, coordinated by the UAB, and it is focused on the different tasks carried out since the project KO, together with the decisions made and the final results obtained by the estimation methodologies implemented and the validation of such results. It also pays special attention to a comprehensive description of the functionalities of the final version of the Web Data Analysis Toolbox, and the FUORE web tool (and the Web Data Analyst), which is actually the main product of the D4 and the whole project.

This project has had two main goals:

1. To build an updated ESPON OLAP Cube (hereafter referred as ESPON Spatial Multidimensional Database[1] or ESPON SMD) with data for FUAs and other selected functional regions.
2. To develop a web tool to exploit the new ESPON SMD and facilitate the analysis of the data and benchmarking of FUAs and selected functional regions.

These goals have been achieved by means of several tasks and milestones which are described and discussed in this document. Those tasks have been organised in five different WP, namely:

- WP 1: Data sources and methods for data estimation
- WP 2: Technological development of the Spatial Multidimensional Database
- WP 3: Validation of the method
- WP 4: Development of the OLAP-based web tool
- WP 5: Coordination and administrative support

After a comprehensive background section, the following chapters correspond to these WP and summarise the methodologies undertaken, all the decisions made regarding source and ancillary datasets, all the results obtained with regard to the estimation of indicators, the validation of the results and the web tools implemented (the FUORE web tool plus the complementary Web Data Analysis Toolbox). After reading the FR, it should be clear how the service provider has faced the following challenges:

- Identifying and obtaining delineations for functional regions different from FUAs.
- Improving former disaggregation methodologies.
- Developing a comprehensive methodology and validating it.
- Dealing with a huge amount of data to be processed and stored.

[1] A database resulting of a disaggregation/reaggregation methodology containing indicator figures by several spatial objects, namely FUAs and/or other functional regions, including long time series.

- Estimate indicators from ESPON DB base indicators at functional area level.
- Optimise the required processing time.
- Engage with different experts for validation.
- Developing a user-friendly web tool and optimising its performance.
- Providing an extra tool for ad hoc indicator estimation.
- Ad hoc requests from users to add new data in the web tool.

To be as concise as possible, different annexes with extra information are included and referred to within the text.

2 Background

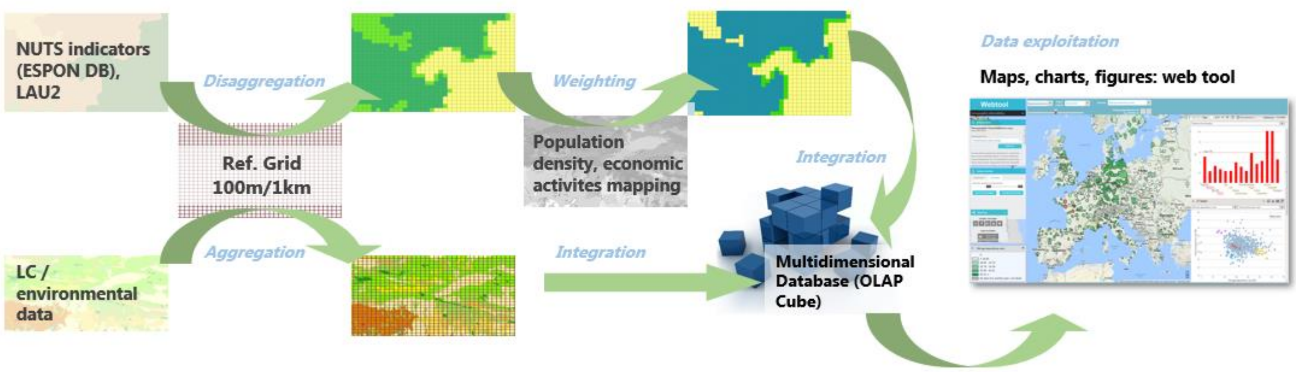
This section is aimed at describing the philosophy behind this project and providing an overview of the methodological approach that has been implemented.

2.1 The philosophy of a Common Data Integrator and OLAP

When it comes to integrating and combining data from different sources that do not match spatially, there are different possibilities. During the M4D project (ESPON, 2015), the analysis of different experiences of data integration worldwide (Deichmann et al, 2001; Gallego, 2001; Nordhaus, 2006) led to the conclusion that the best way to downscale socioeconomic data and make them comparable with other kind of data, was using a regular grid structure, in which each cell takes a figure of the indicator or variable. We assume this idea and propose the 1 km^2 European Reference Grid (ERG), adopted by several European stakeholders at the First Workshop on European Reference Grids (JRC-IES-LMU-ESDI, 2003), as a common data integrator layer. This does not mean that we won't work at higher spatial resolutions (typically 100x100 m) when disaggregating and weighting information, to obtain the final database of figure per each grid cell.

Once all the information is "gridded", i.e. transferred into a grid structure, to facilitate the exploitation of such huge database, Online Analytical Processing (OLAP) technologies can be used. OLAP is a computer-based technique to answer multi-dimensional analytical queries swiftly. OLAP tools enable users to analyse multidimensional data interactively from multiple perspectives. It experienced a strong growth in the late 90s, but it had been mainly applied to business data in the search for business intelligence. The application of such technique to social and geographical information was rather innovative and provided added value to the integration of data by means of a grid. However, the former OLAP Cube will be replaced by the ESPON SMD, a multidimensional aggregated database, as referred in previous section, which will serve the web tool. The main reason is that OLAP tools for spatial data have not experienced further development and have become obsolete to serve a web tool that needs to be fast and reliable. Figure 2.1 shows the overall process of data transformation through a common data layer and the eventual integration in the current SMD (former OLAP DB).

Figure 2.1. General schema of data processing for SMD (former OLAP) integration



Source: ESPON FUORE

2.2 Methodological approach

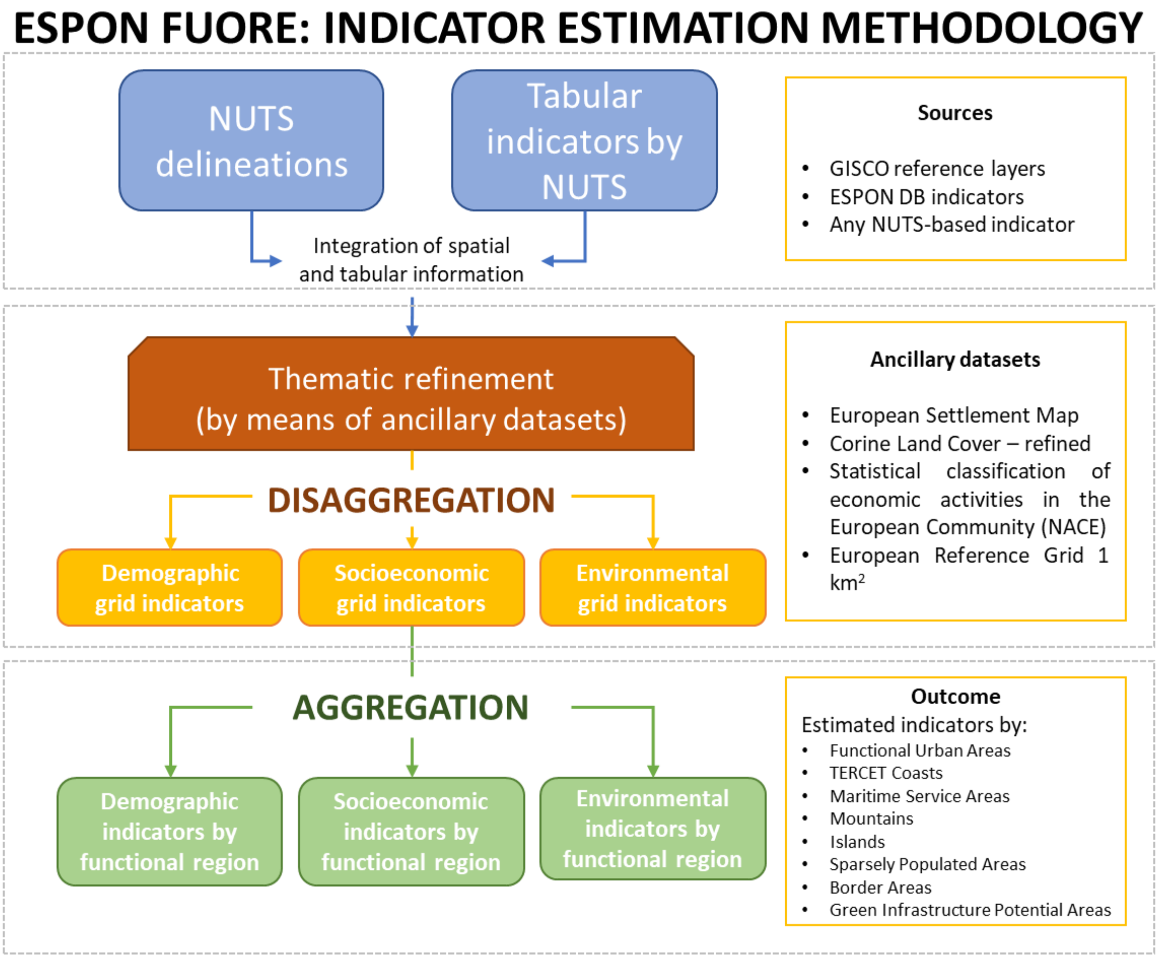
Before entering into the details of the results of the project regarding the disaggregation of indicators, this section provides an overview of the methodology proposed with the objective of improving former results on data disaggregation made by the ESPON M4D project. The ESPON OLAP cube developed within that project was able to generate data estimations of spatial data gaps using the population grid as ancillary data for the disaggregation of NUTS 3 indicators. As population grid data has been used to proportionally disaggregate any kind of indicators, the first generation of OLAP cube had left some researchers sceptical about results. By applying population distribution pattern for a variety of indicators such as economic activities, environment or transport, the OLAP cube is inducing a bias. Indeed, the spatial distribution of disaggregated data is likely to generate various "copies" of the population pattern that is used as a weighting system for all kind of indicators.

Oppositely, one could expect that disaggregation of economic or environmental indicators would follow the spatial distribution pattern of the concerned economic activity or land use category. This is how the service provider understands the objective of the ToR when asking for a new improved methodology on how to disaggregate NUTS 3 or other data.

Over the past couple of years, new spatial data have become available offering great opportunities to improve the disaggregation keys for the SMD. These datasets are listed in Table 3.1: Summary of main ancillary datasets of this document. Making use of a selection of these ancillary data, the service provider has defined **different weighting matrices according to the nature of the indicator** to be disaggregated (demographic, socio-economic or environmental). In Figure 2.2. General schema of the disaggregation/aggregation method, a general schema of the disaggregation/aggregation methodology is shown. For further details, refer to section 3.5.

It is foreseen the usage of a 1 km^2 ERG as the final data integrator. Nevertheless, some of the methodologies of disaggregation that will be tested might use higher resolution (e.g. 100x100m) grids, which will be eventually aggregated at 1 km^2. This intermediate step allows us to maintain the detail of the data, improving the accuracy in which the data is reported at 1 km^2 grid.

Figure 2.2. General schema of the disaggregation/aggregation method



Source: ESPON FUORE

2.2.1 From the ESPON OLAP Cube to the ESPON SMD

The former ESPON OLAP Cube was developed in the framework of the ESPON M4D project. The technical details behind it were published as a paper in the Second ESPON 2013 Scientific Report (Milego et al, 2013). In that paper it is explained how the ESPON OLAP Cube is built by means of a star schema of a gridded fact table, linked to several dictionary tables. That implementation has been replaced by the ESPON SMD, a relational database where aggregated values for all types of functional regions are stored, and all potential data queries are precalculated. Additionally, it is ready to store new indicators that advanced users may upload into the system at NUTS 2 or 3 level to be obtained reported at any of the functional regions of ESPON FUORE.

3 Data sources and methods for data estimation

The present section is about the "input materials" upon which the ESPON SMD is built on. Sub-section 3.1 reviews potential sources that could provide functional regions or functional regions definition for ESPON FUORE. Sub-section 3.2 shortly presents the selected functional regions and how their original delineations, originating from a variety of LAU nomenclatures, have been transferred into the ESPON LAU census 2011 nomenclature in order to gain coherence in spatial delineations and fit the ESPON 2020 Map kit. Sub-section 3.4 provides a concise overview of the selected functional areas. Finally, sub-section 3.5 describes the methodological approach for data estimation, detailing the selection of reference grids, source, target and ancillary datasets, as well as the scripting and processing required for implementing the disaggregation/aggregation process. It includes an assessment of the first results obtained.

3.1 Review of potential sources for the provision of functional regions

Following the ToR, the service provider has reviewed existing initiatives or potential datasets related to the definition of functional regions in Europe, together with former ESPON projects on that topic. Namely, this is the list of references reviewed:

- Functional areas in Member States of the Council of Europe. Preparatory study for the 17th session of the Council of Europe Conference of Ministers responsible for spatial planning (CEMAT) [2].
- Eurostat's Methodological manual on territorial typologies (EUROSTAT, 2018).
- Eurostat's task force on Labour Market Areas (LMAs)[3].
- Eurostat's "Maritime service areas. Highlighting the impact of coastal maritime activities on the hinterland"[4]
- The final reports of eight former ESPON projects, which are detailed in section 3.1.1.

In order to look for potential functional areas and regions to be used in the context of this project, the service provider established the following selection criteria:

- Type of region: the region is not a simple aggregation of NUTS, but a territory based on geographical, socio-economic and/or functional relationships
- Coverage: at least 28 countries of the 28+4 that compose the ESPON space are covered in a consistent manner
- Delineation: clear geometries (ideally based on EBM[5] LAU units) are available and the delineation methodology can be consistently applied over the ESPON space.

The Member States of the Council of Europe describe functional areas in two main definitions: (1) those territories that aggregate around urban centres and concentrate systemic relations, (2) territories delimited by one or several criteria (e.g. geographical or socio-economic features)

---

[2] http://www.mdrap.ro/userfiles/STUDIU_CEMAT-RAPORT_FINAL_EN.pdf

[3] https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en

[4] https://ec.europa.eu/eurostat/en/web/products-statistics-in-focus/-/KS-SF-11-041

[5] EuroBoundaryMap from Eurogeographics: https://eurogeographics.org/products-and-services/ebm/

which determine the cohesiveness and nature of internal and external interactions. This definition has also been considered to propose the set of functional areas and regions for ESPON FUORE.

The service provider has also considered the fact that one main objective of this project is implementing a methodology to disaggregate indicators at NUTS 3 level at grid cell level, and aggregating figures back to different functional delineations. In this regard, as stated in the criteria above, functional regions based on NUTS 3 delineations are not to be considered, because the indicators for such areas can be obtained by simple aggregation of NUTS 3 figures. The availability of spatial data covering most of the ESPON space (as stated in the criteria above as well) has been an important element for the final proposal.

3.1.1 Review of former ESPON projects

A review of the outcomes of several former ESPON projects has been carried out, in search of the definition of functional regions fulfilling the next criteria:

Along with ToR specification, the following projects have been assessed and we summarise below the main outcomes:

ESPON 1.1.1 "Urban areas as nodes in a polycentric development" (ESPON, 2005): Polycentricity is defined by this project based on FUA.

ESPON 1.4.3. "Study on Urban Functions" (ESPON, 2007): This project was also using FUAs.

ESPON Best Metropolises (ESPON, 2013b): This project focused in the metropolitan areas of Paris, Berlin and Warsaw only, and it was using FUAs as delineations for the analysis.

ESPON FOCI (ESPON, 2010a): FOCI was using FUAs also as delineation for their assessments.

ESPON Metroborder (ESPON, 2010b): This project is using FUAs and MUAs. MUAs are analysed in the M4D project point.

ESPON Polyce (ESPON, 2012a): Focused only in 5 metropolises (Bratislava, Budapest, Ljubljana, Prague and Vienna), using FUAs, Core Cities and Metropolitan regions.

ESPON GEOSPECS (ESPON, 2012b): This project had defined different territories with geographical specificities, which have been considered useful by the service provider. For further details, refer to section 3.2.2.

ESPON Ulysses (ESPON,2013c): A project focused in different border and cross-border areas. It was using LAU and NUTS delineations, and FUA also for reporting.

ESPON M4D (ESPON, 2015): Former ESPON Database project, it was mainly collecting data from the different ESPON projects of ESPON 2013 Programme. Nevertheless, it contains two other types of urban objects: MUAs and UMZ. These delineations are morphological boundaries of cities, which do not represent functionality. Considering the existence of FUAs, they would be redundant.

ESPON Spima (ESPON, 2018): This project has defined the so-called 'Metropolitan Development Area' (MDA) for ten stakeholder cities. In some cases, MDA is a legally binding area with fixed borders, while in other cases it has more fluid borders. Some MDAs are based on the extent of the transport infrastructure networks while others represent specific institutional arrangements between regions and municipalities. The method uses GIS tools based on local spatial data and data from European and OECD databases. It allows a breakdown of spatial data at the spatial scales of MDA, FUA and MUA, based on aggregation of LAU (local administrative units). Unfortunately, it was only applied to ten cities (Vienna, Zurich, Prague, Brussels, Brno, Oslo & Akershus, Turin, Terrassa, Lille, Lyon).

Synthesis of the ESPON projects review: As a result of the above analysis, the service provider realises there are no other functional regions different from FUAs that have been developed by any of these projects – i.e. that fulfil the criteria to be taken as reporting units for ESPON FUORE, with the exception of the outcomes from ESPON GEOSPECS.

3.1.2 Conclusion

First, and having analysed the different information sources mentioned above, the **FUAs** are considered as the most common and stable definition of functional regions in the urban environment. Additionally, ESPON GEOSPECS is also considered a good source of functional regions based on geographical specificities. Although in some cases it is mentioned that delineations from GEOSPECS could be updated or improved, it is beyond the scope of this project doing so. Some of the potential functional regions defined by the preparatory study of the European Council do not fulfil the criteria needed for the current project, because they are either:

- based on NUTS delineations;
- too local;
- not available as spatial data;
- inapplicable in a consistent way to all the ESPON space.

Regarding the LMAs, it is considered a very interesting on-going initiative which, unfortunately, is covering few countries yet. It was agreed during the KO that the service provider cannot include the LMAs at this stage of development in the current project.

As for the MSAs as suggested by Eurostat, and in agreement with the ESPON EGTC they have also been included together with the TERCET coastal delineations, as they add some value in terms of functionality.

A final proposal for **nine functional regions** is made in section 3.2.

3.2 Selection of functional regions

The proposed and eventually selected functional regions derive from several territorial typologies from Eurostat and ESPON that fit the three criteria of 1) type of region 2) coverage and 3) delineation mentioned in section 3.1. These delineations, based on LAU aggregation, originate from three main sources that provide complementary territorial perspectives, from urban/territorial functionalities to geographic specificities or Green Infrastructure (GI) potential:

- Territorial typologies from Eurostat (TERCET[6])
- Areas of geographic specificities from ESPON GEOSPECS[7] / ETMS[8] projects
- Network of GI potential from ESPON GRETA[9] project

For the sake of simplicity and consistency with the project name and specifications, the service provider refers to **functional regions** to identify all delineations proposed as reporting units, together with FUAs, although they might contain various types or approximations to functionality (as already stated in the ToR of this project's call for tender), and do not reflect exactly functional linkages.

3.2.1 Objects as defined in the TERCET

TERCET is a legislative initiative launched by Eurostat that is aimed at ensuring a harmonised application of territorial typologies at NUTS 3 level primarily, but LAU level also.

Apart from the **Functional Urban Areas**, which are the central focus of the present project, the LAU **coastal delineation** has also proved relevance for ESPON FUORE. Additionally, although they are not formally part of TERCET, it has been agreed with the ESPON EGTC to include the **MSA** coast from Eurostat, considering their added value in terms of functionality with respect to the official TERCET coasts.

3.2.2 GEOSPECS territories

Some of the LAU delineations of areas of geographical specificities originating from ESPON GEOSPECS (and reviewed by ESPON ETMS) have been deemed to fulfil the above criteria and complement the selection of the functional regions as defined in the TERCET, namely:

- **Islands** (with no fixed link to continent)
- **Mountain massifs**[10]
- **Sparsely populated areas**
- **Borders** (45 and 90 minutes): both will be included as the 45 min threshold stands as proxy for cross-border commuting functionality, whereas the 90 min one provides a larger perspective to border influence.

Additionally, the MSA have been completed with GEOSPECS coast delineation (LAUs 45min. accessibility + contiguous to coast) where gaps occur.

More detailed information on the delineation method and transfer from original nomenclatures to the ESPON LAU Census 2011 can be accessed in "Annex 1: Methodology to delineate FUAs and other functional regions into ESPON LAU Census 2011 nomenclature". Full details on how

[6] https://ec.europa.eu/eurostat/web/nuts/tercet-territorial-typologies

[7]https://www.espon.eu/programme/projects/espon-2013/applied-research/geospecs-geographic-specificities-and-development

[8] https://www.espon.eu/tools-maps/etms-tool-european-territorial-monitoring-system

[9] https://www.espon.eu/green-infrastructure

[10] Improved version of mountain massifs defined by (Price et al, 2019)

these categories have been delineated can be found in the ESPON GEOSPECS final scientific report (ESPON, 2012b).

3.2.3 Territories of high Green Infrastructure (GI) potential

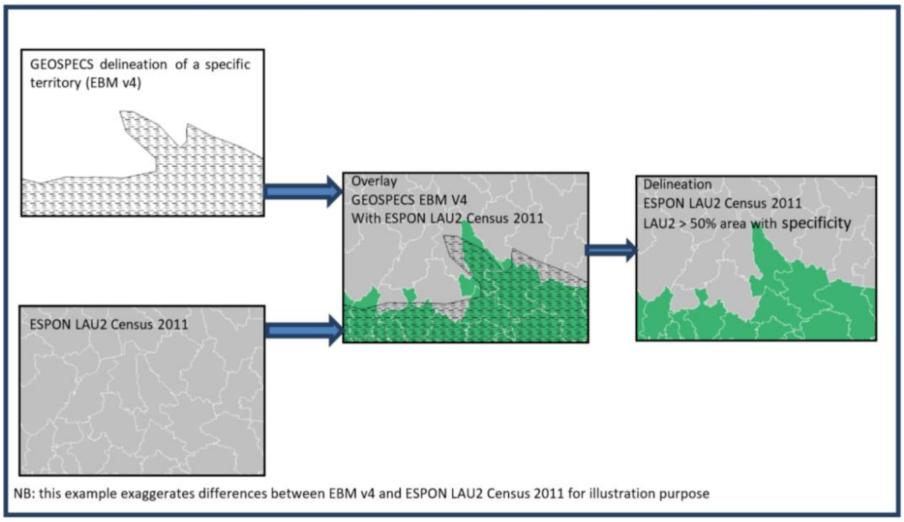
Taking advantage of the outcomes from the ESPON GRETA project, the service provider has delineated the **areas with an elevated potential regarding GI**, to be used as a new functional area / reporting unit. The source data used has been the GI Potential Network map produced by GRETA (ESPON GRETA Final Report – chapter 3.1), which used as source a 100x100m resolution binary raster stating if each cell has GI potential or not. A more detailed methodology description is also included in Annex 1: Methodology to delineate FUAs and other functional regions into ESPON LAU Census 2011 nomenclature. These functional regions complete the GEOSPECS territories, adding an environmental perspective.

3.2.4 Method to deal with a variety of original nomenclatures

Having selected the functional regions for ESPON FUORE, the issue is to bring into coherence a variety of functional/geographical territories that have been originally delineated from various LAUs nomenclatures (FUAs version 2015, Coasts from Eurostat version 2016, GEOSPECS delineations version 2008 and GI PN from GRETA version 2018). As the ESPON 2020 Map kit is based on 2011 nomenclature, a choice has been made to transfer original delineations into the ESPON LAU Census 2011 layer. That choice will also make it easy for ESPON projects to map their results within the ESPON 2020 Map kit.

The transfer method from their original nomenclature into ESPON LAU census 2011 nomenclature is a quite pragmatic one (i.e. the same one used by TERCET to define coastal LAUs). It consists in overlaying the functional/geographic types of territories in their original nomenclature with ESPON LAU Census 2011 layer. As a result, all 2011 LAUs, which over 50% of their area is covered by the functional region, belong to the type of functional specificity. In cases of coasts or borders, an additional contiguity criteria has been adjusted when relevant.

Figure 3.1. Process of updating ESPON LAU Census 2011 layer with delineation of Functional regions



Source: ESPON FUORE

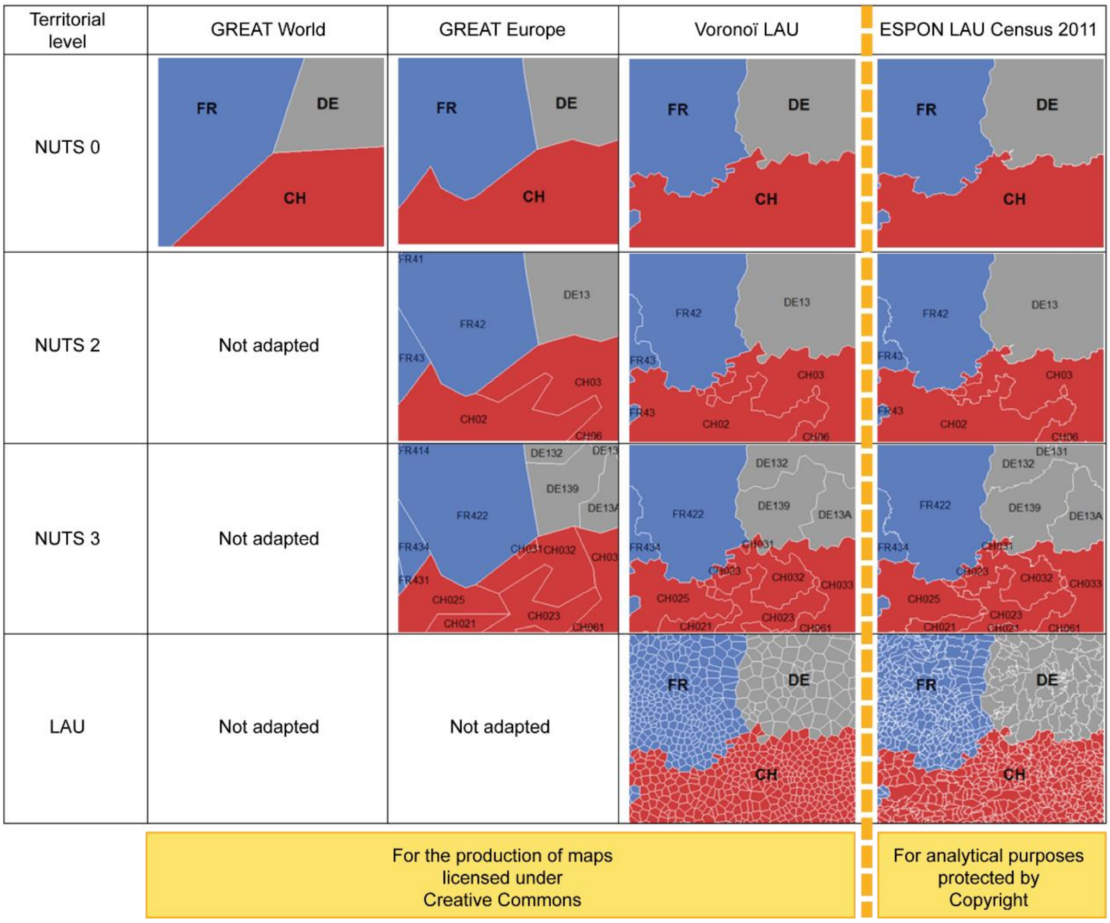
3.3 Creation of analytical and mapping layers

When ESPON launched the Map kits 2020 project with the aim of providing the Programme with a coherent set of mapping layers and layout, reflections highlighted that mapping layers are used for two distinct purposes: analysis and mapping (Michelet JF, Gloersen E, Ysebaert R, Giraut F., 2018):

- For analyses, the key criterion is precision. ESPON LAU Census 2011 Map, based on a collection of various versions of EuroBoundaryMap in order to fit the Census 2011 nomenclature with a resolution corresponding to a scale of 1:100'000, fits perfectly for that purpose.

- For mapping, the key criterion is informational value. The objective is to convey the cartographic message in an efficient way. This implies that 'noise' (i.e. unnecessarily detailed boundaries) must be eliminated. A coherent set of generalised mapping layers has been produced, including a Voronoi "twin" of the ESPON LAU Census 2011 Map. Additionally, low file size in association with Creative Commons licensing, makes 'on the fly' production of maps using interactive mapping portal feasible, even for pan-European maps.

Figure 3.2. Voronoi layer at different levels in comparison to the analytical layer



Source: Michelet & al. 2018

Following this strategy, the service provider has produced:

- Delineation of all functional regions in the analytical/precision layer (ESPON LAU Census 2011) to feed the ESPON SMD for spatial analysis.
- An exact copy of these functional regions in the ESPON Voronoi layer (generalised LAU geometries) for the project's web tool, by using a "table join" on LAU IDs that are 100% compatible with ESPON LAU Census 2011 layer.

Turkey, which does not exist in ESPON 2020 Map kit LAU Census 2011 mapping layer, has required:

1. Creating a "patch" from GEOSPECS analytical mapping layer, so that Turkish LAUs touching the Greek and Bulgarian ESPON LAU Census 2011 units fit seamlessly.
2. Creating of a LAU Voronoi layer, using the same nomenclature and allowing that Turkish functional/geographical regions to be represented in the web tool.

3.3.1 Extent of the selected functional regions

The service provider has tried to cover as much ESPON space as possible in the selection of functional regions. Special effort has been made to include Turkey and the Outermost regions whenever the selected data sources covered them. Figure 3.3. Overview of selected functional regions is providing some details about the few remaining "Issues", knowing that the rule of having at least 28 countries covered is fulfilled in any case:

- Dark green: EU 28 + 4 + Turkey are covered from a single source
- Light green: two different sources (i.e. similar in methodology, but not fully comparable) have been used so as to cover the full area (EU28 + 4 + Turkey)
- Orange: EU 28 + 4 are covered, but either one country (e.g. Turkey) or for part of one country (e.g. French outermost regions) are not covered.

3.4 Schematic overview of selected functional regions

Figure 3.3. Overview of selected functional regions

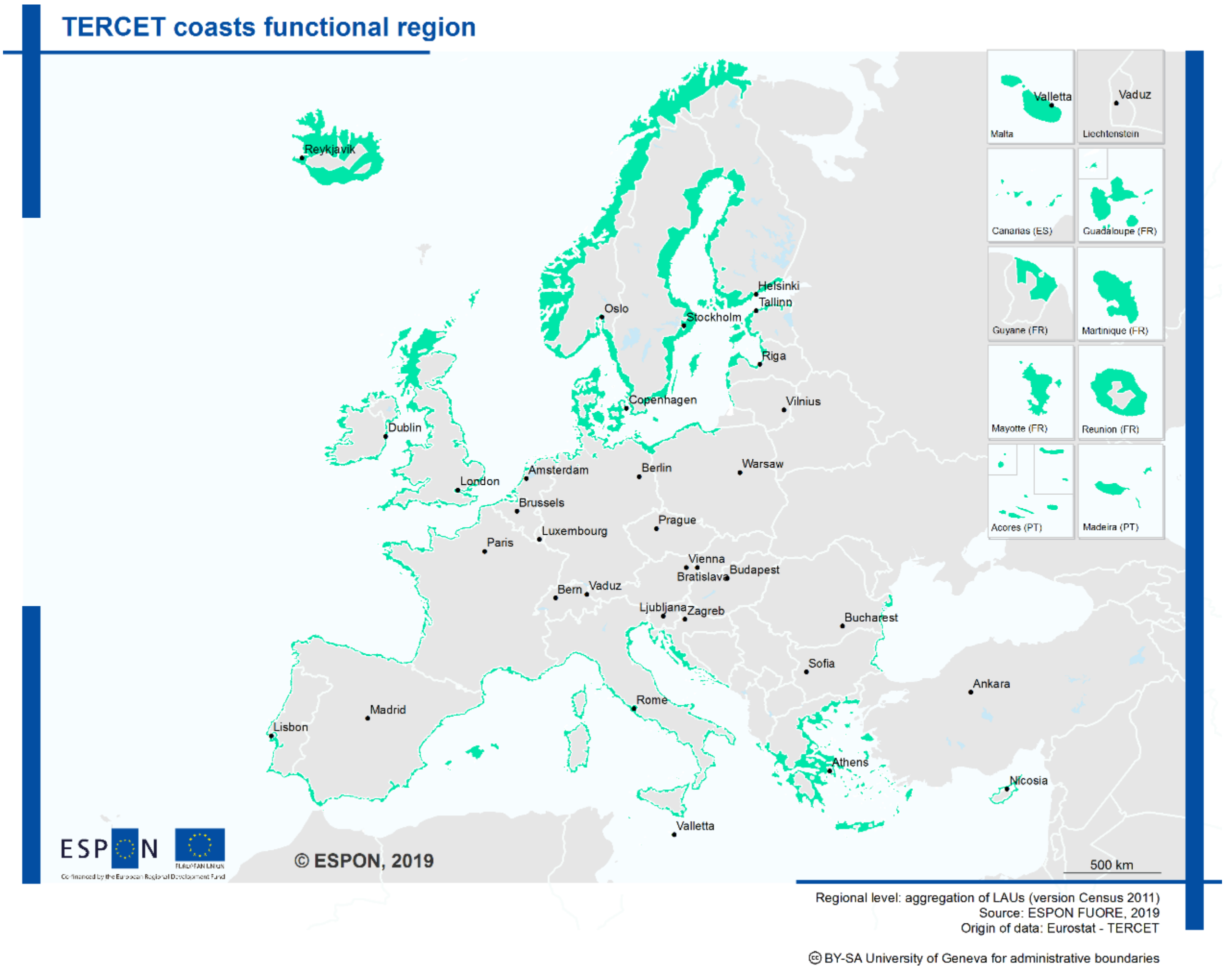
Data Origine	Type of region	Abreviation	Delineation principle	Overlapping	Level 1	Level 2	Level 3	Level 4	Level 5	Issues
TERCET	Coasts	TCOA16	Morphologic Contiguous to coast or 50% area within 10km buffer to coast	no	LAU		NUTS 2 Nb = 149	NUTS 0 *Nb = 25*	EU	TR Missing
TERCET	Functional Urban Areas	FUA	Functional Eurostat-OECD based on commuting movments	yes	LAU		FUA *Nb = 805*	NUTS 0 *Nb = 32*	EU	None
EUROSTAT - Maritime service areas	Coasts	MSA	Functional Areas that can be reached within a given travelling time, starting from a location at the coast and using the existing transport network. Travelling time depends on the points of interactions (large ports, small ports and coastal settlements)	no	LAU		NUTS 2 Nb = 188	NUTS 0 Nb = 26	EU	HR, IS, NO, TR & Mayotte delineations from GEOSPECS = 45min travelling time to coasline + contiguous
ETMS-GEOSPECS	Mountains	MTN	Morphologic 50% area covered by EEA mountain delineation Enclaves and Exclaves excluded	no	LAU	By national part of a massif *Nb = 48*	By massif (transnational) *Nb = 16*	NUTS 0 *Nb = 25*	EU	None
ETMS-GEOSPECS	Islands	ISL	Morphologic Totally surrounded by sea Islands with a fixed link to continent have been excluded	no	LAU	Island *Nb = 240*	NUTS 2 Nb = 55	NUTS 0 *Nb = 19*	EU	None
ETMS-GEOSPECS	Sparsely populated Areas	SPA	Functional Settlement pattern based on threshold of population potential (= how many persons can be reached within 45min driving time)	no	LAU		National sub-type *Nb = 39*	NUTS 0 *Nb = 20*	EU	None
GRETA	Green Infrastructure	GI	Morphologic Green Infrastructure Potential Network, calculated by ESPON GRETA	no	LAU		NUTS 2 Nb = 265	NUTS 0 *Nb= 32*	EU	French OMR missing
GEOSPECS	Border "narrow" (45 minutes)	BDA1	Functional Time-distance to a line: 50% area with 45min driving accessibility to border	yes	LAU	National parts of borders *Nb = 119*	Border (bi-national) *Nb = 73*	NUTS 0 *Nb = 30*	EU	"TR - EU" countries = OK "TR - other countries" = not delineated
GEOSPECS	Border "large" (90 minutes)	BDA34	Functional Time-distance to a line: Contiguous to border or 50% area with 90min driving accessibility to border	yes	LAU	National parts of borders *Nb = 127*	Border (bi-national) *Nb = 80*	NUTS 0 *Nb = 30*	EU	"TR - EU" countries = OK "TR - other countries" = only contiguous

Source: ESPON FUORE

3.4.1 Maps for the selected functional regions

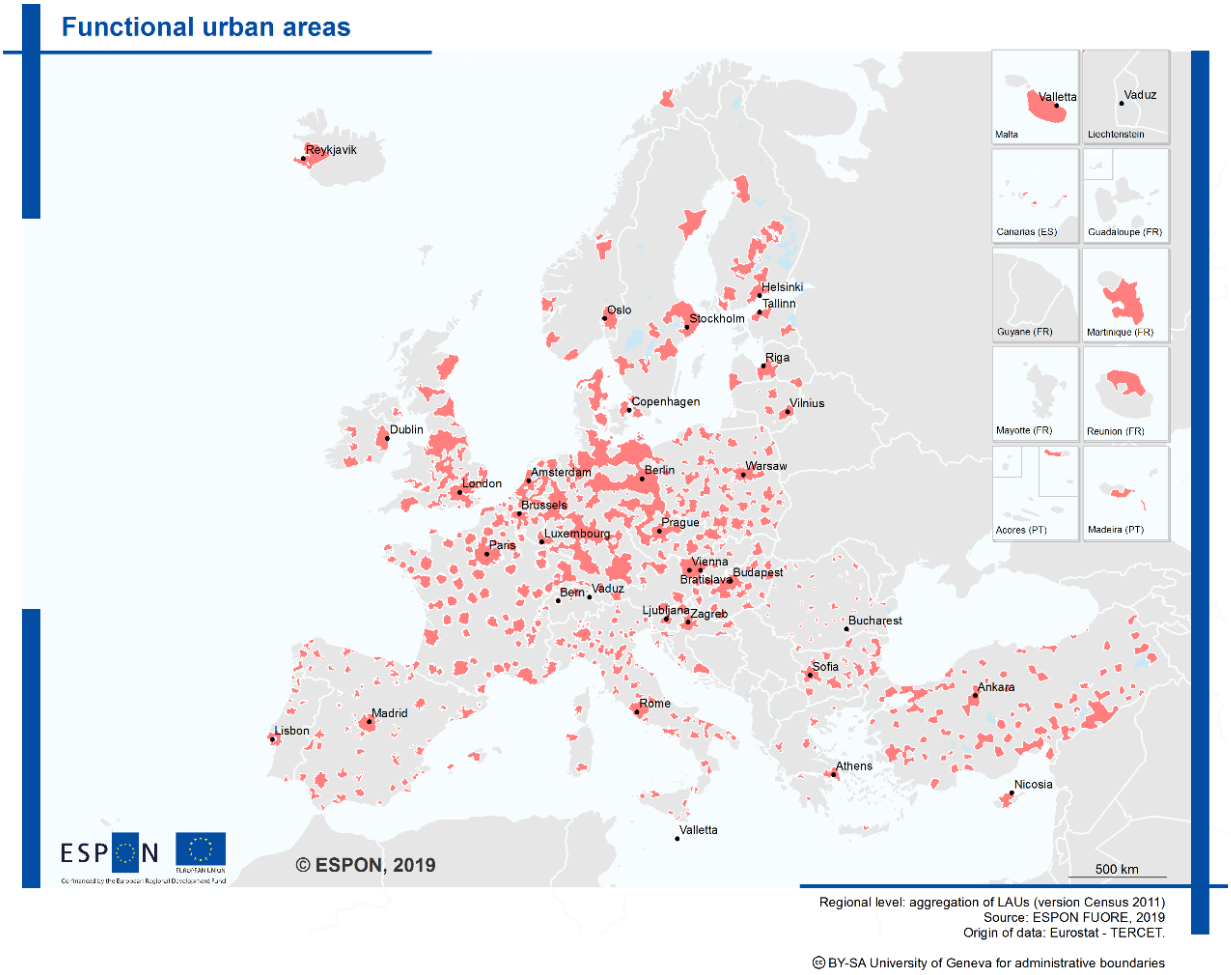
In this section we present the overview maps for the **nine selected functional regions**. They show the areas covered by each functional region. In Annex 1: Methodology to delineate FUAs and other functional regions into ESPON LAU Census 2011 nomenclature, detailed maps of the different reporting units are shown for each type of functional region.

Map 3.1. Coasts functional region as defined in the TERCET



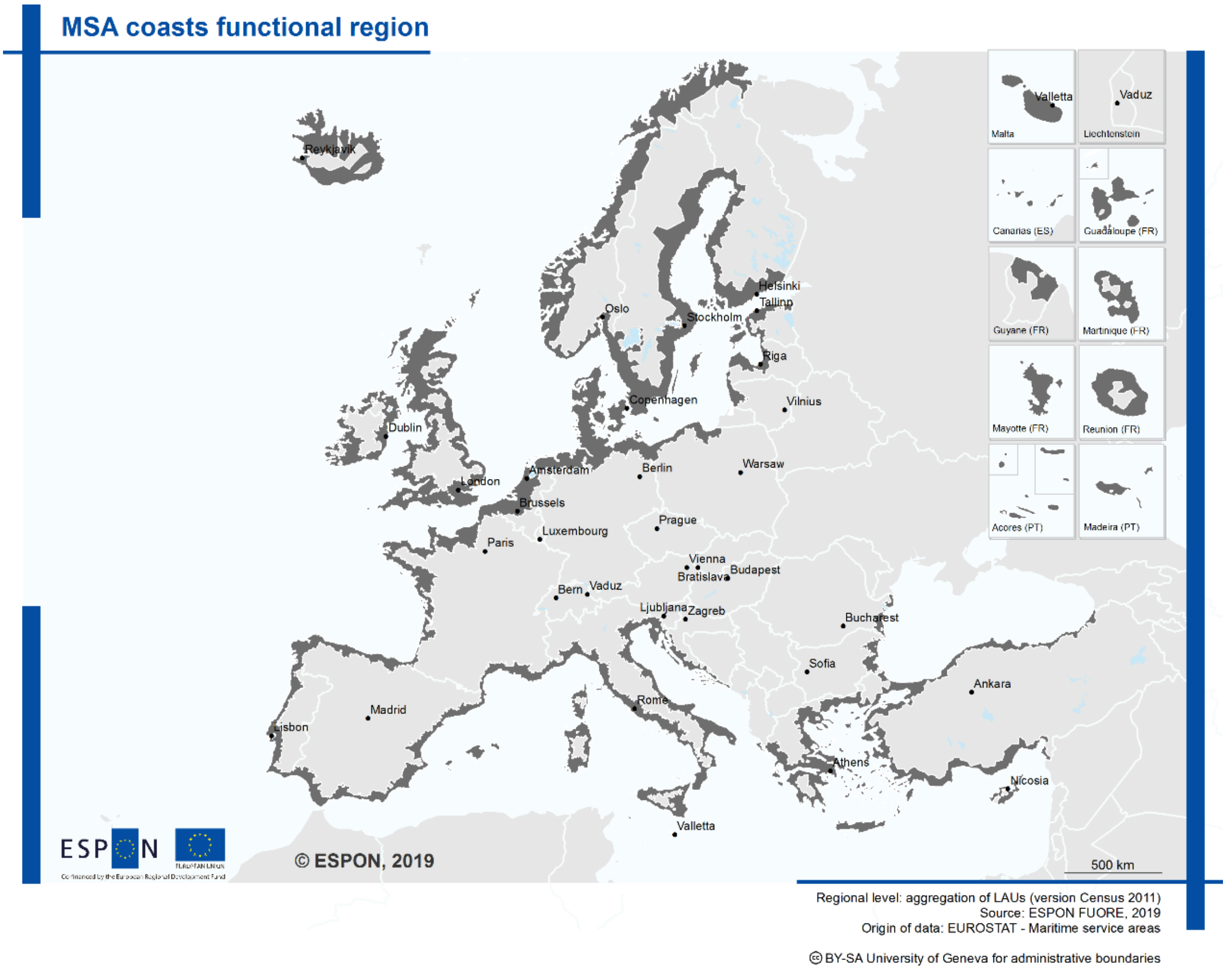
Source: Eurostat – TERCET, ESPON FUORE

Map 3.2. TERCET Functional urban areas functional region



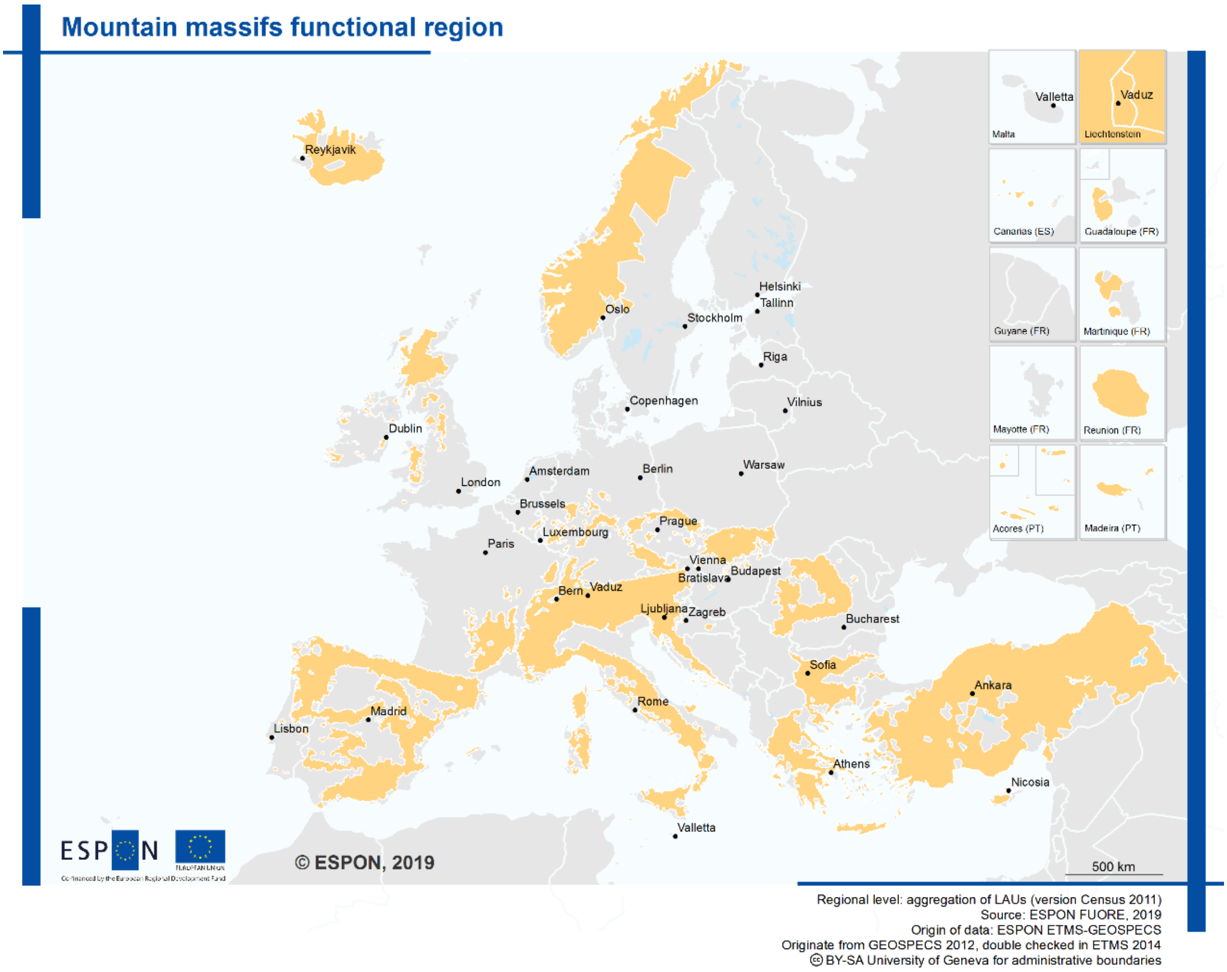
Source: Eurostat – TERCET, ESPON FUORE

Map 3.3. MSA coasts functional region



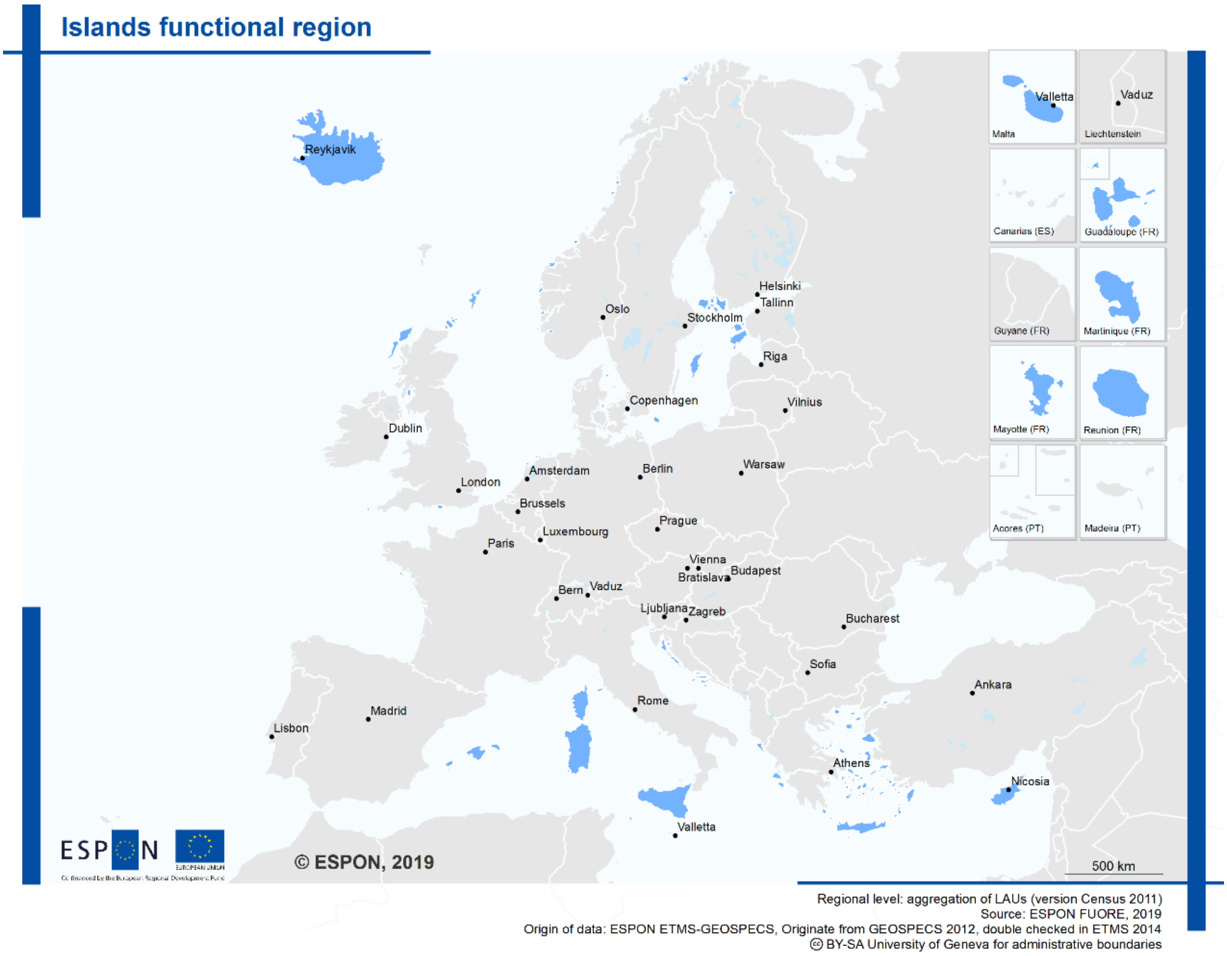
Source: Eurostat – MSA, ESPON FUORE

Map 3.4. Mountain massifs functional region



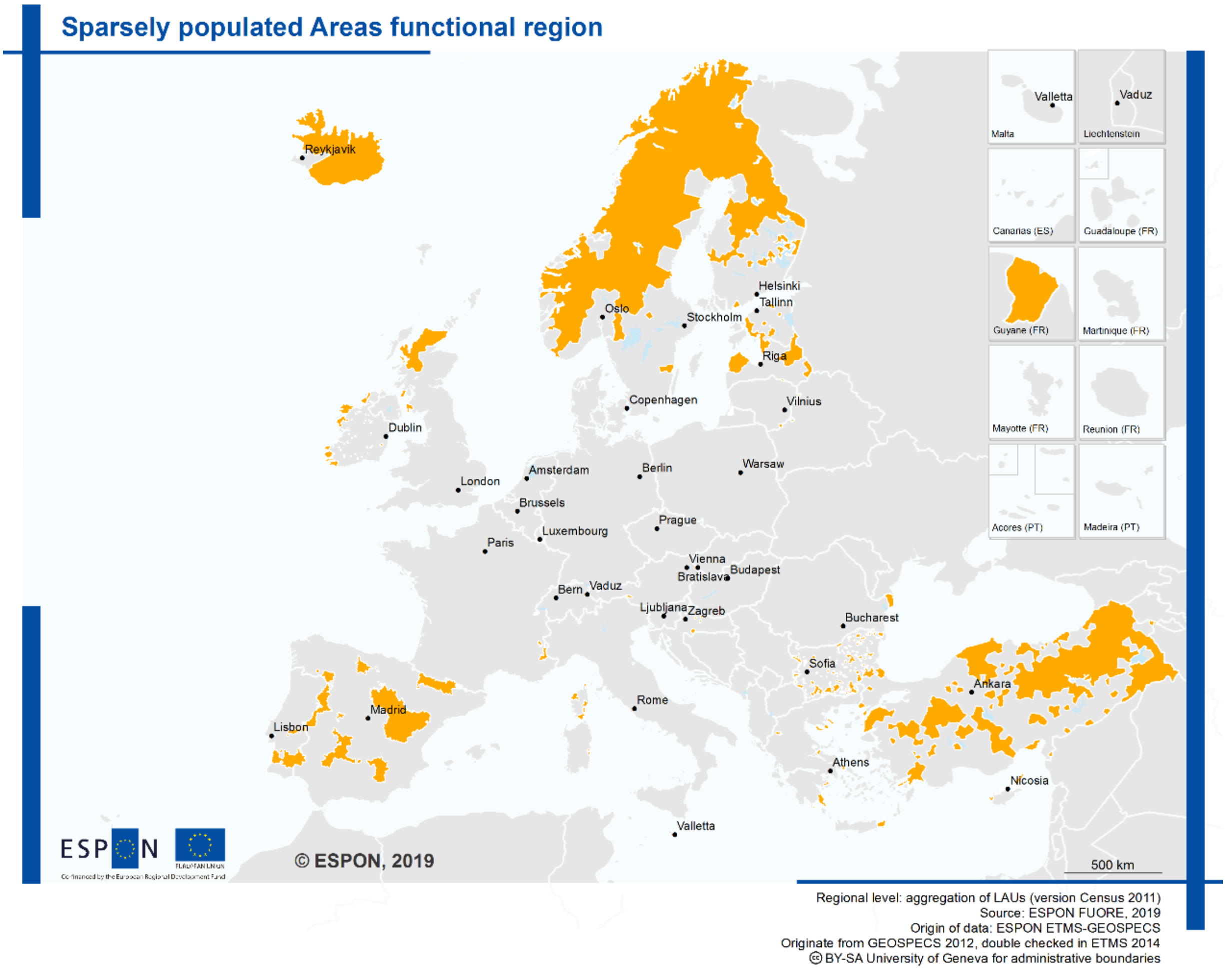
Source: ESPON GEOSPECS/ETMS, ESPON FUORE

Map 3.5. Islands functional region



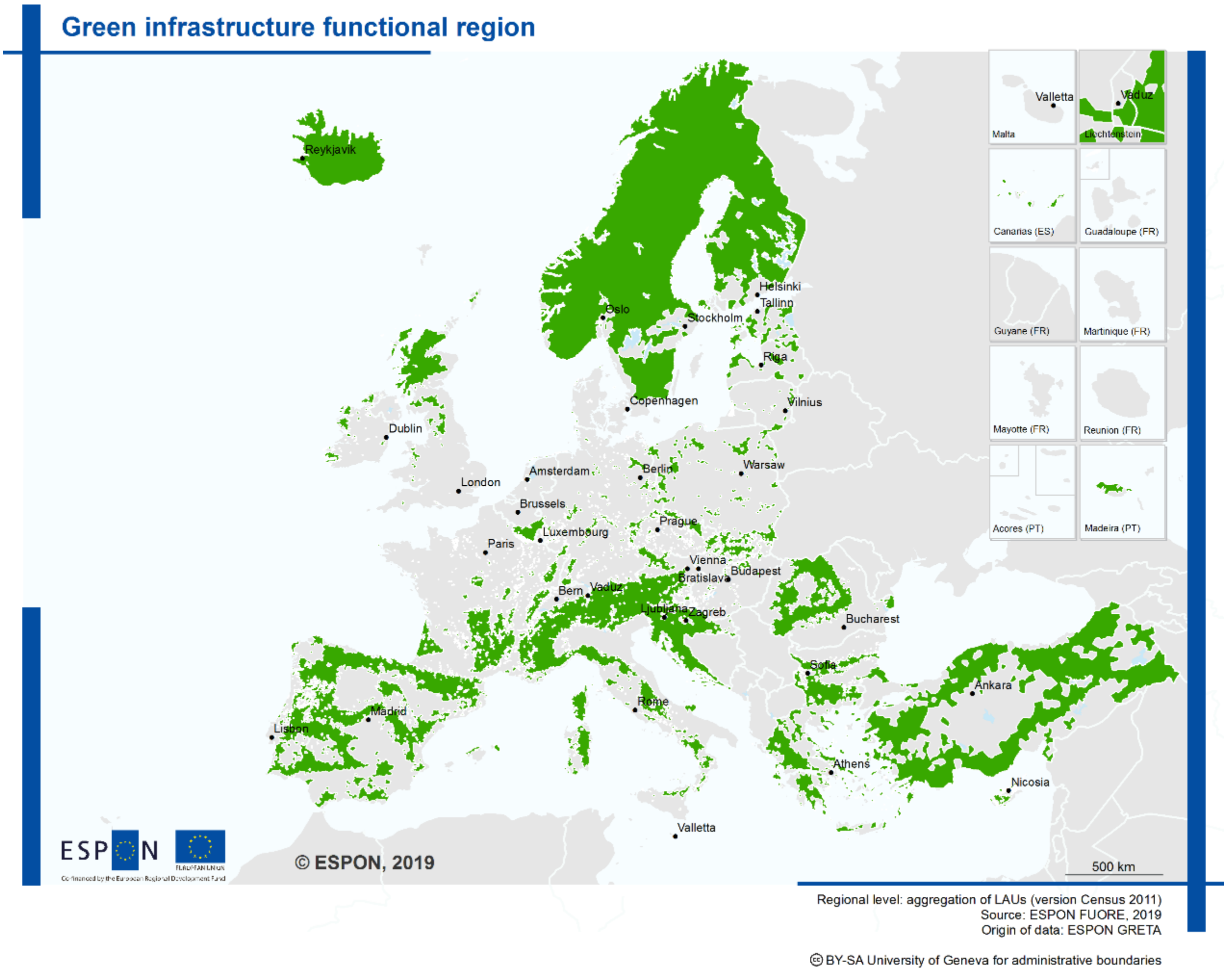
Source: ESPON GEOSPECS/ETMS, ESPON FUORE

Map 3.6. Sparsely Populated Areas functional region



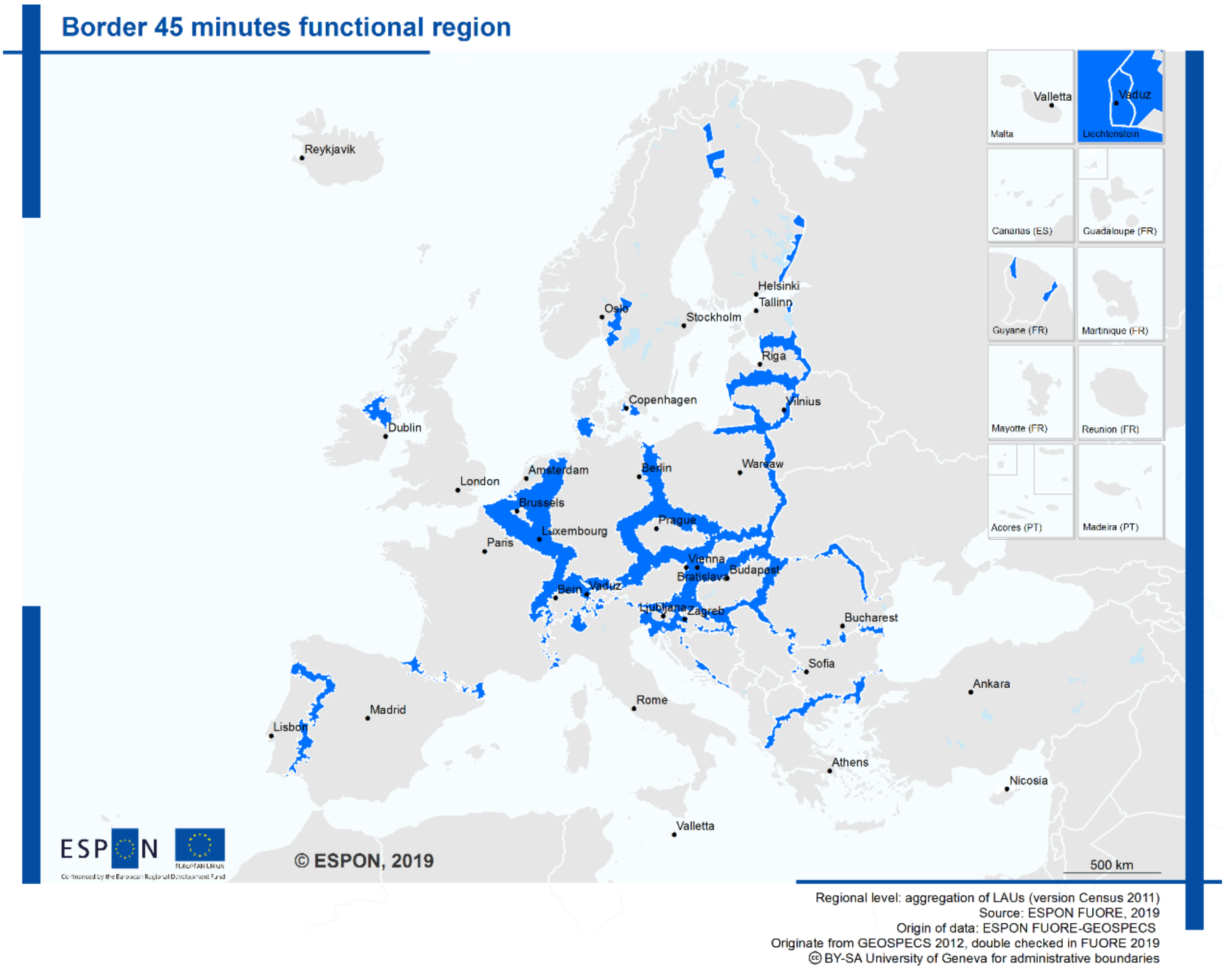
Source: ESPON GEOSPECS/ETMS, ESPON FUORE

Map 3.7. Green infrastructure functional region



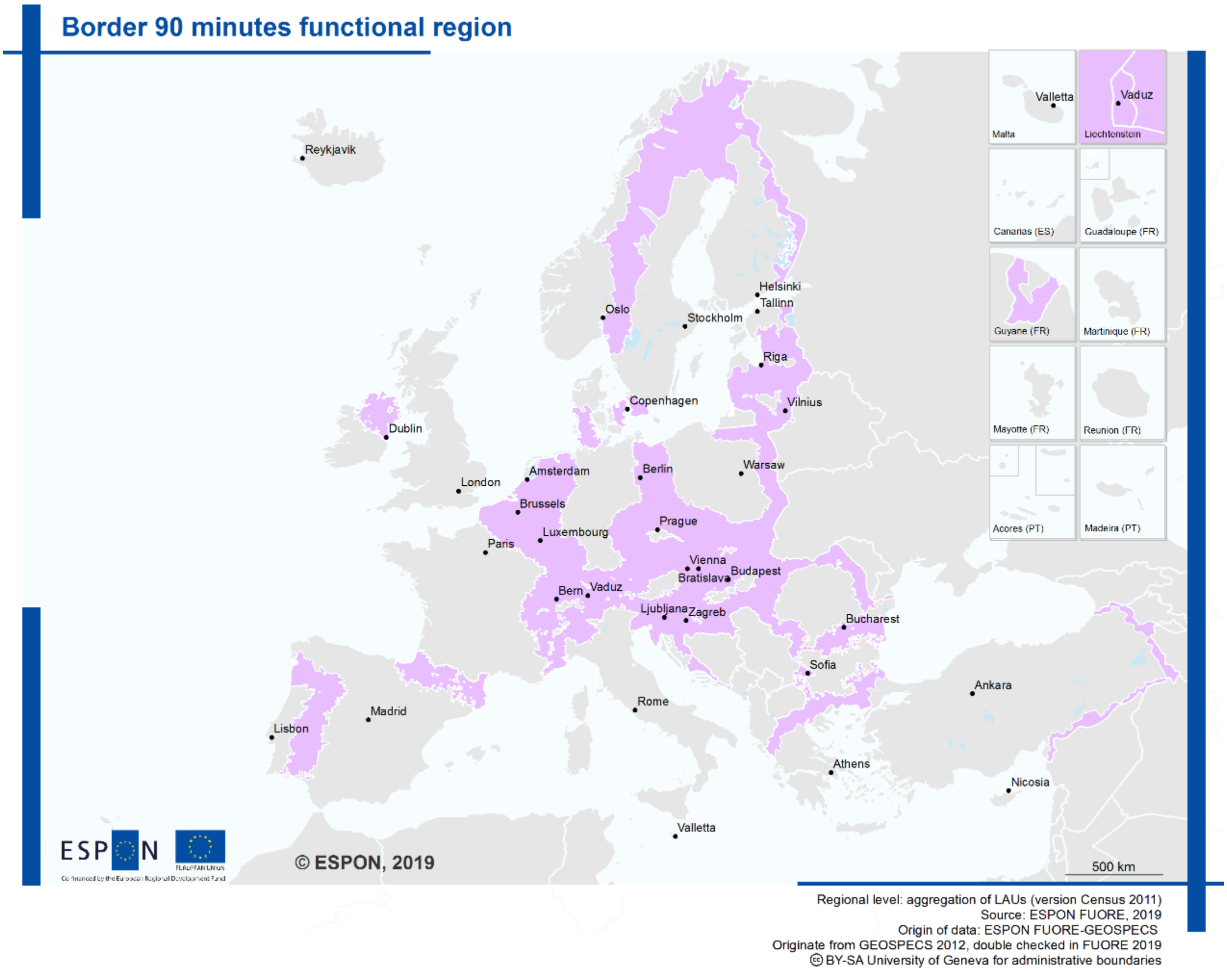
Source: ESPON GRETA, ESPON FUORE

Map 3.8. Border 45 minutes functional region



Source: ESPON GEOSPECS/ETMS, ESPON FUORE

Map 3.9. Border 90 minutes functional region



Source: ESPON GEOSPECS/ETMS, ESPON FUORE

3.5 Data estimation

ESPON FUORE has estimated different indicators (demographic, socio-economic and environmental) contained in the ESPON 2020 Database as base indicators, by FUAs and other functional regions proposed and eventually selected (see sub-section 3.2). Such estimation is done by an areal interpolation procedure by means of the GIS operations detailed below (see sub-section 3.5.2). The outcomes of the procedure have subsequently fed the ESPON SMD (see section 4).

3.5.1 Data sources and ancillary datasets

The source data and supporting layers (target and ancillary data) relevant for the estimation of the indicators by functional regions are detailed under this sub-section.

Source data: the indicators to be recalculated and reported by functional areas and regions have been taken from the set of the ESPON base indicators. The number of indicators eventually integrated in the ESPON SMD depend on some methodological issues, but they cover the different topics in a comprehensive way. The selection of the final set of indicators has been done in agreement with the ESPON EGTC and the PST. The complete list of the ESPON base indicators eventually used is included in Annex 2: List of ESPON 2020 Base Indicators eventually estimated by FUORE. A small set of the ESPON base indicators were discarded for different reasons. For further details, please refer to sub-section 3.5.2.1.

Target data: the set of polygonal entities for which the original values needs to be downscaled to (i.e. disaggregated) before reaggregation into functional regions. This dataset is used to redistribute the information of the source data, typically reported at NUTS level, proportionally to the distribution of ancillary variables (see sub-section 3.5.2). As target feature, reference grids with 1 x 1 km resolution have been used. For Europe, the grid used is based on the INSPIRE Geographical Grid Systems[11] which has been extended to cover Turkey. For the outermost regions, official grids for Martinique and Réunion have been obtained, but a single global grid for all the outermost regions has been prepared to facilitate the disaggregation process for such areas in the WGS84 system.

Ancillary data: spatially explicit data guiding the downscaling process. The degree to which the disaggregation approximates real values is highly dependent on the quality of the ancillary data. Therefore, the finest and the most updated available ancillary datasets have been used. However, for FUAs and other functional regions not covered by these finest and most updated datasets (e.g., outermost regions, Turkey and western Balkans), when available, other ancillary data providing similar information at a global scale have been proposed (for details on ancillary data to be used, see Table 3.1: Summary of main ancillary datasets).

---

[11] https://inspire.ec.europa.eu/Themes/131/2892

3.5.2 Methodology proposal

The combination of source data available at a relatively coarse spatial geometry with covariates of the source indicator (i.e. ancillary data) to generate more detailed maps reporting the source data is known as **‘dasymetric’ mapping** (Tobler (1979)). The method proposed by this project is based on Batista et al. (2013) and Batista & Poelman (2016) and follows a dasymetric approach. For the sake of consistency and clarity, we have defined similar algorithms and we have specified the use of similar datasets, but customizing the weights applied to disaggregate each indicator depending on its particular nature. The downscaling process uses a variety of specific ancillary datasets, also depending on the nature of the indicator. Table 3.1: Summary of main ancillary datasets summarises the source, target and ancillary data layers that are being used in ESPON FUORE.

Table 3.1: Summary of main ancillary datasets

Data type	Name	Reference year	Resolution	Year of release	Coverage	Data provider
Source	ESPON base indicators	1990-2015 (dependent on the indicator)	NUTS 2/3		ESPON countries	ESPON
Target	Grid 1x1 km	2018	1 km^2	2018	ESPON countries	EEA/ETC-UAB
	Grids for outermost regions 1x1 km	2013	1 km^2	2013	Outermost regions	French “Institut National de la Statistique et des études économiques” UNEP/GRID
Ancillary	European Settlement Map (ESM)	2010-2013	10 m	2017	All EU-28	JRC
	Corine Land Cover-Refine (CLC-R)	2012	100 m	2018	33 European EEA member countries and six cooperating countries (EEA39)	JRC
	GHS-BUILT layer produced in the GHSL (Global Human Settlement Layer) framework	2000-2014 (class 3)	38 m	2015	World	JRC
	Statistical classification of economic activities in the European Community (NACE)	2000-2007	LAU 2	2008	ESPON countries	ESPON GEOSPECS, from NSI sources

Specifically, for **demographic indicators**, the main ancillary dataset guiding the downscaling process is the European Settlement Map (ESM) built-up areas (R2017, Ferri et al.), whereas for **socioeconomic and environmental** indicators we take advantage of the Statistical classification of economic activities in the European Community (NACE).

As the different land use/cover classes are also assumed to be related to the variation of the indicators across space (e.g., for demographic indicators, different land uses will have different values of population), the disaggregation process can be refined guided by the ESM and NACE ancillary datasets, using additional ancillary data on land cover. In this way, specific weights can be assigned to the different land cover types which will improve the precision of the final estimates. For all indicators, Corine Land Cover-Refine (CLC-R) is used as additional ancillary dataset in order to refine the downscaling process. In this way, based on specific land use classes (derived from a reclassification which simplifies the original classes contained in CLC-R, different weights can be assigned to different CLC-R classes in order to refine the downscaling process guided by ESM and NACE. The particular weights associated to each land use class is derived from economic information provided by different data sources (i.e., thousands hours worked -Eurostat-; value added estimated from total production for each activity in the European Union -OECD-; Green House Gas emissions by aggregated sector in EU-28 -EEA 2017-). Annex 5 shows how we defined, based on the mentioned economic information, specific weights to each particular CLC-R class.

Figure 3.4 summarizes the different transformations applied to get indicators by functional areas from the ESPON base indicators. Table 3.2: Weighting scheme. Thematic groups of ESPON 2020 Database base indicators and specific ancillary datasets for their disaggregation (see Table 3.1 for a detailed description of each dataset and Table 3.4 for a detail of the CLC classes applied). Additional information on the specific data sources and methods applied for the estimation of the weighting values can be found in Annex 5. details the particular combination of ancillary dataset, CLC classes and the source of the weights applied to refine the estimates of the different main indicator types. Table 3.3 reports the exact values applied for weighting estimates according to different land use classes (see Annex 5 for a detail on the methods applied for deriving weight values). Finally, the specific reclassification of CLC-R classes used for refining the different types of indicators can be found in Table 3.4.

Figure 3.4. From ESPON 2020 NUTS indicators to functional regions. Diagram of the methods applied.

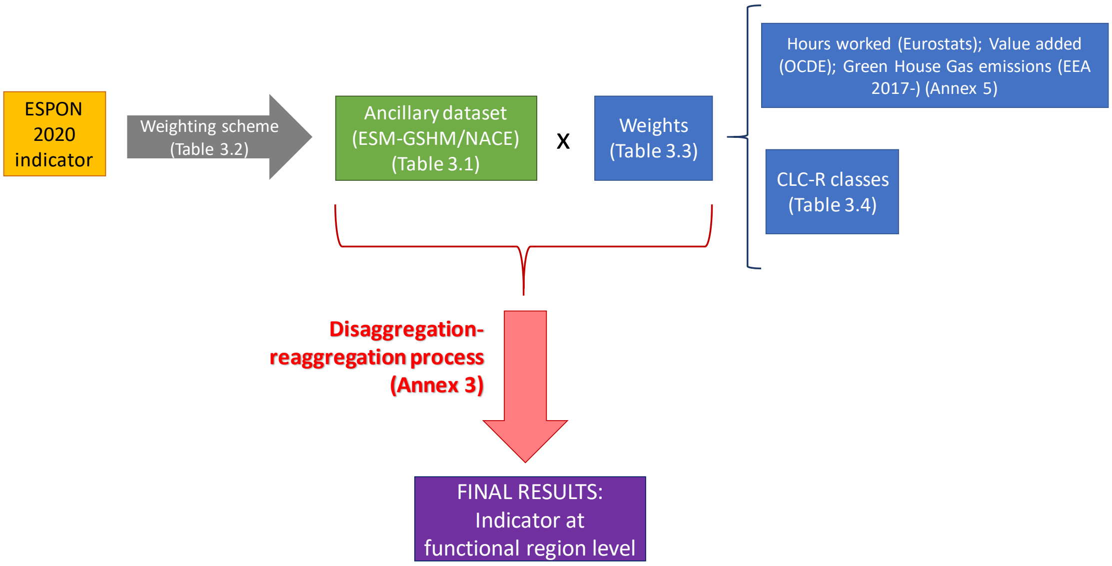


Table 3.2: Weighting scheme. Thematic groups of ESPON 2020 Database base indicators and specific ancillary datasets for their disaggregation (see Table 3.1 for a detailed description of each dataset and Table 3.4 for a detail of the CLC classes applied). Additional information on the specific data sources and methods applied for the estimation of the weighting values can be found in Annex 5.

Indicator type	ancillary dataset	CLC classes	Weight's data source
Demographic, Education, ITS	ESM	Rural, urban, others; class 1	Batista & Poelman 2016
Employment, Society	NACE total employed population	Main activity sectors; class 2	Eurostat
Economy, R&D	NACE employed population by type of activity	Main activity sectors; class 2	OECD
Energy	NACE employed population by type of activity	Main activity sectors, Eurostat, GHG emissions; class 3	EEA Report No 8/2017; Eurostat

The disaggregation of ESPON indicators covering outermost regions has been subjected to the availability of ancillary datasets for these areas. GHS-BUILT R2015B version (see Table 3.1: Summary of main ancillary datasets) had to guide the disaggregation of demographic indicators for these regions instead of the European Settlement Map (ESM). After an assessment of the datasets, the available land cover layers initially identified in the Inception phase of the project for those regions (Modis Global Land Cover 5' -NASA- and Copernicus Globe Cover 100m Africa -European Commission/VITO-) did not offer enough resolution in terms of land use classes as to refine GHS-BUILT dataset. Therefore, for outermost regions, the weights required in the disaggregation process could only be based on GHS-BUILT. NACE is available for the Outermost regions, except from Mayotte and Saint-Martin. Nevertheless, due to the differences in source and ancillary data availability, and time constraints, the disaggregation of indicators for such regions have been left for a future phase of the project.

Overall, the particular weights derived from the ancillary variables (ESM, NACE, CLC-R, GHS-BUILT) have been used to populate each target 1x1 km grid cell with an estimate of the ancillary variable, which has then been used in the source (i.e., ESPON indicator) downscaling process. This assumes a positive relationship between the ancillary variable and the specific indicator. The results have been refined using additional ancillary datasets weighting these estimates (Table 3.1: Summary of main ancillary datasets).

To redistribute values from the source (ESPON indicator) to the target spatial units (1x1 km), the source and target 1x1 km reference grid layers have been firstly intersected geometrically by means of a GIS, resulting in a third layer which has been referred to as 'transitional' geometry. This transitional geometry contains unique combinations between all overlapping source polygons and 1x1km polygons. To estimate the population for each polygon of the transitional geometry the next formula has been applied:

$$P_i' = P_s * \left(\frac{A_i * W_i}{\sum_i^n A_i * W_i} \right)$$

where:

P_i' corresponds to the estimated value in a given polygon *i* of the transitional geometry;

P_s is the known value in the source polygon *S;*

A_i is the ancillary value within polygon *i* (determined using the so-called 'zonal statistics' function from the GIS software) between the transitional geometry and the ancillary dataset geometry (ESM, NACE or GHS-BUILT, depending on the particular indicator and region, see below);

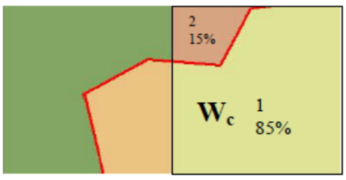
W_i is the ancillary dataset weighting the estimates (CLC-R);

n corresponds to the number of transitional polygons within each source polygon.

For **demographic indicators**, A_i is the total built-up area (derived from ESM or GHS-BUILT datasets) within polygon *i*;

For **all other indicators**, A_i is determined using the following approach, where Wc is the NACE population value:

Figure 3.5. Calculation of ancillary value for non-demographic indicators



Cell value = W_c Σ (V_i * $Share_i$)

Where: V_i = Value of unit i

$Share_i$ = Share of unit i within the cell

W_c = weight assigned to cell c

In the example: W_c * (V_1 * 0.85 + V_2 * 0.15)

Source: Milego et al. 2014.

V_i is the CLC-R re-classified value (see Table 3.4. Corine Land Cover (CLC-R) reclassification for weighting purposes) for weighting the estimates.

Specifically, for **Economy, Energy and R&D related indicators**, the estimation of A_i has been done by separated activities ($A_{i,g}$, weights for all other classes=0) and then summing up all the results from the different activities $\sum^{g} A_{i,g}$. (*g* corresponds to the possible types of economic activity in NACE).

In all the cases, the obtained P_i' is a continuous numerical value, thus for Economy and Energy related indicators, no subsequent action is required. However, regarding demographic indicators, for groups of indicators that must be disaggregated based on NACE total population and for R&D indicators (see Table 3.2), P_i' is a continuous numerical value but population is a discrete number. Therefore, a procedure based on Silva & Poelman (2016) to transform these continuous numerical values into discrete ones has been implemented. Simply rounding the numbers is not appropriate because it would likely result in a total population for the functional region different (even if slightly) from the original total.

The applied procedure starts by rounding down the initial estimates of population for each transitional polygon *i*:

$$P_i''= floor(P_i')$$

This leaves a remainder population *R* in each polygon *i*:

$$R_i= P_i'' - P_i'$$

R_{FUA} is then calculated as the sum of *R* in all *i* polygons within the functional region.

$$R_{FUA} = \sum_{i}^{d} R_i$$

where *d* is the total number of *i* polygons within the functional region.

Because R_i is smaller than 1 in all *i* polygons, R_{FUA} is necessarily smaller than *d.* Therefore, if a R_{FUA} number of *i* polygons is selected and an additional inhabitant is warranted, it is ensured that the sum of population values in all *i* polygons corresponds to the original total population in the functional region. In order to make the selection, the R_i values are firstly sorted in a descending sequence (from highest remainder population to lowest). The sequence is stored as variable O_i:

$$O_i = sort(R_i), with\ O_i \in \{1{:}\,d\}$$

Subsequently, one additional inhabitant is granted in polygons *i* ranked higher than the $R_{FUA} - th$ position (i.e. the highest R_i remainders):

$$P_i''' = \begin{cases} P_i'' + 1\ if\ 0_i \leq R_{FUA} \\ P_i'' \end{cases}$$

This results in an integer number of people P_i'' where $\sum_i^d P_i'''= P_{FUA}$

Finally, the estimated population for each functional region polygon t is simply:

$$P_t = \sum_{i}^{j} P_i'''$$

where *j* corresponds to the number of transitional polygons (1 x 1 km grid) within each target polygon (functional region). Total population for P_t can be obtained:

$$P_t = \sum_{g}^{k} P_t'$$

Table 3.3. Weighting values assigned to different thematic indicators according to specific CLC-R classes (see Table 3.4 below). See also Table 3.2 for a reference on the weighting data source and Annex 5 for additional information on the data sources and methods applied for the estimation of the weighting values.

CLC-R classes	**Demographic, Education, ITS**	**Employment, Society**	**Economy, R&D**	**Energy**
1	1	0.10421773	0.01202494	0.28571429
2	0.5	0.63336144	0.20552212	0.33333333
3	0	1	1	1
4				0.66666667
5				0.57142857

Table 3.4. Corine Land Cover (CLC-R) reclassification for weighting purposes

CLC-R legend	**CLC level3**	**class 1**	**class 2**	**class 3**
Urban fabric dense (>50% built-up)	111	1	3	2
Urban fabric medium density (30-50% built-up)	112	1	3	2
Urban fabric low density (10-30% built-up)	112	1	3	2
Urban fabric very low density and isolated (<10% built-up)	112	2	3	2
Production facilities	121	2	2	3
Commercial service facilities	121	2	3	2
Public facilities	121	2	3	2
Road and rail networks and associated land	122	2	3	5
Major railway stations	122	2	3	5
Port areas	123	2	3	5
Airport areas	124	2	3	5
Airport terminals	124	2	3	5
Mineral extraction sites	131	3	2	4
Dump sites	132	3	2	4

CLC-R legend	CLC level3	class 1	class 2	class 3
Construction sites	133	1	2	4
Green urban areas	141	2	3	2
Sport and leisure facilities	142	2	3	2
Leisure and touristic built-up	142	2	3	2
Non-irrigated arable land	211	1	1	1
Permanently irrigated land	212	1	1	1
Rice fields	213	1	1	1
Vineyards	221	1	1	1
Fruit trees and berry plantations	222	1	1	1
Olive groves	223	1	1	1
Pastures	231	1	1	1
Annual crops associated with permanent crops	241	1	1	1
Complex cultivation patterns	242	1	1	1
Land principally occupied by agriculture, with significant areas of natural vegetation	243	1	1	1
Agro-forestry areas	244	1	1	1
Broad-leaved forest	311	3	1	1
Coniferous forest	312	3	1	1
Mixed forest	313	3	1	1
Natural grasslands	321	3	1	1
Moors and heathland	322	3	1	1
Sclerophyllous vegetation	323	3	1	1
Transitional woodland-shrub	324	3	1	1
Beaches, dunes, sands	331	3	1	1
Bare rocks	332	3	1	1
Sparsely vegetated areas	333	3	1	1
Burnt areas	334	3	1	1
Glaciers and perpetual snow	335	3	1	1
Inland wetlands	410	3	1	1
Salt marshes	421	3	1	1
Salines	422	3	1	1
Intertidal flats	423	3	1	1
Water courses	511	3	1	1
Water bodies	512	3	1	1
Coastal lagoons	521	3	1	1
Estuaries	522	3	1	1
Sea and ocean	523	3	1	1

3.5.2.1 Particularities of some indicators

Besides the overall methodology and the different methodologies explained above by indicator types, there are some particularities and specific cases that have lead to a different way of calculating the figures by functional areas or that have prevented us from calculating them at all. These particularities are detailed in this section.

Derived indicators

Amongst the ESPON database demographic indicators there is a subset that has been derived as a result of simple mathematical operations based on other base indicators (e.g. difference, ratios…). A list of these indicators is shown in Table 3.5. Derived ESPON base indicators.

Table 3.5. Derived ESPON base indicators

code	obj_id	Description
pop_t_oldage	223	Quotient between Total population older than 65 years of age and between 15-65 years of age, measured on 1 January
pop_f_oldage	224	Quotient between Female population older than 65 years of age and between 15-65 years of age, measured on 1 January
pop_m_oldage	225	Quotient between Male population older than 65 years of age and between 15-65 years of age, measured on 1 January
f_total	220	Percentage of female population respect the total, measured on 1 January
m_total	221	Percentage of male population respect the total, measured on 1 January
Ag_In_t	215	Ageing index of female population calculated as the number of persons 60 years old or over per hundred persons under age 15
Ag_In_F	216	Ageing index of female population calculated as the number of persons 60 years old or over per hundred persons under age 15
Ag_In_M	217	Ageing index of male population calculated as the number of persons 60 years old or over per hundred persons under age 15
natural_pop_change	311	Natural population change is the difference between the number of live births and the number of deaths. If natural change is positive, then it is often referred to as a natural increase.
net_migration_total	312	The net migration plus adjustment is calculated as the difference between the total change and the natural change of the population.

For the sake of accuracy, the disaggregation products of these indicators have been obtained from the corresponding indicators previously disaggregated by the different functional regions. In this way, these indicators have been calculated as shown by Table 3.6. Calculation of derived ESPON base indicators. The correspondence between the obj_ids and the indicators is available in Annex 2: List of ESPON 2020 Base Indicators eventually estimated by FUORE.

Table 3.6. Calculation of derived ESPON base indicators

code	obj_id	Required indicators (*obj_id*) and calculations to be performed
pop_t_oldage	223	(obj_id 156 / obj_id 155) * 100
pop_f_oldage	224	(obj_id 149 / obj_id 148) * 100
pop_m_oldage	225	(obj_id 152 / obj_id 151) * 100
f_total	220	(sum (obj_id 147: obj_id 149) / obj_id 314) * 100
m_total	221	(sum (obj_id 150: obj_id 152) / obj_id 314) * 100
Ag_In_t	215	sum(obj_id 208: obj_id 213) / sum(obj_id 196: obj_id 198)
Ag_In_F	216	sum(obj_id 171: obj_id 176) / sum(obj_id 159: obj_id 161)
Ag_In_M	217	sum(obj_id 189: obj_id 194) / sum(obj_id 177: obj_id 179)
natural_pop_change	311	obj_id 313 - obj_id 315
net_migration_total	312	(obj_id 314 [in year t] - obj_id 314 [in year t-1]) - obj_id 311

Life expectancy related indicators

Within the ESPON database there is a set of life expectancy demographic indicators measuring the "mean number of years still to be lived by a person who has reached a certain exact age, if subjected throughout the rest of his or her life to the current mortality conditions (age-specific probabilities of dying)". To our knowledge, there is no available dataset that could guide the disaggregation of such kind of indicators. For this reason, the service provider proposes not to disaggregate life expectancy indicators in the framework of the FUORE project.

Indicators expressed as ratio/percentages

Some of the ESPON Base indicators are expressed as a ratio or percentage (%). In order to be able to disaggregate those indicators and calculate the corresponding figures by functional area, it is needed to transform them to stock values first, carry out the disaggregation/aggregation process and, eventually, recalculate the percentages.

In very few cases, the calculation of stock values has not been possible due to the lack of background information. In such cases, the indicators have not been disaggregated and are not included in the ESPON FUORE project. Table 3.7. ESPON base indicators expressed as ratio or percentage and not disaggregated by FUORE shows a list of those indicators.

Table 3.7. ESPON base indicators expressed as ratio or percentage and not disaggregated by FUORE

code	Name	obj_id
Households_broad_access	Households with broadband access (percentage of households)	318
%_RES	Share of energy from renewable sources (in gross final energy consumption)	324
TOTAL_INTRAMURAL_R&D_EXPEDITURE	Total intramural R & D expenditure (GERD) by sectors of performance (Percentage of GDP)	332

3.5.3 Scripting and processing

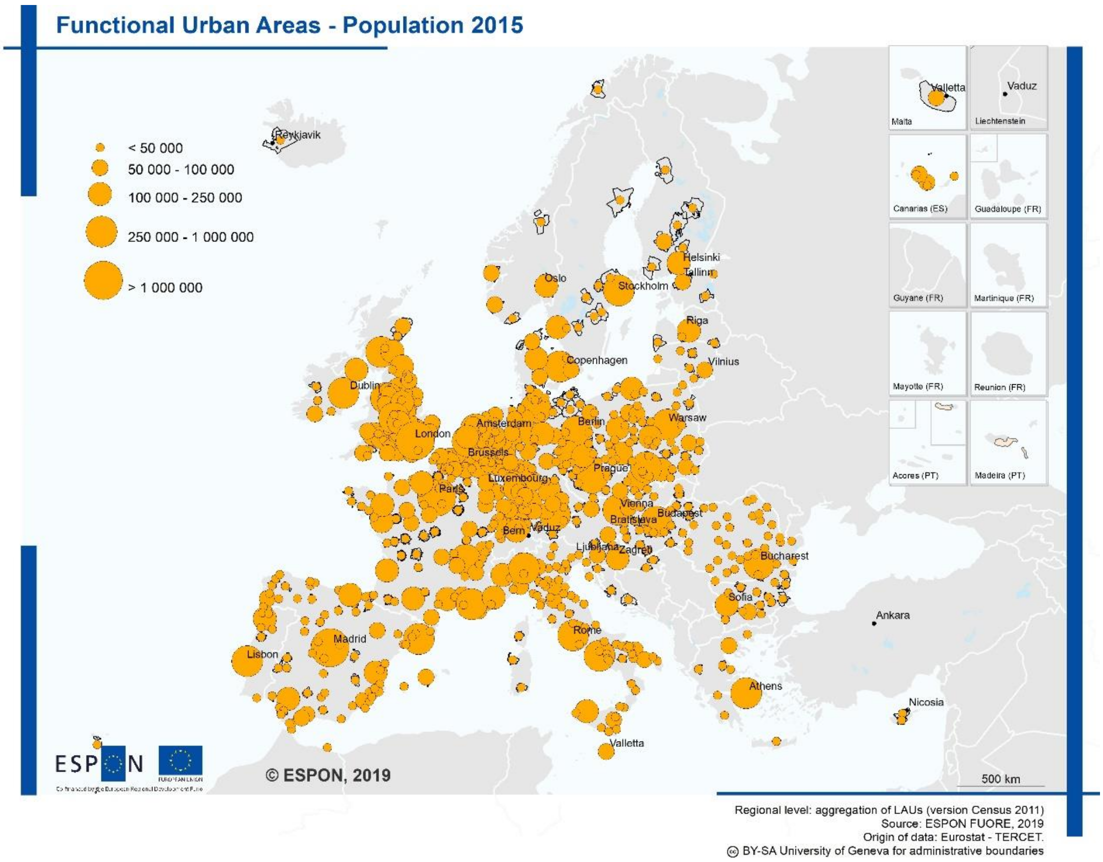
The ESPON base indicators have been selected and extracted automatically by means of the existing web services implemented by the ESPON 2020 Database team. The entire processing is made using different Python libraries, whereas a POSTGIS database is used for storing indicators and final outputs. In order to automate all the methodological steps previously described (sub-section 3.5.2), the programming routines minimise the computation time by applying multiprocessing and parallel computing.

3.5.3.2 Assessment of alternative scripting approaches

In order to assess the best option in terms of accuracy and computation time, two disaggregation processes, based on raster and vector format datasets, respectively, have been tested. Nowadays, array-based processing with the current available libraries has turned out to be really fast, and promising results have been obtained in terms of timing by using rasterised versions of datasets, more specifically in the computation of zonal statistics, disaggregation data and joining tables. However, the accuracy of the resulting outputs is a serious objection to choose the raster approach as the most suitable estimation method because rasterised versions of involved geometries will over/under-estimate values at reporting-unit borders, producing biased results. Although processing vector layers involves larger computation time due to the high spatial resolution of the delineations which we are working with (aggregations of LAU units), their use allow to obtain more accurate disaggregation results.

Raster-based scripts provided disaggregation results in shorter time than vector-based scripts for some steps of the processing. However, the optimization of the scripts processing vector layers allowed to reduce largely the computing time while keeping the spatial precision of the original delineations of the functional regions and the ESPON indicators at NUTS3 (for a detailed report, please see Table A3.5 in Annex 3: Disaggregation/reaggregation process for demographic indicators). For this reason, **vector-based scripts were finally preferred** and have been applied for the disaggregation of indicators.

Map 3.10. Estimated total population by *Functional Urban Areas across Europe in year 2015. Results from disaggregation and re-aggregation* of *the population values contained in the ESPON* DB *(*by *NUTS3).*



A complete report on the scripts, processing and computation can be found in Annex 3: Disaggregation/reaggregation process for demographic indicators.

4 Technological development of the Spatial Multidimensional Database

4.1 The Spatial Multidimensional Database and metadata

The service provider has built an ESPON SMD which includes all the indicators by the different reporting units (nine functional regions, see sub-section 3.2). The design and implementation of the spatial multidimensional database has required the spatial representation of the datasets, the identification of the structure of the different tables and the relationships among the datasets, the definition of the integrity rules and the workflow. For each dataset the service provider has produced a metadata file with the relevant information about the data. For this purpose, the project adopts the INSPIRE Directive metadata model using the ISO 19115/19139 standards. The ESPON SMD generated is fully operational and integrated as part of the web tool (see section 5). The service provider has defined the database schema in which all the data is integrated, and it has stored all the indicators following that schema. That schema is flexible enough to integrate the different functional regions, taking into consideration the characteristics of each dataset, the different spatial levels, and the spatial relationships among them. A detailed report on the database schema and its tables is provided in Annex 3.

4.2 The Web Data Analysis Toolbox

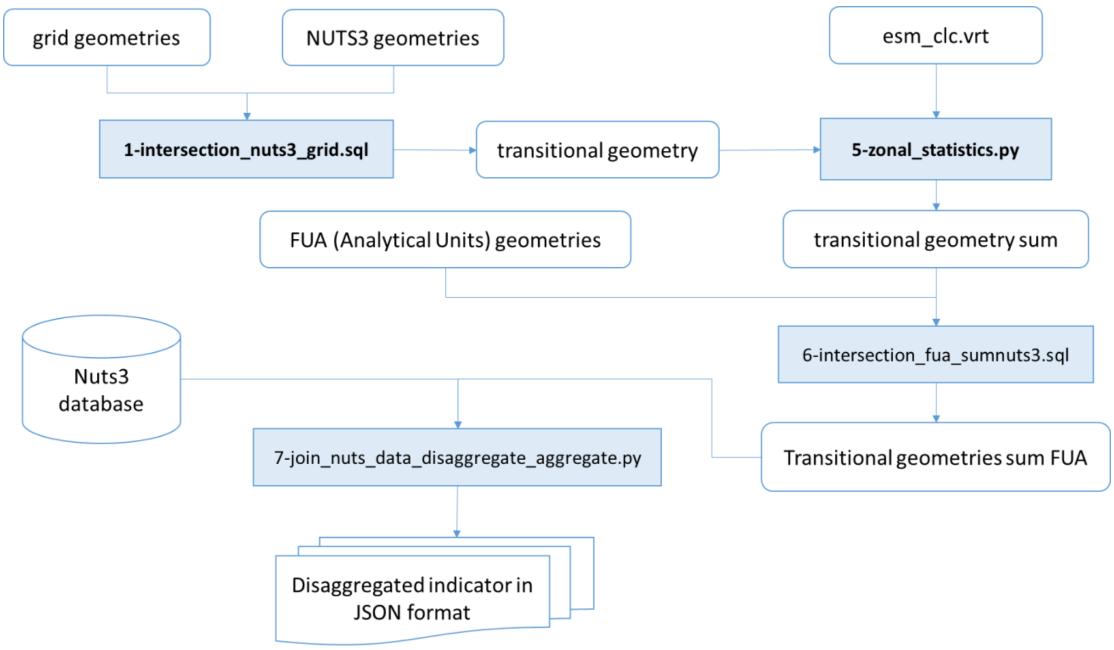
ESPON advanced users will be able to interact with the ESPON Spatial Multidimensional Database and perform deferred computation of a specific NUTS2 or NUTS3 indicator (version 2013) by the different reporting units through the Web Data Analysis Toolbox. This analytical tool is embedded in the web tool. The Analysis Toolbox allows the users to use as data source the ESPON base indicators copied in the SMD, but also providing any indicator the user might be interested in (following a specific input data format), in a new analysis, which will produce new estimation, since the algorithm and processing described in sub-section 3.5.2 and Annex 3, respectively, are fully integrated in the Data Analysis Toolbox. The user-generated data is eventually added as new data to the web tool after internal validation.

The Data Analysis Toolbox implements an innovative technical solution based on Jupyter Hub, a multiuser server for Jupyter notebooks (http://jupyter.org/hub) to be able to run the whole data processing in the cloud. The Jupyter Notebooks contain a user-friendly version of the scripts described in sub-section 3.5.2. This solution allows users to reduce computing time and scale up the system allocating more hardware resources and/or executing Python scripts in a multicore processing or cluster environment.

The service provider has provided technical documentation describing in detail how to run the Web Data Analysis Toolbox to the ESPON administrator but also to the ESPON advanced users, and how data is eventually integrated into the SMD. It is also describing how the data is converted and transformed, imported and exported. Finally, the tool has been tested to verify the production, performance and scalability. The service provider is taking into consideration the characteristics of each dataset, the different spatial levels (NUTS, functional regions and grid) and the spatial relationships among them.

From the user perspective, the Data Analysis Toolbox allows to easily run the disaggregation-aggregation process as well as to visually check and validate the outputs resulting from this process. The Analysis Toolbox offers the possibility to run the scripts interactively to the users, always guided and supported by the technical documentation provided by the tool itself. The workflow diagram with all the scripts developed for the disaggregation process can be found in Figure 4.1.

Figure 4.1. Schema of the disaggregation process used in the Data Analysis Toolbox.



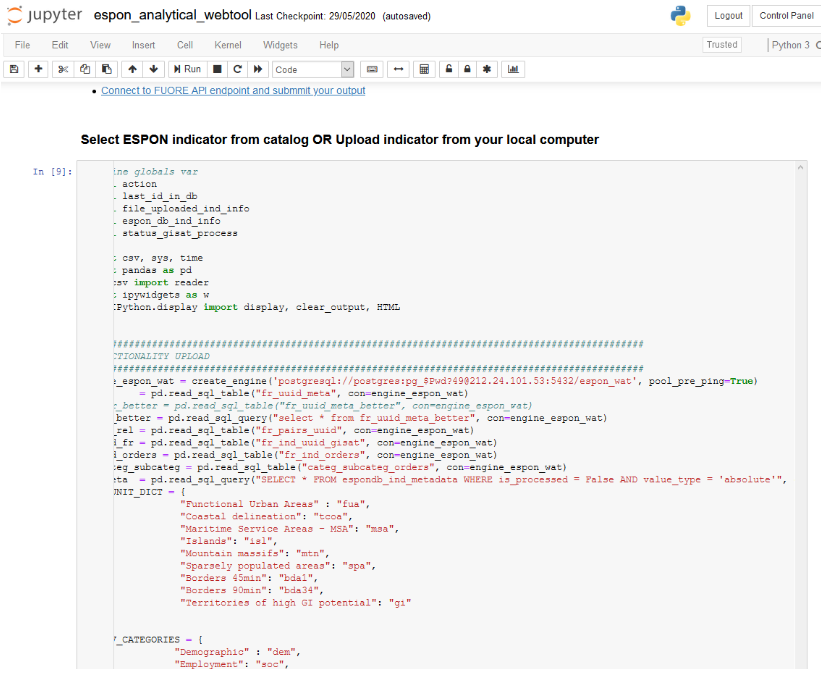
Users are allowed to access the Toolbox using a specific username and a password. Only users with the credentials can log in and use the advance toolbox.

For general users, the Notebook includes a Graphical User Interface (GUI), as shown in Figure 4.2. The GUI eases the user experience enabling to select the specific inputs and analysis using a drop-down menu instead of the code.

Figure 4.2. Overview of the welcome message from the Data Analysis Toolbox.

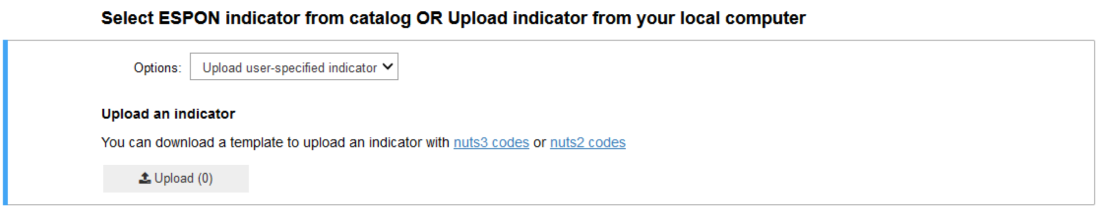
For advanced users, the visualization of the scripts behind the tool can be enabled, as well as the possibility to modify them, as Figure 4.3 is showing.

Figure 4.3. Overview of the Data Analysis Toolbox: advanced users, inspection of the code.



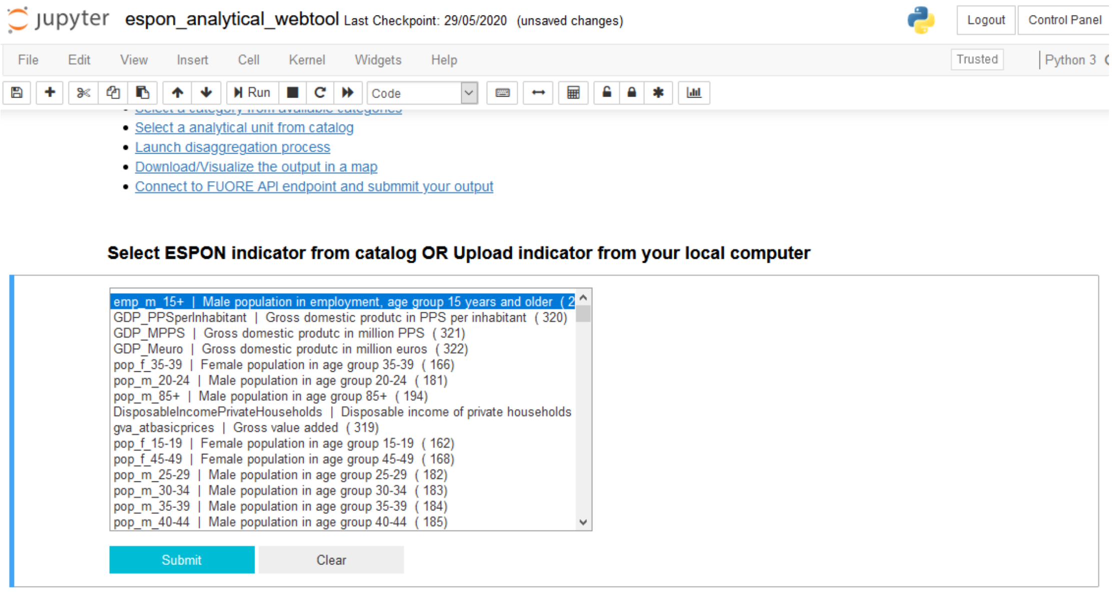
Users are able to process any indicator from the existing base indicators catalogue or upload their own indicators of interest based on a template (for NUTS2 or NUTS3 version 2013), which can be downloaded from the Notebook (see Figure 4.4. Overview of the Data Analysis Toolbox: Indicator upload functionality. below).

Figure 4.4. Overview of the Data Analysis Toolbox: Indicator upload functionality.



If the user aims at processing an ESPON base indicator, he/she can inspect the available indicators in the catalogue from the Toolbox itself (Figure 4.5)

Figure 4.5. Overview of the Data Analysis Toolbox: ESPON base indicators list.

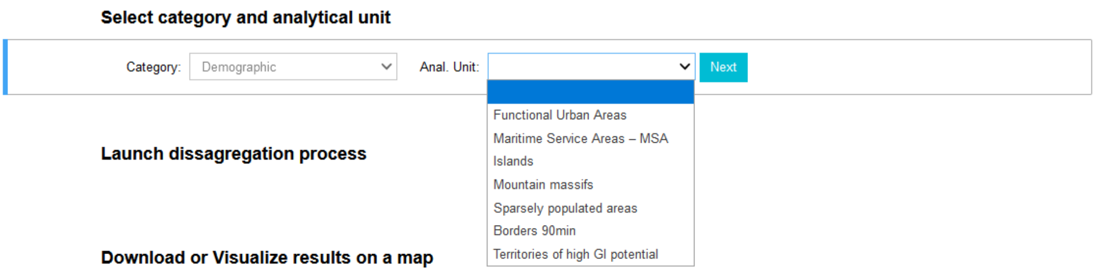


In case the user wishes to process his/her own indicator, he/she should upload the data using the provided templates by means of the "Upload" button shown in Figure 4.4. Overview of the Data Analysis Toolbox: Indicator upload functionality..

Once the user has specified the indicator to be processed, the Data Analysis Tool applies the disaggregation methodology on the basis of the type of indicator (demographic or socioeconomic), which is automatically detected in the case of the ESPON base indicators, and selected by the user in the case of an *ad hoc* processing.

The user should also select the output functional region (Functional Urban Areas, Coastal delineation, Maritime Service Areas, Islands, Mountain Massifs, Sparsely populated areas, Territories of high GI potential) for the disaggregation (Figure 4.6).

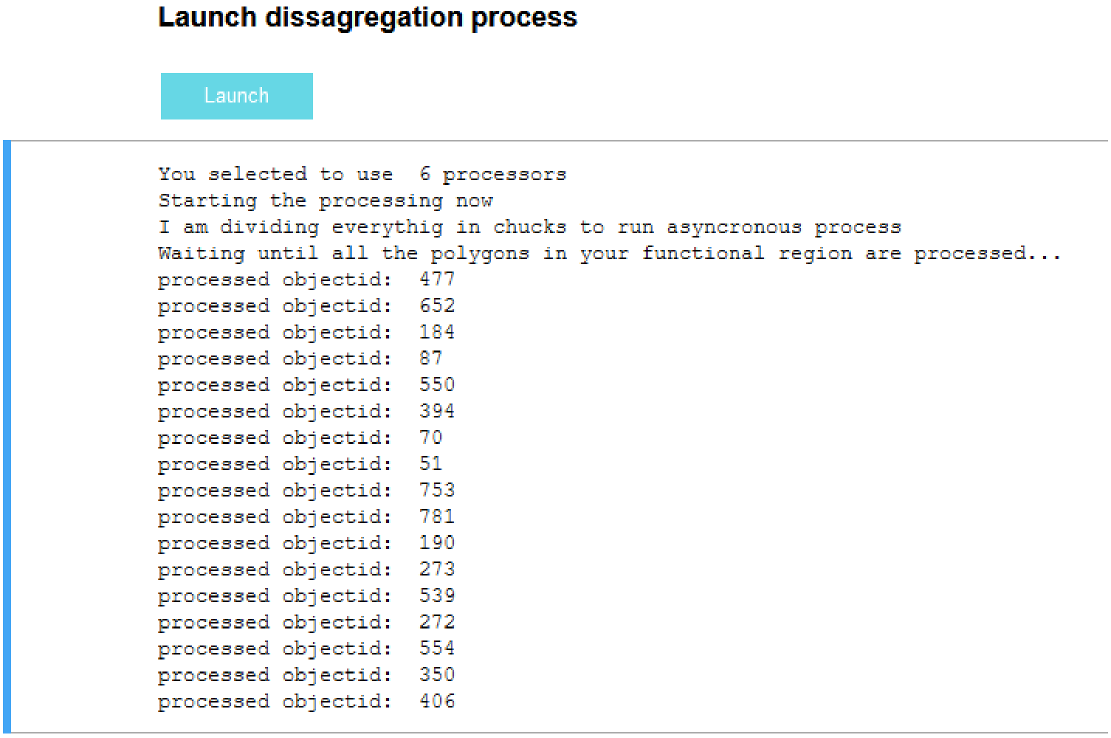
Figure 4.6. Overview of the Data Analysis Toolbox: choosing the functional region.



Although in the framework of the present contract the Tool is offering a limited and pre-defined number of functional regions for the disaggregation, the functionalities of the Data Analysis Tool could be extended thus other functional regions (i.e., other geometries) could be used in the future.

The processing takes few minutes and the user can follow the progress of the processing from the Data Analytical Tool, a shown in Figure 4.7.

Figure 4.7. Overview of the Data Analysis Toolbox: processing.



Although the main objective of the Tool is to provide the user with a processing tool, the Data Analysis Toolbox also allows for a visual inspection of the outputs derived from the processing. In this way, the user can check the results before they are uploaded into the web tool (see section 6). The visualisation can be done in two different ways: 1) spatially through maps, and 2) temporally through charts (see Figure 4.8 to Figure 4.12 below).

Figure 4.8. Overview of the Data Analysis Toolbox: visualisation of the outputs.

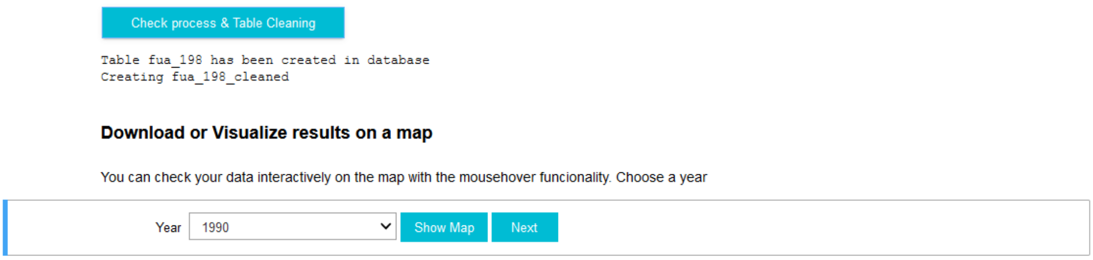
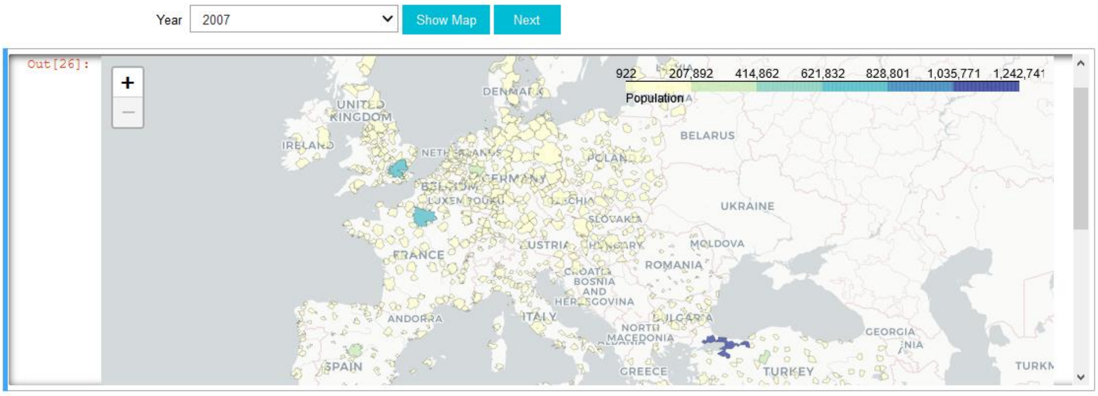
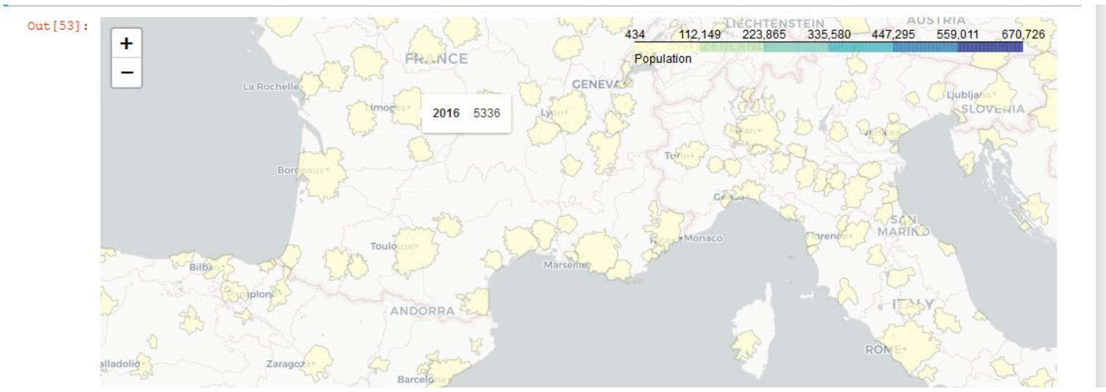


Figure 4.9. Overview of the Data Analysis Toolbox: visualisation of the outputs, maps.



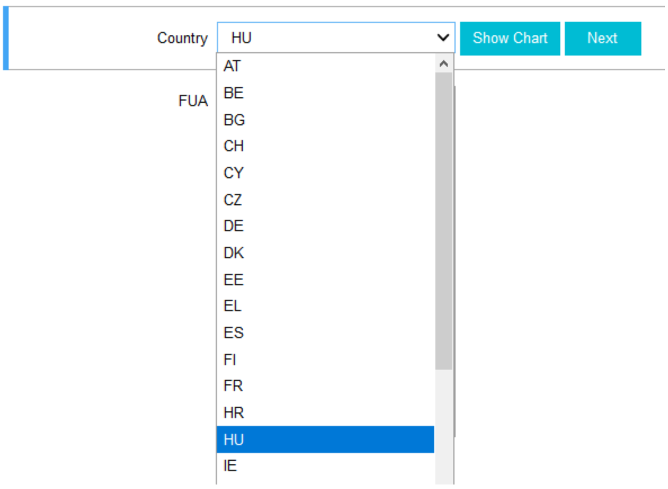
The user can scroll through the map, zoom in and zoom out, and to interactively explore the results with the mouse hover functionality.

Figure 4.10. Overview of the Data Analysis Toolbox: visualisation of the outputs, maps.



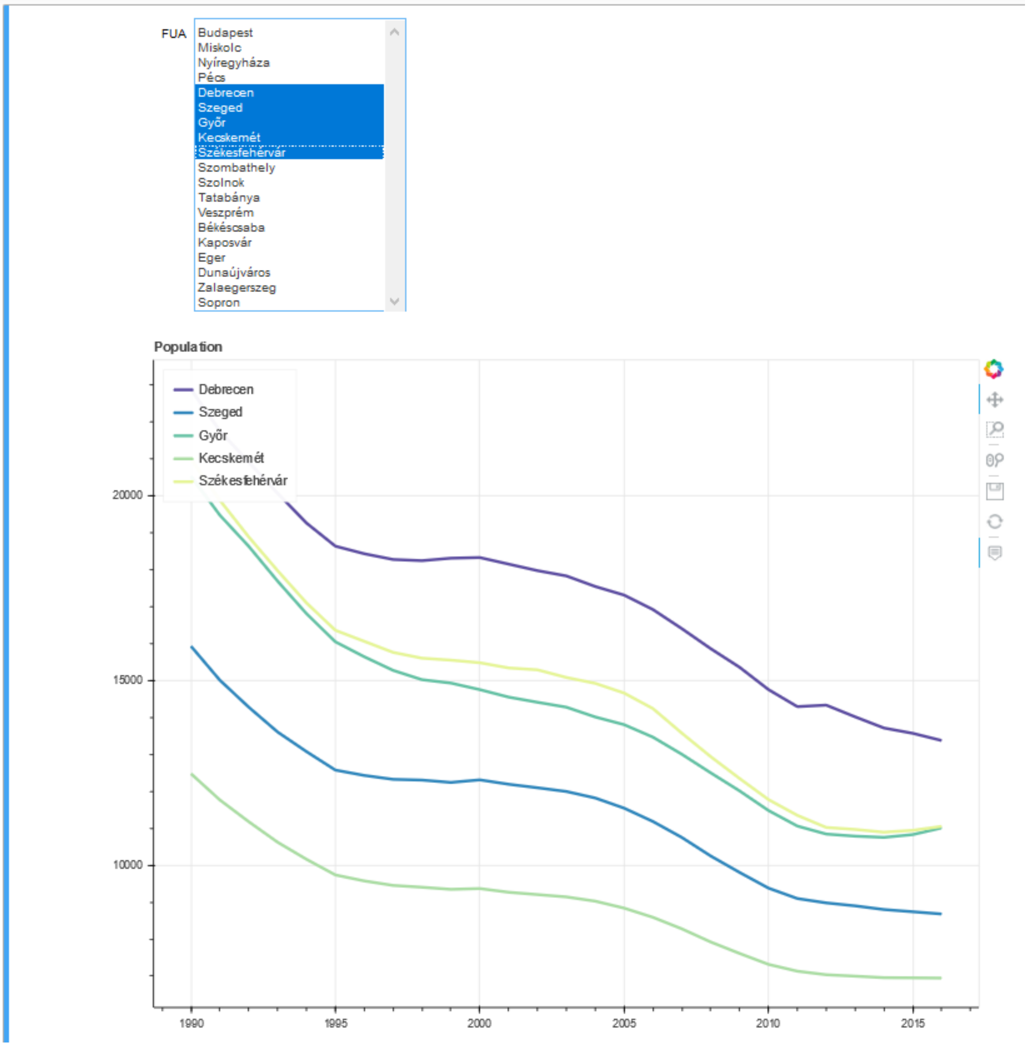
The user is also able to inspect the evolution of the disaggregated indicator over time for specific countries and particular functional regions within the country by visualising the corresponding chart in the Data Analysis Tool:

Figure 4.11. Overview of the Data Analysis Toolbox: visualisation of the outputs, charts.



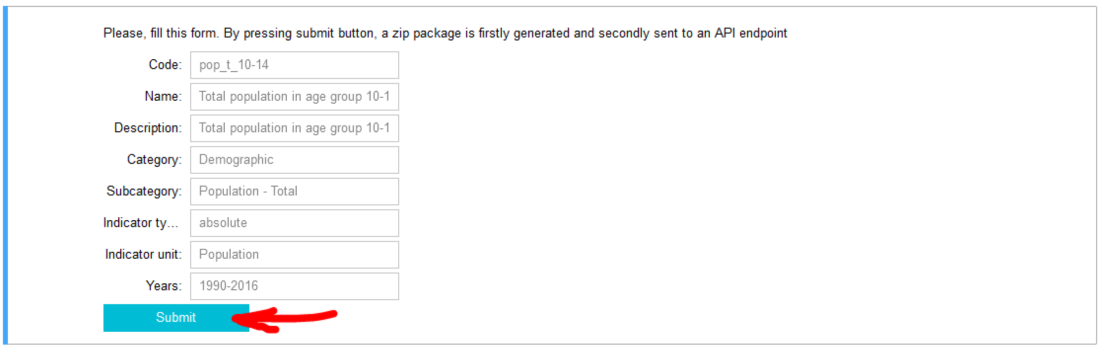
Like the map, the chart is also interactive thus the user can check the different values over the period by clicking with the pointer on the plot line (see Figure 4.12).

Figure 4.12. Overview of the Data Analysis Toolbox: visualisation of the outputs, charts.



After using the multiple and interactive ways offered by the Data Analysis Toolbox for the inspection of the outputs generated by the user, he/she can verify or fill in some metadata, before sending the results to the web tool (see Figure 4.13 below).

Figure 4.13. Overview of the Data Analysis Toolbox: exporting the outputs.



Summarizing, the following functionalities are implemented in the Data Analysis Toolbox:

- Selection of the functional region and the particular indicator to be computed using code or a Graphical User Interface (GUI);
- Disaggregation-aggregation of indicators provided by the users (using specific NUTS2 or NUTS3 templates);
- Exploration of the output results through a multi selection chart;
- Visualization of the results on an interactive map;
- Eventual submission of results to the FUORE web tool.

The required format of the outputs for the web tool, as well as a detailed report on the requirements of the user environment for a successful processing can be found in Annex 4: User requirements for the Data Analysis Toolbox.

The entire software environment has been set up as part of the existing cloud infrastructure for this project. As the Data Analysis Tool runs in the cloud, it is much more efficient than other traditional software solutions such as Geographic Information Systems. It provides the user with a highly interactive and flexible tool which allows the visualisation of the outputs obtained, both spatially and temporally. This greatly facilitates the inspection of the results and the decision making of the user on the suitability of the outputs obtained. But the user is not a mere passive subject visualizing data. Thanks to the Data Analysis Tool, he/she has the power to generate his/her own data and to share them with other users through the web tool, since both Data Analysis and web tools are connected. Moreover, advanced-expert users are allowed to modify the code and to do improvements or adjustments in the processing.

However, the current capabilities of the Data Analysis Tool could be even improved and extended. For example, the disaggregation process could be applied to different kind of geometries introduced by the user and not being restricted to a limited number of functional regions as it is now.

Currently, the set of ESPON base indicators available are the same ones that the FUORE web tool is providing, but they can be updated in line with the ESPON DB updates, and the user would be generating new estimations from the available indicators from the ESPON DB at any time. Furthermore, this Data Analysis Tool could be an interesting way to easily update indicators (or to filling the gaps in their time series) for other ESPON projects in such a way that the process would become more transparent to the users.

5 Validation of the methodology

The process of validation of the methodology involves making data estimations for the different functional regions using the ESPON SMD. The idea is to determine whether the methodology used by ESPON FUORE produces data that is close enough to officially collected data. Thus, the aim of the validation task has consisted in verifying the methodology and identifying any gap, if existing. The validation process included comparisons with the official statistics and one expert validation workshop during which the representatives of Eurostat, NSI and other experts on data collection had to validate the data provided by the ESPON SMD and the ESPON FUORE web tool. Nevertheless, the COVID-19 crisis prevented the workshop from being celebrated. The results might be eventually presented and discussed in the next ESPON Seminar.

The following sections summarise the different steps of the validation process.

5.1 Statistical validation of estimated data by the ESPON SMD

The service provider has validated the indicators estimated by the SMD against official data. This validation has checked the data quality and completeness of the data.

5.1.1 Data quality

The quality of the results is checked by carrying out an outlier control on the indicators produced by the project. This control is performed using the methodology proposed by the ESPON M4D project, which implements the Mahalanobis distance test. This test calculates, for each value in the dataset, the distance between a data point and a multivariate space's centroid. Thus, it is a measure of how far away a particular case is from the remaining cases. Whenever the Mahalanobis test score exceeds a "critical value" for a particular case, it is considered an outlier.

The Mahalanobis distance test is implemented through the commercial statistical package SPSS.

The starting point of the testing consists of the preparation of the data. This part of the validation has been carried out for 290 individual indicators with data available for the time span 1990-2016 at different statistical units:

Table 5.1. Indicators which have gone through the outlier testing

Indicators (classes)	Themes	Time span	Statistical units	Individual values
290	• Basic demography (deaths, births) by gender and age group • Total population by gender and age group • Derivate population indicators (ageing	1990-2019	• Border narrow • Border large • Functional urban areas • Green infrastructure • Islands	1.7 million

	index, population change, net migration) • Employment by gender and age group (in number and %) • Unemployment by gender and age group (in number and %) • Long-term unemployment • Employment by economic activity		• MSA coasts • Mountains • Sparsely populated areas • TERCET coasts	

The data have been processed from the individual json files for each indicator, and set up in a way that can be fed to SPSS automatically. This has required setting up a MS Access macro that ultimately generates a table with all the values to be tested.

Given that many of the indicators are part of the same kind of measurement, that is different classes of a multi-indicator (according to the definitions of ESPON Database Portal 2020), the Mahalanobis test applies to all of them at once and there is no need to test each individual class.

The SPSS Regression command can save the squared Mahalanobis Distance for each case from the centroid of the predictor variables. The dependent variable for the regression does not affect the calculation of Mahalanobis Distance, so any numeric variable outside the predictor list can be used as the dependent variable if the calculation of Mahalanobis Distance values is the sole reason for running the Regression procedure.

Next, the results of the test are presented for the different indicator groups clustered by statistical unit, with boxplots for each cluster showing the unit codes for the outliers:

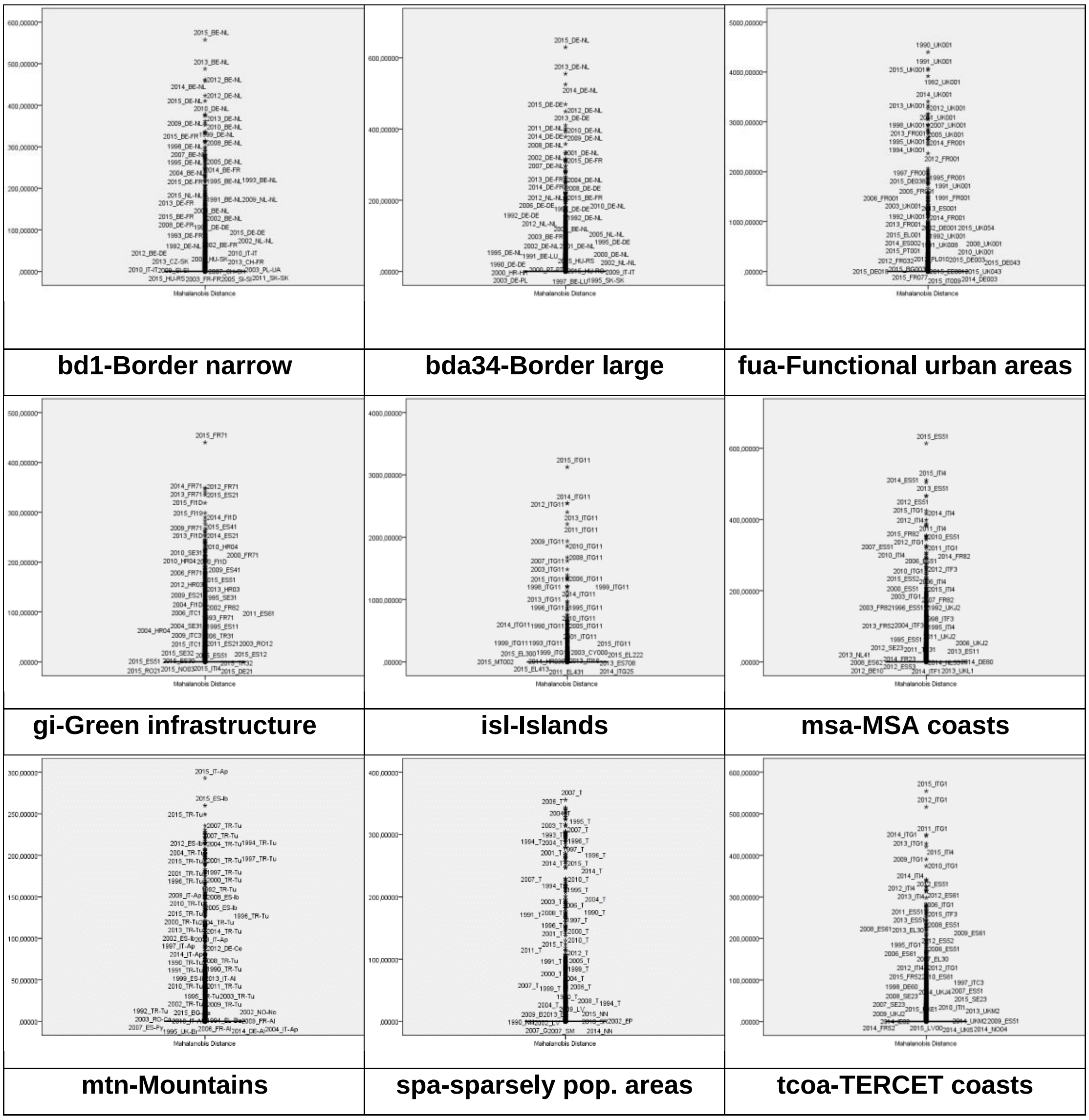
5.1.1.1 Basic demography (deaths, births) by gender and age group

Indicator definition and contents

This group of indicators encompasses the absolute number of births and deaths by gender and age group measured in 5-year intervals.

Outliers test

Figure 5.1. Boxplot for the basic demographic indicators by functional areas.



Analysis

In general, the test identifies outliers for regions with biggest total population, so this is not an error but a measure of how far away these regions are, compared to the rest of regions in terms of births and deaths. For instance, the FUAs of London, Paris or Madrid are way above the other units. Turkish regions in mountains and SPA are clear outliers, but due to the fact that these typologies almost encompass the whole country.

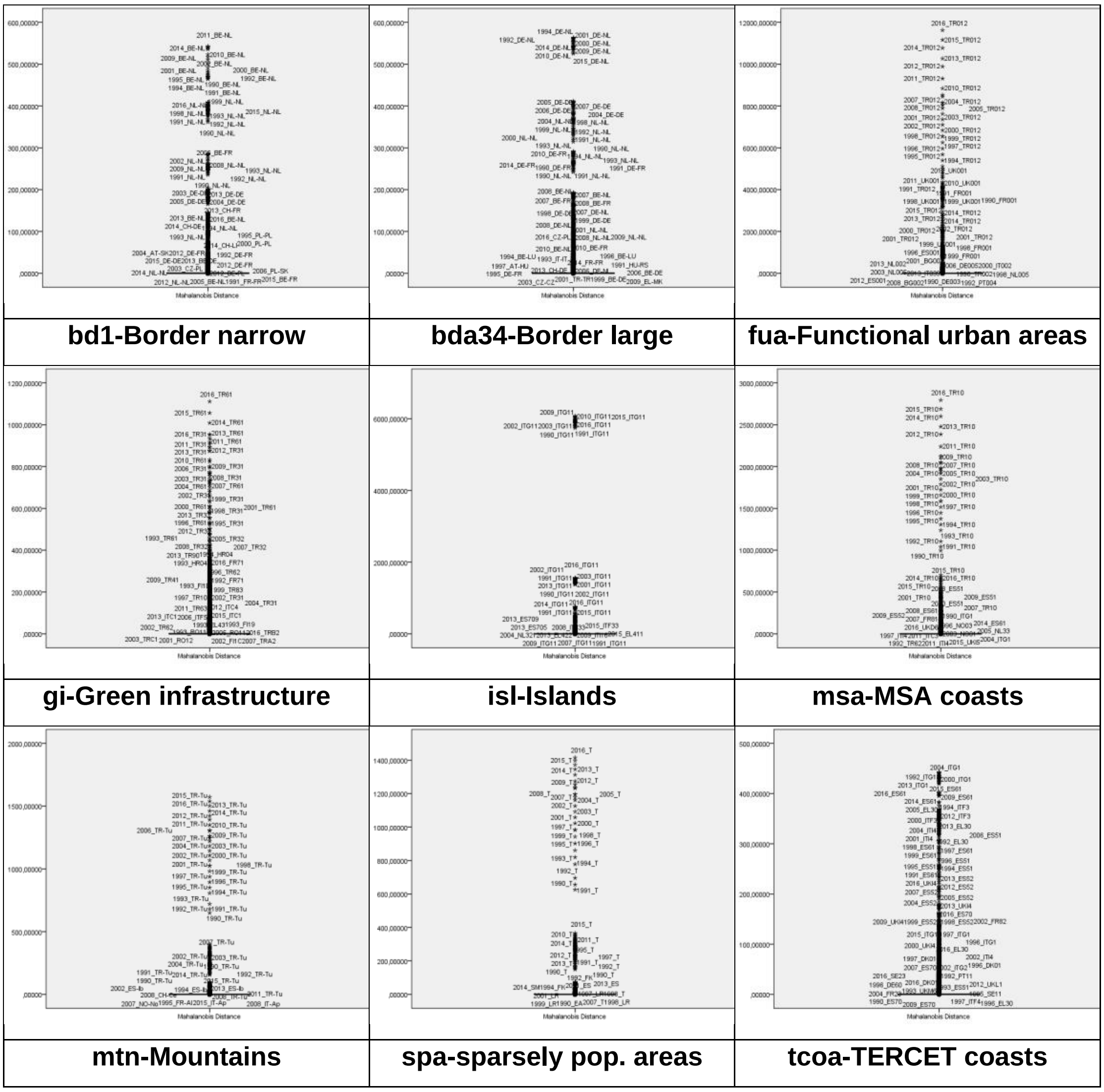
5.1.1.2 Total population

Indicator definition and contents

This indicator measures the total population at 1st January by gender and age group in 5-year intervals.

Outliers test

Figure 5.2. Boxplot for total population by functional areas.



Analysis

The outlier analysis identifies different areas that outstand in terms of total population, but a careful examination allows seeing these are in fact reflecting the underlying distribution of population. For instance, in border areas the border between BE, DE and NL we find the biggest population concentration. On other typologies like mountains, SPA, GI and MSA, Turkey stands out due to the concentration of population in the two big agglomerations of Istanbul and Ankara.

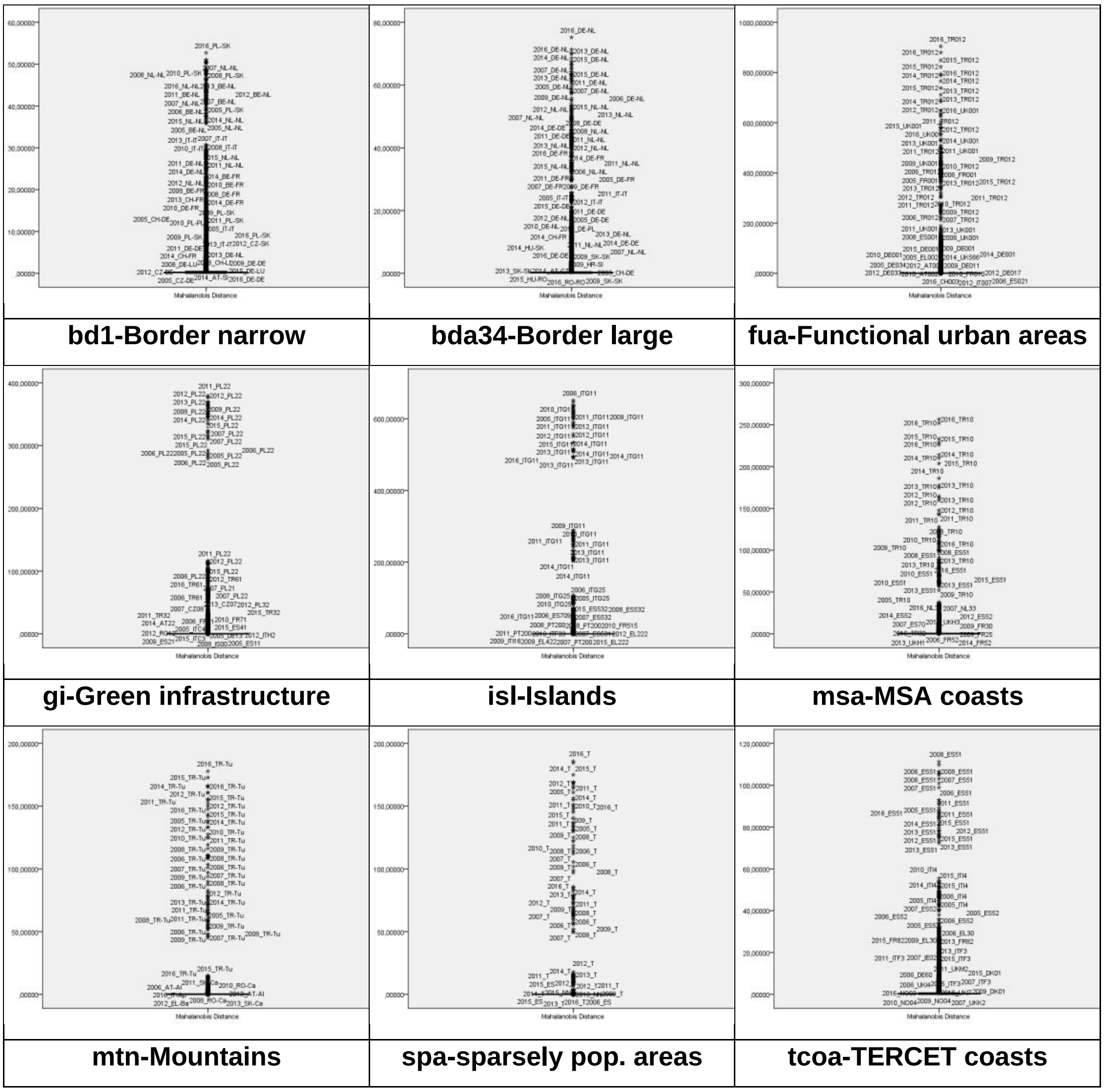
5.1.1.3 Employment by gender and age group

Indicator definition and contents

This group of indicators measures the number of employed people in absolute terms (number of people), differentiating gender and broad age groups (+15, +65, 15-24, 15-64, 20-64, 55-64)

Outliers test

Figure 5.3. Boxplot for employment by gender and age group by functional areas.



Analysis

As in previous cases, Turkish regions stand out most in mountain, SPA, MSA and FUA due to the geography of the functional units and the relation with population. Again, this is a normal result and there are no errors here. Polish regions also stand out in this case for border regions and GI, mainly due to the high concentration of employed people around the zone of Katowice-Częstochowa in the most recent years.

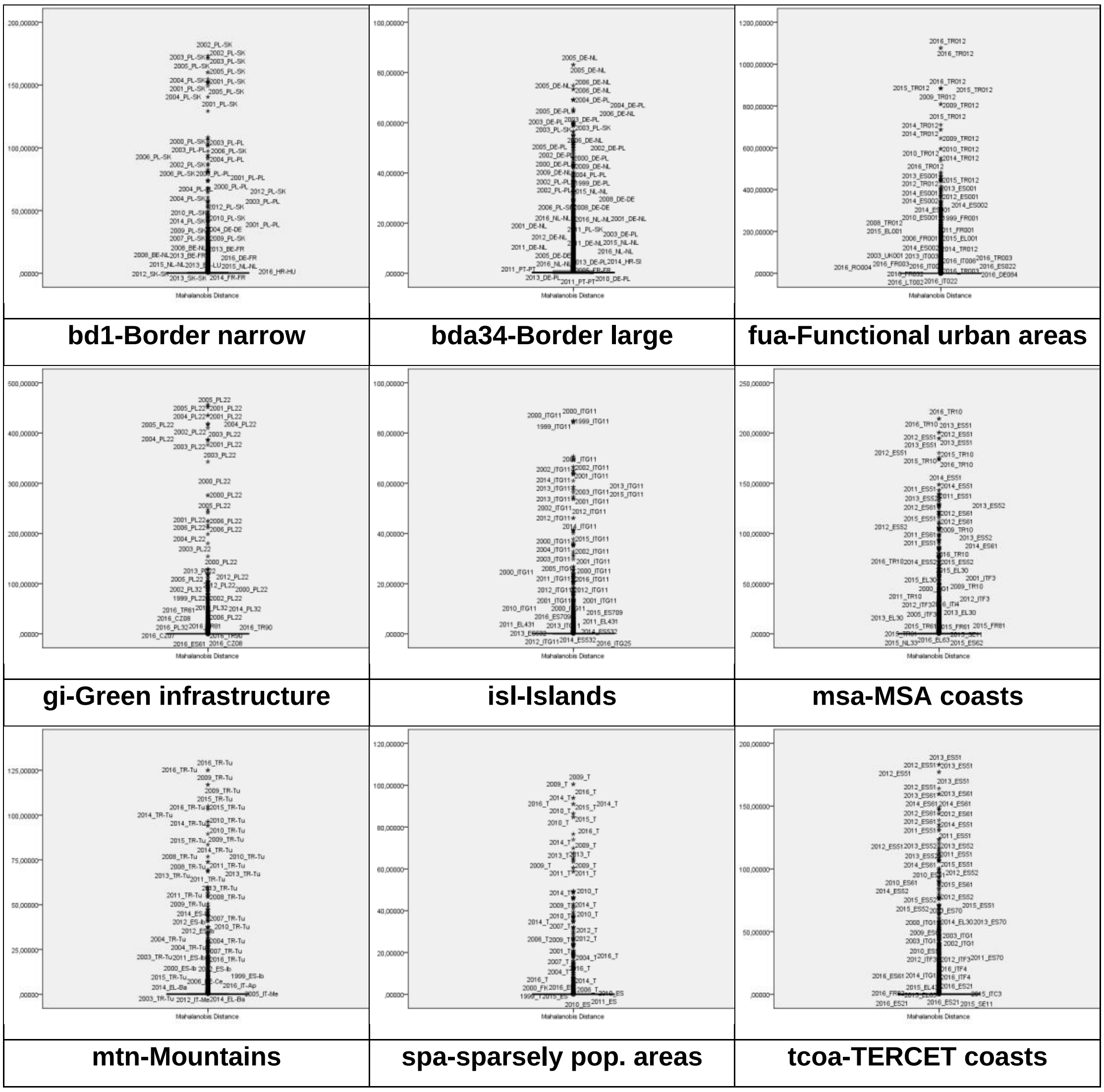
5.1.1.4 Unemployment by gender and age group

Indicator definition and contents

Unemployment indicators measure the number of unemployed people in absolute value differentiating gender and broad age groups (+15, +65, 15-24, 15-64, 20-64, 55-64)

Outliers test

Figure 5.4. Boxplot for unemployment by gender and age group by functional areas.



Analysis

Polish regions are clear outliers on border typologies for the years 2000-2004, as this area had lots of unemployment at the time. Once again Turkey stands out as an outlier for FUA, mountains, and SPA for a wide range of years as there is a high concentration of unemployed people. Some Spanish regions also stand out for coastal typologies in the years after the economical crisis, when unemployment was very high in Spain.

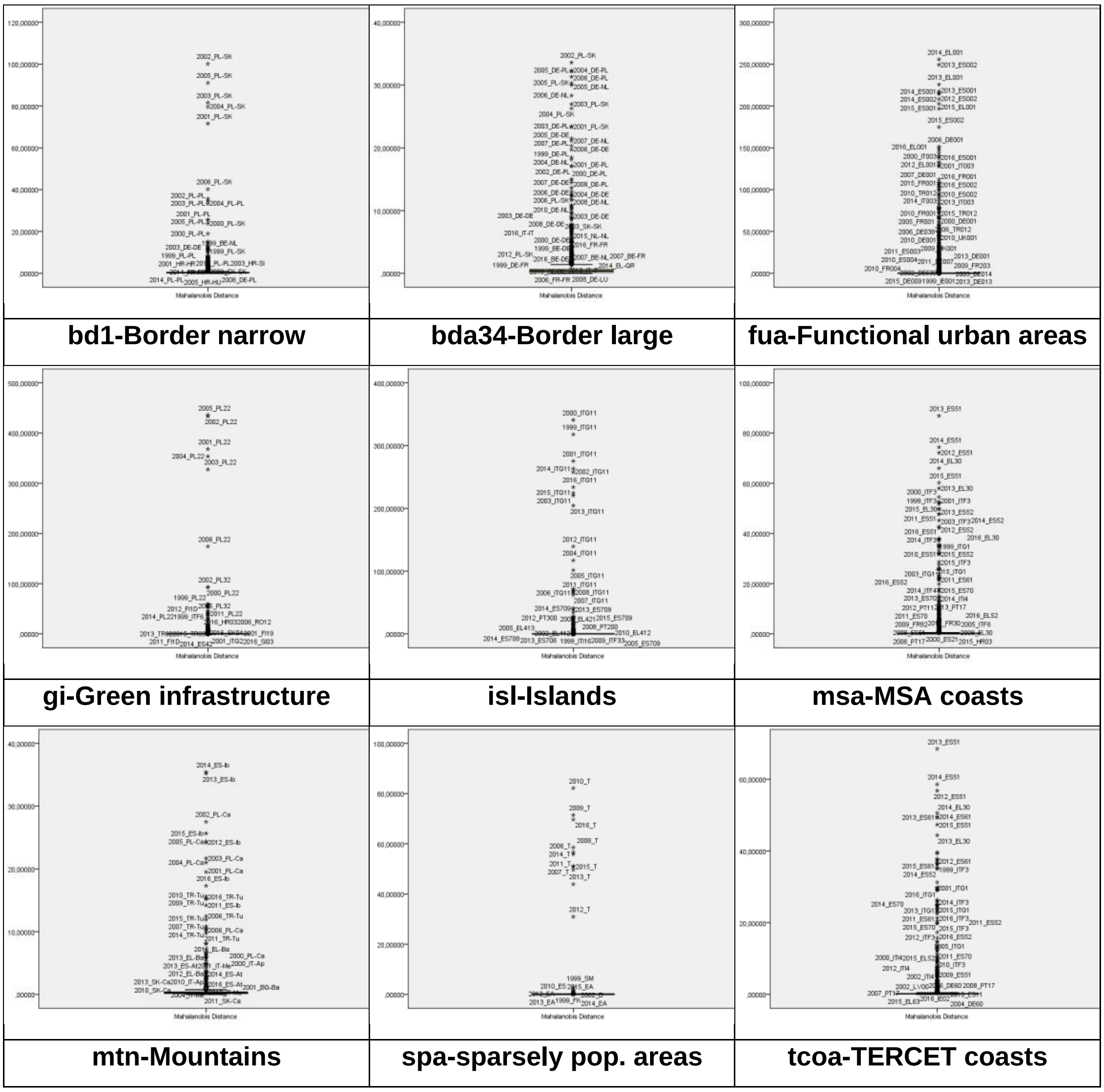
5.1.1.5 Long-term unemployment

Indicator definition and contents

Long-term unemployment measures the number of people who have been actively seeking employment for more than one year.

Outliers test

Figure 5.5. Boxplot for long-term unemployment by functional areas.



Analysis

On border areas, Poland stands out from the rest for the 2000-2004 period, as unemployment was high at the time. For FUA, coasts and mountains Spanish regions are the bigger outliers, showing the effects of the 2008 economic crisis in terms of high unemployment of long duration.

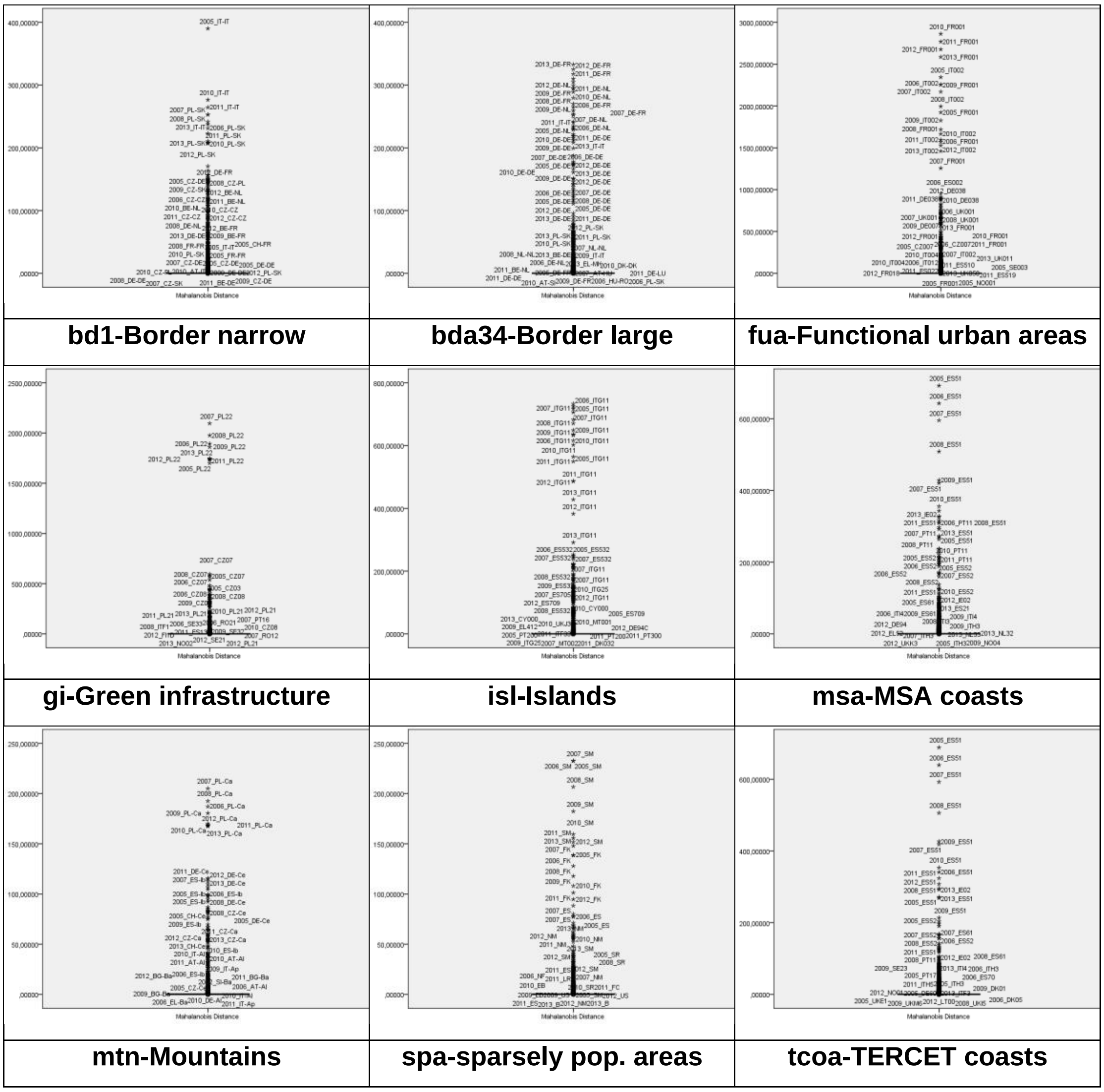
5.1.1.6 Employment by economic activity

Indicator definition and contents

These indicators measure the number of employed people in several specific economic sectors according to NACE classification (it includes sections manufacturing, construction and Accommodation and food service activities).

Outliers test

Figure 5.6. Boxplot for employment by economic activity by functional areas.



Analysis

In this case, the outlier test identifies for FUAs the big economic poles of Europe, such as Paris, Milan or Barcelona. This is logical as these regions concentrate a higher number of employed people on the analysed economic sectors. The Mediterranean coast of Spain is also an outlier but again this makes sense when looking at the data and it is not an error of the model.

5.1.2 Goodness of fit with official data

This task has consisted of assessing the goodness of fit of the SMD indicators against observations from official databases. The assessment produces a summary of the discrepancy between observed values and the outcomes of the model.

This assessment has been done against two different data sources:

- FUA and NUTS3 data from EUROSTAT
- Data from National Statistical Offices

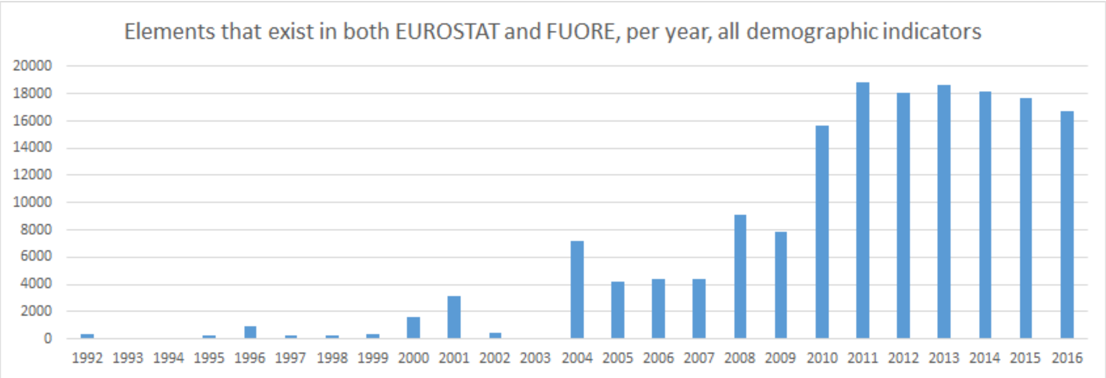
5.1.2.1 Goodness of fit with EUROSTAT data

FUA demography

For demography indicators we have taken the data available in EUROSTAT at FUA level. This means total population by sex and age group for the time period 1990-2016.

However, EUROSTAT data series are not complete, especially for older years, thus when comparing the original data with the FUORE estimations the amount of data points that can be checked is as follows:

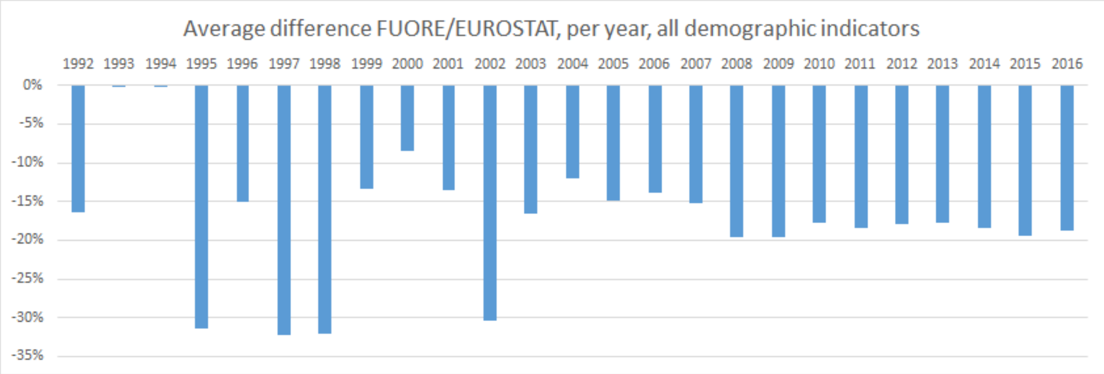
Figure 5.7. Number of data points per year for FUA values existing in EUROSTAT and FUORE for the selected demographic indicators.



This highlights the usefulness of the FUORE methodology, which will allow filling in all the gaps from 1990-2009 period with estimations of population by sex and age group at FUA level.

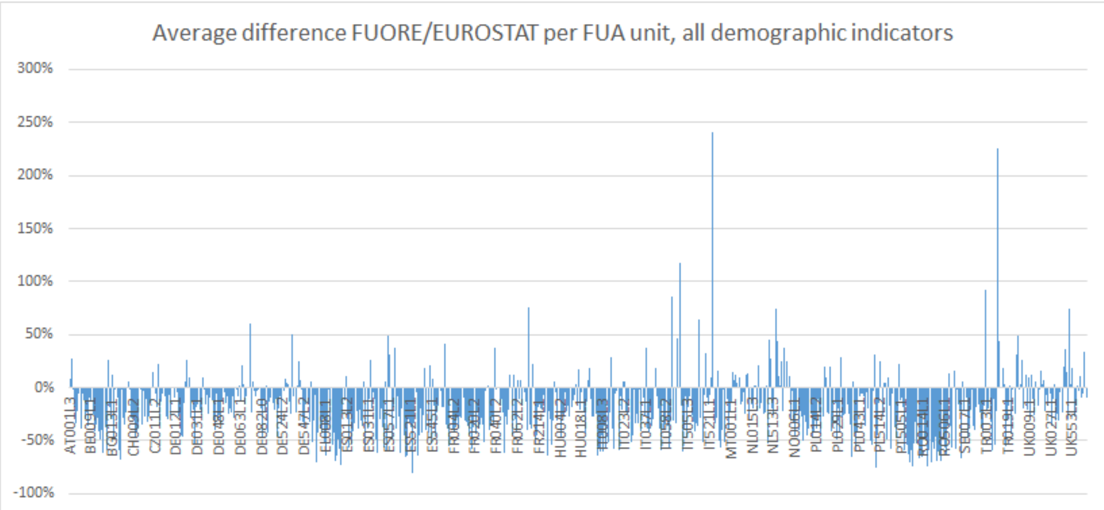
When looking at the differences between the EUROSTAT data and FUORE outputs, aggregating all age groups and sex by year, it is evident that there is a consistently lower amount of population in the FUORE data, showing that the inputs are different. However, this does not invalidate the methodology, simply highlights the need of making further analysis to determine the difference in total amount of population.

Figure 5.8. Average difference FUORE/EUROSTAT FUA values for demographic indicators



With 668 unique FUA elements, it is not simple to display the fit of the model to the EUROSTAT values in detail. The next graphic shows the relative difference per FUA unit aggregating all years, sex and age groups:

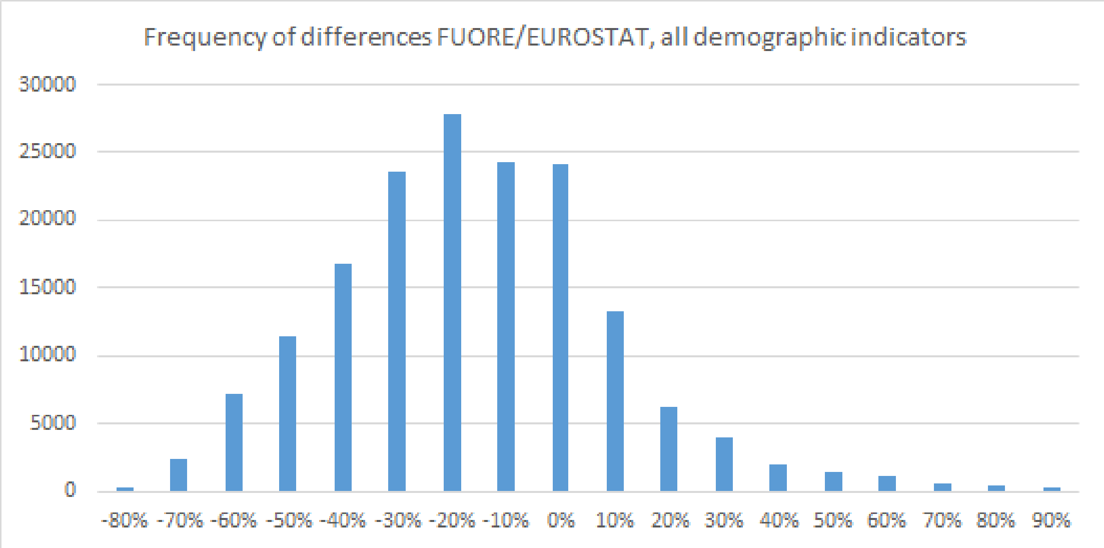
Figure 5.9. Average difference FUORE/EUROSTAT per FUA unit for demographic indicators



Many territorial units display a value around –25%, consistent with the overall smaller amount of total population seen in the aggregated chart. Very few units show a more significant variation.

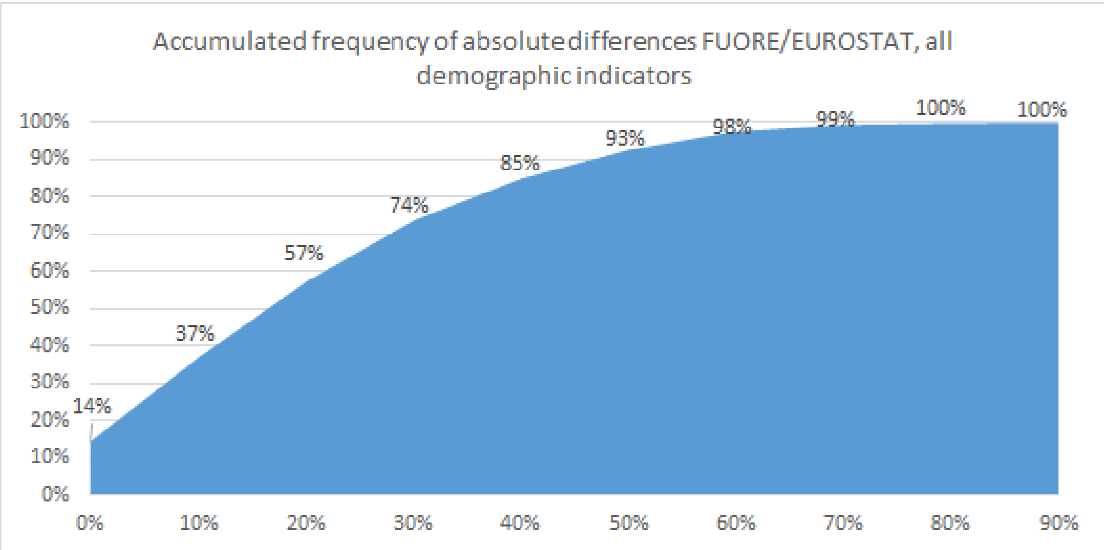
When looking in detail the relative difference for each individual data point, we see again this bias in the total population towards smaller values for the FUORE data but the amount of values that are within an error margin of ±30% is quite significant, indicating that the model produces good results.

Figure 5.10. Frequency of differences FUORE/EUROSTAT in FUAs for demographic indicators



Plotting the absolute relative difference FUORE/EUROSTAT of each value in an accumulated graphic, we can see that 74% of the total 167034 individual data points have a difference of less than 30%, highlighting the good fit of the FUORE model.

Figure 5.11. Accumulated frequency of differences FUORE/EUROSTAT in FUAs for demographic indicators



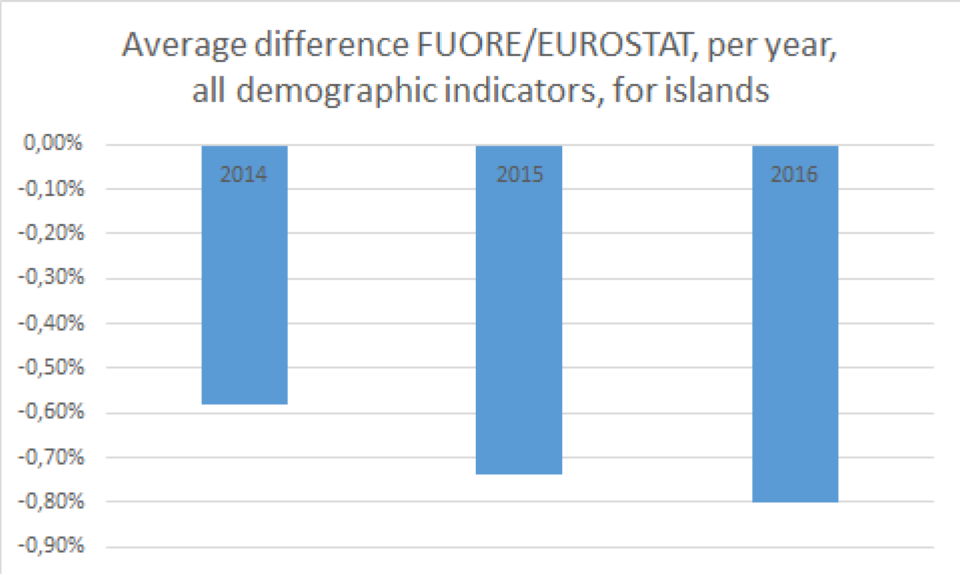
These results could have a better fit if the starting data series would have the same total population amounts as the comparison vectors, but in any case, the methodology produces correct results.

ISLANDS demography

An additional check has been carried out for the island typology, as many islands coincide with NUTS3 level units and data is then directly comparable with basic EUROSTAT demography indicators. In this case then we take the data available in EUROSTAT at NUTS3 level for relevant regions (Cyprus, Greek Aegean islands, Baleares, Canaries, Madeira, Azores, Iceland, Malta and Aland). This means total population by sex and age group for the time period 2014-2016.

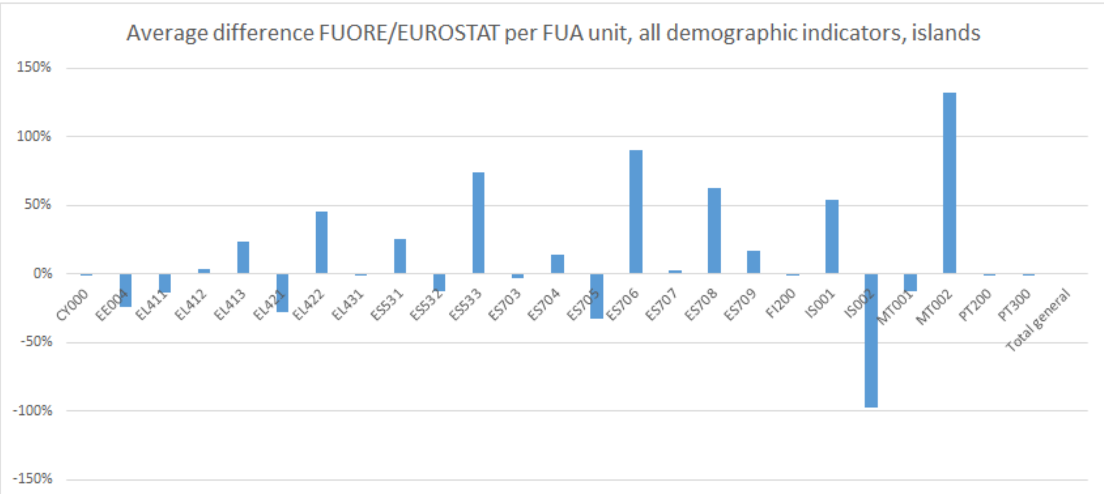
When looking at the differences between the EUROSTAT data and FUORE outputs, aggregating all age groups and sex by year for the selected territorial units we see a very small difference of less than 1%, meaning that the data source for FUORE coincides with the one for EUROSTAT.

Figure 5.12. Average difference FUORE/EUROSTAT Islands values for demographic indicators



The next graphic shows the relative difference per NUTS3 (island) unit aggregating all years, sex and age groups:

Figure 5.13. Average difference FUORE/EUROSTAT per Island unit for demographic indicators

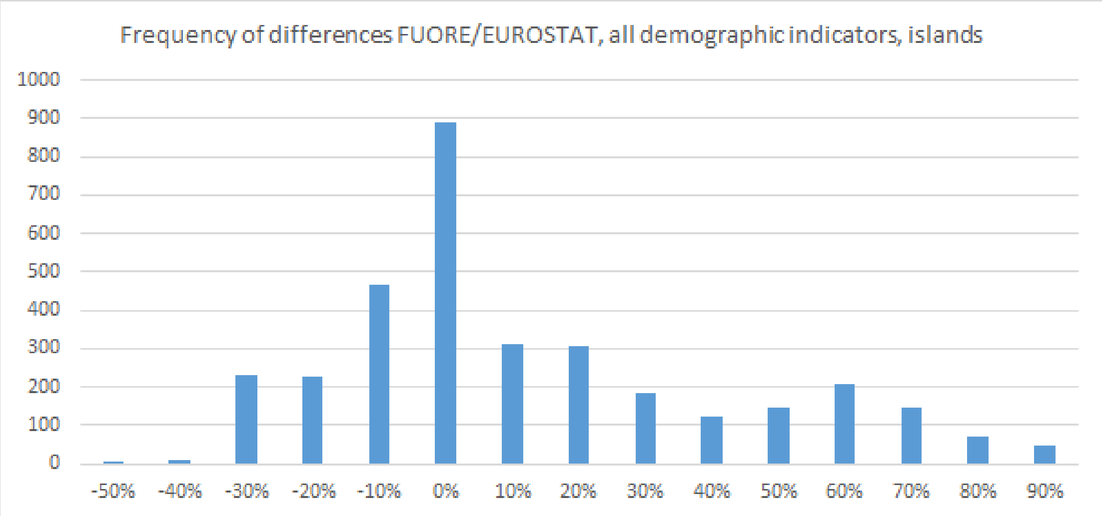


Many territorial units display a value around 0%, thus showing an almost perfect fit to the statistics. However, a few spikes are noticeable but can be explained as follows:

- Iceland and Malta are broken in two NUTS3 regions each, and the FUORE tool distributes the population differently than the statistics.

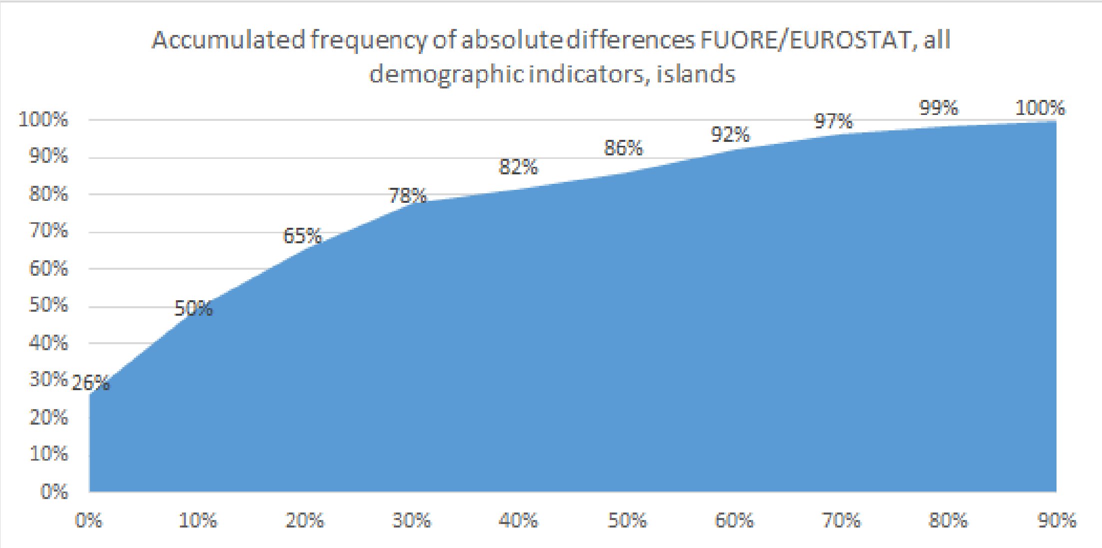
- Baleares and Canaries in Spain have the correct total amount of population, but the tool splits the values across islands differently than the statistics, with some variations of up to 90% for smaller islands.
- In Greek islands there are also some deviations in Aegean islands (EL41, EL42) where FUORE distributes the values with variations of up to 46% for some cases.

Figure 5.14. Frequency of differences FUORE/EUROSTAT in Islands for demographic indicators



Plotting the absolute relative difference FUORE/EUROSTAT of each value in an accumulated graphic, we can see that 78% of the total 3358 individual data points have a difference of less than 30%, highlighting the good fit of the FUORE model.

Figure 5.15. Accumulated frequency of differences FUORE/EUROSTAT in Islands for demographic indicators



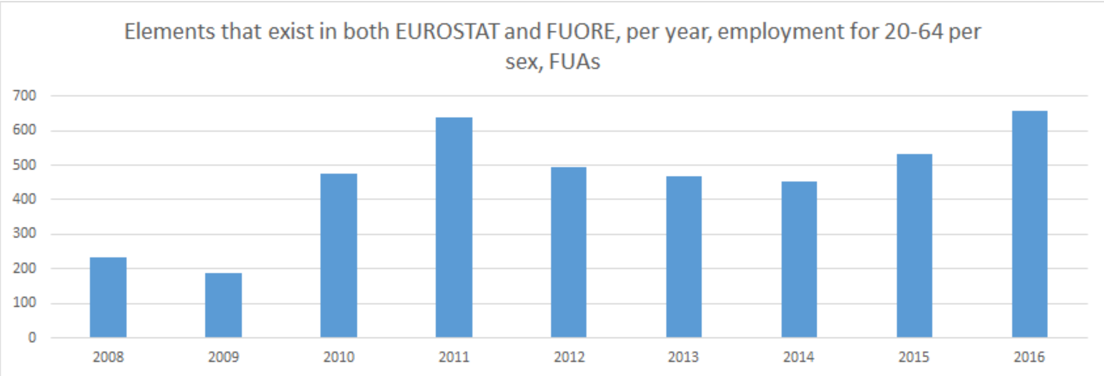
The methodology produces overall good results for islands, but some issues occur at lower detail levels, with important variations in individual islands of smaller extension. This could be improved in the future with a better underlying grid of population for the disaggregation/reaggregation procedure of FUORE.

FUA employment for 20-64 age group by sex

EUROSTAT offers data of number of employed people by sex for the 20-64 age group. These indcators are also available in FUORE and can be checked for the period 2008-2016.

However, EUROSTAT data series are not complete, especially for older years, thus when comparing the original data with the FUORE estimations the amount of data points that can be checked is as follows:

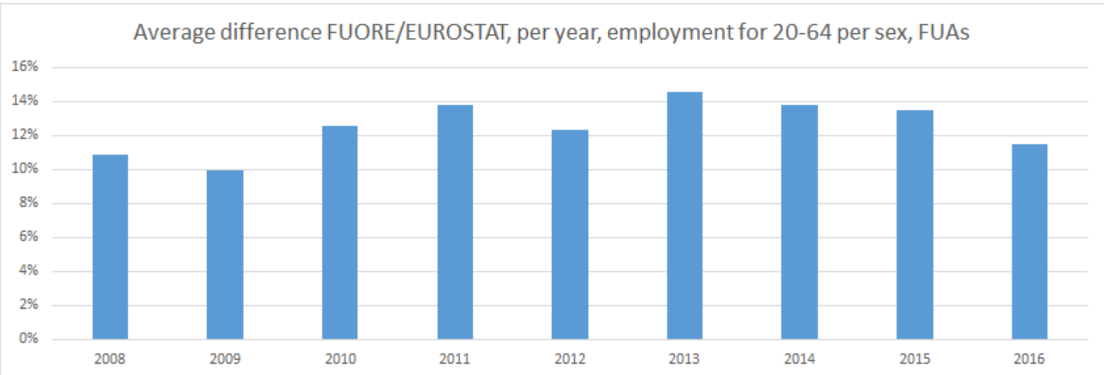
Figure 5.16. Number of data points per year for FUA values existing in EUROSTAT and FUORE for the selected employment indicators.



This highlights the usefulness of the FUORE methodology, which will allow filling in all the gaps from 2008-2016 period with estimations of number of employed people by sex at FUA level.

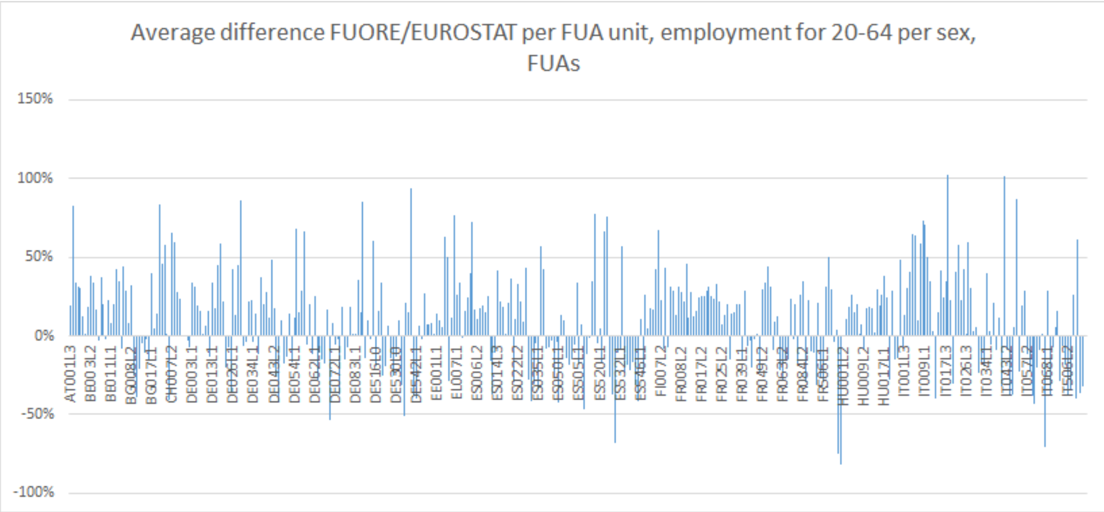
When looking at the differences between the EUROSTAT data and FUORE outputs, aggregating sex by year, it is evident that there is a consistently higher amount of employed people in the FUORE data, showing that the inputs are different. However, this does not invalidate the methodology, simply highlights the need of making further analysis to determine the difference in total amount employment.

Figure 5.17. Average difference FUORE/EUROSTAT FUA values for employment indicators



With 668 unique FUA elements, it is not simple to display the fit of the model to the EUROSTAT values in detail. The next graphic shows the relative difference per FUA unit aggregating all years and sex:

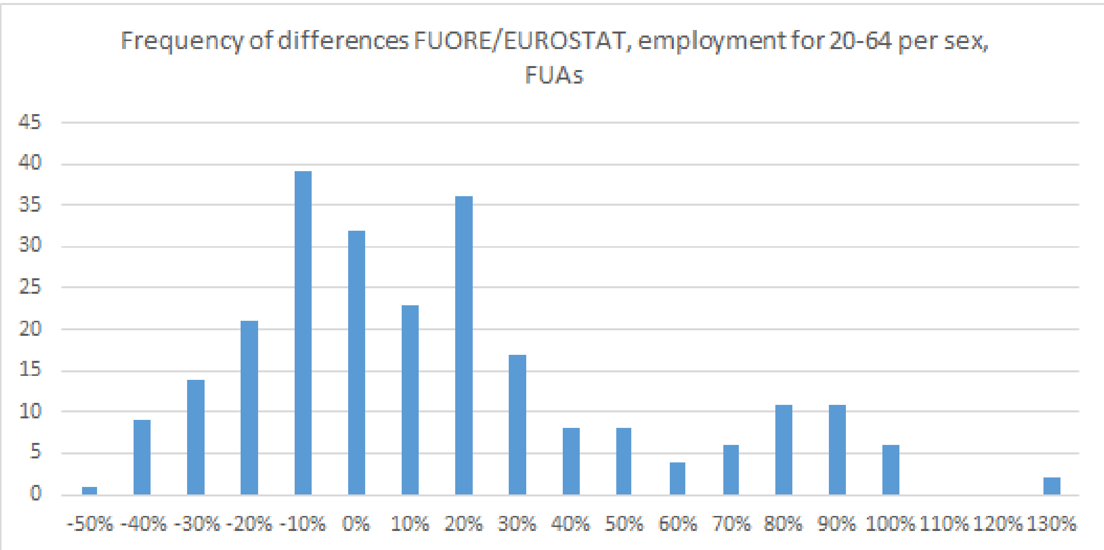
Figure 5.18. Average difference FUORE/EUROSTAT per FUA unit for employment indicators



Many territorial units display a value around +20%, consistent with the overall higher amount of employment seen in the aggregated chart. Several units show a more significant variation.

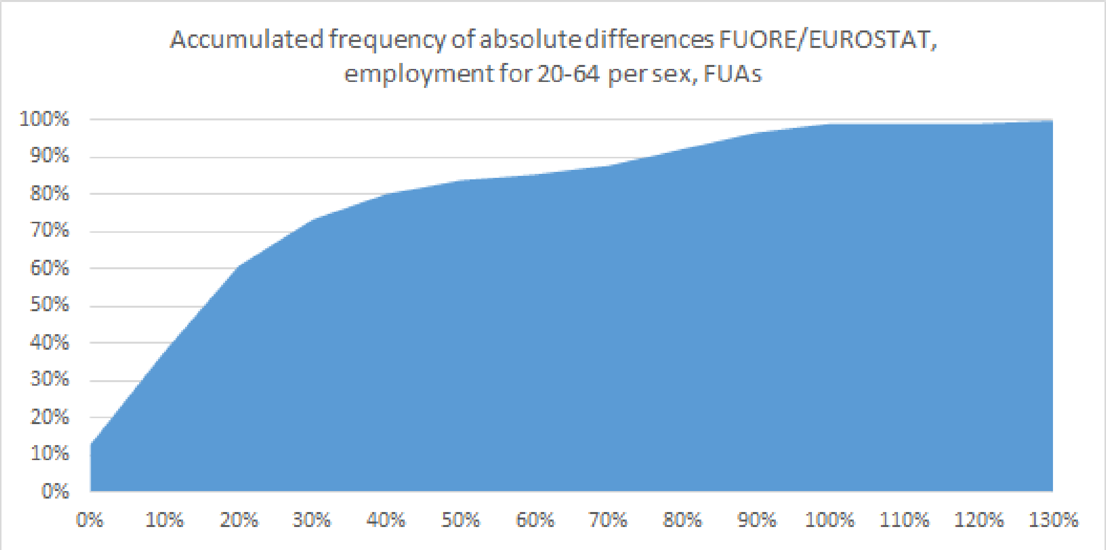
When looking in detail the relative difference for each individual data point, we see again this bias in the employment figures towards higher values for the FUORE data but the amount of values that are within an error margin of ±30% is quite significant, indicating that the model produces good results.

Figure 5.19. Frequency of differences FUORE/EUROSTAT in FUAs for employment indicators



Plotting the absolute relative difference FUORE/EUROSTAT of each value in an accumulated graphic, we can see that 73% of the total 248 individual data points have a difference of less than 30%, highlighting the good fit of the FUORE model.

Figure 5.20. Accumulated frequency of differences FUORE/EUROSTAT in FUAs for employment indicators

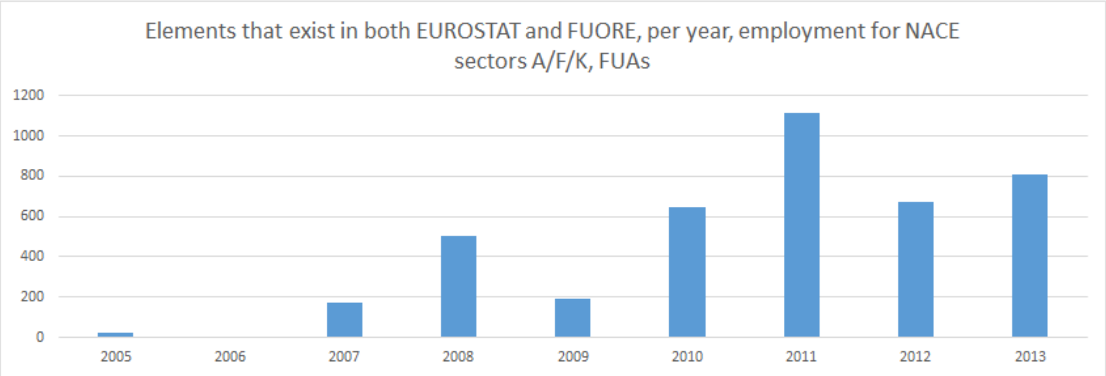


FUA employment for NACE groups A, F, K

EUROSTAT offers data of number of employed people by NACE activity. These indicators are also available in FUORE for categories A, F, and K and can be checked for the period 2005-2013.

However, EUROSTAT data series are not complete, especially for older years, thus when comparing the original data with the FUORE estimations the amount of data points that can be checked is as follows:

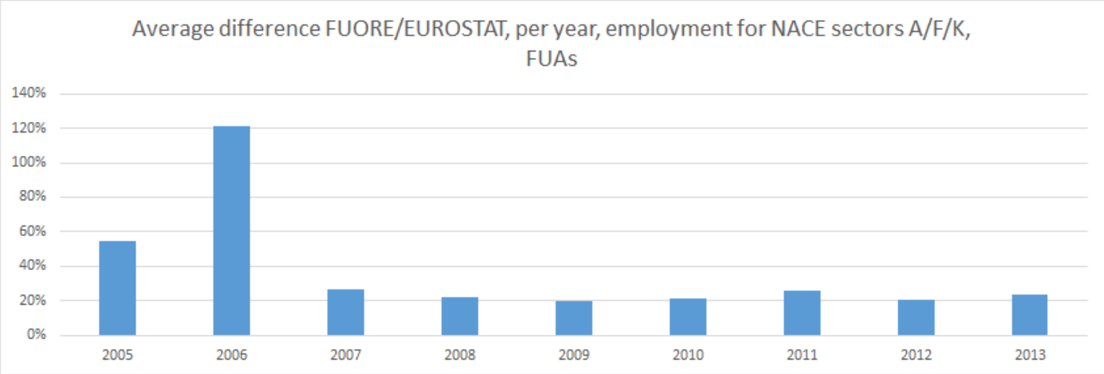
Figure 5.21. Number of data points per year for FUA values existing in EUROSTAT and FUORE for the selected employment indicators by sector.



This highlights the usefulness of the FUORE methodology, which will allow filling in all the gaps from 2005-2013 period with estimations of number of employed people by sector at FUA level.

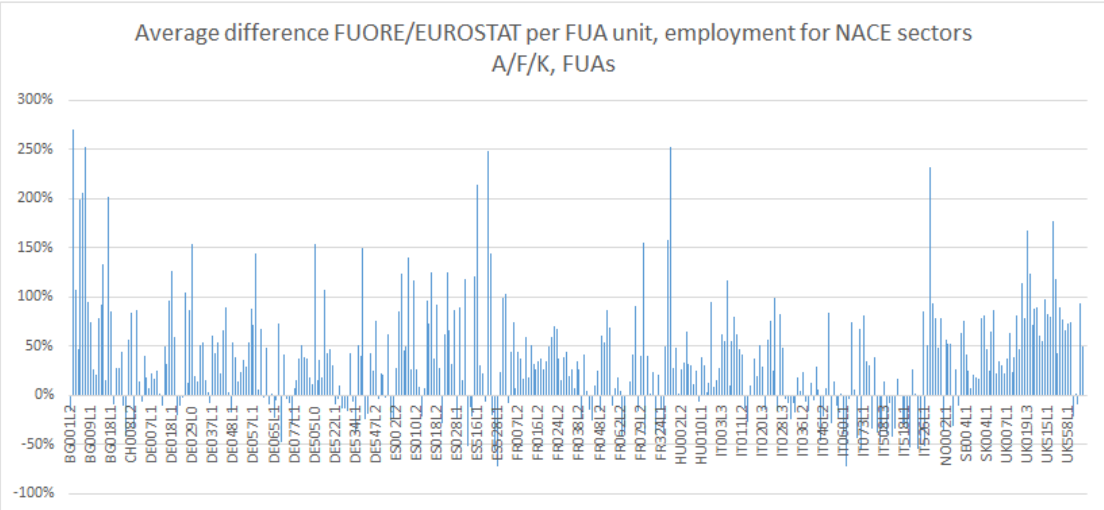
When looking at the differences between the EUROSTAT data and FUORE outputs, aggregating sex by year, it is evident that there is a consistently higher amount of employed people in the FUORE data (around +20%), showing that the inputs are different. However, this does not invalidate the methodology, simply highlights the need of making further analysis to determine the difference in total amount employment.

Figure 5.22. Average difference FUORE/EUROSTAT FUA values for employment indicators per sector



With 668 unique FUA elements, it is not simple to display the fit of the model to the EUROSTAT values in detail. The next graphic shows the relative difference per FUA unit aggregating all years and NACE categories:

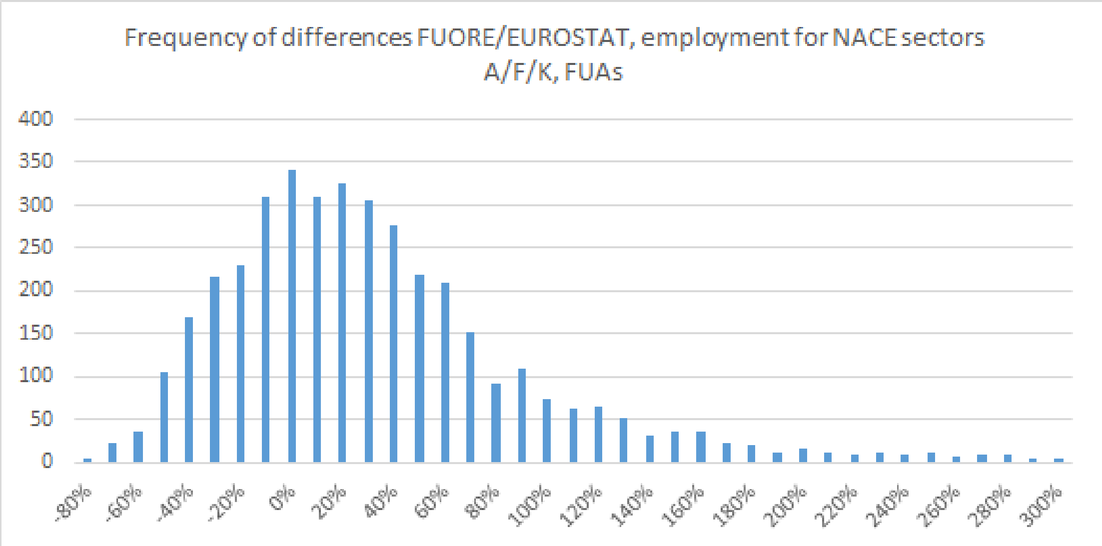
Figure 5.23. Average difference FUORE/EUROSTAT per FUA unit for employment indicators per sector



Many territorial units display a value around +20%, consistent with the overall higher amount of employment seen in the aggregated chart. Several units show a more significant variation, indicating that in this case the fit of the FUORE model to the EUROSTAT statistics is not so good.

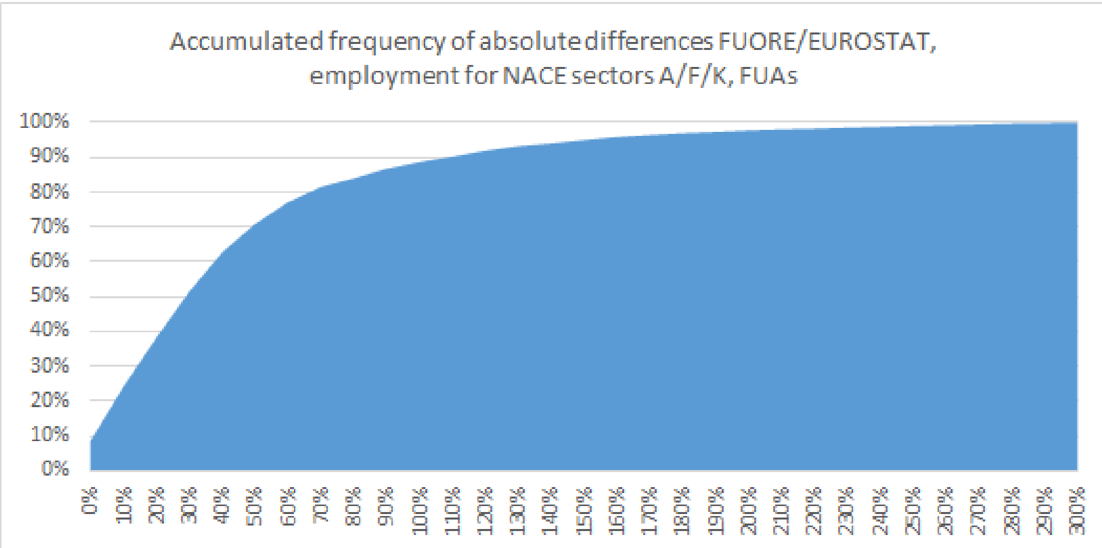
When looking in detail the relative difference for each individual data point, we see again this bias in the total population towards higher values for the FUORE data but also more spread in the amount of errors, indicating a worse fit to the statistics.

Figure 5.24. Frequency of differences FUORE/EUROSTAT in FUAs for employment indicators per sector



Plotting the absolute relative difference FUORE/EUROSTAT of each value in an accumulated graphic, we can see that 71% of the total 3957 individual data points have a difference of less than 50%, highlighting a higher degree of dispersion in the FUORE model as compared to the predictions of other data sets.

Figure 5.25. Accumulated frequency of differences FUORE/EUROSTAT in FUAs for employment indicators per sector



5.1.2.2 Goodness of fit with National Statistical Offices data

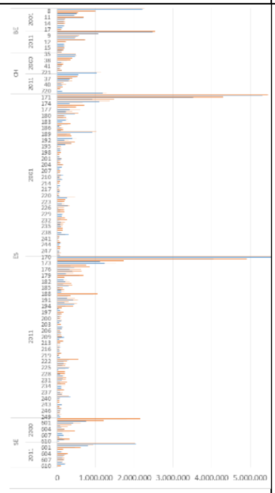
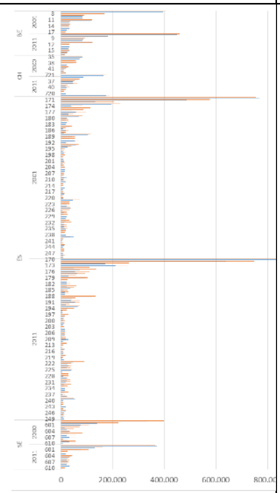
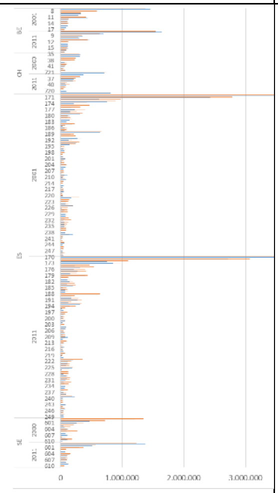
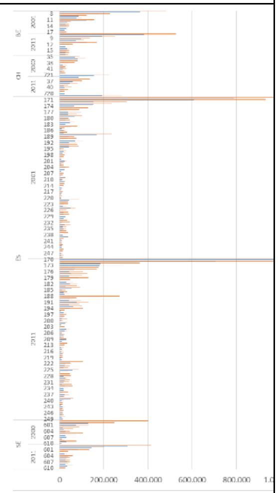
For the preparation of this analysis, the following steps have been carried out:

- Collection of statistical information on population at LAU level from different National Statistical Offices:
    - Data for BE, CH, ES, SE, in order to cover geographical and dimensional heterogeneity.
    - Population segregated by 5-year age groups for 2000/2001 and 2011

- Aggregation of LAU statistics at the level of the functional typologies of the ESPON FUORE project (bda1, bda34, fua, gi, isl, msa, mtn, spa, tcoa)
    - Generation of a centroid of each LAU polygon
    - Spatial join of the LAU centroids against each functional typology to determine the association of each centroid to the different functional units they are included in
- Comparison of the aggregated LAU values by functional unit with the outputs of SMD

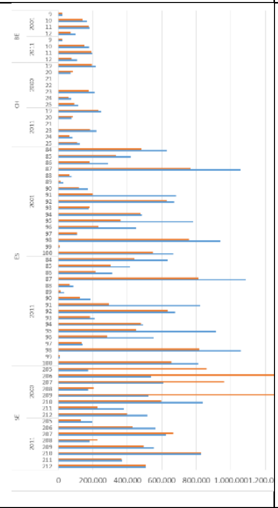
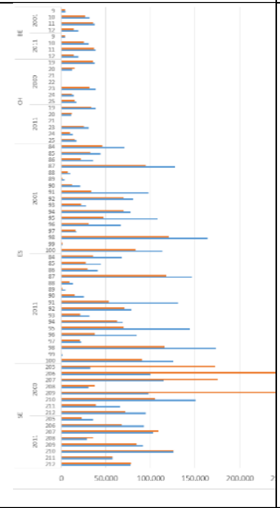
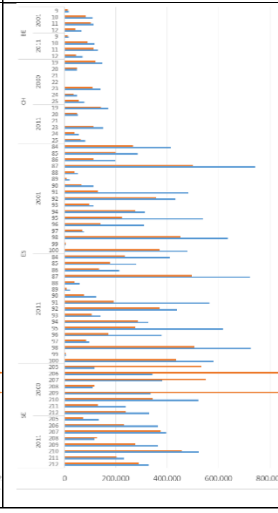
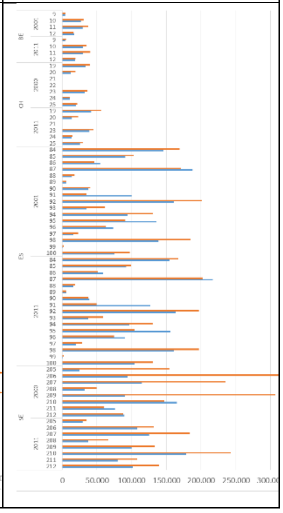
The following graphics show the results for the four more aggregated indicators of population, that is Total_pop, pop <15 years, pop from 15 to 64 years, pop > 65 years with statistical values derived from LAU data in blue and FUORE model values in orange

Figure 5.26. Comparison of FUORE indicators (orange) and LAU aggregations (blue) for FUAs.

fua Total_pop	fua <15 years	fua 15 to 64 years	fua >65 years
			

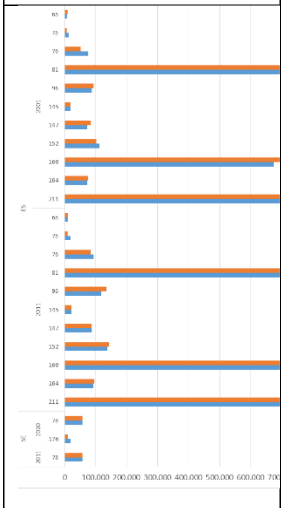
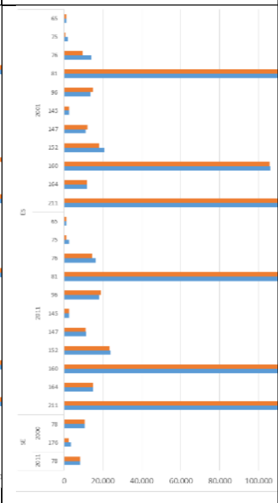
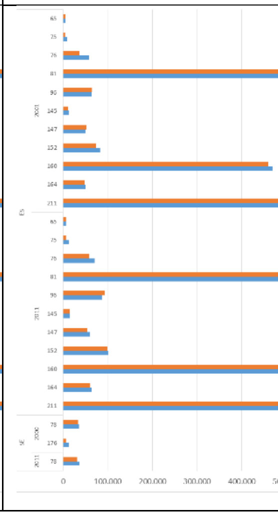
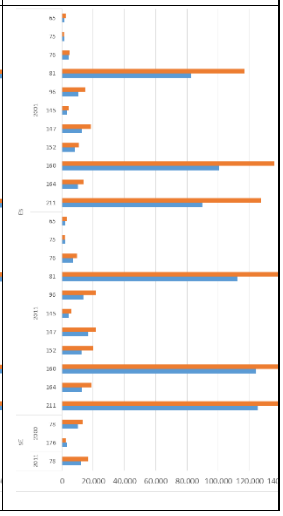
For fua most of the territorial units have deviations of less than 15% compared to the real statistical value, so in general the FUORE model is working fine in this case. There are a few outliers though with deviations over 50, due to the different way in which the statistical data and the FUORE calculation are computed. This can cause to leave outside a given FUA or to include within it certain LAU elements with bigger population concentration, resulting in bigger differences between the two methods.

Figure 5.27. Comparison of FUORE indicators (orange) and LAU aggregations (blue) for GI.

gi Total_pop	gi <15 years	gi 15 to 64 years	gi >65 years
			

Regarding the Green Infrastructure (gi) regions, the results are very good in the case of Belgium and Switzerland with differences under 15%. Figures in Spain and Sweden deviate more in some cases. For Spain the main reason is the shape of GI areas, with jagged edges and many holes, that make the match with the LAU difficult to replicate in the FUORE calculation. For Sweden, the differences are explained because of the shape of the GI areas, which is also a bit jagged and with holes in the south of the country, whereas in the northern GI areas, which are smoother, the values of the methodology match much better those of the statistical calculation.

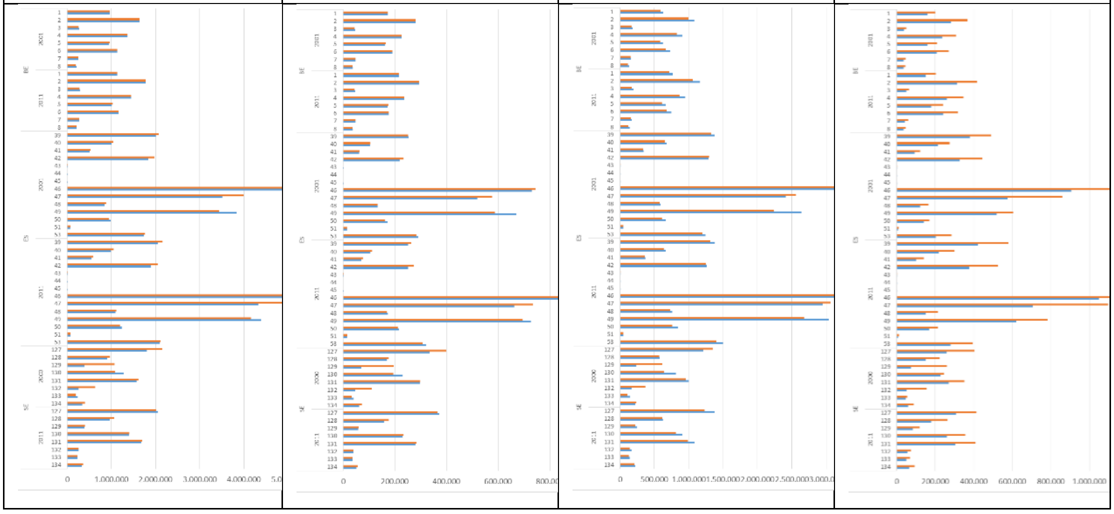
Figure 5.28. Comparison of FUORE indicators (orange) and LAU aggregations (blue) for ISL.

isl Total_pop	isl <15 years	isl 15 to 64 years	isl >65 years
			

The Island typology presents very good results, with almost all units showing differences of less than 5%

Figure 5.29. Comparison of FUORE indicators (orange) and LAU aggregations (blue) for MSA.

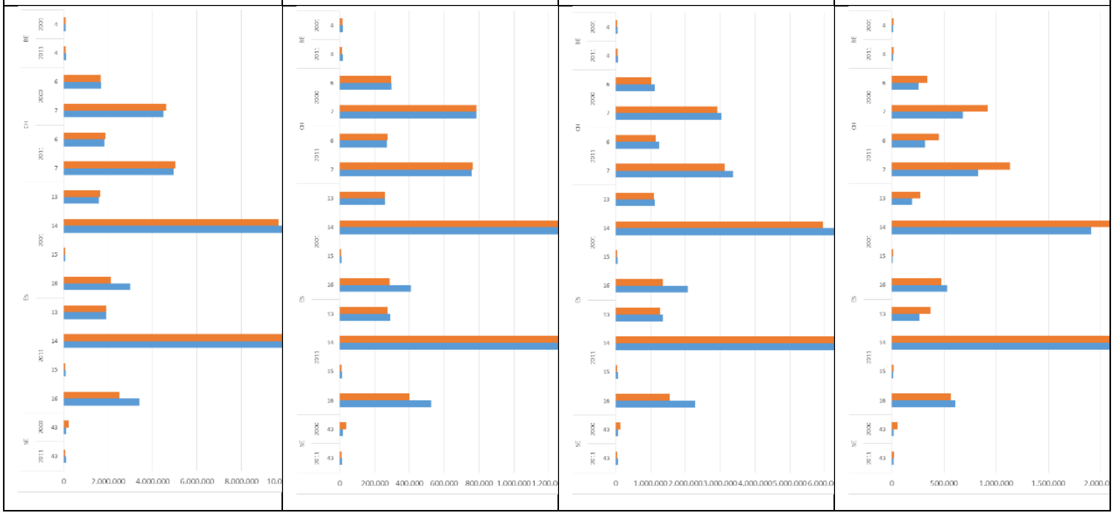
msa Total_pop	msa <15 years	msa 15 to 64 years	msa >65 years



The MSA coastal typology shows almost a perfect fit with the LAU statistics. The FUORE calculation yields in general higher values of population for the >65 class due to differences on the total volume of the datasets being compared, but the relative differences across units are maintained, and thus the fit is very good.

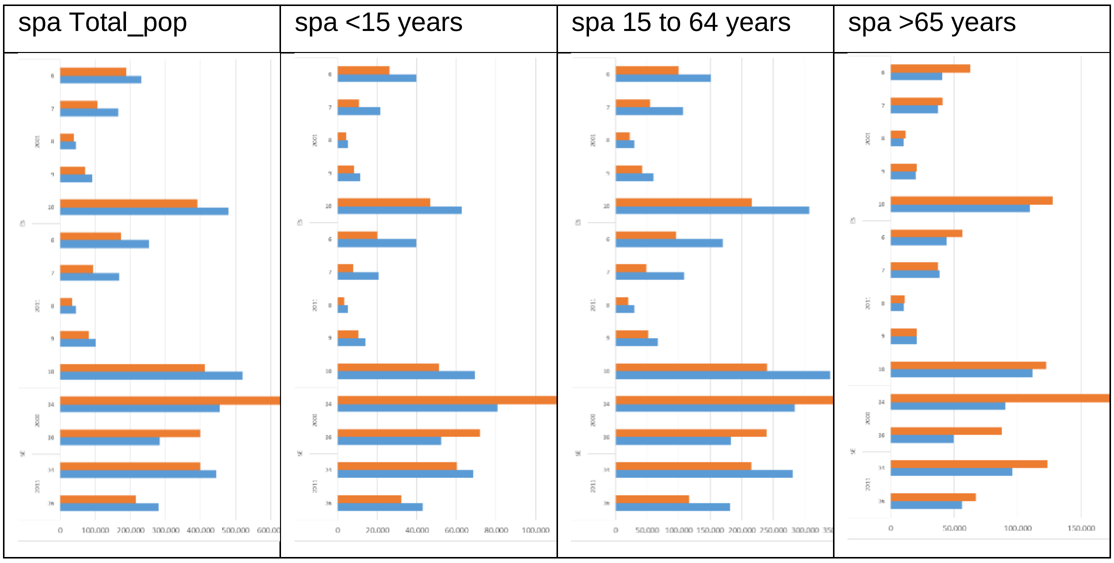
Figure 5.30. Comparison of FUORE indicators (orange) and LAU aggregations (blue) for MTN.

mtn Total_pop	mtn <15 years	mtn 15 to 64 years	mtn >65 years



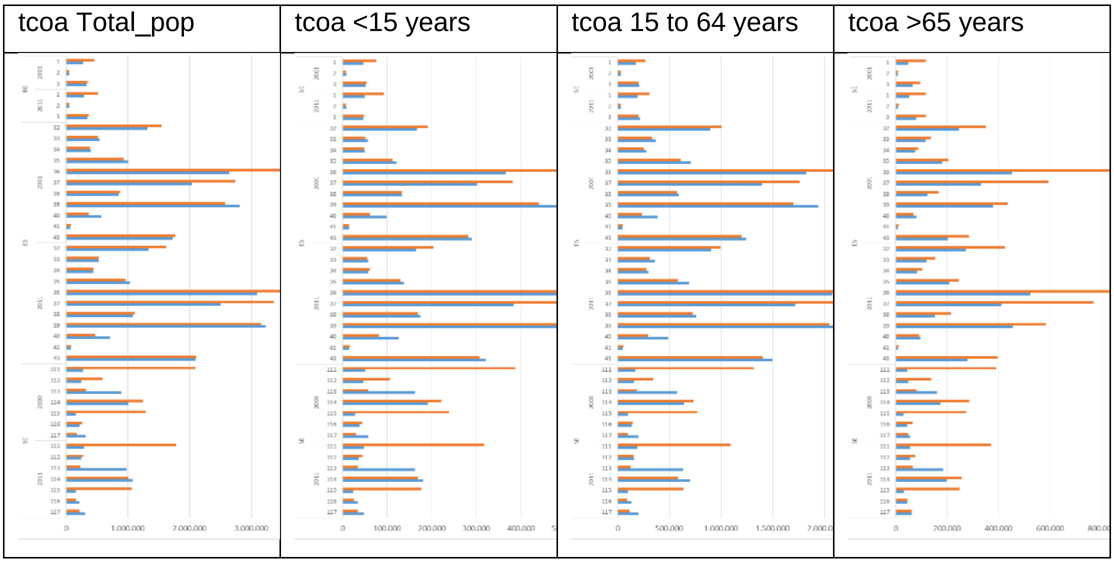
The MTN typology shows a very good fit for all units except for Spain, where there are deviations of up to 40% caused again by the shape of the territorial units, with lots of jagged edges that cause differences in the process of automatic inclusion/exclusion of LAU units into the aggregated values when calculating the statistics.

Figure 5.31. Comparison of FUORE indicators (orange) and LAU aggregations (blue) for SPA.



On Sparsely Populated Areas (SPA) results show average deviations of 20% that can be explained by the total volume differences in the datasets being compared.

Figure 5.32. Comparison of FUORE indicators (orange) and LAU aggregations (blue) for TCOA.



As for TERCET coasts (tcoa) we find several units with a good fit (deviations under 5%) but there are several values with deviations exceeding more than 50% in Sweden caused by a transference of values among neighbouring regions. Besides this, we can see again the general overestimation of the >65 class caused by the global difference in volumes between the datasets being compared.

Conclusion

In general, the FUORE estimations for the demographic indicators are quite good, but it is usually the case that, when a territorial typology unit involves many LAU, then the predicted value tends to deviate much more from the LAU data aggregation. The explanation is the fact that we are using LAU data for validation and the source data used in the FUORE methodology is NUTS3 based. Also, the disaggregation/aggregation procedure for the FUORE methodology and the calculation of statistical data from LAU values yields differences when the territorial units involved are not very smooth in shape, as this causes some LAU centroids to be or not included within a certain territorial unit depending on how detailed in scale is the definition of such unit. It is worth mentioning that some deviation is always expected, as the FUORE methodology is an estimation tool not aimed at replacing official statistics when they exist, but being able to estimate indicators at functional areas level, whenever they do not exist from official sources.

When comparing demography indicators with EUROSTAT data at FUA level, FUORE produces values that are very close to the statistics in terms of distribution among units even if the totals are not the same because of the way FUORE creates the FUA values. For the same indicators but looking at Islands typology, the total values in FUORE and EUROSTAT coincide but FUORE distributes them a bit differently among individual islands in the case of archipelagos. The result though can still be considered good, as the deviation is not so big.

For employment indicators at FUA level, the FUORE data yields very good results when looking at total employment per sex, although the total aggregated values are generally higher than in the EUROSTAT series. When looking at employment per sector the results are not as good and differences with the EUROSTAT statistics are bit more noticeable.

In summary though, FUORE produces results that are good enough for its purpose and will provide a sound and consistent method to generate statistical information for different territorial units when no disaggregated data exists.

5.2 Expert validation

The expert validation workshop had to occur at a meeting at ESPON EGTC premises on March 25, at which representatives of Eurostat, DG-REGIO or the JRC and other experts on data collection were to attend.

However, due to the ongoing COVID-19 crisis in Europe at the time, the workshop had to be cancelled.

The objective of that activity was presenting the web tool, the methodology and a few outcomes of the tests performed during the validation phase. Next, there would have been a discussion with the participants to assess the goodness of fit of results with official statistics, to discuss the functionalities of the web tool, and to debate challenging results to increase the level of understanding of the implemented techniques.

Finally, ESPON EGTC has replace the workshop by a written procedure. Feedback from the different stakeholders is already integrated in this report and the rest of FUORE outcomes.

6 The FUORE web tool

This chapter provides a description of the FUORE web tool, which is one of the main outcomes of the project. It serves for interactive and user-friendly online presentation of the project results. It provides many interactive features like map, charts, filter, query and various exporting tools.

The data presented in the web tool are estimated indicators by the nine selected functional regions in Europe (see section 3.4). Therefore, the figures shown by the FUORE web tool do not replace official statistics whenever they exist, but they provide an added value for indicators not reported at a functional regional level. Additionally, the embedded web data analysis toolbox allows for the integration of ad hoc indicators in the web tool (for advanced users).

The following sections detail how the web tool has been implemented and which components and functionalities offers to the users.

6.1 Technological framework and deployment

6.1.1 Web tool fully based on open source components

The web tool exploits technologies based on open source components. The internal component (Backend) of the application is written in JavaScript. It communicates with client applications through REST API. The visualisation of geo-data layers in the web map is realized through GeoServer and datasets handled and displayed in the tool are stored in a PostgreSQL database. The tool is optimized for the newest versions of Google Chrome and Mozilla Firefox.

6.1.2 Deployment of the tool to the ESPON infrastructure

During most of the duration of the project, the tool has been temporarily deployed on GISAT´s internal servers’ infrastructure. The exploitation of the open source components-based technology should ensure that no additional licences will be necessary when the tool is eventually deployed to the ESPON servers. This internal version of the tool has undertaken continuous development and testing process, and it should be relatively easily transferred to ESPON servers/premises. This is ensured by the compatibility of the server environments between ESPON and GISAT. A communication between GISAT and ESPON IT experts has been done and it has been agreed that the web tool will be deployed in one Docker container, which needs to be installed on ESPON servers. The requirements on the infrastructure were provided by GISAT to ESPON IT experts (are also specified in the Technical report), wo they can prepare appropriate infrastructure. The next step to be undertaken after the final version of the web tool is approved by the ESPON EGTC, is to deploy the tool on ESPON servers. This will be done remotely by GISAT´s IT experts. Also, GISAT will provide all relevant instructions and guidance materials regarding the technical specification and operation of the tool and management of its data content to the ESPON experts.

6.2 The landing page

The integral part of the web tool is the landing, or introductory, page. This page represents the first interface to the end-user and therefore provides the user with a short information about the web tool. This landing page should be made accessible directly from the ESPON toolbox.

For the landing page, as well as for the web tool itself, the service provider has implemented an up-to-date UX design to make the application attractive for the users. However, it still respects the ESPON corporate identity and contains the logo, colours etc., to become part of the ESPON Toolbox. The landing page of the web tool is available at: fuore.eu**.**

Besides the provision of a short information about the tool and its purpose, the most important function of this landing page is to provide the user with easy access to the web tool itself. This access is enabled through multiple interactive tiles - each of them representing one of the reporting units integrated into and presented by the tool (the nine functional regions). For each tile, an illustrative figure is provided, showing the delineation of that specific type of the functional region in a map view. Once the user clicks on the tile dedicated to that one particular type of functional region, he is re-directed to a page providing access to all indicators which are linked (and available) for that specific type of region (see Figure 6.1. Final version of the landing page of the web tool (available at fuore.eu)).

Figure 6.1. Final version of the landing page of the web tool (available at fuore.eu)
Source: ESPON FUORE web tool

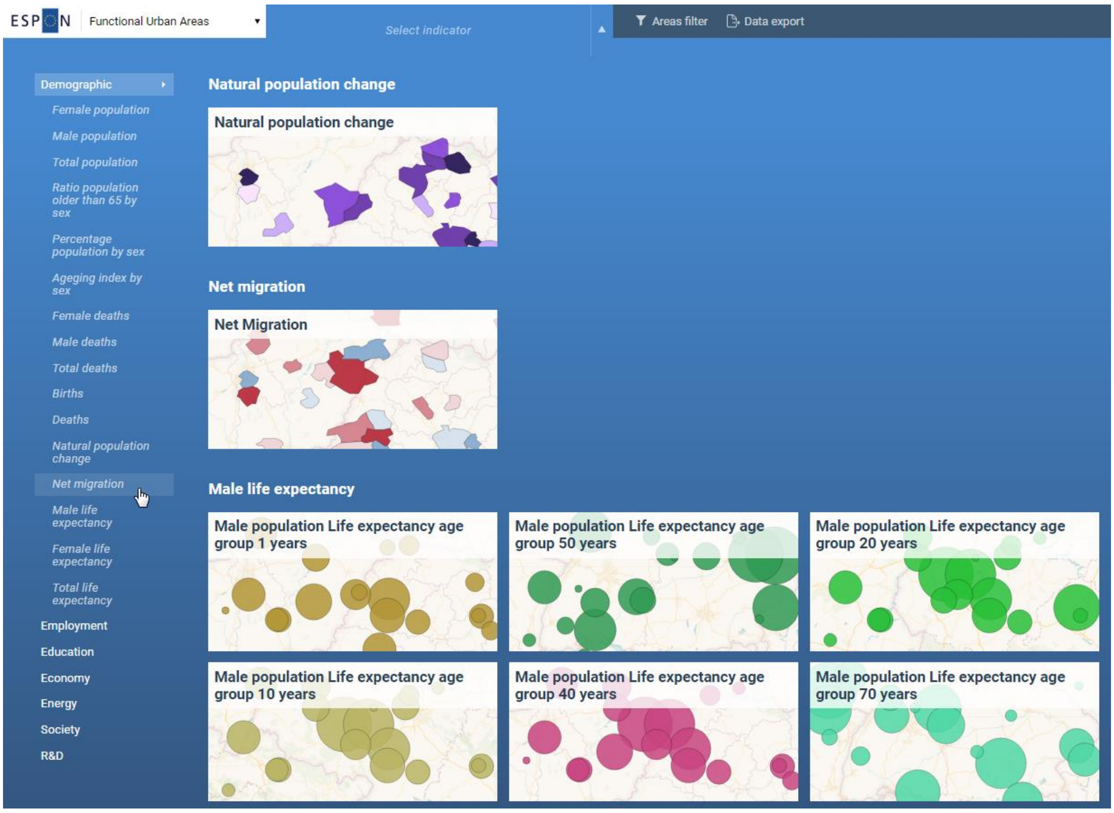
The landing page also provides the access to the:

- official project information (project web page on ESPON programme portal)
- methodology of the disaggregation process
- description of delineation methodology for each type of functional region
- GUI of the Web Data Analysis Toolbox (“Custom disaggregation tool” hyperlink), which allows the user to disaggregate their own indicators and visualise the results on the functional regions´ level in the web tool.
- YouTube channel with guidance videos on how to use the tool in praxis
- Final report of the project

6.3 Access to the data and information in the tool

For each type of functional region, a specific **set of indicators** is integrated and presented in the web tool by illustrative tiles (Figure 6.2). This complete set of indicators is divided into few thematic groups **- categories**, and **sub-categories** so the user will start with the **selection of the category or sub-category** first and, after that, he/she will select the **indicator**.

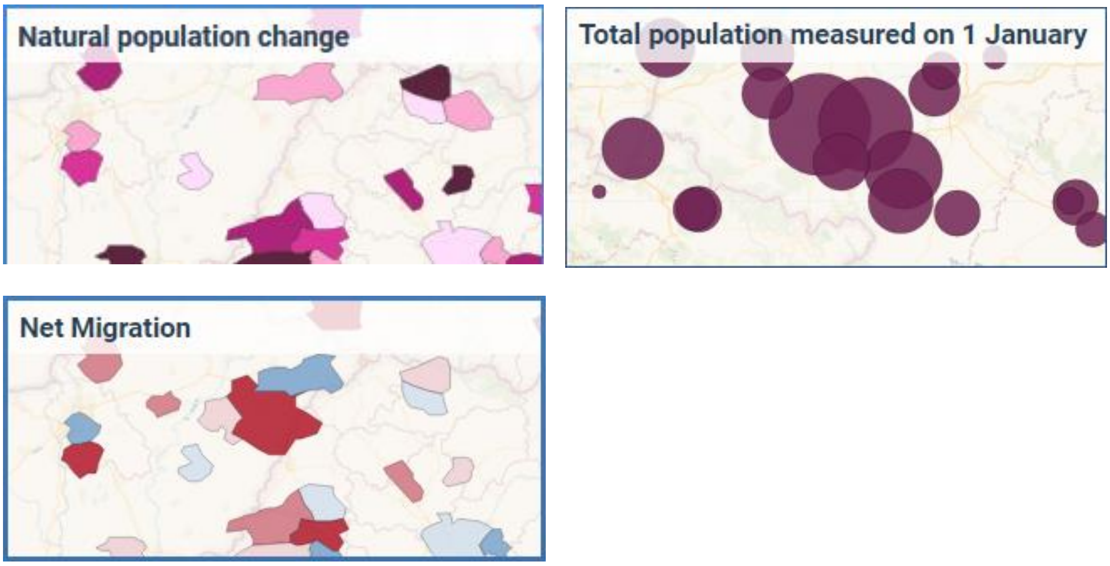
Figure 6.2. Access to data and information presented in the web tool – for each type of functional region, a set of indicators is integrated and presented, divided into thematic categories and sub-categories.



Source: ESPON FUORE web tool

The various indicator types – stock or ratio - are automatically reflected in figures on the corresponding tiles. Following the basic cartography rules, ratio figures are shown as choropleth maps, whereas, for stock values, a carthodiagram is used (see Figure 6.3. *Different* *types of figures on the tiles representing ratio and stock indicators*). For that specific case when the data series for an indicator contains both positive and negative values, diverging colour scheme is applied – with warm colours (red, dark brown representing positive values) and cold colour (blue or green) representing negative values. The latest, and also its colour, is automatically generated, based on the input metadata. The same type of map visualisation and colour shown on the tiles is also used in the Map View of the tool.

Figure 6.3. Different types of figures on the tiles representing ratio and stock indicators

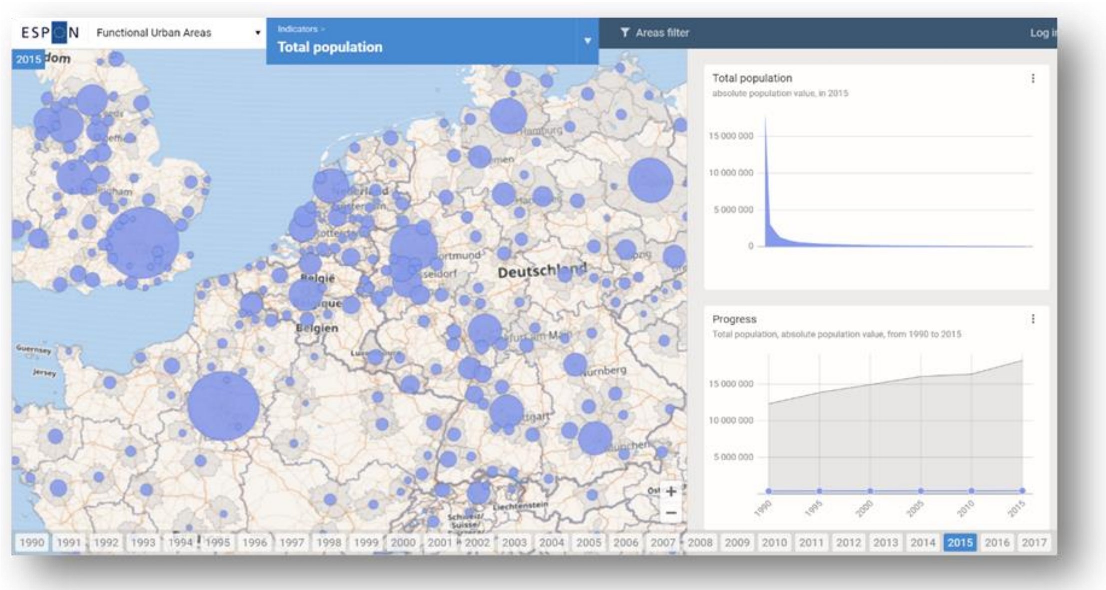


Source: ESPON FUORE web tool

6.4 Web tool Map View

Once the user selects the indicator of his/her interest (by clicking on the respective tile), the Map View of the web tool opens (Figure 6.4. Typical Map View of the web tool – a pre-defined visualisation presenting the selected indicator for the selected type of functional regions, with interactive map, charts and time axis.), with a pre-defined visualisation dedicated to the selected indicator. This Map View contains the **interactive map window**, showing a **thematic map** and a **panel with interactive charts** presenting the values of selected indicator for the reporting units. At the bottom of this Map View, an **interactive time axis** is available, enabling the selection of the year or multiple years of interest, for which the values are displayed in the map and charts.

Figure 6.4. Typical Map View of the web tool – a pre-defined visualisation presenting the selected indicator for the selected type of functional regions, with interactive map, charts and time axis.



Source: ESPON FUORE web tool

Each Map View is dedicated to one (selected) indicator and presents its values in: interactive map window; column chart (status values for selected year) and line chart (showing development of indicator values in the time-series). Once inside such Map View, dedicated to one selected indicator, the user can always **change the indicator**, thematic **category or sub-category** or even **type of functional region** of his interest, through the interactive menu at the top of the Map View.

6.4.1 Interactive components and functionalities of the web tool

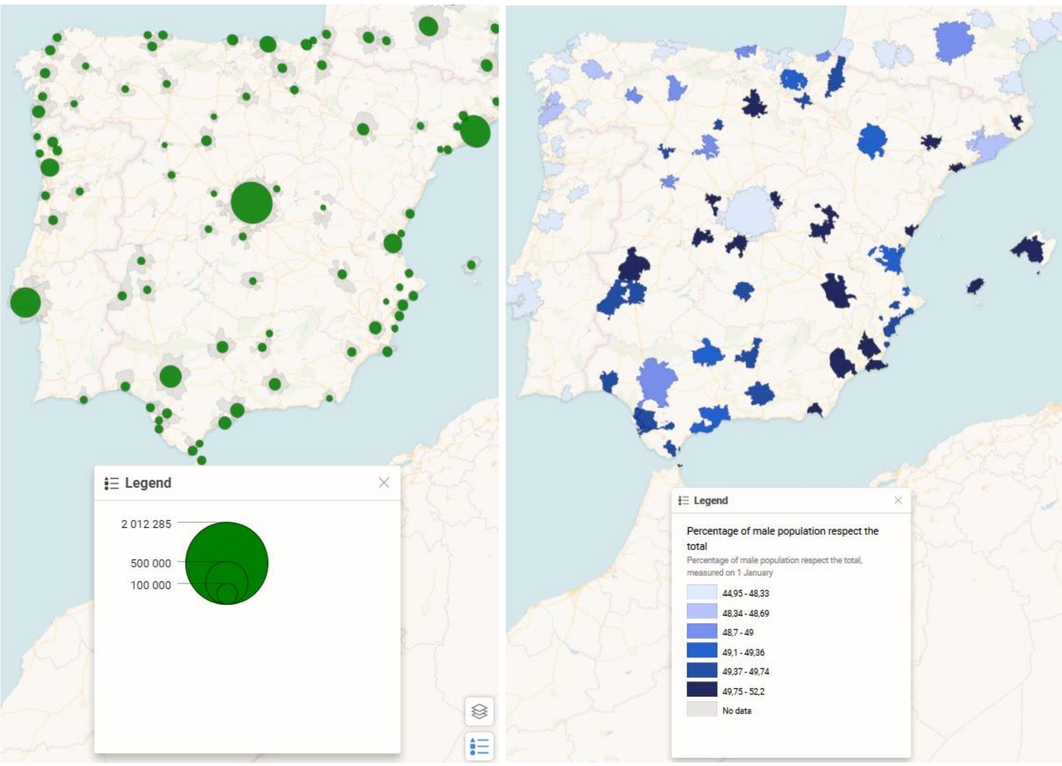
The web tool provides few interactive components and functionalities which are crucial to enable the user effective analysis of the integrated data and information. The following types of interactive functionalities and components are included:

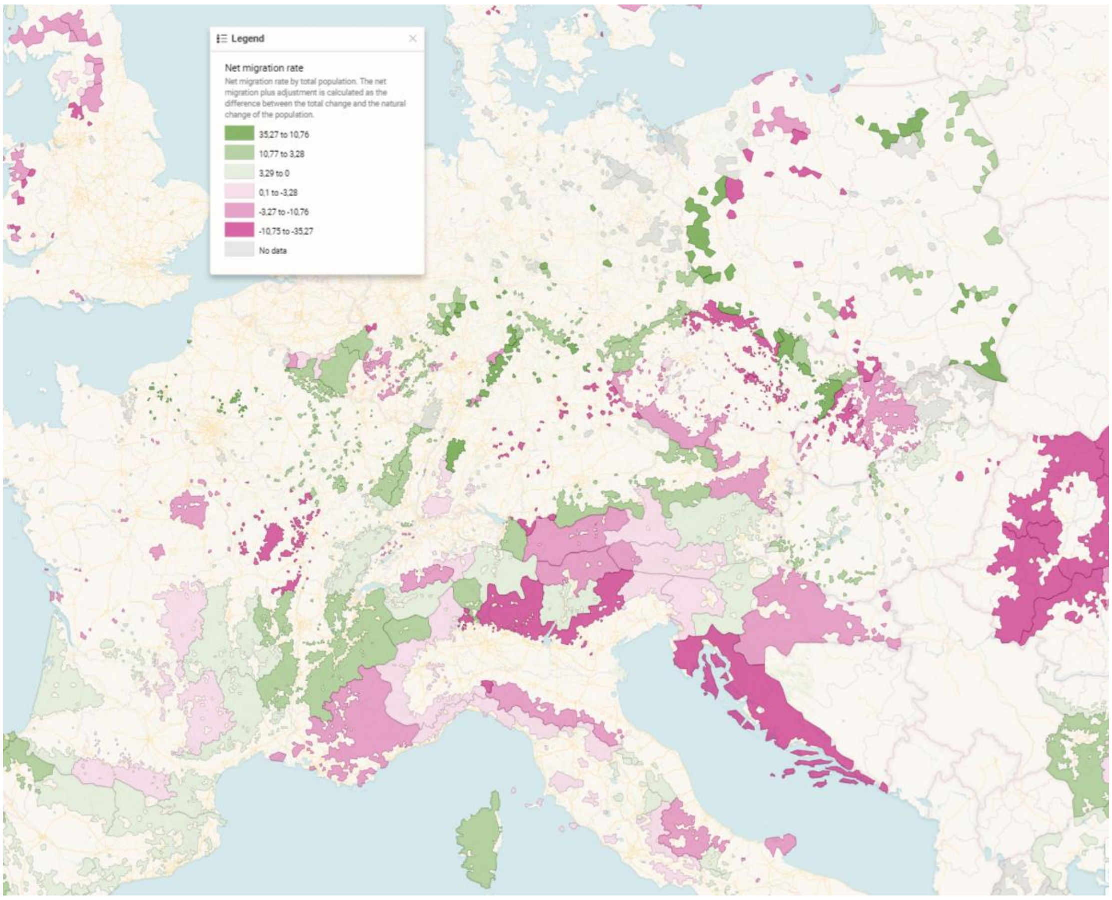
Interactive map window

Considering the spatial character of the data to be presented, the interactive map window is a very important component of the web tool. In the map, data layers stored on Geoserver are displayed through WMS service. A generalized version of the geometries is used for the purpose of visualisation of analytical units in the map, which rapidly improves the performance (speed) of the map layers rendering. The user can interactively **zoom in** and **out** in the map or **pan the map** in any direction. There are **two ways in which the indicators can be displayed in the map**, depending on the type of the indicator (see Figure 6.5. Two types of cartographic visualisation in the map – carthodiagram for stocks (left)/choropleth map for ratios (right), diverging colour scheme for combination of positive and negative values, all with corresponding legends.). For the indicators with absolute values (stocks), a **carthodiagram** is available. For the indicators with the relative values showing ratios (%), a **choropleth** map is used. **Diverging**

colour scheme is applied in case an indicator shows both positive and negative values. For each type of map visualisation, a widget showing the legend is available, which can be opened using a “legend” button in the lower right corner of the map window.

Figure 6.5. Two types of cartographic visualisation in the map – carthodiagram for stocks (left)/choropleth map for ratios (right), diverging colour scheme for combination of positive and negative values, all with corresponding legends.

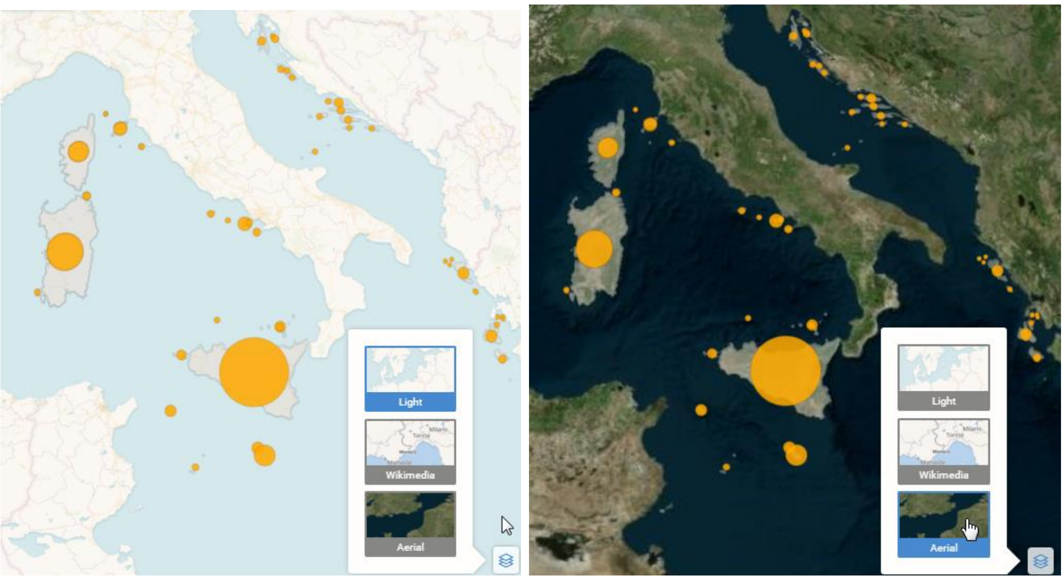




Source: ESPON FUORE web tool

For the background map in the Map Window, a switcher is available (a button in the lower right corner of the map window. As shown by Figure 6.6. *Various types of background map can be switched on in the map window*, this allows to switch between different types of background maps, including various open-source maps (like Wikimedia or CARTO) or aerial map (Bing).

Figure 6.6. Various types of background map can be switched on in the map window

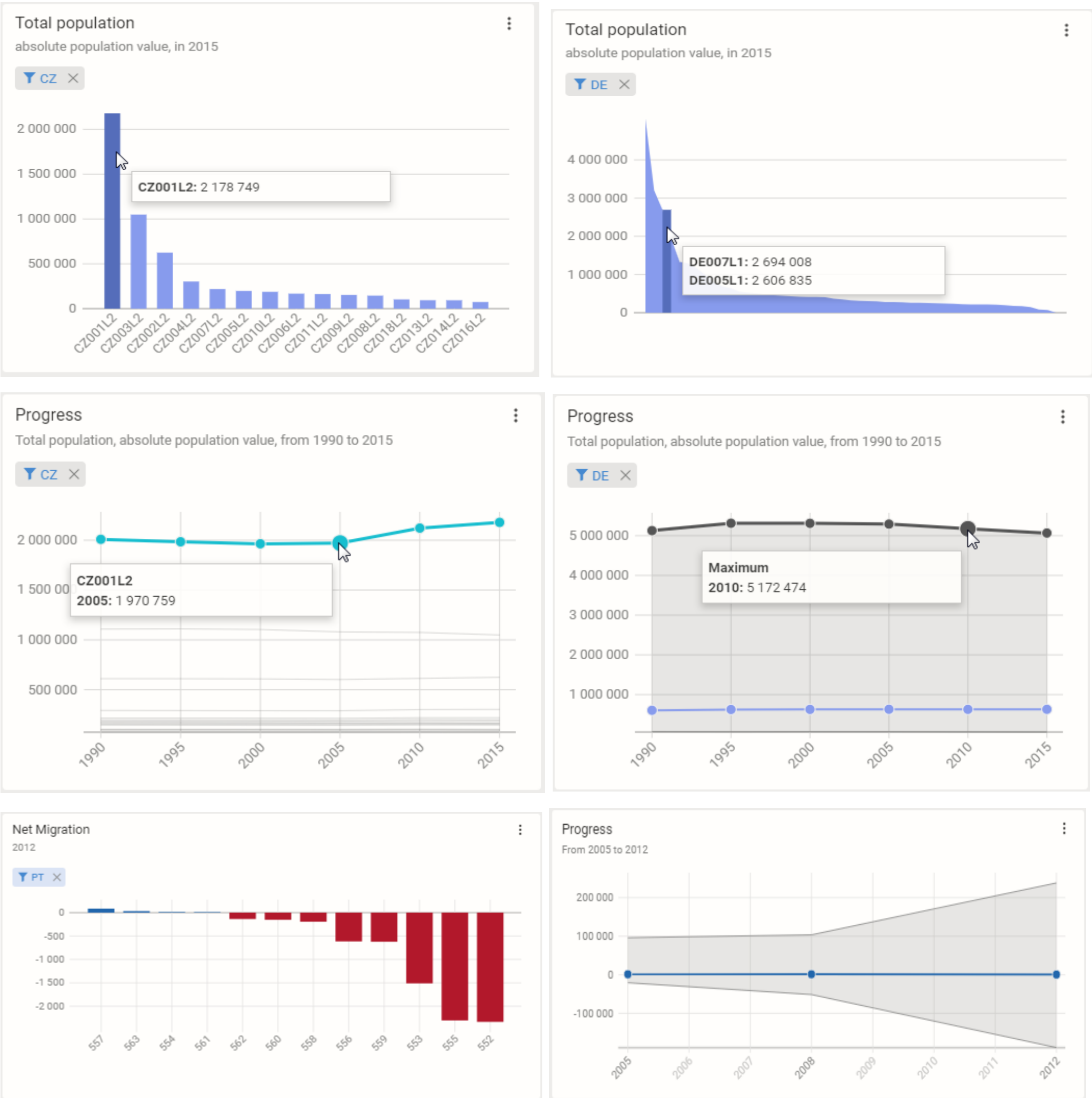


The connection with this background maps from external sources is realized through OGC compliant Web Map Service (WMS). This functionality allows to connect and display any type of background map which is exposed as WMS.

Interactive charts

Two types of interactive charts are available by default in each indicator-related Map View (see Figure 6.7. Charts available in the web tool (column bar – top, line chart – middle, diverging scheme both column and line - bottom).). These charts are situated in the charts panel on the right side of the Map View – **column bar** to display the status values of indicators for one (selected) year and **line chart** to display the five-early time-series between 1990 and 2015. In both types of charts, the user can interactively display the information about the value of indicator for specific selected unit, which is then highlighted in the chart. In case a large number of units is displayed in the charts, the units are grouped in the column chart and only the maximum and average values for the whole set of units are presented in the line chart. In the column charts, the reporting units are automatically sorted in descending way, based on the indicator´s values. In the beta version of the tool, the full names of functional regions are used as a labels.

Figure 6.7. Charts available in the web tool (column bar – top, line chart – middle, diverging scheme both column and line - bottom).



Source: ESPON FUORE web tool

Displaying data in time-series

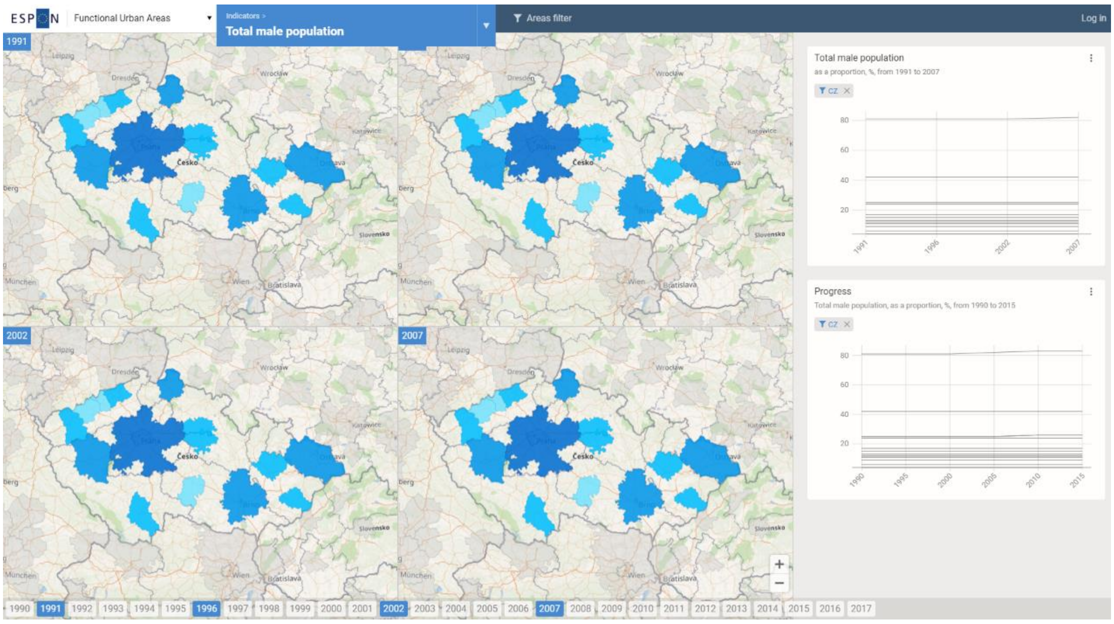
The user is enabled to display the data for one or **more different years** in the Tool. The **multiple-year mode** is enabled through the interactive axis (Figure 6.8. Axis for selection of the year(s) of interest. Years with available data for selected indicators are visible. Years to be displayed in the map and charts are highlighted in colour blue.), located at the bottom of the map application window.

Figure 6.8. Axis for selection of the year(s) of interest. Years with available data for selected indicators are visible. Years to be displayed in the map and charts are highlighted in colour blue.

Source: ESPON FUORE web tool

This axis offers all years for which the selected indicator is available. The user can display (through clicking on the respective button) the data for selected year (or multiple years at once) in the map and charts (Figure 6.9. Multiple-year mode of the web tool, showing values for more years in map and charts). For the map, the number of years displayed at once is limited up to 9, to assure good readability of the information displayed in the map. Once these nine years are selected, the possibility to select additional year on the axis is disactivated. In the line chart, the user can display complete time-series of the indicator (typically from 1990 to 2016).

Figure 6.9. Multiple-year mode of the web tool, showing values for more years in map and charts



Source: ESPON FUORE web tool

Filtering and benchmarking functionalities

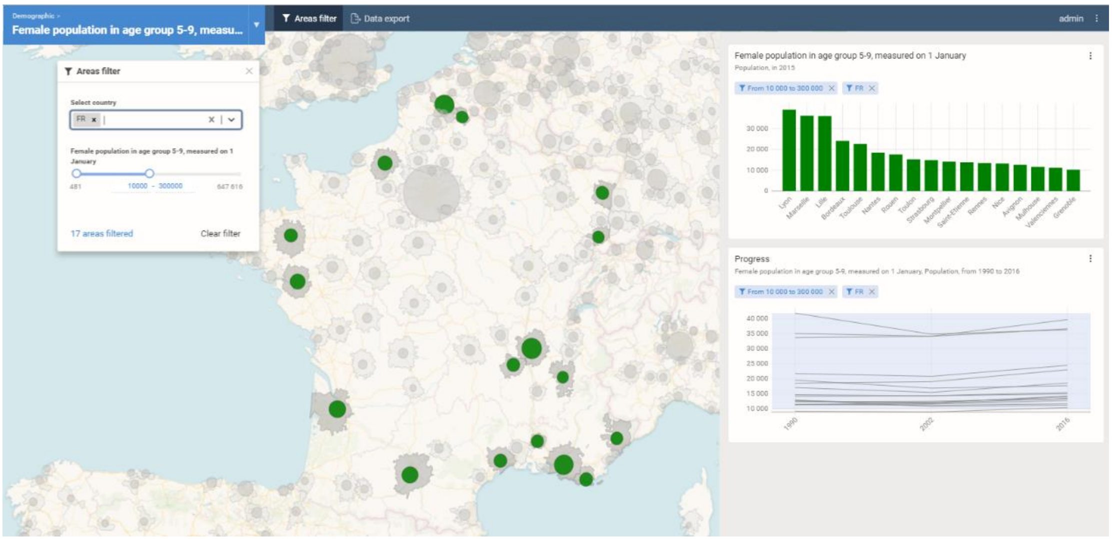
Interactive functionalities for filtering and benchmarking of reporting units (based on indicators´ values or other properties) are integrated into the web tool. Those functionalities provide the users the possibility to interactively filter the reporting units based on one or the combination of different attributes, and, through this, to create temporal selections of units of interest, which could be also exported into tabular or GIS formats.

The main filtering functionality is represented by the **Areas filter** (see Figure 6.10. Filtering FUAs based on combination of country (FR) and indicator value – result in the map and charts.), which is the interactive filtering widget integrated into the tool. It enables the user to filter the units based on:

- country
- Indicator values
- Combination of both

The user can select one or multiple countries of interest and combine the country filter with a filter of indicator´s values. Once the filter is applied, only the selected units are displayed in the map and charts. This selection can be quickly modified or deactivated by the user.

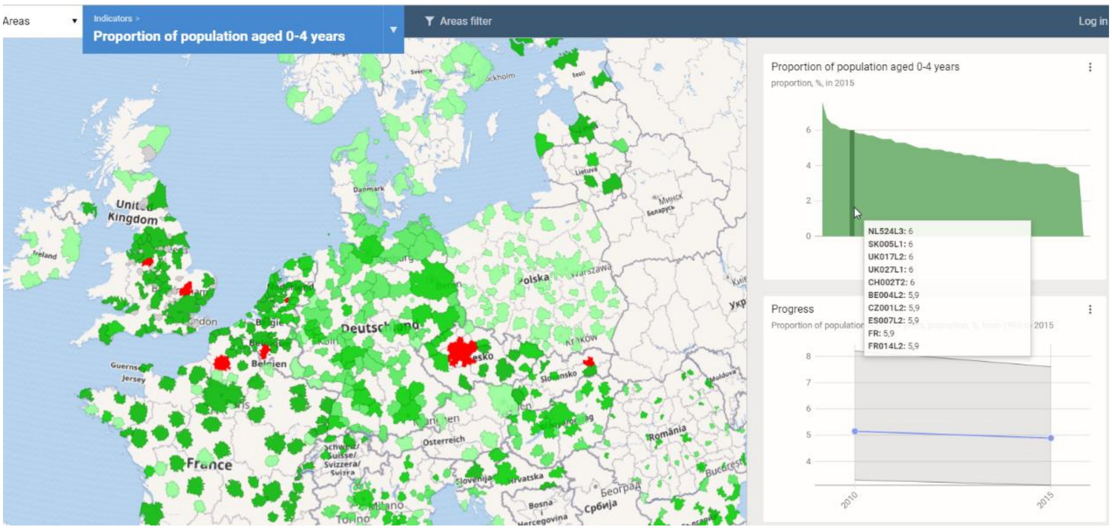
Figure 6.10. Filtering FUAs based on combination of country (FR) and indicator value – result in the map and charts.

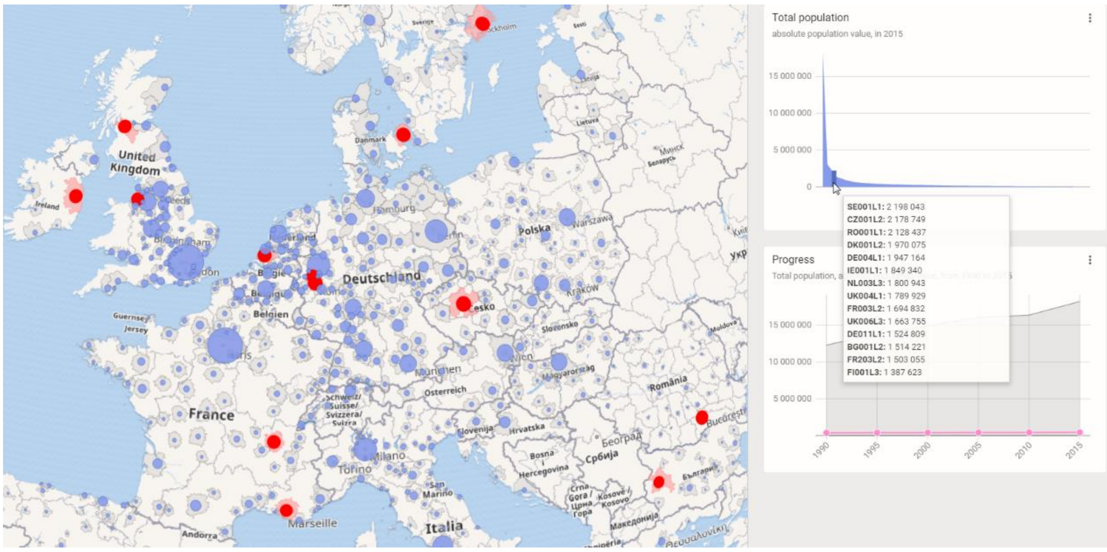


Source: ESPON FUORE web tool

For the interactive data selection and benchmarking, there is a **hover functionality** integrated (Figure 6.11. Hover functionality – units selected in column chart are highlighted in map.). It is an option to interactively select a unit (or groups of units with similar indicator values) in a chart and such selection is immediately highlighted in the map window, so the user can check the geographical distribution of the selected reported units. The same hover functionality works also vice-versa – which means the user can start the selection of the reporting unit in the map window and this selection is reflected in the charts.

Figure 6.11. Hover functionality – units selected in column chart are highlighted in map.



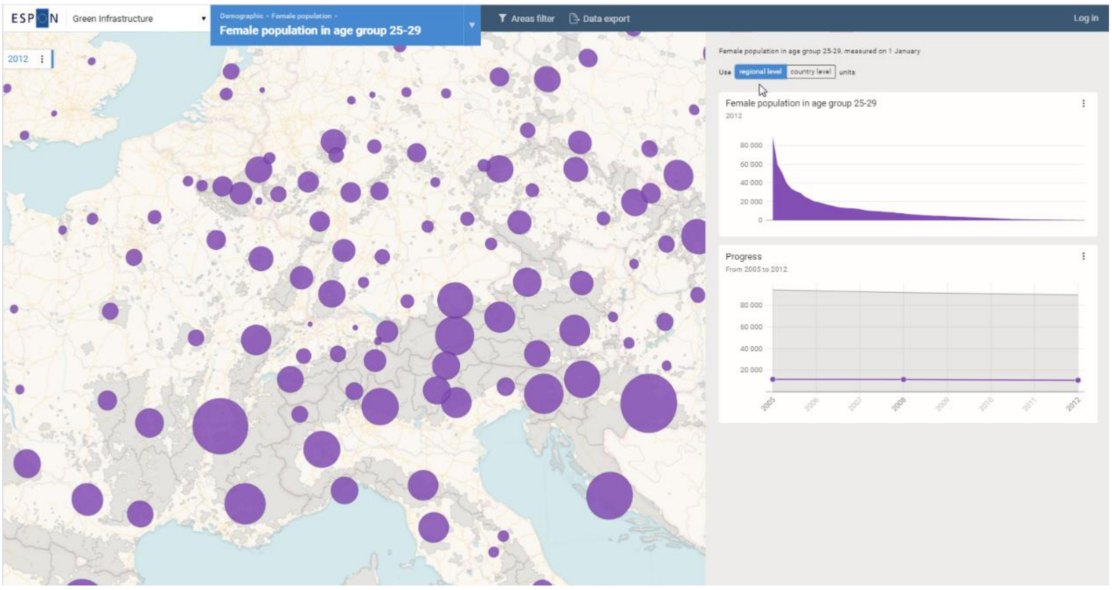


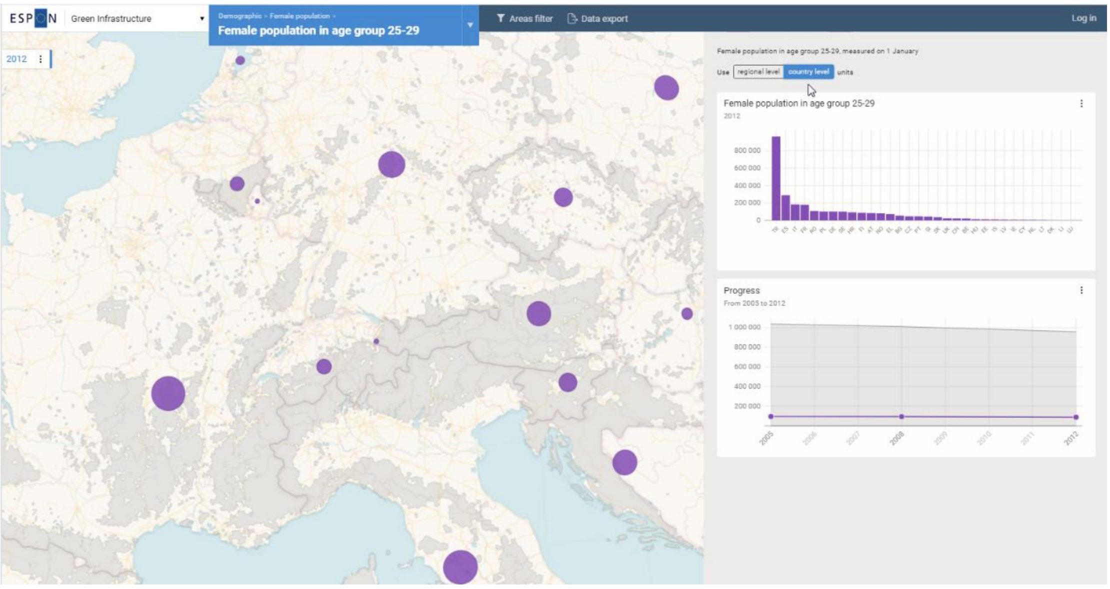
Source: ESPON FUORE web tool

Hierarchical levels of analytical units

For each type of functional region, two hierarchical levels of analytical units are presented in the tool. By default, the “regional” level is switched on. In case the user wants to display values aggregated (for that particular type of functional region) at country level, he/she just needs to switch to “country” level on the interactive switcher in the tool (as demonstrated in Figure 6.12. *Switching between hierarchical levels of analytical units (regional level – top, country level – bottom).* below).

Figure 6.12. Switching between hierarchical levels of analytical units (regional level – top, country level – bottom).



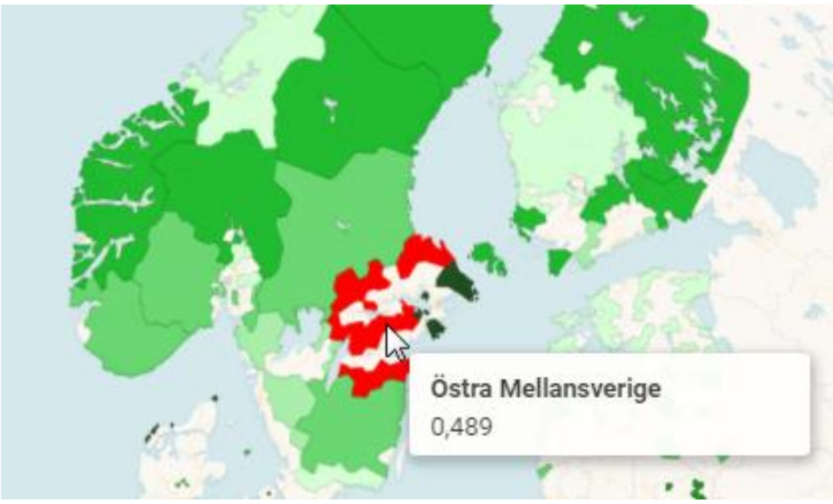


Source: ESPON FUORE web tool

Querying, exporting and sharing data

A functionality to interactively query data in the map is implemented (based on Web Feature Services). Once the user puts the cursor over an analytical unit in the map, the name of the unit is displayed, together with value of indicator for this selected unit, as shown by Figure 6.13. *Querying the indicator value for analytical unit in the map.*.

Figure 6.13. Querying the indicator value for analytical unit in the map.



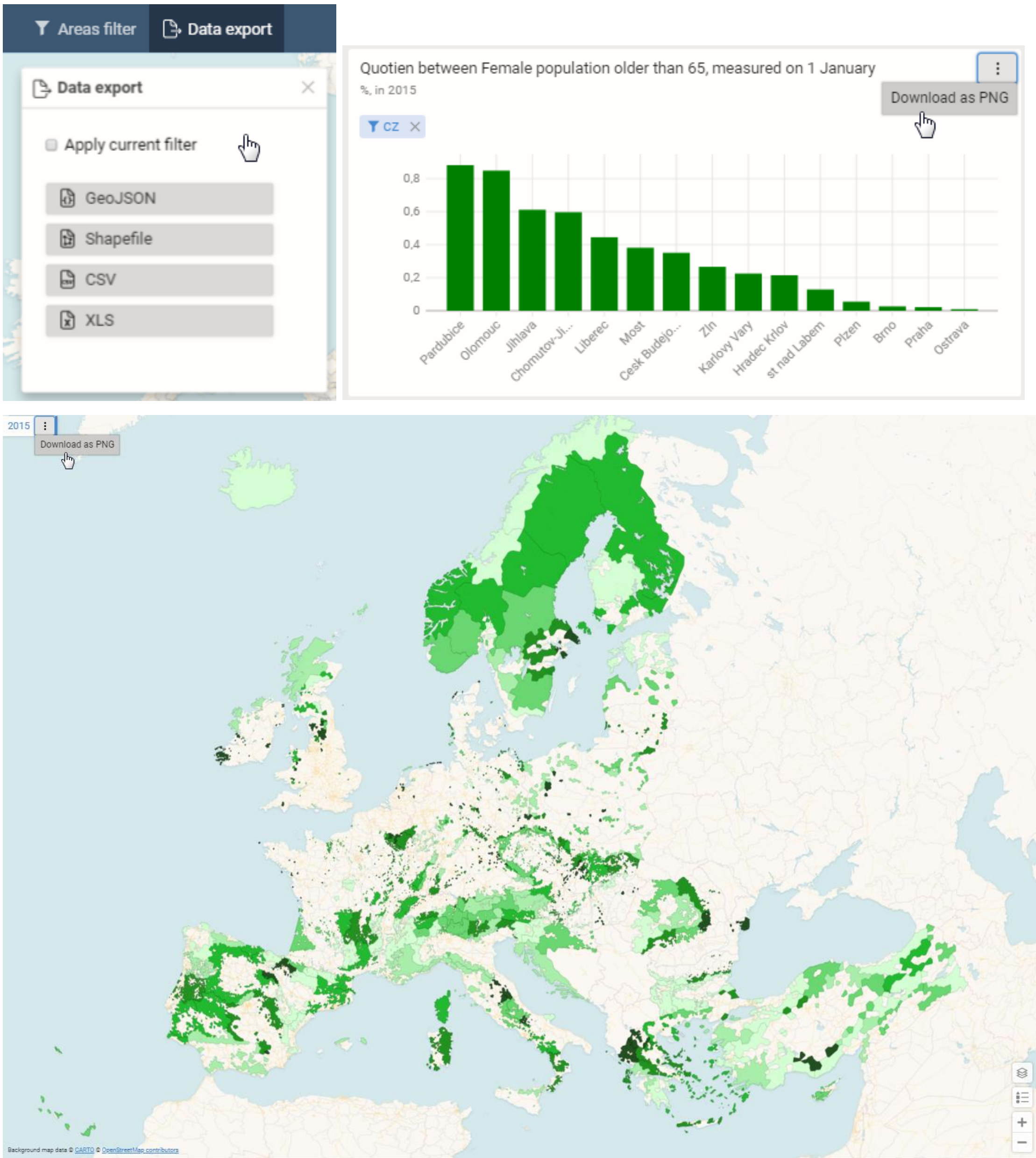
Source: ESPON FUORE web tool

Various exporting functionalities are implemented in the beta version of the Tool, in order to enhance interactivity of the tool and to support the user-driven analysis (see Figure 6.15). These includes:

- Export map as PNG
- Export chart as PNG

- Data Export (a set of units selected in the Area filter or all units can be exported):
    - As GIS layer: GeoJSON or Shapefile
    - As table: CSV or XLS

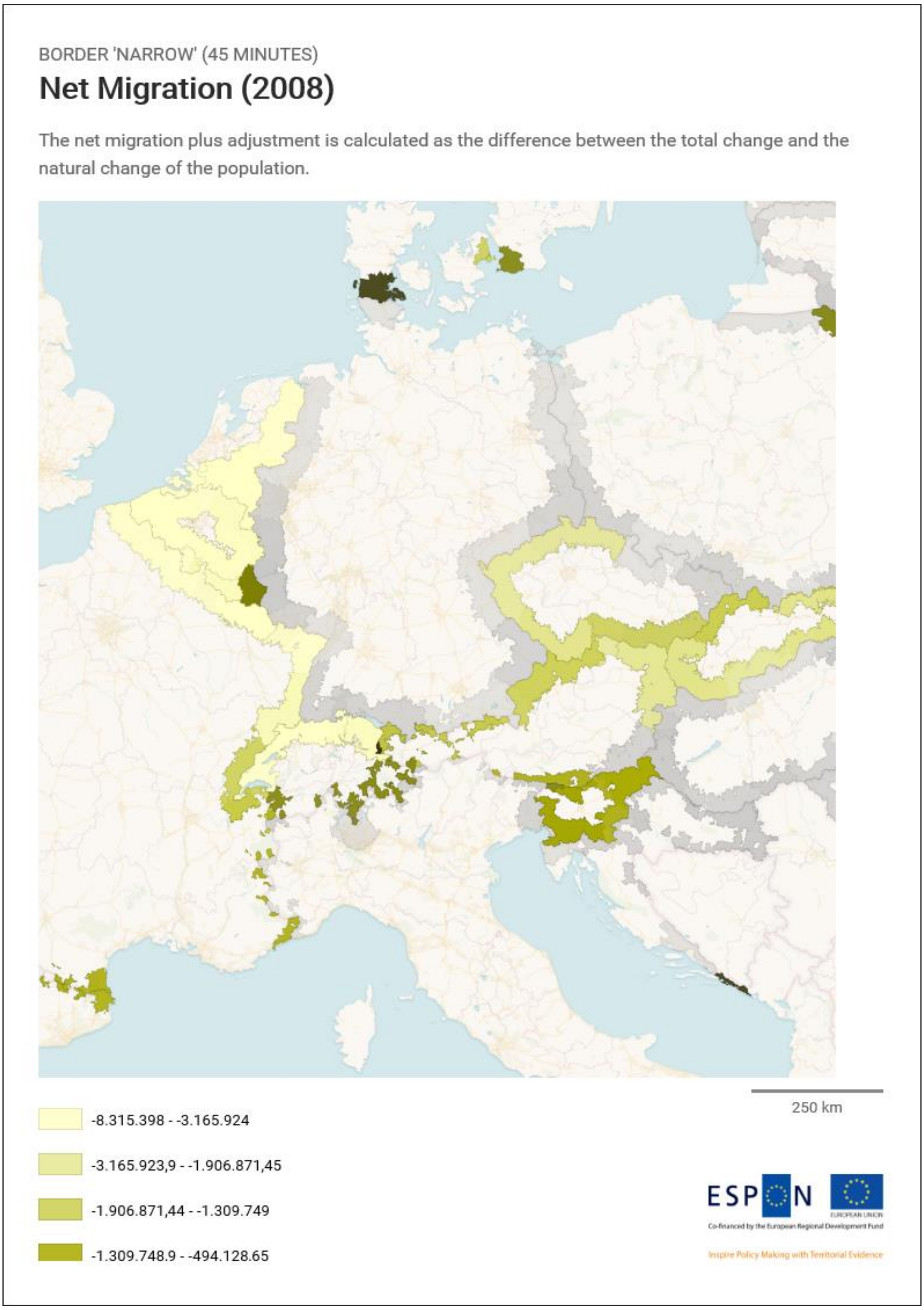
Figure 6.14. Exporting tools (data export, graphical elements export).



Source: ESPON FUORE web tool

For the map export, a complete Map Layout is exported, containing the map window with all layers displayed in the map, accompanied with graphical scale and legend, title of the map and the ESPON logo. A draft layout of such exported map is presented in Figure 6.15. This layout could be further adjusted, based on ESPON preferences.

Figure 6.15. Map Layout as PNG, including title, legend and ESPON logo, as a result of the Data Export functionality of the web tool.



Source: ESPON FUORE web tool

6.5 Integration of the web tool with the ESPON SMD and automatic update of the data in the web tool

The crucial aim in the project is the synchronization of the data and information displayed in the web tool with the one generated by the disaggregation scripts and stored in the ESPON SMD. For that purpose, a sophisticated system has been developed for the integration of the disaggregated indicators into the tool, which also assures automatic update of data content of the Web tool, in case the result of disaggregation is updated.

To secure this, the **results of the disaggregation/aggregation process**, as generated by the corresponding scripts, are produced in a **standardized structure**, which is appropriated for the web tool. This includes few JSON and GeoJSON files containing:

- Metadata (defining analytical units and attributes)
- Data tables (geometries and attribute data with indicator values)

The tables with metadata and data are stored in a single zip package. This system allows to update easily and very quickly the data content of the tool. This update is realized via a REST API interface. This approach allows to:

- Add new regional typology (provided all the processing needed is previously carried out).
- Add new indicator (e.g. indicator disaggregated by the user)
- Add data for additional years for already integrated typologies and indicators
- Update already existing regional typologies or indicators (both metadata description or numeric values)
- Remove any data from the tool

The ESPON SMD stores all the base indicators already aggregated for all the functional regions, plus the indicators calculated *ad hoc* from the input of external users, in a way that the aggregations are not carried out on the fly, speeding up the visualisation and performance of the web tool.

All this process is documented to facilitate future updates once the FUORE web tool and the SMD are deployed on ESPON's infrastructure.

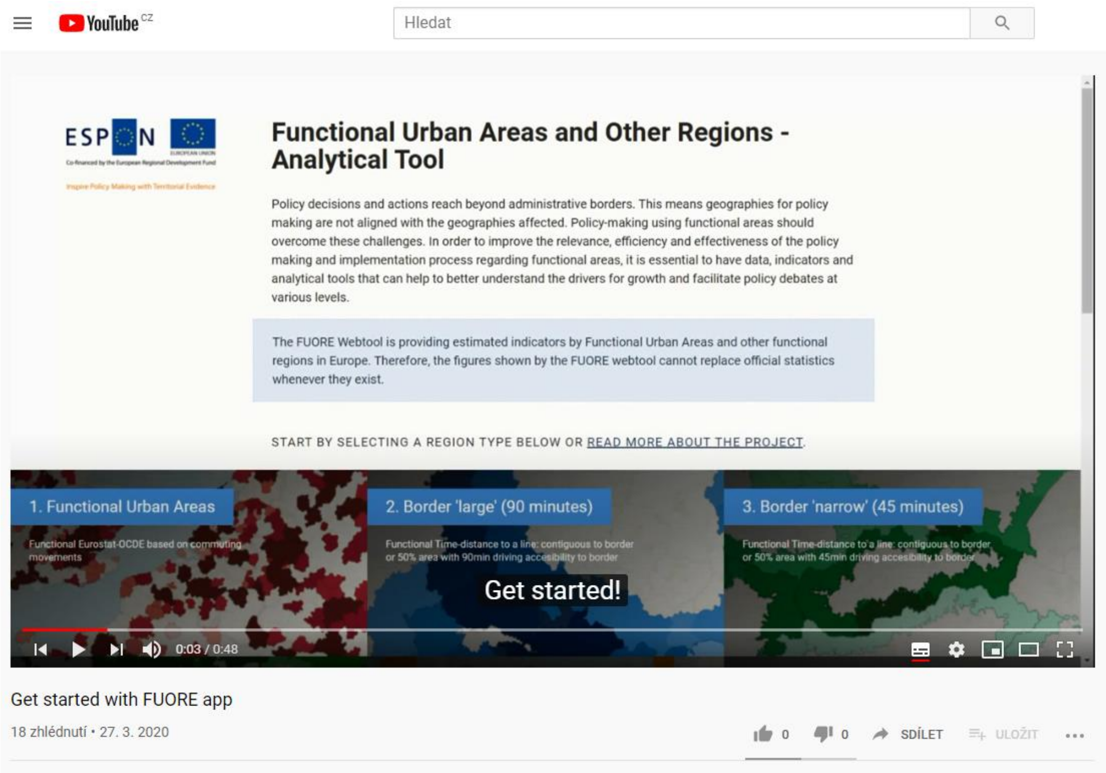
User-driven disaggregation and visualisation of a new indicator

For the advanced user, there is a specific functionality implemented (see section 4.2), enabling the upload of data for a new indicator (at NUTS3 level), from which the figures can be disaggregated and reaggregated back to any of the functional areas available in the web tool. A hyperlink is available on the landing page of the web tool, providing access to this functionality. This link re-directs the advanced user (who needs to log-in first) to the GUI of the Web Data Analysis Toolbox (WDAT). In this Toolbox, the user is able to parametrize, run, and validate his own disaggregation process, as explained in section 4.2. As a final step, the user can upload the results of the disaggregation process into the web tool and explore and analyse them interactively in the interactive Map View of the web tool, in interactive maps and charts, or export them as graphics, GIS data or tables.

6.6 Guidance materials

With the final version of the web tool, well-structured guidance materials have been provided. The main guidance material is represented by the Guidance manual, which presents all the components and functionalities of the tool and provides the user with detailed guidance on how to use the too. Practical examples on how the user can exploit the tool for different types of analysis are presented in the leaflet. In addition, illustrative videos showing the non-technical user how to easily operate the web tool and analyse the integrated data (see Figure 6.16. Video guidance tour showing the web tool and its functionalities.) are available. A YouTube channel has been established for this purpose. In such way, the guidance videos can be easily updated and shared between the users. The videos can be eventually moved to the ESPON YouTube channel if the ESPON EGTC agrees on it.

Figure 6.16. Video guidance tour showing the web tool and its functionalities.



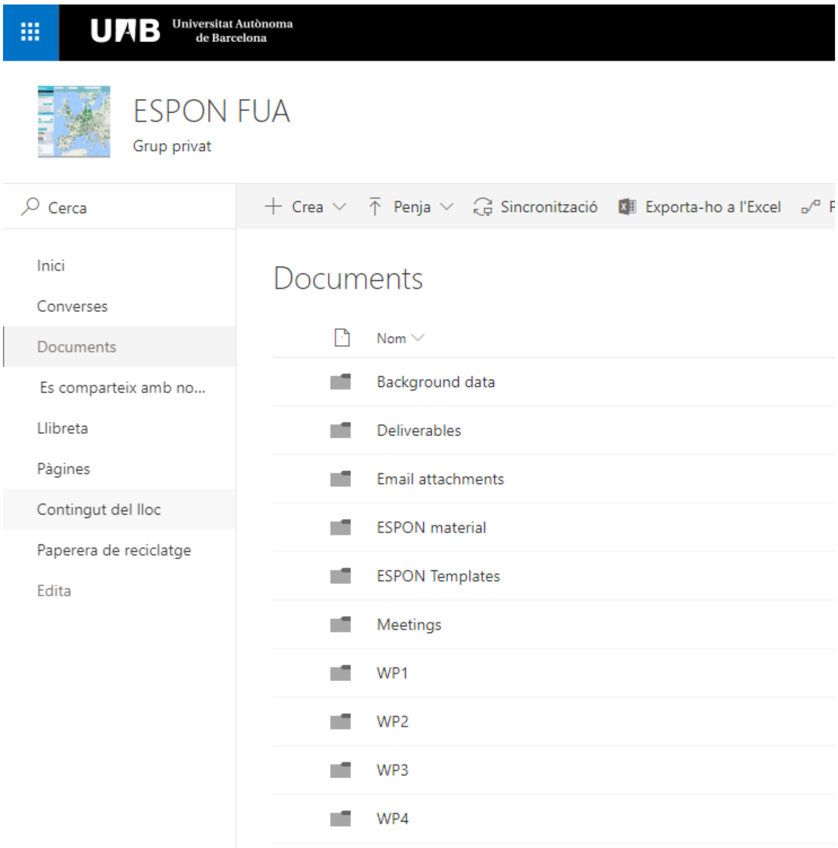
Source: ESPON FUORE

7 Management

The UAB has been the partner managing the project. The project was kicked off in Luxembourg on October 19th, 2018 in a physical meeting attended by the UAB, the ESPON EGTC and one member of the PST, plus a representative of Eurostat. Since then, we summarise hereafter the activities undertaken regarding all the managing aspects:

- Creation of a shared environment to share documents amongst all the project partners: a specific environment using Microsoft SharePoint infrastructure has been set up to share documents and data within the project. It allows editing documents both online and offline.

Figure 7.1. Folder structure under the ESPON FUORE SharePoint group



Source: screenshot, ESPON FUORE

- Assessment of cloud-computing services to be used by our project: a simple cost-benefit analysis has been carried out on different service providers and it is likely that we use the services of clouding.io or Time4VPS. The summary of the assessment is attached as "Annex 6: Analysis of cloud service providers".
- Selection and set-up of a cloud-computing environment, where the SMD and the WDAT are running.
- Physical meetings organised:

  - First service provider Meeting held in Barcelona on 29-30 October 2018.
  - Second service provider Meeting, held in Vienna on 4 December 2018, taking advantage of the ESPON Seminar.
  - Third service provider Meeting, held in Iasi on 20 June 2019, taking advantage of the ESPON Seminar.
  - Fourth and last service provider Meeting, held in Helsinki on 28 November 2019, taking advantage of the ESPON Seminar.
- Other physical meetings attended:
  - KO meeting held in Luxembourg on 19 October 2018.
  - ESPON Seminar, held in Vienna on 5-6 December 2018.
  - Partnership meeting of ongoing ESPON projects, held in Vienna on 6 December 2018.
  - ESPON Seminar, held in Iasi on 19-20 June 2019. The service provider gave a presentation and participated in "Session B1: Functional urban areas and links to declining areas". A video of the presentation can be watched on the project's website (https://www.espon.eu/functional-urban-areas-tool).
  - Partnership meeting of ongoing ESPON projects, held in Iasi on 20 June 2019.
  - ESPON Seminar, held in Helsinki on 27-28 November 2019. The service provider participated in the networking session named "ESPON open data and tools for functional areas and sustainable development goals", giving a presentation of the project outcomes and discussing the results with different stakeholders. Some of the opinions gathered there have been used in the improvement and finalisation of the project.
- Virtual meetings: During all the project, several teleconferences have been organised to coordinate between partners. Besides a regular monthly teleconference amongst all partners, regular coordination meetings have occurred weekly. During the COVID-19 crisis, those kind of contacts were intensified, also including feedback meetings with the ESPON EGTC.
- The Draft Final Report was delivered on time, and a reviewed version, including feedback from ESPON EGTC, the ESPON MC and the PST, was produced. It has already been published on ESPON's website (https://www.espon.eu/functional-urban-areas-tool).

7.1 Challenges and Implemented solutions

In the project proposal, the service provider collected already several challenges and potential difficulties, and how to overcome them. This is a summary of the main challenges that were defined for this project and some of the solutions adopted by the awarded team:

- Identifying and obtaining delineations for functional regions different from FUAs. Solution: establishing a clear criterion for the selection and definition of delineations.
- Missing data for some areas (e.g. Outermost regions, Turkey, western Balkans). Solution: looking for alternative sources whenever existing.
- Improvement of former disaggregation methodologies. Solution: making use of the existence of better ancillary datasets with the experience of the service provider and all the team members working for institutions like the JRC or the EEA.
- Huge amount of data to be stored and processed. Solution: implementation of innovative technical solutions based on Jupyter notebook servers together with cloud processing.
- Linkages with the ESPON DB indicators. Solution: the service provider is also involved in the ESPON DB and knows very well the details behind the database.
- Huge processing time expected. Solution: cloud processing and optimised scripts have been applied.
- Engagement of experts for validation. Solution: Good networking together with the ESPON EGTC. Unfortunately, the workshop with experts could not eventually take place.
- Performance of the web tool. Solution: Same database schemas and precomputed aggregations have been implemented.
- Ad hoc requests from users to add new data in the webtool. Solution: users can provide NUTS 2 or 3 data on a predefined template and can choose the method of disaggregation and target functional region. When the processing is done, the user is notified.

References and literature used

Batista e Silva, F. & Poelman, H. (2016) Mapping population density in Functional Urban Areas - A method to downscale population statistics to Urban Atlas polygons. JRC Technical Report no. EUR 28194 EN. doi:10.2791/06831.

Batista e Silva, F., Poelman, H., Martens, V. & Lavalle, C. (2013) Population estimation for the Urban Atlas Polygons. JRC Technical Report no. EUR 26437 EN. doi:10.2788/54791.

Deichmann, U. et al (2001) Transforming Population Data for Interdisciplinary Usages: From census to grid. CIESIN, Columbia University. http://sedac.ciesin.columbia.edu/gpw-v2/GPWdocumentation.pdf

EEA (2017) Analysis of key trends and drivers in greenhouse gas emissions in the EU between 1990 and 2015. EEA Report, No 8/2017. ISSN 1977-8449. Available at: https://www.eea.europa.eu/publications/analysis-of-key-trends-and (accessed 13 November 2019)

ESPON (2005) ESPON 1.1.1 “Urban areas as nodes in a polycentric development”. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/fr-1.1.1_revised-full_0.pdf (accessed 14 January 2019)

ESPON (2007) ESPON 1.4.3. “Study on Urban Functions”. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/fr-1.4.3_April2007-final.pdf (accessed 14 January 2019)

ESPON (2010a) ESPON FOCI. Executive summary. Available at: https://www.espon.eu/sites/default/files/attachments/FOCI_FinalReport_ExecutiveSummary_20110310.pdf (accessed 14 January 2019)

ESPON (2010b) ESPON Metroborder. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/METROBORDER_-_Final_Report_-_29_DEC_2010.pdf (accessed 14 January 2019)

ESPON (2012a) ESPON Polyce. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/POLYCE_FINAL_MAINREPORT.pdf (accessed 14 January 2019)

ESPON (2012b) ESPON Geospecs. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/GEOSPECS_Final_Report_v8___revised_version.pdf

ESPON (2013a) Science in support of European Territorial Development and Cohesion, Second ESPON 2013 Scientific Report, December 2013. ISBN 978-2-919777-53-2. https://www.espon.eu/sites/default/files/attachments/ESPON_SCIENTIFIC_REPORT_2.pdf (accessed 4 March 2019)

ESPON (2013b) ESPON Best Metropolises. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/BestMetropolises_FR_Executive_SummaryxMain_Report.pdf (accessed 14 January 2019)

ESPON (2013c) ESPON Ulysses. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/Ulysses_Final_Report_2013_01_25.pdf (accessed 14 January 2019)

ESPON (2015) ESPON M4D. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/M4D_FinalReport_MainReport.pdf (accessed 14 January 2019)

ESPON (2017) Handbook for Service providers of ESPON 2020. Version 26/01/2017.

ESPON (2018) ESPON Spima. Final report. Available at: https://www.espon.eu/sites/default/files/attachments/SPIMA%20Final%20Report.pdf (accessed 14 January 2019)

Eurostat (2018) Methodological manual on territorial typologies. https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-008 (accessed 14 February 2019)

Gallego J., Peedell S. (2001) Using CORINE Land Cover to map population density. Towards Agri-environmental indicators, Topic report 6/2001 EEA, Copenhagen, pp. 92-103. http://reports.eea.eu.int/topic_report_2001_06/en

JRC-IES-LMU-ESDI (2003), Short Proceedings of the 1st European Workshop on Reference Grids, Ispra, 27-29 October 2003. http://eusoils.jrc.ec.europa.eu/projects/alpsis/Docs/ref_grid_sh_proc_draft.pdf

Michelet JF, Gloersen E, Ysebaert R, Giraut F. (2018) Producing a simplified and harmonized map of European local administrative units (LAU): when “less” offers “more”…. ESPON Scientific Conference 2018. London. Available at: https://www.espon.eu/scientific-report-2019 (accessed 13 November 2019)

Milego, R., Ramos, M.J. & Martínez, C (2013). Spatial disaggregation of socio-economic data and combination with spatial data by means of OLAP technologies. In Science in support of European Territorial Development and Cohesion, Second ESPON 2013 Scientific Report, December 2013. Available at: https://www.espon.eu/topics-policy/publications/scientific-reports/second-espon-2013-scientific-report (accessed 13 November 2019)

Milego, R., Ramos, M.J., Domingues, F. & Martínez, C. (2014). OLAP technologies applied to the integration of geographic, thematic and socioeconomic data in the context of ESPON. Technical report. ESPON M4D- Multi dimensional database design & development. European Union. Available at: https://www.espon.eu/sites/default/files/attachments/M4D-DFR_TR-OLAP_draft_20140630.pdf (accessed 13 November 2019)

Nordhaus, W. (2006) New Metrics for Environmental Economics: Gridded Economic Data. OECD. http://www.oecd.org/environment/cc/37117455.pdf

Pesaresi, Martino; Ehrilch, Daniele; Florczyk, Aneta J.; Freire, Sergio; Julea, Andreea; Kemper, Thomas; Soille, Pierre; Syrris, Vasileios (2015): GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014). European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_built_ldsmt_globe_r2015b

Price, M.F., Arnesen, T., Gløersen, E. et al. J. Mt. Sci. (2019) Mapping mountain areas: learning from Global, European and Norwegian perspectives16: 1. https://doi.org/10.1007/s11629-018-4916-3

Rosina, K., Batista e Silva, F., Vizcaino-Martinez, Pl, Marín Herrera M.A., Freire, S. & Schiavina, M. (2018). An improved European land use/cover map derived by data integration AGILE 2018 – Lund, June 12-15, 2018

Stefano FERRI, Alice SIRAGUSA, Filip SABO, Maria PAFI, Matina HALKIA; The European Settlement Map (2017) Release. Methodology and output of the European Settlement Map; EUR 28644 EN; doi:10.2760/780799

Tobler, W.R. (1979). `Smooth pycnophylactic interpolation for geographical regions.'. Journal of the American Statistical Association 74(367):519-530.

List of Annexes

Annex 1: Methodology to delineate FUAs and other functional regions into ESPON LAU Census 2011 nomenclature

Due to the large number of maps in this annex, a map content list is provided hereafter:

1. Introduction

As the ESPON 2020 Map kit is based on 2011 LAU nomenclature, choice has been made to transfer original delineations into the ESPON LAU Census 2011 layer. The issue is to bring into coherence a variety of functional region and geographical territories that have been originally delineated from various LAU nomenclatures (FUA 2015-2018, Coasts 2016, MSA 2001; GEOSPECS 2008 and GRETA 2011). That choice will also make it easy for ESPON projects to map their results within the ESPON 2020 Map kit and will provide the ESPON LAU Census 2011 mapping layers with functional/geographical delineations attributes.

The transfer method from their original nomenclature into ESPON LAU census 2011 nomenclature is a pragmatic one. It consists in overlaying functional/geographic types of territories in their original LAU nomenclature with ESPON LAU Census 2011 layer. As a result, all 2011 LAUs, which over 50% of their area (i.e. the same threshold used by TERCET to define coastal LAUs) is covered by the functional/geographical type, belong to the type of territory concerned. In cases of coasts or borders, contiguity criteria have been adjusted when relevant.

The present annex is divided into three sections, as methodologies applied to non-overlapping areas, overlapping and territories with Green potential differ.

2. Non-overlapping types of territories

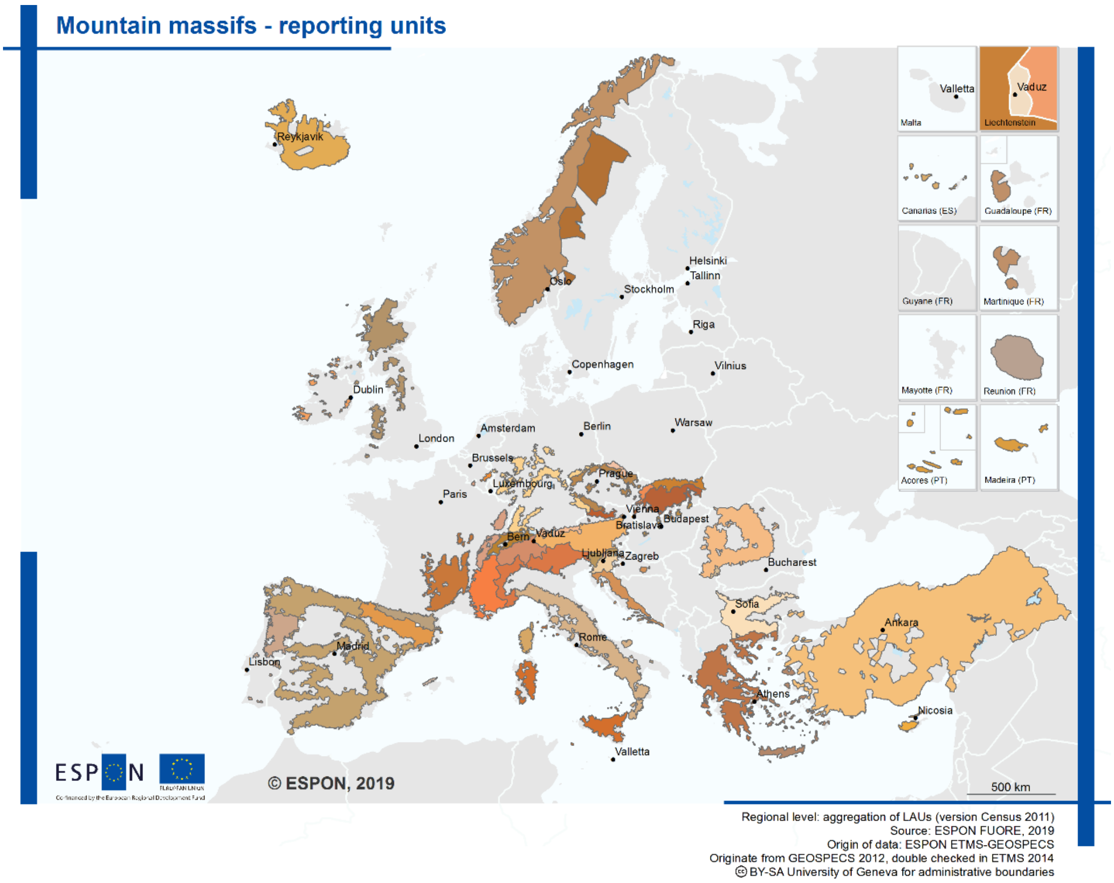
Mountains, islands and sparsely populated areas from GEOSPECS; coasts, as delineated by TERCET; and maritime service areas (MSA) from Eurostat derived from GRETA, are non-overlapping areas/objects. Transfer method to pass them from their original nomenclature into ESPON LAU census 2011 nomenclature can be summarised as follow:

1. Select all ESPON LAU Census 2011 units that are touching FUA/other types of territories as delineated in their original mapping layer.
2. “Clip (intersecting)” selected ESPON LAU Census 2011 units with the original delineation.
3. Dissolve all objects in ESPON LAU Census 2011, using Cens_ID. Indeed, LAUs have been exploded in a multitude of small parts as result of the clipping (due to topological differences between the two versions of the EuroBoundaryMap).
4. Divide the original area of the ESPON LAU Census 2011 units by their area resulting from the clipping.
5. Erase all LAU where percentage resulting from step 4 is < 50%.
6. Create “centroids (inner)” out of selected units (step 5) from ESPON LAU Census 2001.
7. Pass attributes of FUA/territories from original delineation into centroids of selected ESPON LAU Census 2011 by “spatial overlay”.
8. Pass the typology and delineation attributes into original ESPON LAU Census 2011 layer by making a “table join” with the ESPON LAU Census 2011 centroids.
9. Dissolve ESPON LAU Census 2011 units into the desired functional and other territories.

2.1. Mountains

GEOSPECS mountain massifs are based on topographic delineation from the European Environmental Agency (EEA 2010). This set of grid cells with mountainous topography was approximated to 2008 LAU boundaries by considering that LAU units with more than 50% mountainous terrain should be considered mountainous. Continuous mountain areas of less than 100 km2 were then identified, and designated as exclaves, which were excluded from the mountain delineation (except on islands of less than 1000 km2). Similarly, non-mountainous groups of LAU units of less than 200 km2 surrounded by mountain areas were identified as enclaves and included in the mountain delineation. Mountain areas have been grouped into 16 massifs.

Annex Map 1 : Mountain massifs, as delineated in ESPON LAU Census 2011 nomenclature

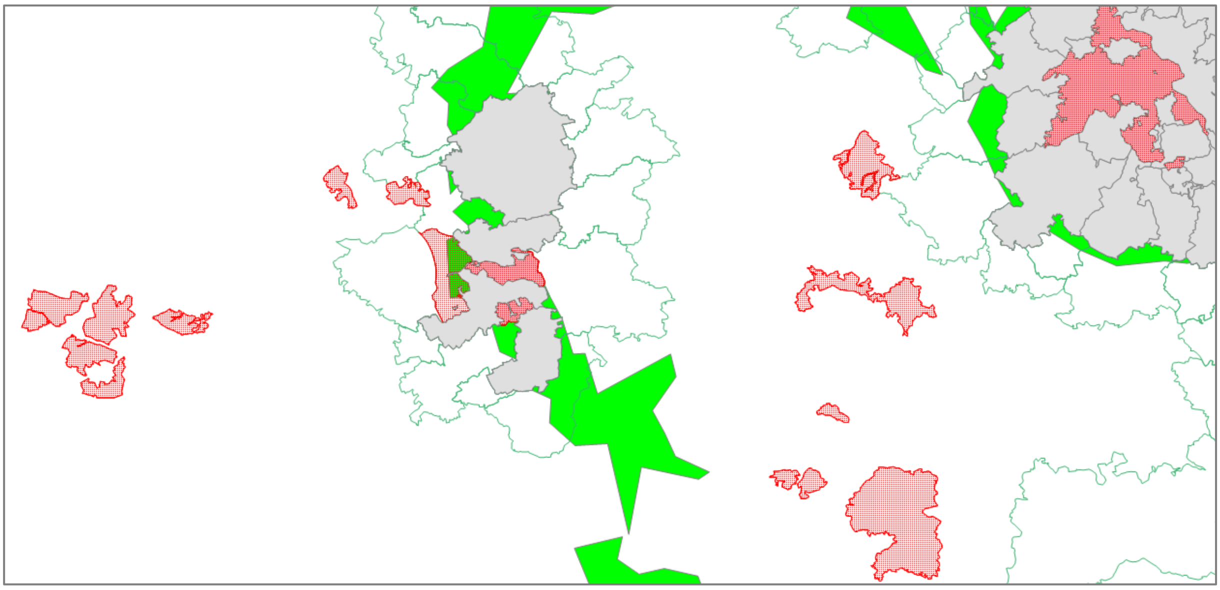


Delineation issues

In Germany, forest conservation areas (= false LAU, in red Map 2) with mountainous and non-mountainous parts have been merged into a single administrative unit for the making of the

ESPON LAU Census 2011 layer. As mountainous part (in grey Map 2), is less than 50% of the merged unit, it has been coded as non-mountainous.

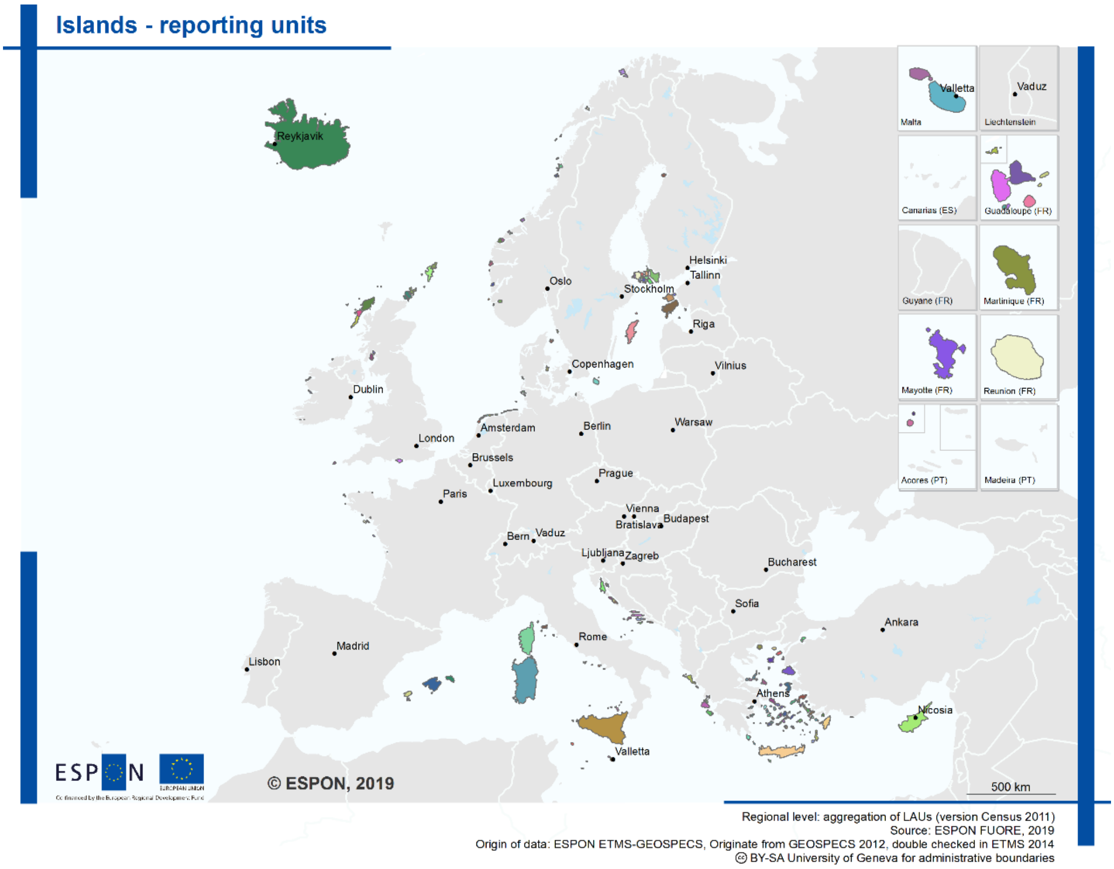
Annex Map 2 : Mountain delineation, Issues for two forest conservation areas



2.2. Islands

GEOSPECS Islands are a given morphological delineation: it is about a LAU (or group of LAUs) that is surrounded by sea. Islands with a fixed link to continent have been excluded. Three French islands have been added since GEOSPECS project as the result of changes in their administrative status: Saint Martin & Saint-Barthelemy in the Caribbean Sea and the 17 LAU making the island of Mayotte in the Indian Ocean.

Annex Map 3 : Islands, as delineated in ESPON LAU Census 2011 nomenclature



Delineation issues

None

2.3. Sparsely populated Areas (SPA)

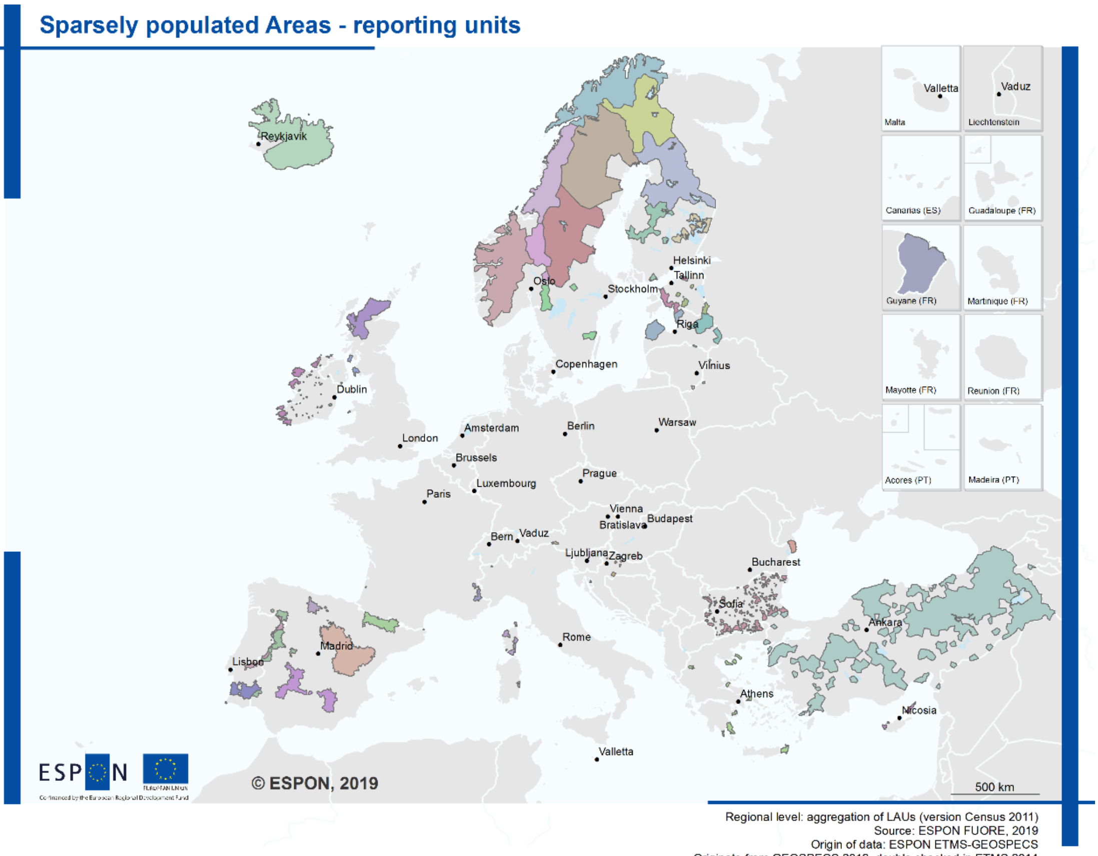
GEOSPECS SPAs are delineated based on low settlement pattern - i.e. the number of persons that are within a reasonable commuting distance of each “point” in Europe. The threshold is fixed at 100’000 persons. Sparsity is calculated with a combination of two complementary models:

- Isotropic model: 1*1 km grid cells in Euclidian distance of 50 km (i.e. as the crow flies)
- Directed model: 1*1 km grid cells in 45min driving time (isochrones) using detailed road network modelling.

Cells have then been aggregated at LAU level. Only municipalities with over 90% sparsely populated area (either from isotropic or directed model) have been selected as SPA. In a second step, a total of 39 Sparse Territories have been identified, based on geographic contiguity and proximity and close socio-cultural proximity of sparse LAU units.

The GEOSPECS SPA as delineated in their original LAU 2008 nomenclature have been transferred into the ESPON LAU Census 2011 layer, using the 50% threshold.

Annex Map 4 : SPA, as delineated in ESPON LAU Census 2011 nomenclature



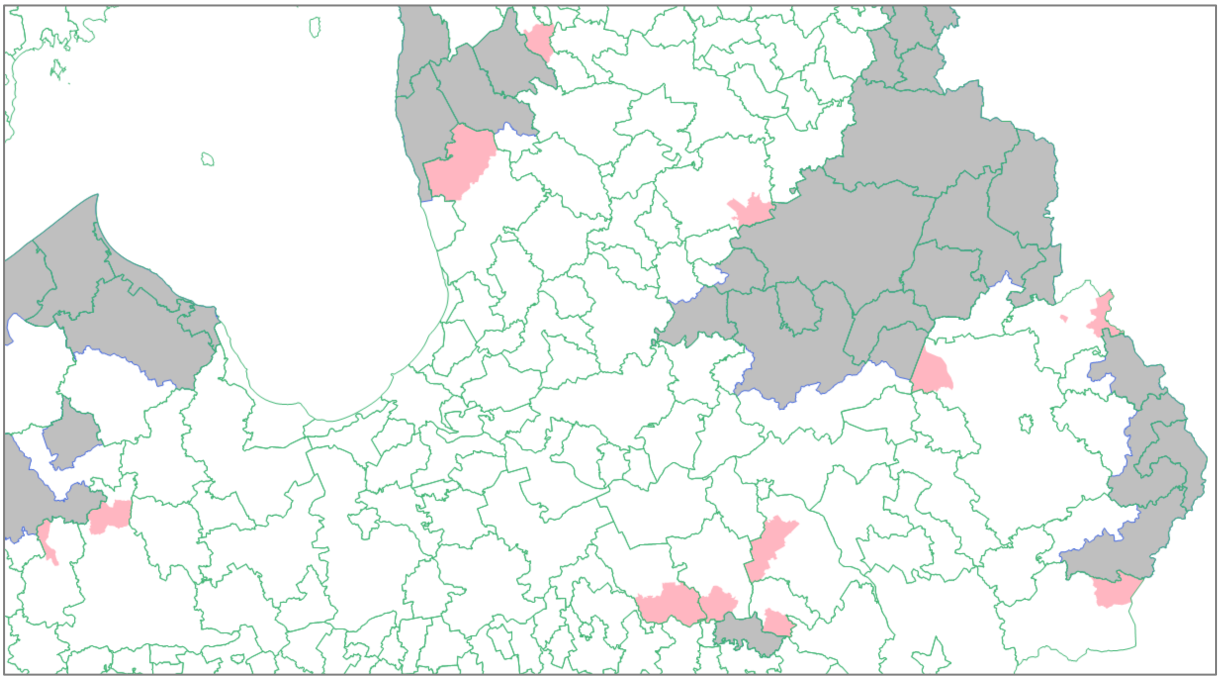
Regional level: aggregation of LAUs (version Census 2011)
Source: ESPON FUORE, 2019
Origin of data: ESPON ETMS-GEOSPECS
Originate from GEOSPECS 2012, double checked in ETMS 2014
© BY-SA University of Geneva for administrative boundaries

Delineation issues

SPA delineation remains consistent over ESPON space.

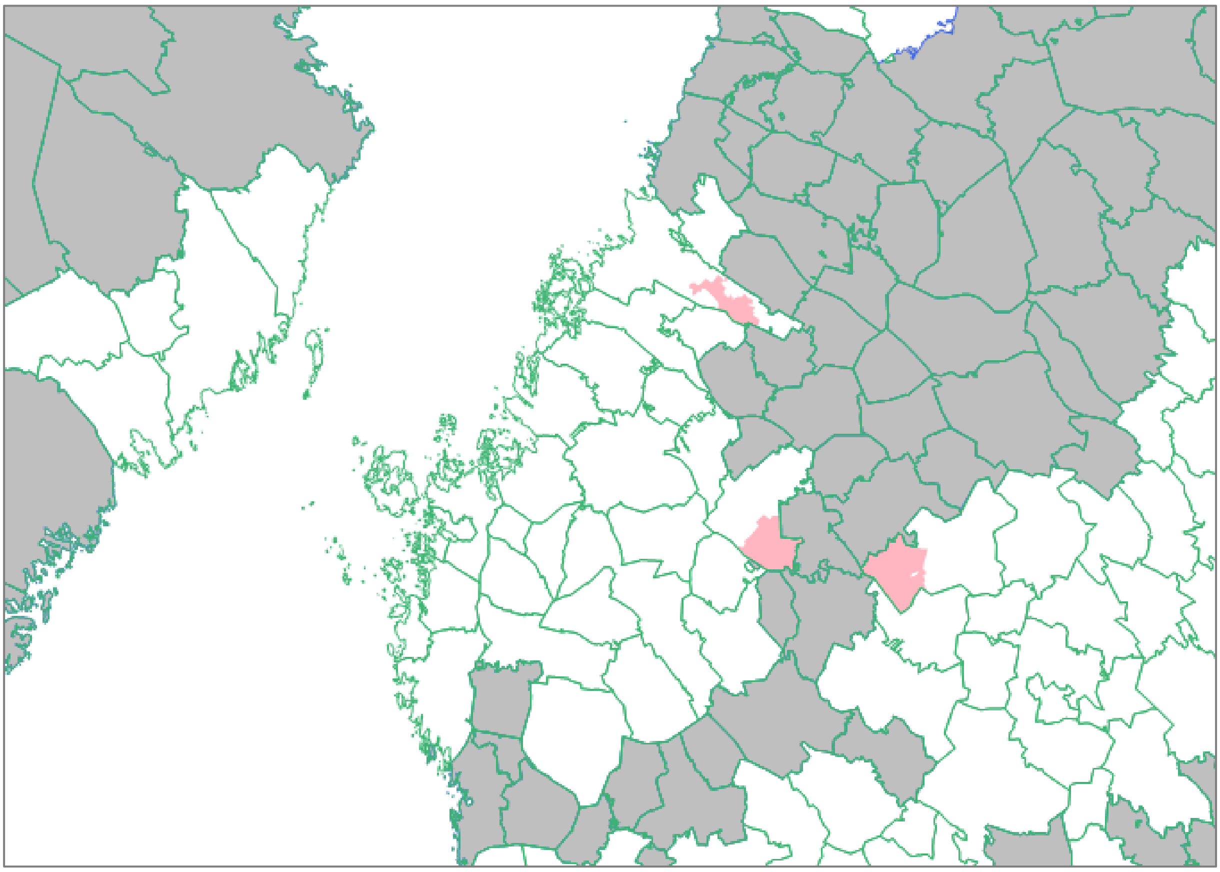
However, a change in LAU boundaries/nomenclature (merging of municipalities into larger entities) induces some reduction in SPA area in Latvia and Finland.

Annex Map 5 : SPA changes in delineation: Latvia



To a lesser extent, similar changes happen in Finland.

Annex Map 6 : SPA changes in delineation: Finland



2.4. Coast as defined in the TERCET

TERCET provides a morphological delineation of coasts whit two criteria a LAU must fill:

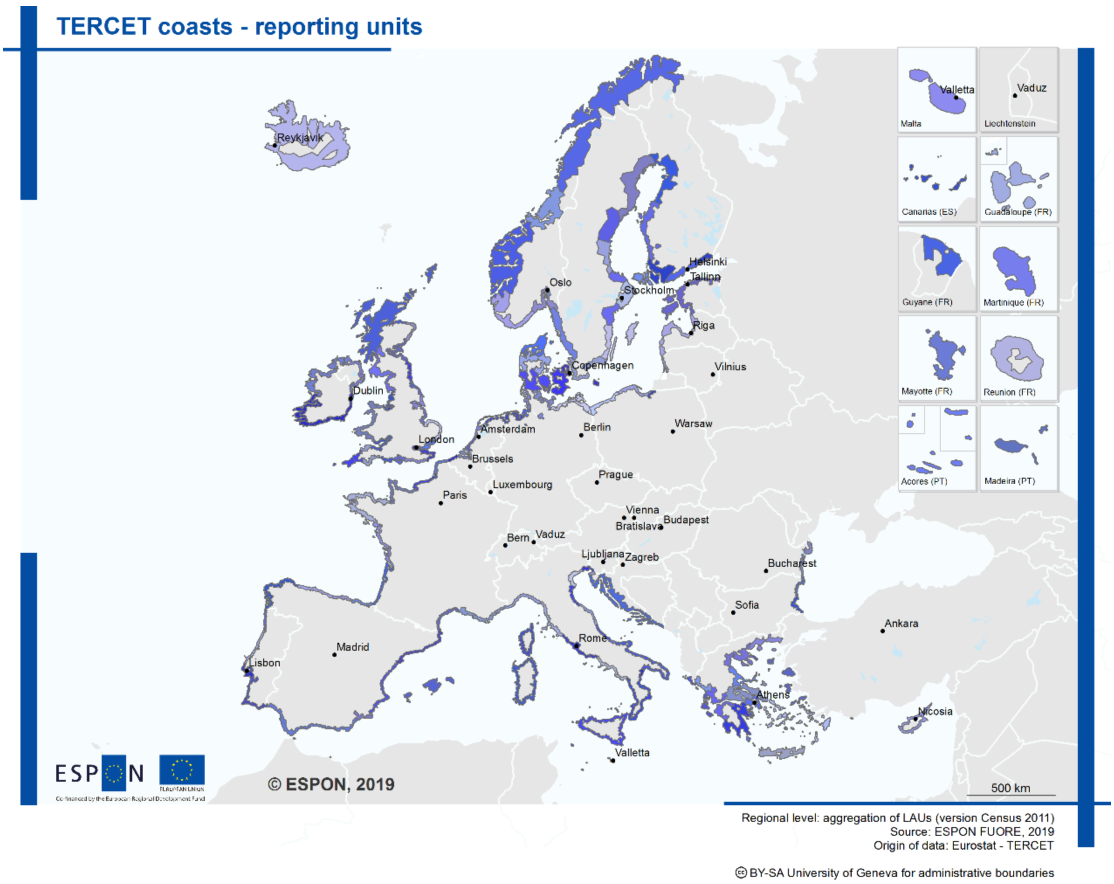
1) Being contiguous to coast;
2) And/or having ≥ 50 % of surface area within 10 km of the coastline.

Additionally, a LAU can be added/retracted at request from a member state. 2016 and 2017 LAU coastal nomenclatures have been produced, but only the 2016 has been yet validated by member state. For e.g., the LAU of Rotterdam (largest Harbour in EU, contiguous to the sea) has been retracted from coastal typology by Netherlands, because its size would distort the Dutch coastal tourism data. As a result, Eurostat strongly suggested using the 2016 version.

Additional criteria pledge for 2016:

- The fit is perfect in EL, while in 2017 EL has introduced a new LAU nomenclature (much smaller units) that induces important inconsistencies with ESPON LAU Census 2011 map.
- In UK, the 2016 units are much smaller than 2017 (4-5 times). Resulting coast delineation is therefore much closer from a morphological definition it is based on. This is especially obvious in Scotland.
- Overall fit with ESPON Census 2011 LAU layer is almost 100%.

Annex Map 7 : Coasts from TERCET, as delineated in ESPON LAU Census 2011 nomenclature

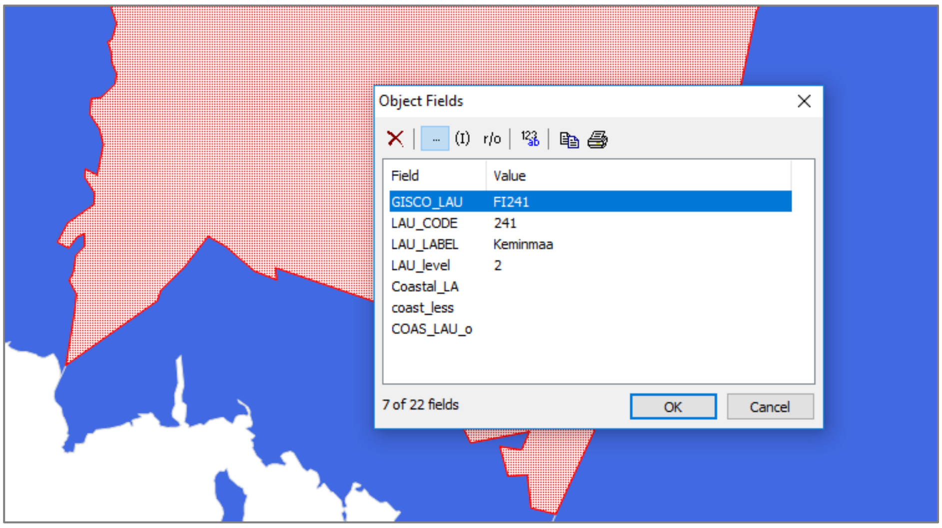


Delineation issues

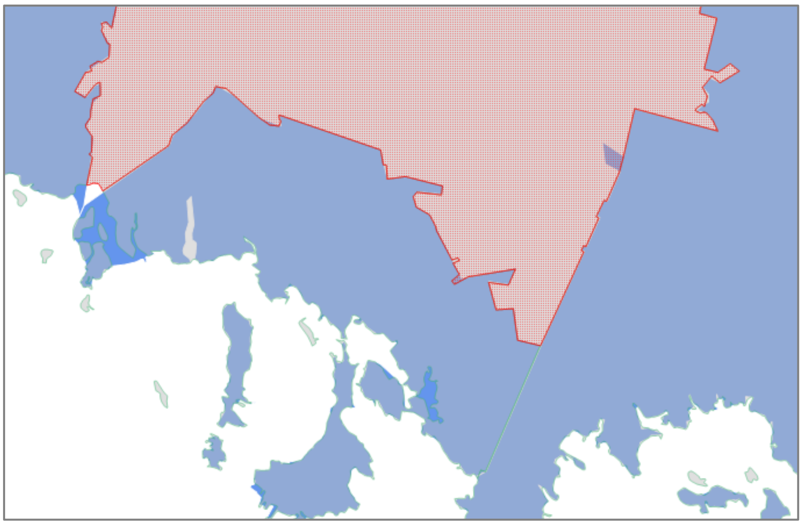
The French island of St-Barthelemy was added in both 2016 and 2017 TCOA.

Insufficiently detail LAU map used to produce the TERCET coast ends up into ignoring LAUs that are contiguous to coast. However, as TERCET coastal classification 2016 has been review by national authorities, it has been strictly respected for these LAUs.

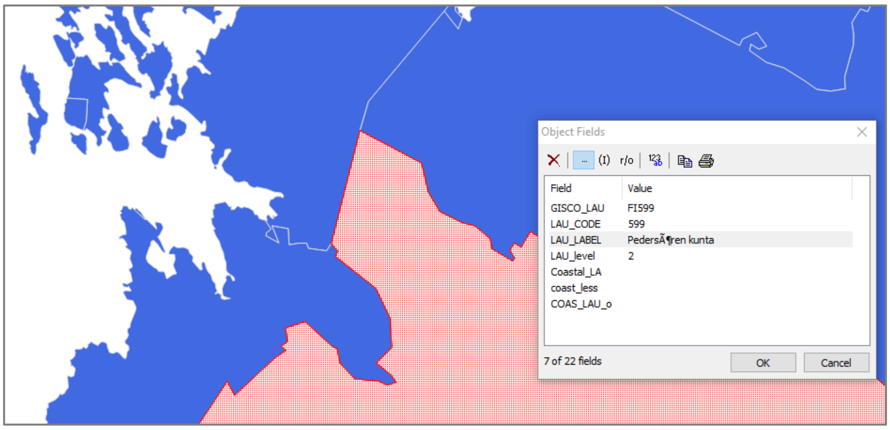
Annex Map 8 : Keminmaa in FI (TERCET mapping layer)



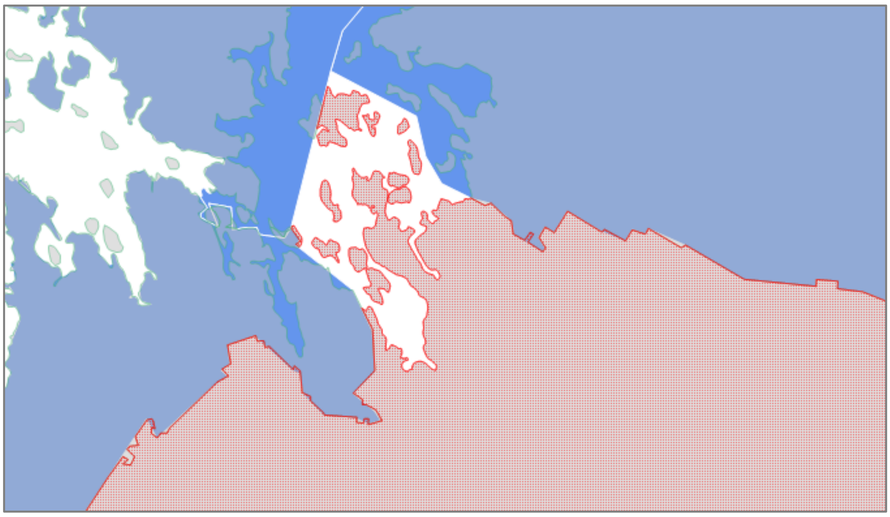
Annex Map 9: Keminmaa (ESPON Census 2011 mapping layer) is contiguous to coast



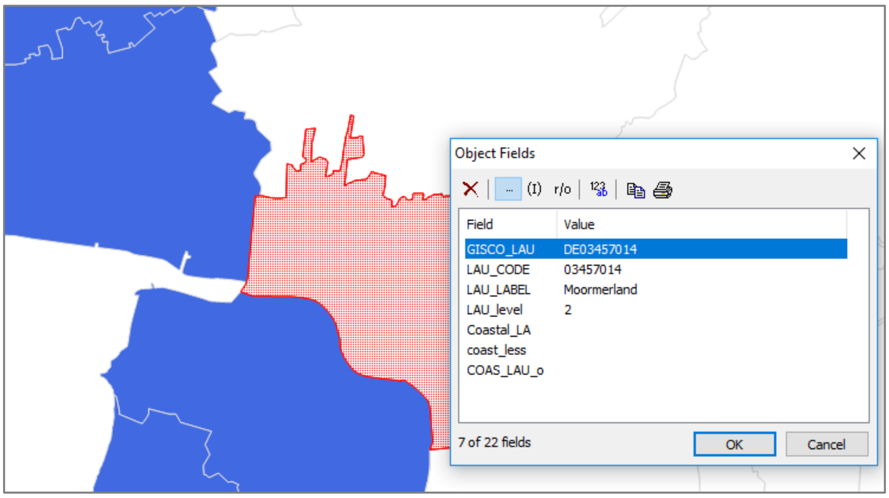
Annex Map 10: Pedersören kunta in FI (TERCET mapping layer) is not contiguous



Annex Map 11 : Pedersören kunta (ESPON Census 2011 mapping layer) is contiguous to coast



Annex Map 12 : Moormerland in DE is contiguous to coast in both layers (TERCET & ESPON Census 2011) but not coded as coast



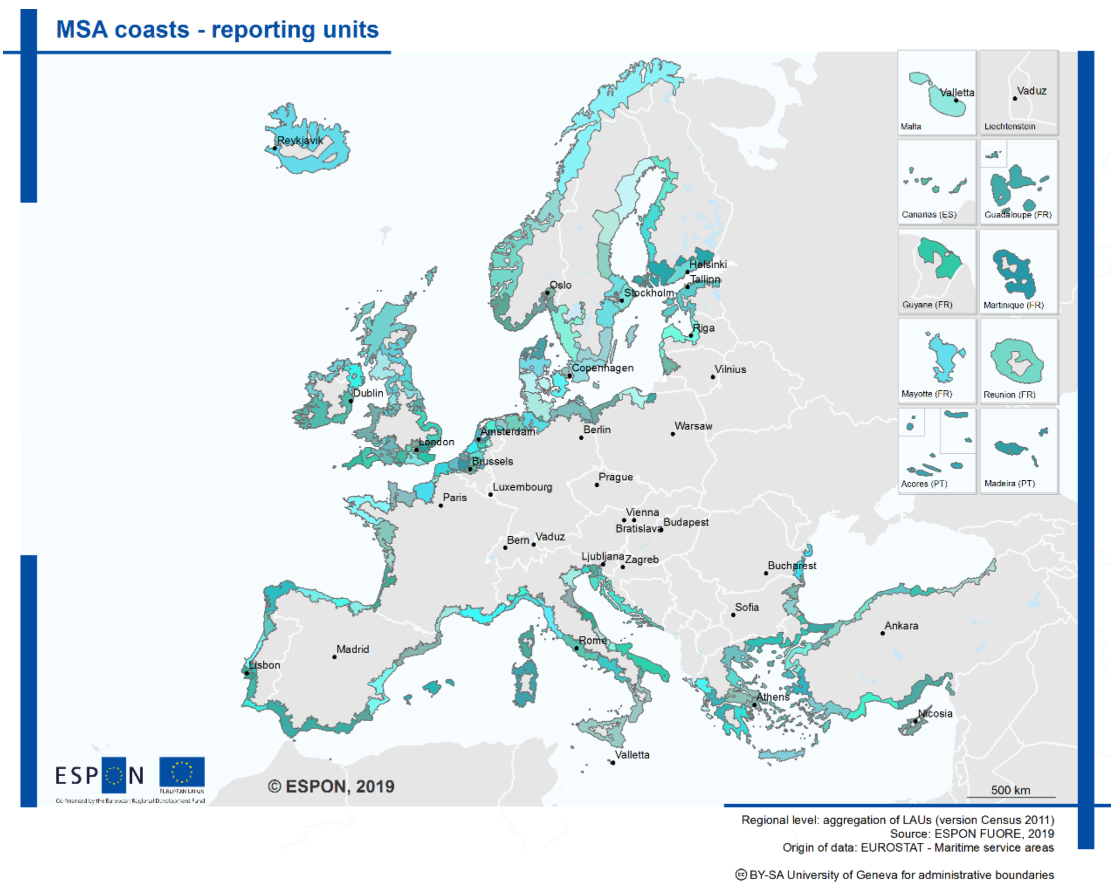
2.5. Maritime service areas (MSA)

MSA are the areas that can be reached within a given travelling time, starting from a location at the coast and using the existing transport network. Travelling time depends on the points of interactions on the shoreline (large ports, small ports and coastal settlements). The complete coverage of service areas from any port was calculated for each LAU. This resulted in a set of 26 700 LAU with more than 5 % of their surface area covered by a service area.

However, Croatia, Island, Norway, Turkey and the French island of Mayotte are not delineated in the MSA file. For these countries MSA file has been completed with the GEOSPECS coastal LAUs based on 45 minutes driving time-distance to coast (more than 50% of their surface accessible in less than 45 minutes) + contiguous. While the methodology is based on commuting-time and respects the 45 minutes commuting time to large ports used in the MSA, it must be kept in mind that despites results are similar for these countries, they must be compared with caution.

MSA provide therefore a complementary information on costal functionalities (population, activities located within commuting time from coast), whereas TERCET coast is based on a geographic buffer of 10 kilometres.

Annex Map 13 : MSA completed with GEOSPECS coasts, as delineated in ESPON LAU Census 2011 nomenclature

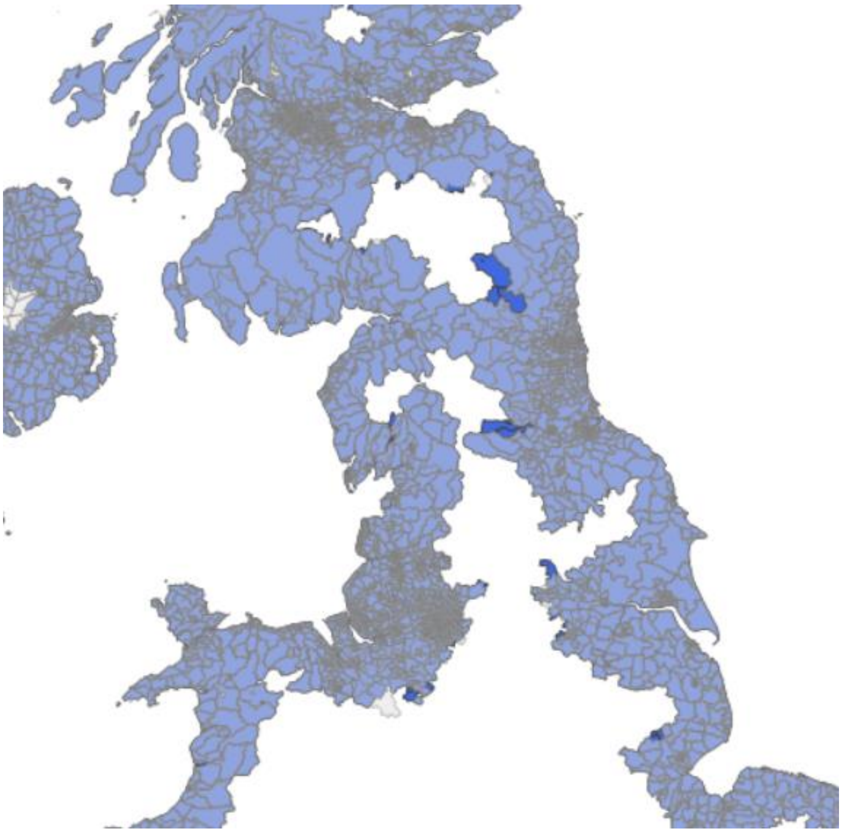


Delineation issues

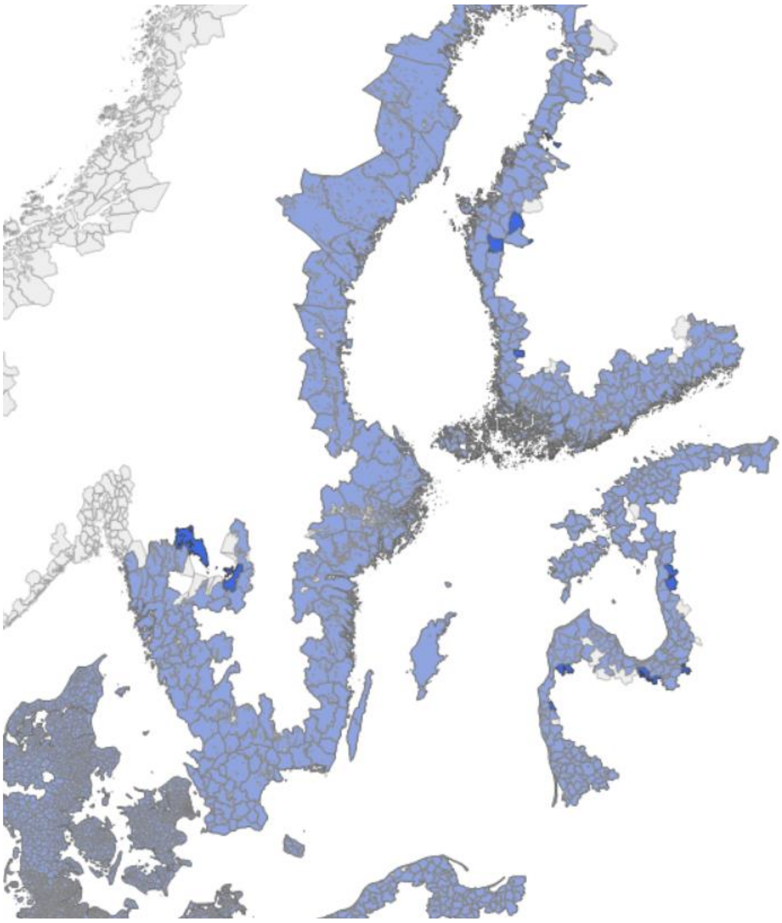
No changes in delineation in most countries.

Few and minor issues here and there resulting from changes in LAU boundaries. Major changes are reported in the next three maps.

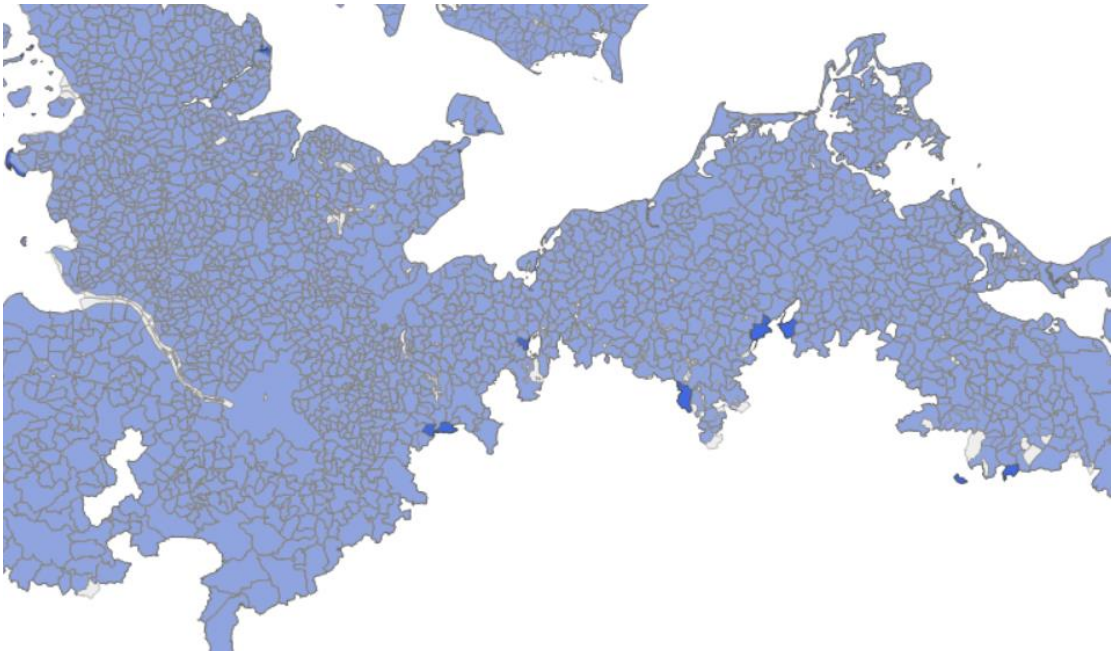
Annex Map 14 : MSA delineations in central UK



Annex Map 15 : MSA delineations around the Baltic Sea



Annex Map 16 : MSA delineation in Northern Germany



3. Overlapping types of territories

GEOSPECS borders and TERCET FUA are overlapping areas, as a result that territories delineated on functionality (e.g. commuting, accessibility) are likely to provide overlapping delineations.

In cases where a LAU belongs to two (or more) FUAs, coasts or borders, it has received each of the codes of the functional units it belongs to. These codes are separated by a ",". In order to re-aggregate LAUs to produce all the overlapping functional regions, following steps must be implemented:

1. Follow steps 1-8 presented for non-overlapping entities (cf. section 1), to transfer delineations from as delineated in their original nomenclature into ESPON LAU Census 2001 nomenclature.
2. Step 9: "Select containing" all LAU units having the code of the of the functional unit
3. Step 10 "Copy and paste" selected LAUs into a dedicated shapefile.
4. Step 11 "Dissolve" them to obtain the FUA/territory
5. Repeat steps 9-11 for all units composing the Border or FUA layer.

3.1. Border Areas (BDA)

ESPON GEOSPECS borders are based on driving time-distance to border, taking account therefore of the road accessibility to delineate various buffer zones of mutual influence. As borders are of multidimensional nature, with a variety of associated border effects, they must be understood as "interface areas" of variable geometries. Considering multi-dimensional territorial perspectives of Interreg post 2020 presented by Nathalie Verschelde from DG REGIO (ESPON seminar, Vienna), service provider has decided to select two dimensions among the variety of GEOSPECS border delineations:

- 45 minutes / max. 90km: corresponding to a daily potential commuting area, proximity cooperation.
- 90 minutes / max. 180km + contiguity: corresponding to larger potential border influence effects.

Delineation of borders is generated at LAU level, then aggregated into "border areas" (BDA). It is about zones made of LAUs which (individually) have >50% area located within 45/90 minutes accessibility to a border. Driving time-distances have been calculated from each border. As some LAUs can be within accessible time distance of more than one national border, borders are consequently overlapping objects.

Measures of time-distance are based on the "friction surface" of cells (grids from 250*250m to 5*5km). Friction is defined by the average travel time required to cross a cell in all directions, taking into account road and off-road travel speeds.

GEOSPECS used the EuroRegionalMap V30-31 road model (TRAN_Road_LI_250K) for all countries with the exception of Western Balkans, Bulgaria, Turkey and all outermost regions for which they were missing. In these countries, the Eurogeographics road model has been replaced by Open Street Map. The limited number of secondary roads in Open Street Map, however, creates some bias in the analyses that must be kept in mind when interpreting results. Time-distance and the resulting “isochrone areas” covering the entire GEOSPECS space are used in the calculation of accessibility to borders.

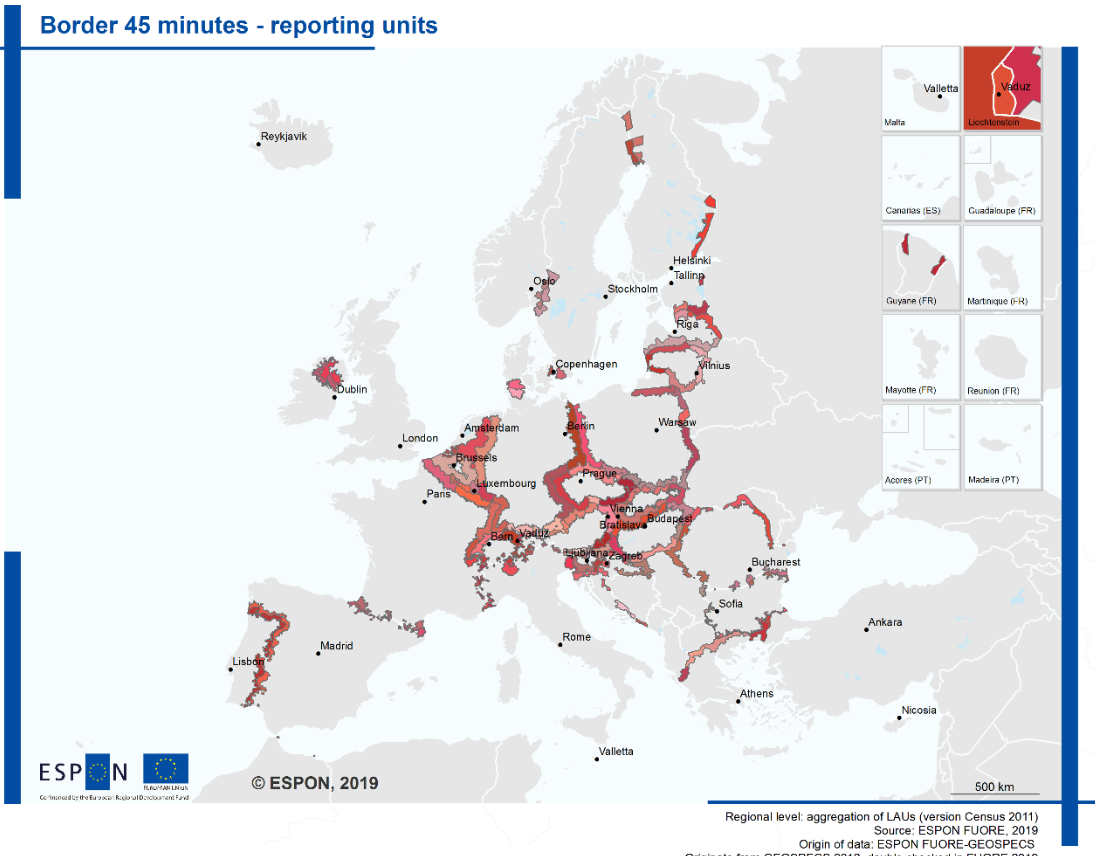
BDA 45 minutes

Delineation methodology:

- Areas made of the aggregation of LAUs where more than 50% of LAU surface is accessible within 45 min driving time from a border (= ±1 hour on Google Map).

It stands as proxy for cross-border commuting functionality. In that sense BDA 45 shows gaps over mountain massifs and other border areas with poor road connection.

Annex Map 17 : 45 min accessibility to border, as delineated in ESPON LAU Census 2011 nomenclature

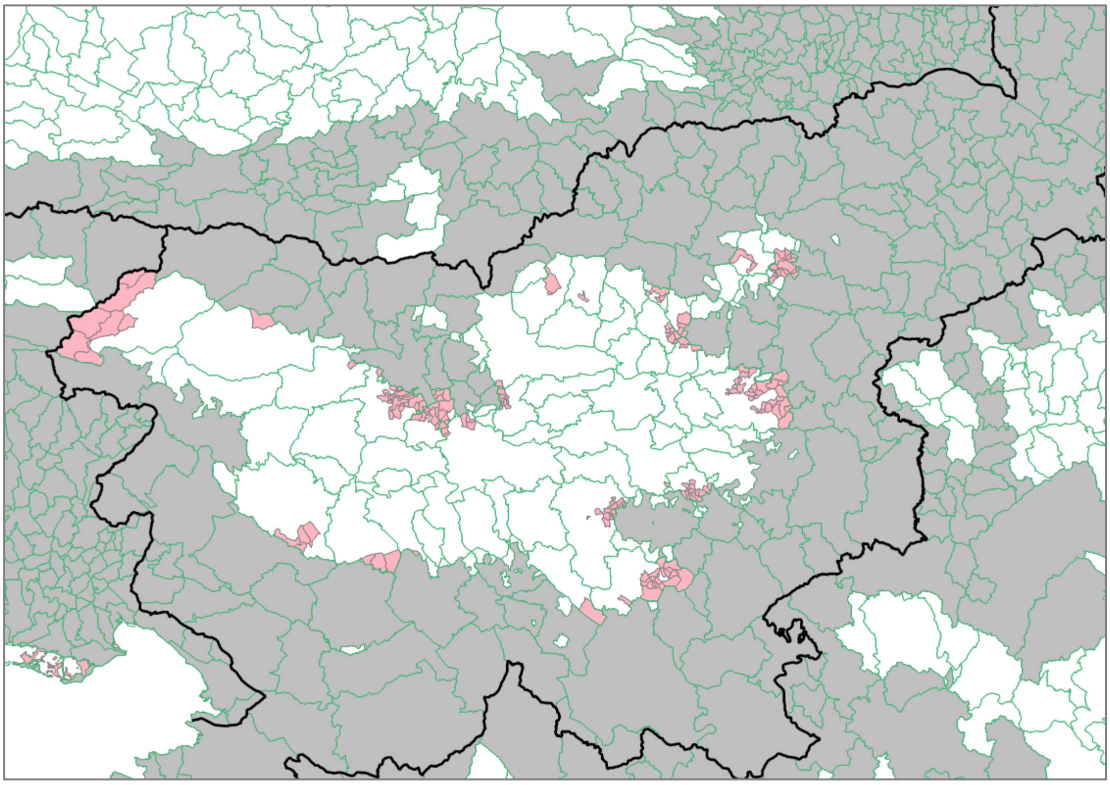


Regional level: aggregation of LAUs (version Census 2011)
Source: ESPON FUORE, 2019
Origin of data: ESPON FUORE-GEOSPECS
Originate from GEOSPECS 2012, double checked in FUORE 2019
© BY-SA University of Geneva for administrative boundaries

Delineation issues

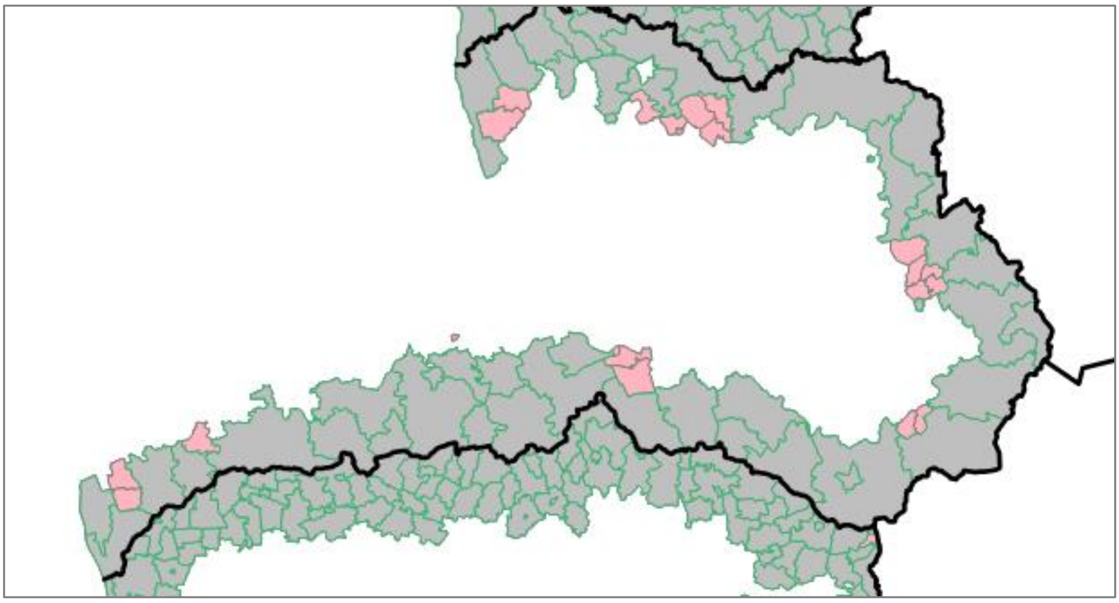
In Slovenia the fact the reduction of the number of LAUs from 6023 units by the time of GEOSPECS to only 210 units by 2011 nomenclature induces some reduction of BDA 45 area.

Annex Map 18 : BDA45 changes in delineation: Slovenia



In Latvia, a change in LAU boundaries/nomenclatures (merging of municipalities into larger entities) induces some reduction of BDA 45 area.

Annex Map 19 : BDA45 changes in delineation: Latvia



BDA 90 minutes + contiguous

Delineation methodology:

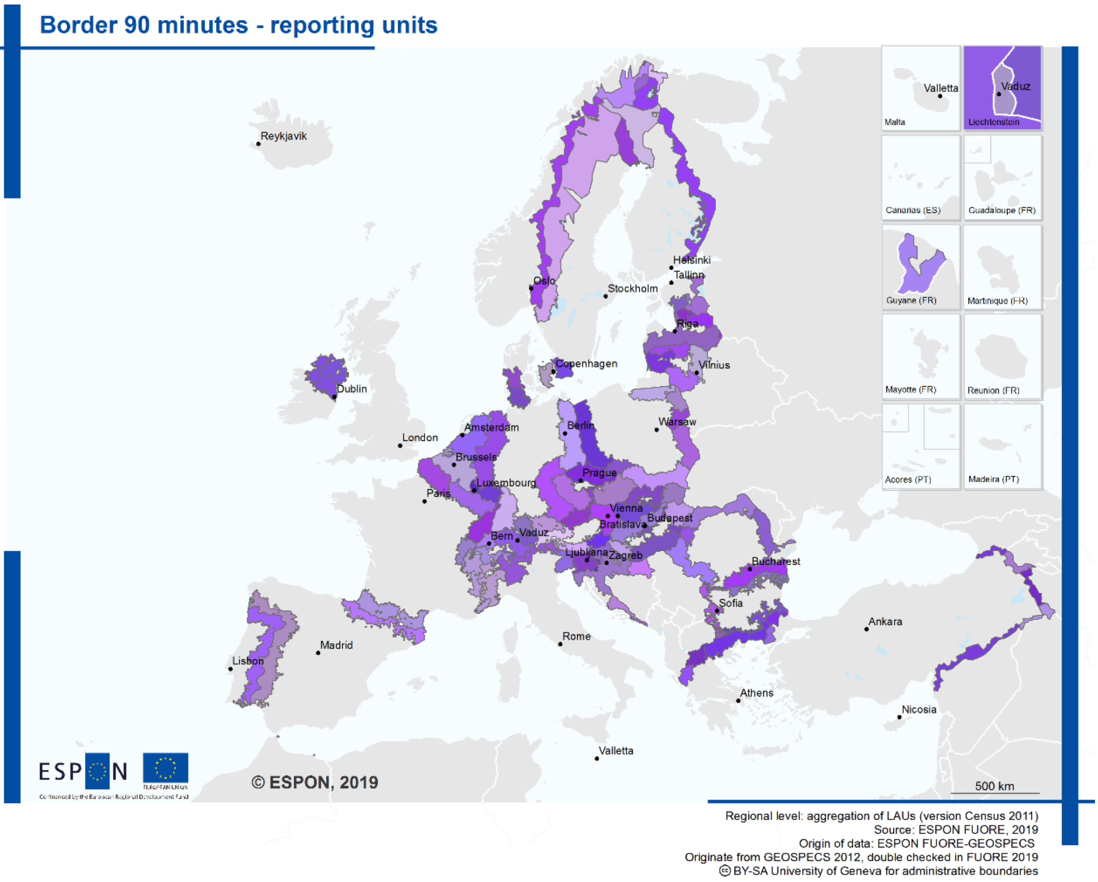
- Areas made of the aggregation LAUs where more than 50% of LAU surface is accessible within 90 min driving from a border (= ±2 hour on Google Map).
- LAUs where less than 50% of their surface is accessible within 90 min driving from a border, but that are contiguous to that border, have been added.

BDA 90 provide a larger perspective to potential border influence. All LAUs contiguous to a border have been added to this category, as contiguity (even in the case of insufficient road accessibility to border) can lead to cooperation on the example of cross-border nature conservation areas.

LAUs have been coded manually as BDA90 with the following method:

- Upload border lines
- Select LAUs “touching” border lines
- Select from previous selection “LAU not coded as BDA90”
- Code these LAU as BDA90

Annex Map 20 : 90 min accessibility + contiguity to border, as delineated in ESPON LAU Census 2011 nomenclature

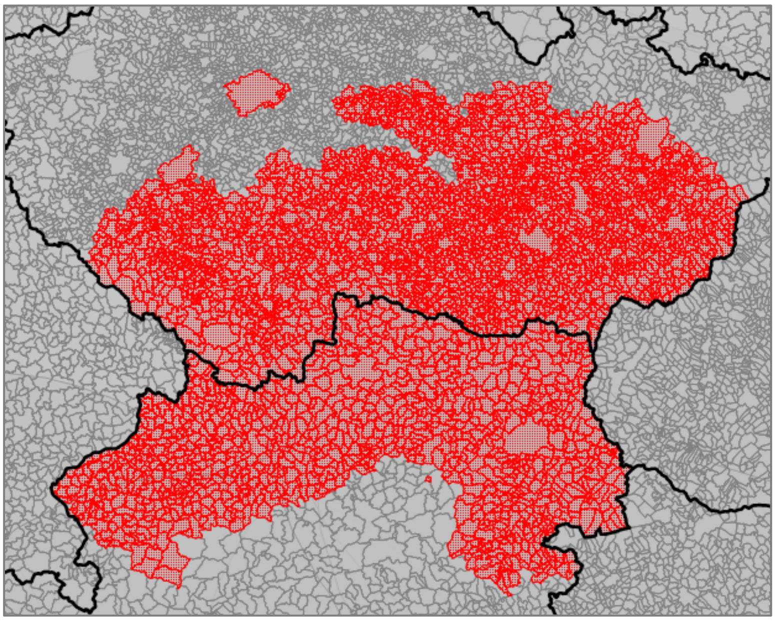


Delineation issues

They are a few LAUs that seems obviously miscoded in the original GEOSPECS typology. In these cases, basic geographical expertise on the situation, completed with travel time estimation via Goole map, has been carried on.

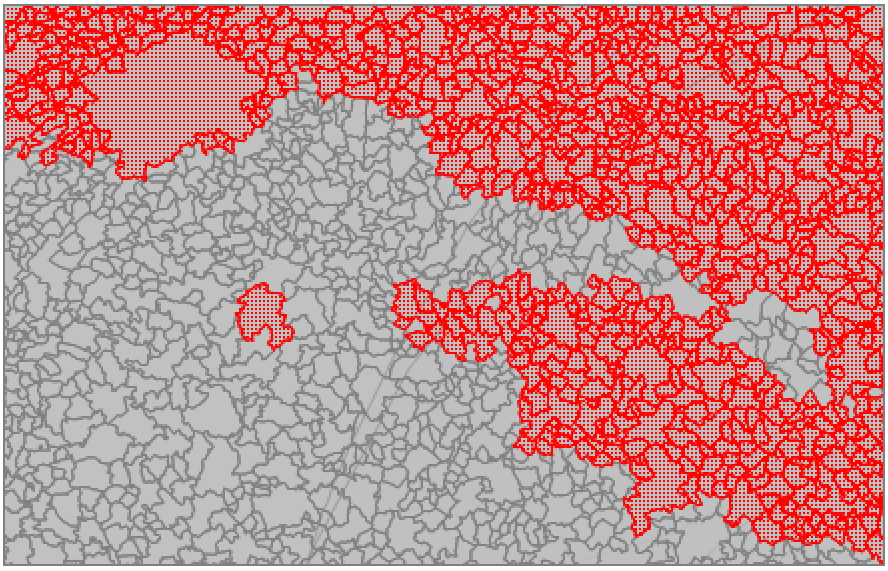
The case of Prague that is included in Czech-Austrian 90 minutes border is obviously not consistent. The LAU appears as an "island", far away from the other BDA 90 LAUs. It has been retracted from the selection of BDA 90.

Annex Map 21 : BDA 90, issue of Prague



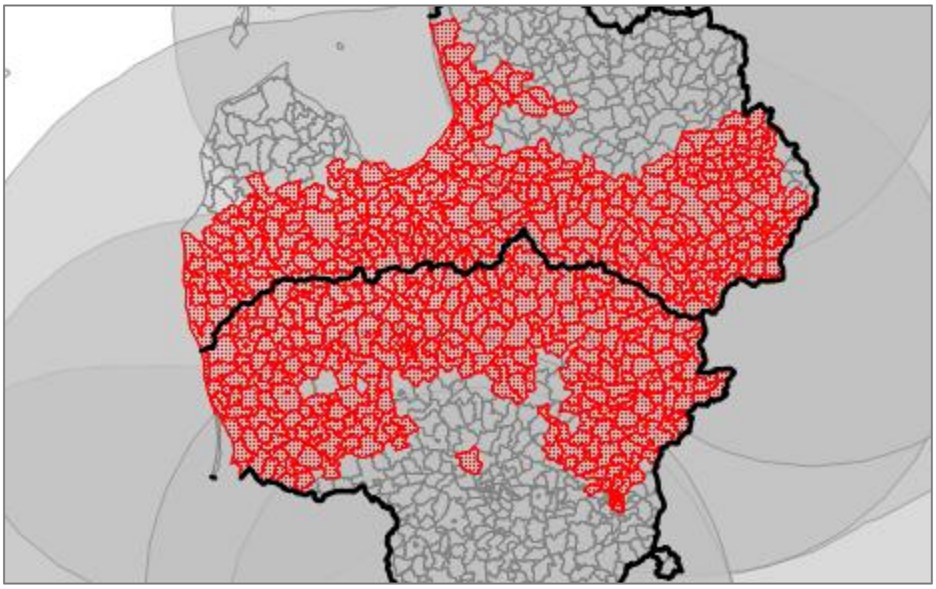
The case of Bystrice that is included in Polish-Czech BDA 90r is obviously not consistent. The LAU appears as an "island", far away from the other BDA 90 LAUs. It has been retracted from the selection of BDA 90.

Annex Map 22 : BDA 90, issue of Bystrice



The case of Babtai that is included in Latvian-Lithuanian 90 minutes border is obviously not consistent. The LAU appears as an “island”, far away from the other BDA 90 LAUs. It has been retracted from the selection of BDA 90.

Annex Map 23 : BDA 90, issue of Babtai

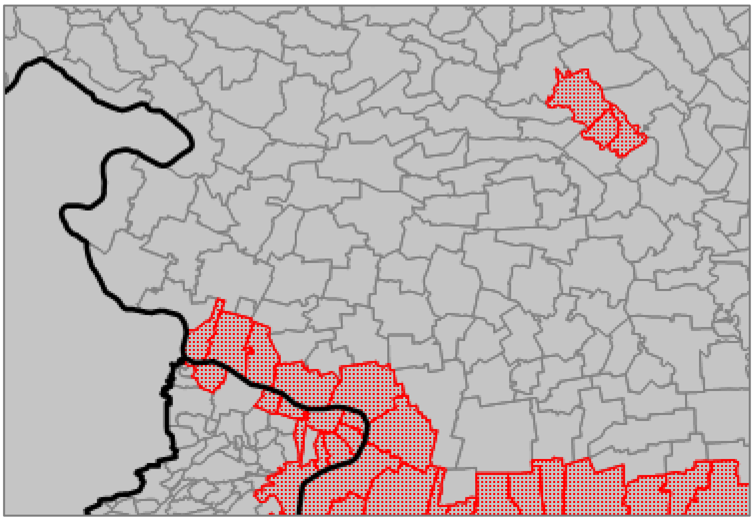


The case of three Romanian LAUs, located on the E79 road and included in Romanian-Bulgarian 90 minutes border is obviously consistent (cross-checked with Google Map itineraries).

CS_LAU2-2008_Su_BDA-fixed Table *

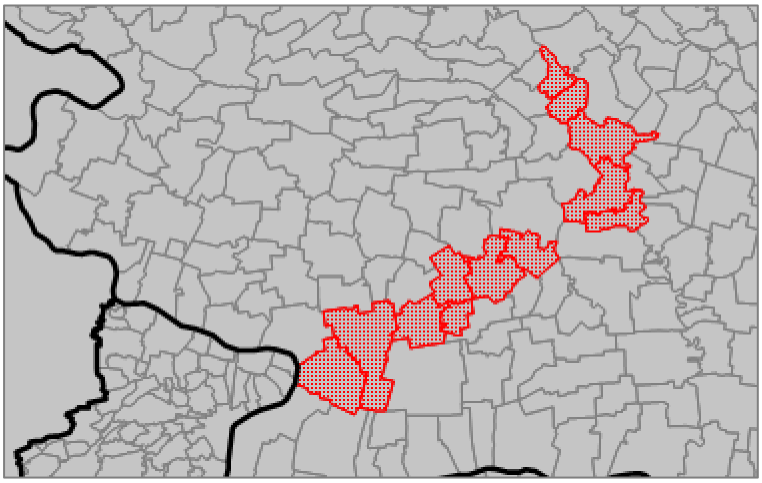
CNT...	COMM_ID	SHAPE_Area	NAME_ASCI
RO	RO28074875	0.002729	Cotofenii din Fata
RO	RO28071055	0.007104	Bradesti
RO	RO28070566	0.003124	Almaj

Annex Map 24 : BDA 90, issue of Romanian-Bulgarian border



Therefore, LAU located along this road and closer to the border should be integrated into the BDA 90 as well. After checking access time with Google Map, they have been included in the selection of BDA 90.

Annex Map 25 : BDA 90, issue of Romanian-Bulgarian border: integration of 12 LAUs



3.2. Functional urban areas (FUA)

According to TERCET, a FUA consists of a city (urban centre of more than 50'000 inhabitants) and its commuting zone whose labour market is highly integrated with the city. In some places with high urban concentration, FUA can be overlapping. FUA delineation evolve over time. The latest version "2015-2018 Urban Audit delineations" (based on the census commuting data from 2011 and onwards) has been used for the present project.

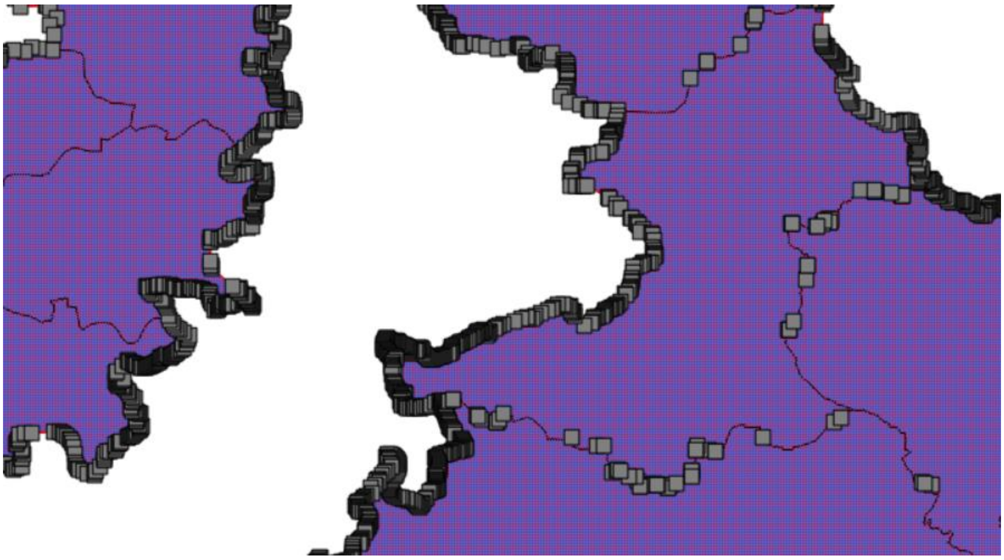
As the delineation of FUA 2015-2018 is made with the same precision as the ESPON LAU Census 2011 layer (1:100:000) the initial attempt was to use directly the Eurostat FUA mapping layer (= no transfer into ESPON LAU Census 2011 nomenclature) as input for ESPON Spatial Multidimensional DB. The Eurostat layer is however not seamless when dissolved into a single object[12]. A significant number of topological errors result into branched objects that make the cleaning task too lengthy and arduous.

[12] cf. map 26 next page: when dissolved into a single object, FUAs (FUA_RG_100K_2015_2018) create many gaps and overlaps between contiguous FUA.

The risk (a slight risk though) is that a cell from a ancillary grid data would not be counted as it would fall into one of these gap

More generally non-seamless shapefile do not allow unit aggregation or creation of cartograms when making cartographic processing

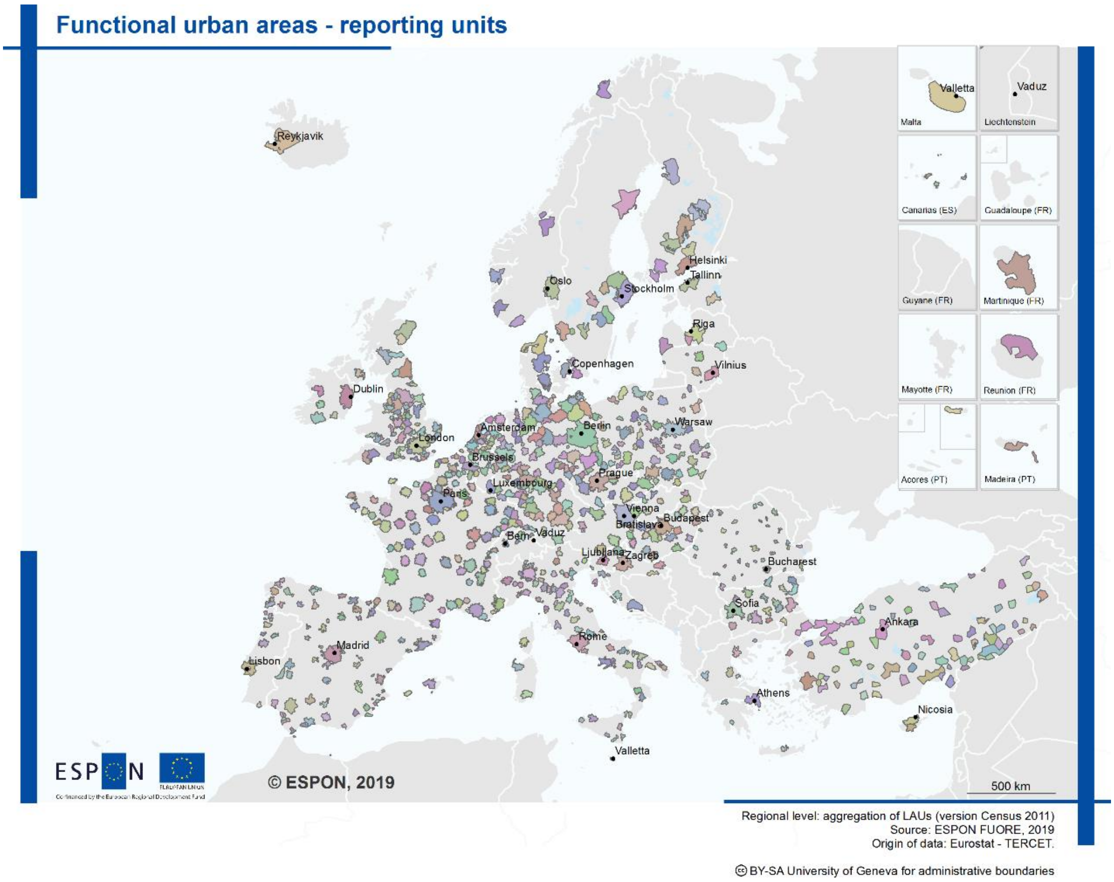
Annex Map 26 : Sliver polygons hidden in the Eurostat FUA 2015-2018 shapefile



Therefore, decision was made to pass FUA codes into ESPON LAU Census 2011 map that is seamless. The result is a delineation of FUA 2015-2018, translated into LAU nomenclature 2011 with the following methodology:

1. Implement steps 1-8 presented for non-overlapping entities (cf. section 1), to transfer delineations from Eurostat FUA 2015-2018 to into ESPON LAU Census 2001 nomenclature.
2. Step 9: Make a table join from the ESPON LAU Census 2011 centroids to the original ESPON LAU Census 2011 polygonal layer:
   - ⇨ Repeat step 9 three times for:
     1) LAUs belonging to non-overlapping FUA
     2) LAUs belonging to overlapping FUA (first overlap)
     3) LAUs belonging to overlapping FUA (for cases where there are two overlaps, for the second overlap)
   - ⇨ Code the LAUs with FUA_codes (separated by “, “in cases of one or more overlaps)

Annex Map 27 : FUA 2015-2018, as delineated in ESPON LAU Census 2011 nomenclature



Delineation issues

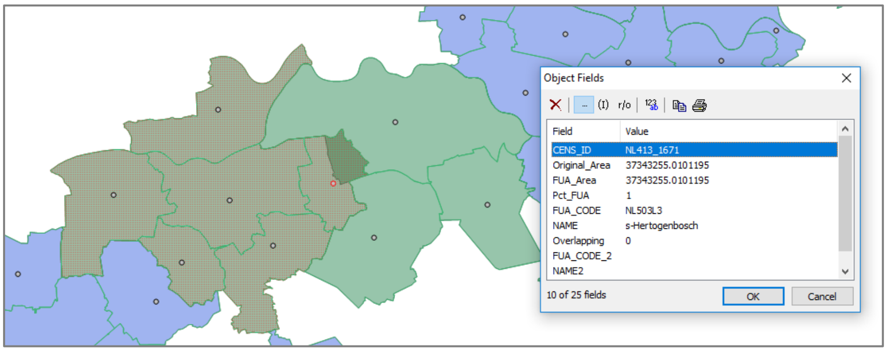
There is only issue concerns the Dutch FUA of “s-Hertogenbosch (NL503L3)” that overlaps with the FUA “Oss (NL521L3)”.

The LAU of Maasdonk belongs to both FUAs:

- 1/3 of its area is located on “Oss” FUA
- 2/3 of its area is located on “s-Hertogenbosch” FUA

⇨ Therefore, it has been coded as “s-Hertogenbosch (NL503L3)” only, as it is impossible in that case to tile the LAU while preserving the ESPON LAU Census 2011 nomenclature.

Annex Map 28 : FUA 2015-2018, issue Dutch LAU of Maasdonk

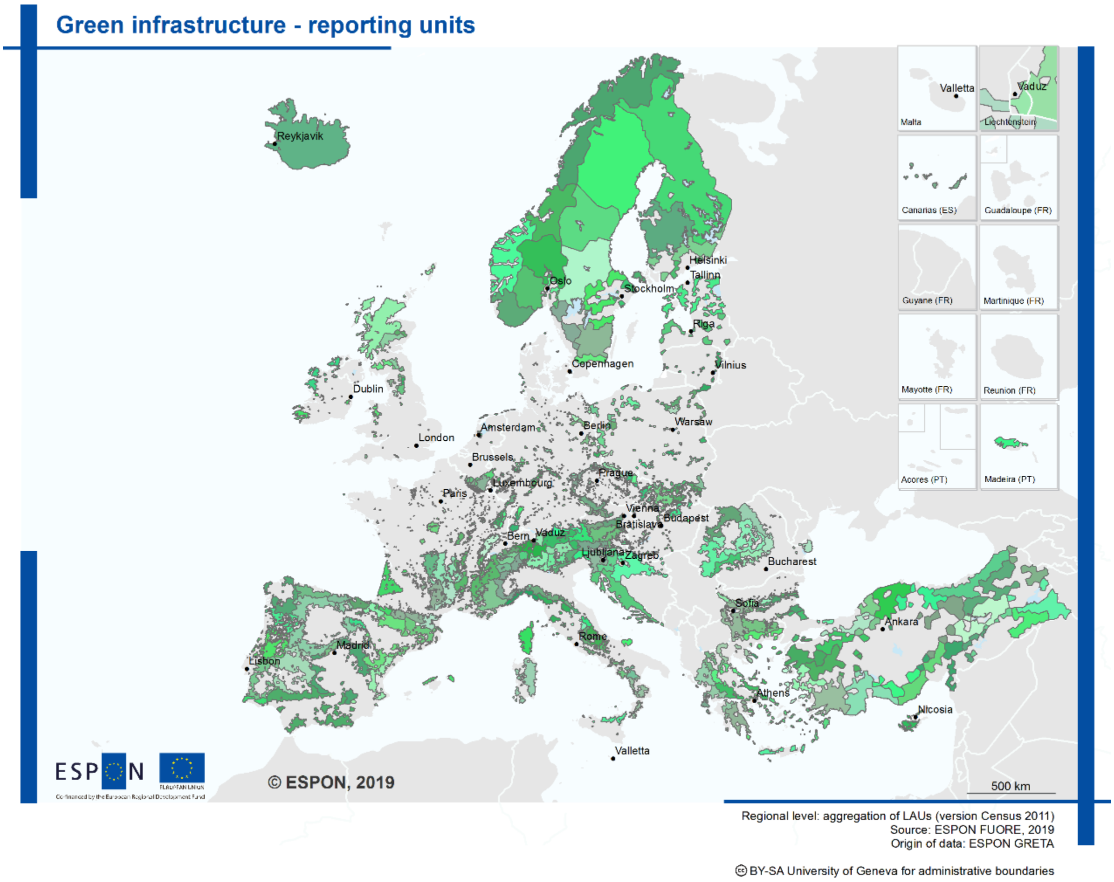


4. Regions of high Green Infrastructure (GI) potential

Taking advantage of the outcomes from the ESPON GRETA project, service provider has delineated areas with an elevated potential regarding GI to be used as a new reporting unit. The source data used is the “GI Potential Network” map produced by GRETA, which is a 100x100m resolution binary raster stating if the cell has GI potential or not. The methodology to transfer the grid data into ESPON LAU census 2011 can be summarised as follow:

- Calculation of GI Potential Network average (mean) value by LAU.
- Definition of a > 70% threshold for the average value to classify LAUs as high GI potential.
- Dissolving selected LAUs by country to create functional regions with high GI potential by country.

Annex Map 29 : GRETA green infrastructure, as delineated in ESPON LAU Census 2011 nomenclature



5. Conclusion

The above detailed delineation tasks have allowed service provider to reach three objectives of whom two are directly linked to the ESPON Spatial Multidimensional DB and the third one to wider capacity in ESPON LAU mapping capacities.

The first objective was to bring into coherence a variety of functional regions delineated in various LAU nomenclatures into a single harmonised nomenclature: the ESPON LAU Census 2011. The precision of the associated mapping layer will allow the disaggregation/re-aggregation of NUTS indicators into the ESPON SMD.

As the result of the first objective, the project has built an attribute table containing all delineations at LAU level for the ESPON LAU Census 2011 nomenclature. Based on this file, the next objective will be to create shapefiles for the FUAs and other functional regions in the ESPON LAU Voronoi layer. Simplified representation will then feed with dedicated mapping layers the web-tool of the project.

Finally, the third objective is somehow a collateral effect of the job accomplished in the framework of the present project. Indeed, ESPON LAU mapping layers have gained a comprehensive set of delineation attributes on several functional, geographical and administrative objects. Included in ESPON database, these delineations are likely to allow future ESPON projects to produce perspectives on FUA and other functional regions.

Annex 2: List of ESPON 2020 Base Indicators eventually estimated by FUORE

Obj_id	Code	Name
3	d_f_0-4	Female deaths in age group 0-4
4	d_f_5-9	Female deaths in age group 5-9
5	d_f_10-14	Female deaths in age group 10-14
6	d_f_15-19	Female deaths in age group 15-19
7	d_f_20-24	Female deaths in age group 20-24
8	d_f_25-29	Female deaths in age group 25-29
9	d_f_30-34	Female deaths in age group 30-34
10	d_f_35-39	Female deaths in age group 35-39
11	d_f_40-44	Female deaths in age group 40-45
12	d_f_45-49	Female deaths in age group 45-49
13	d_f_50-54	Female deaths in age group 50-54
14	d_f_55-59	Female deaths in age group 55-59
15	d_f_60-64	Female deaths in age group 60-64
16	d_f_65-69	Female deaths in age group 65-69
17	d_f_70-74	Female deaths in age group 70-74
18	d_f_75-79	Female deaths in age group 75-79
19	d_f_80-84	Female deaths in age group 80-84
20	d_f_85+	Female deaths in age group 85+
21	d_m_0-4	Male deaths in age group 0-4
22	d_m_5-9	Male deaths in age group 5-9
23	d_m_10-14	Male deaths in age group 10-14
24	d_m_15-19	Male deaths in age group 15-19
25	d_m_20-24	Male deaths in age group 20-24
26	d_m_25-29	Male deaths in age group 25-29
27	d_m_30-34	Male deaths in age group 30-34
28	d_m_35-39	Male deaths in age group 35-39
29	d_m_40-44	Male deaths in age group 40-44
30	d_m_45-49	Male deaths in age group 45-49
31	d_m_50-54	Male deaths in age group 50-54
32	d_m_55-59	Male deaths in age group 55-59
33	d_m_60-64	Male deaths in age group 60-64
34	d_m_65-69	Male deaths in age group 65-69
35	d_m_70-74	Male deaths in age group 70-74
36	d_m_75-79	Male deaths in age group 75-79
37	d_m_80-84	Male deaths in age group 80-84
38	d_m_85+	Male deaths in age group 85+
40	d_t_0-4	Total deaths in age group 0-4
41	d_t_5-9	Total deaths in age group 5-9
42	d_t_10-14	Total deaths in age group 10-14
43	d_t_15-19	Total deaths in age group 15-19

Obj_id	Code	Name
44	d_t_20-24	Total deaths in age group 20-24
45	d_t_25-29	Total deaths in age group 25-29
46	d_t_30-34	Total deaths in age group 30-34
47	d_t_35-39	Total deaths in age group 35-39
48	d_t_40-44	Total deaths in age group 40-44
49	d_t_45-49	Total deaths in age group 45-49
50	d_t_50-54	Total deaths in age group 50-54
51	d_t_55-59	Total deaths in age group 55-59
52	d_t_60-64	Total deaths in age group 60-64
53	d_t_65-69	Total deaths in age group 65-69
54	d_t_70-74	Total deaths in age group 70-74
55	d_t_75-79	Total deaths in age group 75-79
56	d_t_80-84	Total deaths in age group 80-84
57	d_t_85+	Total deaths in age group 85+
60	d_f_0-4	Female deaths in age group 0-4
61	d_f_5-9	Female deaths in age group 5-9
62	d_f_10-14	Female deaths in age group 10-14
63	d_f_15-19	Female deaths in age group 15-19
64	d_f_20-24	Female deaths in age group 20-24
65	d_f_25-29	Female deaths in age group 25-29
66	d_f_30-34	Female deaths in age group 30-34
67	d_f_35-39	Female deaths in age group 35-39
68	d_f_40-44	Female deaths in age group 40-45
69	d_f_45-49	Female deaths in age group 45-49
70	d_f_50-54	Female deaths in age group 50-54
71	d_f_55-59	Female deaths in age group 55-59
72	d_f_60-64	Female deaths in age group 60-64
73	d_f_65-69	Female deaths in age group 65-69
74	d_f_70-74	Female deaths in age group 70-74
75	d_f_75-79	Female deaths in age group 75-79
76	d_f_80-84	Female deaths in age group 80-84
77	d_f_85+	Female deaths in age group 85+
78	d_m_0-4	Male deaths in age group 0-4
79	d_m_5-9	Male deaths in age group 5-9
80	d_m_10-14	Male deaths in age group 10-14
81	d_m_15-19	Male deaths in age group 15-19
82	d_m_20-24	Male deaths in age group 20-24
83	d_m_25-29	Male deaths in age group 25-29
84	d_m_30-34	Male deaths in age group 30-34
85	d_m_35-39	Male deaths in age group 35-39
86	d_m_40-44	Male deaths in age group 40-44
87	d_m_45-49	Male deaths in age group 45-49
88	d_m_50-54	Male deaths in age group 50-54

Obj_id	Code	Name
89	d_m_55-59	Male deaths in age group 55-59
90	d_m_60-64	Male deaths in age group 60-64
91	d_m_65-69	Male deaths in age group 65-69
92	d_m_70-74	Male deaths in age group 70-74
93	d_m_75-79	Male deaths in age group 75-79
94	d_m_80-84	Male deaths in age group 80-84
95	d_m_85+	Male deaths in age group 85+
97	d_t_0-4	Total deaths in age group 0-4
98	d_t_5-9	Total deaths in age group 5-9
99	d_t_10-14	Total deaths in age group 10-14
100	d_t_15-19	Total deaths in age group 15-19
101	d_t_20-24	Total deaths in age group 20-24
102	d_t_25-29	Total deaths in age group 25-29
103	d_t_30-34	Total deaths in age group 30-34
104	d_t_35-39	Total deaths in age group 35-39
105	d_t_40-44	Total deaths in age group 40-44
106	d_t_45-49	Total deaths in age group 45-49
107	d_t_50-54	Total deaths in age group 50-54
108	d_t_55-59	Total deaths in age group 55-59
109	d_t_60-64	Total deaths in age group 60-64
110	d_t_65-69	Total deaths in age group 65-69
111	d_t_70-74	Total deaths in age group 70-74
112	d_t_75-79	Total deaths in age group 75-79
113	d_t_80-84	Total deaths in age group 80-84
114	d_t_85+	Total deaths in age group 85+
147	pop_f_0-14	Female population with age group 0-14
148	pop_f_15-64	Female population in age group 15- 64
149	pop_f_65+	Female population in age group 65+
150	pop_m_0-14	Male population with age group 0-14
151	pop_m_15-64	Male population in age group 15- 64
152	pop_m_65+	Male population in age group 65+
154	pop_t_0-14	Total population with age group 0-14
155	pop_t_15-64	Total population in age group 15- 64
156	pop_t_65+	Total population in age group 65+
159	pop_f_0-4	Female population in age group 0-4
160	pop_f_5-9	Female population in age group 5-9
161	pop_f_10-14	Female population in age group 0-16
162	pop_f_15-19	Female population in age group 15-19
163	pop_f_20-24	Female population in age group 20-24
164	pop_f_25-29	Female population in age group 25-29
165	pop_f_30-34	Female population in age group 30-34
166	pop_f_35-39	Female population in age group 35-39
167	pop_f_40-44	Female population in age group 40-44

Obj_id	Code	Name
168	pop_f_45-49	Female population in age group 45-49
169	pop_f_50-54	Female population in age group 50-54
170	pop_f_55-59	Female population in age group 55-59
171	pop_f_60-64	Female population in age group 60-64
172	pop_f_65-69	Female population in age group 65-69
173	pop_f_70-74	Female population in age group 70-74
174	pop_f_75-79	Female population in age group 75-79
175	pop_f_80-84	Female population in age group 80-84
176	pop_f_85+	Female population in age group 85+
177	pop_m_0-4	Male population in age group 0-4
178	pop_m_5-9	Male population in age group 5-9
179	pop_m_10-14	Male population in age group 10-14
180	pop_m_15-19	Male population in age group 15-19
181	pop_m_20-24	Male population in age group 20-24
182	pop_m_25-29	Male population in age group 25-29
183	pop_m_30-34	Male population in age group 30-34
184	pop_m_35-39	Male population in age group 35-39
185	pop_m_40-44	Male population in age group 40-44
186	pop_m_45-49	Male population in age group 45-49
187	pop_m_50-54	Male population in age group 50-54
188	pop_m_55-59	Male population in age group 55-59
189	pop_m_60-64	Male population in age group 60-64
190	pop_m_65-69	Male population in age group 65-69
191	pop_m_70-74	Male population in age group 70-74
192	pop_m_75-79	Male population in age group 75-79
193	pop_m_80-84	Male population in age group 80-84
194	pop_m_85+	Male population in age group 85+
196	pop_t_0-4	Total population in age group 0-4
197	pop_t_5-9	Total population in age group 5-9
198	pop_t_10-14	Total population in age group 10-14
199	pop_t_15-19	Total population in age group 15-19
200	pop_t_20-24	Total population in age group 20-24
201	pop_t_25-29	Total population in age group 25-29
202	pop_t_30-34	Total population in age group 30-34
203	pop_t_35-39	Total population in age group 35-39
204	pop_t_40-44	Total population in age group 40-44
205	pop_t_45-49	Total population in age group 45-49
206	pop_t_50-54	Total population in age group 50-54
207	pop_t_55-59	Total population in age group 55-59
208	pop_t_60-64	Total population in age group 60-64
209	pop_t_65-69	Total population in age group 65-69
210	pop_t_70-74	Total population in age group 70-74
211	pop_t_75-79	Total population in age group 75-79

Obj_id	Code	Name
212	pop_t_80-84	Total population in age group 80-84
213	pop_t_85+	Total population in age group 85+
215	Ag_In_t	Ageing index of total population
216	Ag_In_F	Ageing index of female population
217	Ag_In_M	Ageing index of male population
220	f_total	Percentage of female population respect the total
221	m_total	Percentage of male population respect the total
223	pop_t_oldage	Percentage between Total population older than 65 years of age and between 15-65 years of age, measured on 1 January
224	pop_f_oldage	Percentage between Female population older than 65 years of age and between 15-65 years of age, measured on 1 January
225	pop_m_oldage	Percentage between Male population older than 65 years of age and between 15-65 years of age, measured on 1 January
228	unempl_males_15-24	Unemployment male population in age group 15-24
229	unempl_males_15-74	Unemployment male population in age group 15-74
230	unempl_males_15-O	Unemployment male population in age group 15-0
231	unempl_males_20-64	Unemployment male population in age group 20-64
232	unempl_males_25-O	Unemployment male population in age group 25-0
233	unempl_females_15-24	Unemployment female population in age group 15-24
234	unempl_females_15-74	Unemployment female population in age group 15-74
235	unempl_females_15-O	Unemployment female population in age group 15-0
236	unempl_females_20-64	Unemployment female population in age group 20-64
237	unempl_females_25-O	Unemployment female population in age group 25-0
239	unempl_total_15-24	Unemployment total population in age group 15-24
240	unempl_total_15-74	Unemployment total population in age group 15-74
241	unempl_total_15-O	Unemployment total population in age group 15-0
242	unempl_total_20-64	Unemployment total population in age group 20-64
243	unempl_total_25-O	Unemployment total population in age group 25-0
246	empl_males_15-24	Male employment rate for age group 15-24
247	empl_males_25-64	Male employment rate for age group 25-64
248	empl_males_65+	Male employment rate for age group 65+
249	empl_females_15-24	Female employment rate for age group 15-24
250	empl_females_25-64	Female employment rate for age group 25-64
251	empl_females_65+	Female employment rate for age group 65+
253	empl_total_15-24	Total employment rate for age group 15-24
254	empl_total_25-64	Total employment rate for age group 25-64

Obj_id	Code	Name
255	empl_total_65+	Total employment rate for age group 65+
258	emp_m_15+	Male population in employment, age group 15 years and older
259	emp_m_65+	Male population in employment, age group 65 years and older
260	emp_m_15-24	Male population in employment, age group 15 to 24 years
261	emp_m_15-64	Male population in employment, age group 15 to 64 years
262	emp_m_20-64	Male population in employment, age group 20 to 64 years
263	emp_m_55-64	Male population in employment, age group 55 to 64 years
264	emp_f_15+	Female population in employment, age group 15 years and older
265	emp_f_65+	Female population in employment, age group 65 years and older
266	emp_f_15-24	Female population in employment, age group 15 to 24 years
267	emp_f_15-64	Female population in employment, age group 15 to 64 years
268	emp_f_20-64	Female population in employment, age group 20 to 64 years
269	emp_f_55-64	Female population in employment, age group 55 to 64 year
271	emp_t_15+	Total population in employment, age group 15 years and older
272	emp_t_65+	Total population in employment, age group 65 years and older
273	emp_t_15-24	Total population in employment, age group 15 to 24 years
274	emp_t_15-64	Total population in employment, age group 15 to 64 years
275	emp_t_20-64	Total population in employment, age group 20 to 64 years
276	emp_t_55-64	Total population in employment, age group 55 to 64 years
279	emp_A_LFS	Number of employed people in NACE Rev. 2 section A
280	emp_F_LFS	Number of employed people in NACE Rev. 2 section F
281	emp_K_LFS	Number of employed people in NACE Rev. 2 section K
284	25-64_M_0-2	Male educational attainment levels 25-64, 0-2
285	25-64_M_3-4	Male educational attainment levels 25-64, 3-4
286	25-64_M_5-8	Male educational attainment levels 25-64, 5-8
287	30-34_M_0-2	Male educational attainment levels 30-34, 0-2
288	30-34_M_3-4	Male educational attainment levels 30-34, 3-4
289	30-34_M_5-8	Male educational attainment levels 30-34, 5-8
290	25-64_F_0-2	Female educational attainment levels 25-64, 0-2
291	25-64_F_3-4	Female educational attainment levels 25-64, 3-4
292	25-64_F_5-8	Female educational attainment levels 25-64, 5-8
293	30-34_F_0-2	Female educational attainment levels 30-34, 0-2

Obj_id	Code	Name
294	30-34_F_3-4	Female educational attainment levels 30-34, 3-4
295	30-34_F_5-8	Female educational attainment levels 30-34, 5-8
297	25-64_T_0-2	Total educational attainment levels 25-64, 0-2
298	25-64_T_3-4	Total educational attainment levels 25-64, 3-4
299	25-64_T_5-8	Total educational attainment levels 25-64, 5-8
300	30-34_T_0-2	Total educational attainment levels 30-34, 0-2
301	30-34_T_3-4	Total educational attainment levels 30-34, 3-4
302	30-34_T_5-8	Total educational attainment levels 30-34, 5-8
304	Part_Rate_Education_T	Total Part Rate Education
305	Part_Rate_Education_M	Male Part Rate Education
306	Part_Rate_Education_F	Female Part Rate Education
308	early_leavers_18-24_T	Total Early Leavers 18-24
309	early_leavers_18-24_M	Male Early Leavers 18-25
310	early_leavers_18-24_F	Female Early Leavers 18-26
311	natural_pop_change	Natural population change
312	net_migration_total	Net Migration
313	births	Live births
314	pop_t	total population
315	deaths	Total deaths
316	long-term_unempl	Long-term unemployment
317	DisposableIncomePrivateHouseholds	Disposable income of private households
319	gva_atbasicprices	Gross value added
320	GDP_PPSperInhabitant	Gross domestic produtc in PPS per inhabitant
321	GDP_MPPS	Gross domestic produtc in million PPS
322	GDP_Meuro	Gross domestic produtc in million euros
325	pec_MTOE	Primary energy consumption
327	ghg_mton	Emissions of greenhouse gases
328	People_risk_poverty_social_exclusion	Poverty and social exclusion risk
335	TOTAL_R&D_PERSONNEL_RESEARCHERS	Description: Total R&D personnel and researchers by sectors of performance, sex
340	Unempl_rate_T	Total unemployment rate
341	Unempl_rate_M	Male unemployment rate
342	Unempl_rate_F	Female unemployment rate
733	emp_F_SBS	Employment by economic activity F
734	emp_F41_SBS	Employment by economic activity F41
735	emp_F42_SBS	Employment by economic activity F42
736	emp_F43_SBS	Employment by economic activity F43
738	emp_C_SBS	Employment by economic activity C
739	emp_C10_SBS	Employment by economic activity C10
740	emp_C11_SBS	Employment by economic activity C11
741	emp_C12_SBS	Employment by economic activity C12

Obj_id	Code	Name
742	emp_C13_SBS	Employment by economic activity C13
743	emp_C14_SBS	Employment by economic activity C14
744	emp_C15_SBS	Employment by economic activity C15
745	emp_C16_SBS	Employment by economic activity C16
746	emp_C17_SBS	Employment by economic activity C17
747	emp_C18_SBS	Employment by economic activity C18
748	emp_C19_SBS	Employment by economic activity C19
749	emp_C20_SBS	Employment by economic activity C20
750	emp_C21_SBS	Employment by economic activity C21
751	emp_C22_SBS	Employment by economic activity C22
752	emp_C23_SBS	Employment by economic activity C23
753	emp_C24_SBS	Employment by economic activity C24
754	emp_C25_SBS	Employment by economic activity C25
755	emp_C26_SBS	Employment by economic activity C26
756	emp_C27_SBS	Employment by economic activity C27
757	emp_C28_SBS	Employment by economic activity C28
758	emp_C29_SBS	Employment by economic activity C29
759	emp_C30_SBS	Employment by economic activity C30
760	emp_C31_SBS	Employment by economic activity C31
761	emp_C32_SBS	Employment by economic activity C32
762	emp_C33_SBS	Employment by economic activity C33
763	emp_I_SBS	Employment by economic activity I

Annex 3: Disaggregation/reaggregation process for demographic indicators

OVERALL PROCESSING ARCHITECTURE

The global process of disaggregation/ aggregation of indicators at NUTS3 is Python-based. The process follows these steps:

- Downloading indicators from the ESPON database
- Data cleaning
- Disaggregation-aggregation implementation

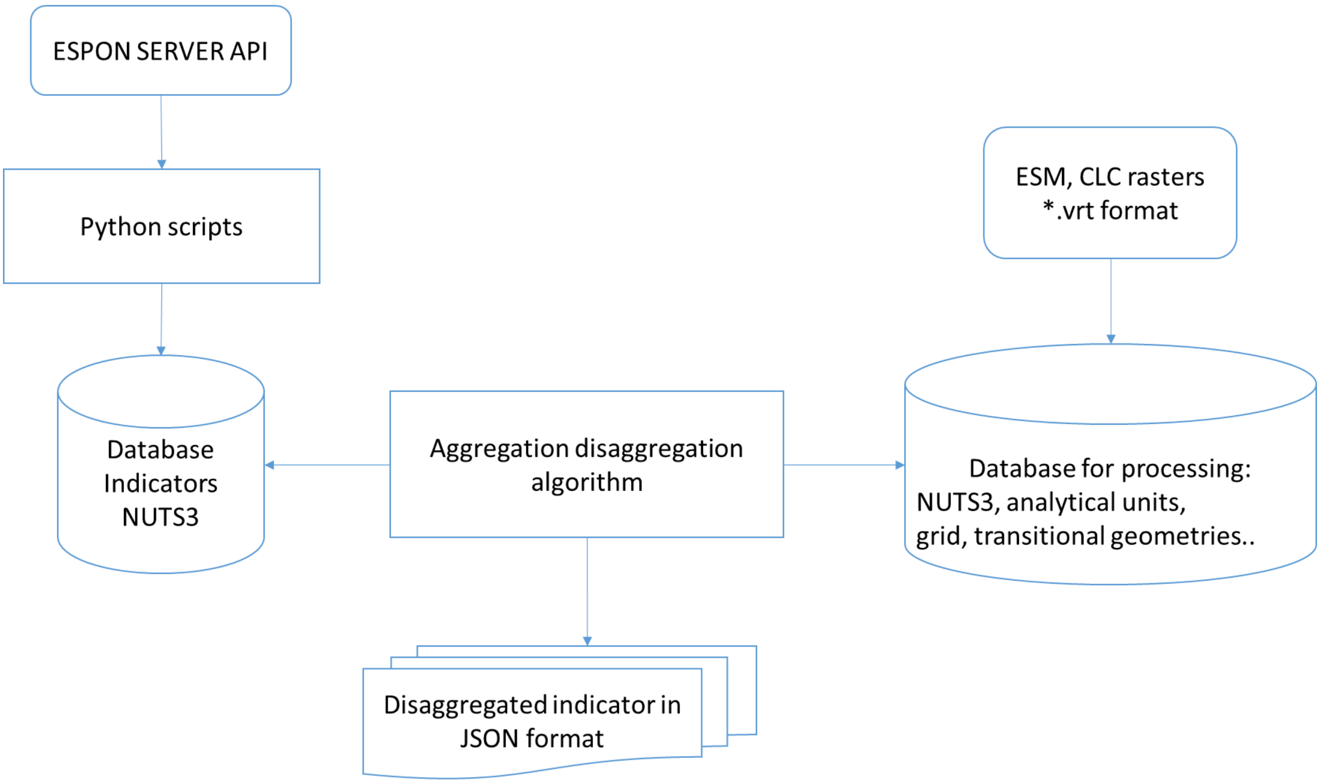


Figure A1- Schematic description of the disaggregation-aggregation process.

A3.1. Download indicators

Indicators are fetched from ESPON API Database and stored on a PostgreSQL database.

Data are downloaded from the API ESPON SERVER at:

https://edp-test.unepgrid.ch/api/public/indicator-data

Downloaded indicators are stored on the Indicator NUTS3 database (see Table A3.2 below).

A3.2. Data cleaning

The structure of the downloaded indicators is reported in Table A3.1. For disaggregation processing it is necessary that all the indicators have the same structure, as expressed in Table A3.2.

Names of downloaded indicators' tables are classified on field **code** in table **Metadata** (see Table 3).

Indicators ready to be disaggregated, output of data cleaning, have the following names:

*proc_'**code**'*

Table 1 summarise method and algorithm used for disaggregating the indicators.

Table A3.1. Elements within the disaggregation-aggregation processing

METHOD		
COLUMN NAME	TYPE	DESCRIPTION
index	bigint	Table's index
Type	text	Describing the type of indicator
ESPON indicator	text	Indicator name, same as metadata.name
Ancillary data	text	Data used as input for mask of weighting data
Weighting data	text	Data used for weight calculation
CLC classes	text	Table reference of CLC classes
Algorithm	double precision	Number of algorithm used

Table A3.2- Structure of the indicator tables downloaded from the ESPON server

DOWNLOADED INDICATORS		
COLUMN NAME	TYPE	DESCRIPTION
index	bigint	table index as reported in pandas
obj	bigint	indicator id
Id	bigint	id referred to the single feature
gid_id	bigint	id referred to the single feature
version	bigint	reference year

nomenclature	text	territorial level name
geom	double precision	geometries related to that feature this is an empty field
nuts_code	text	nutsx code
nuts_name	text	name of the nutsx feature
value	text	values of the indicator
year_end	double precision	ending year the indicator value is referring to
year_start	double precision	starting year the indicator values is referring to

Table A3.3- Metadata table containing information about ESPON indicators

METADATA		
COLUMN NAME	**TYPE**	**DESCRIPTION**
index	bigint	Table's index
id	bigint	indicator id
url	text	indicator URL
code	text	code indicator name
name	text	indicator name
type	text	it the indicator is Dimension/ Class/Multi
years	text	time series of available data
set_type	text	if the indicator is key indicator
is_core	boolean	boolean values with true and false
sequence	bigint	indicator id used for automatic download
status	text	if it is published or not
downloaded	text	if the indicator has been downloaded or not

nuts3	text	if data are available at nuts3 level

Table A3.1 – Structure of the indicator table after data cleaning

INDICATOR TABLE AFTER DATA CLEANING		
COLUMN NAME	TYPE	DESCRIPTION
nuts3_code	character varying(254)	Nuts3 code
nuts3_id	bigint	Nuts3 id
obj	bigint	**Indicator id**. The same of **sequence** on metadata table
yyyy	text	Time series of indicator values from 1990 to 2016

A.3.3. Disaggregation-aggregation implementation

Scripts to be run just once for each weighting factor:

1-intersection_nuts3_grid.sql

2-clc_r_processing.py

3-esm_values.py

4-esm_clc_r_calculation.py

5-zonal_statistics.py

Scripts to be only run for all the functional regions:

6-intersection_fua_sumnuts3.sql

Scripts to be run for each indicator:

7-join_nuts_data_disaggregate_aggregate.py

COMPUTATIONAL TIME

A detail of the computational time for each processing step is reported in Table A3.5.

Table A.3.5- Computational time

Script	Computational time
1-intersection_nuts3_grid.sql	7 hours and 43 minutes
2-clc_r_processing.py & 3-esm_values.py & 4-esm_clc_r_calculation.py	~20 minutes
5-zonal_statistics.py	8 hours
6-intersection_fua_sumnuts3.sq	1 hour
7-join_nuts_data_disaggregate_aggregate.py	~6 minutes

WORKFLOW

Figure A3.2 shows the processing work flow required for the preparation of the ESM layer in order to compute weights on the transitional geometry (see section 3.5). Process involves the following scripts and steps:

- **2-clc_r_processing.py**: creating a raster mask from Corine Land Cover (CLC-R) layer according to values for class 1 in Table 3.
- **3-esm_values.py:** converting ESM values ranging from 0-1 into 0-100
- **4-esm_clc_r_calculation.py:** combination of CLC-R and ESM raster layers.

The final output is a virtual raster, which allows applying zonal statistics algorithm using an asynchronous process.

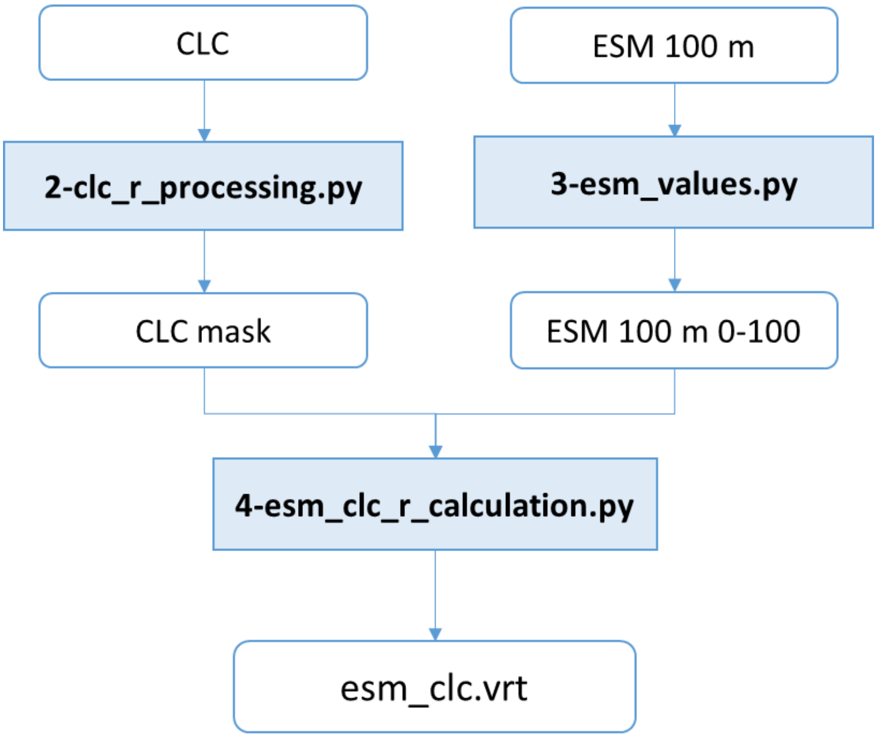


Figure A3.2- Schematic work flow of the weighting computation for demographic indicators

Figure A3.3 below shows the second part of the processing which is composed by:

- **1-intersection_nuts3_grid.sql:** code for the creation of the transitional geometries layer code (intersection with NUTS3 geometry) for each polygon. This latterly allows joining on a second stage the ESPON indicator information.
- **5-zonal_statistics.py:** applying the zonal statistics to the transitional geometry layer. The multiprocessing is executed at FUA (or any other functional region) level.
- **6-intersection_fua_sumnuts3.sql:** creates a spatial intersection between the transitional geometries and FUA layer (any other functional region) and calculates the sum of weight factors for the NUTS3 polygons of the indicator.
- **7-join_nuts_data_disaggregate_aggregate.py:** joins ESPON indicator with the final transitional geometry, disaggregates the indicators and aggregates the results at FUA (or any other functional region) level. The final output is a JSON table.

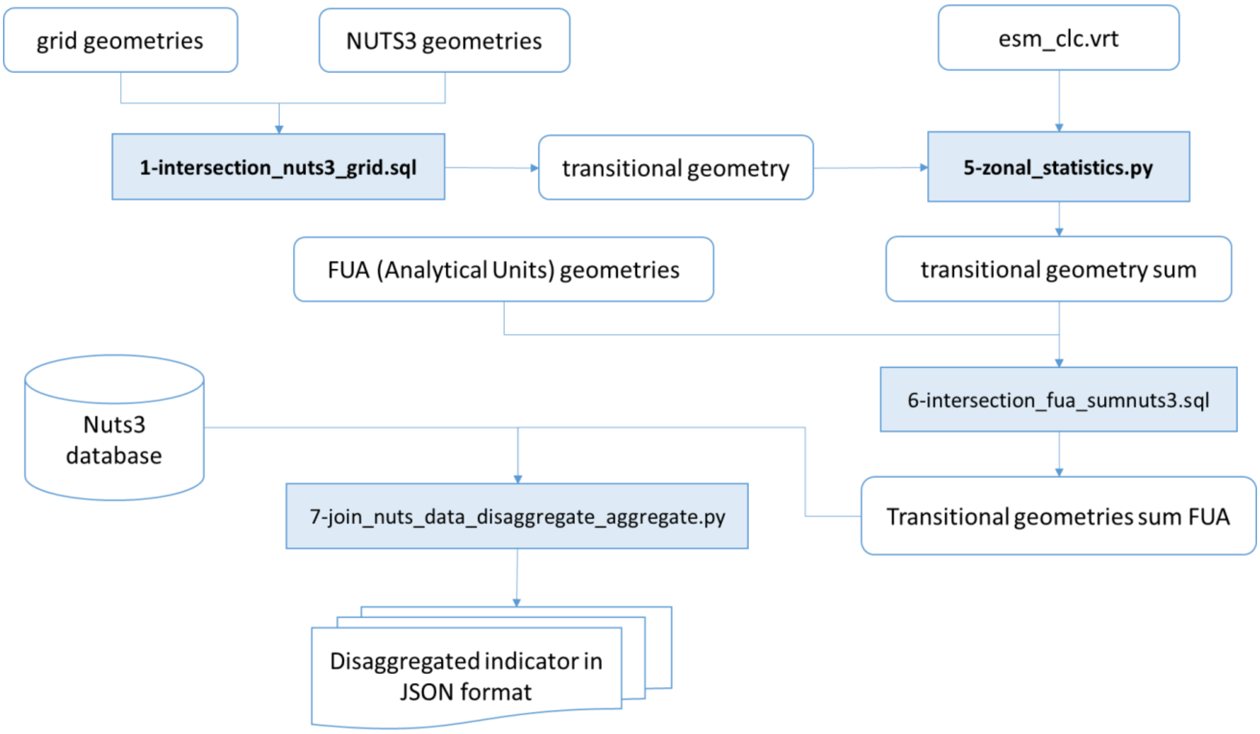


Figure A3.3 - Schema of the disaggregation process for demographic indicators

DATA MODEL

Functional regions (i.e., analytical units) have been geometrically corrected and the field "objected" has been added to all of them. This field is required in the disaggregation process since it is the minimum unit in which the multiprocessing is based on.

Table A3.6. Analytical units (i.e., functional regions) descriptive table

ANALYTICAL_UNITS		
COLUMN NAME	**TYPE**	**DESCRIPTION**
id	integer	Analytical unit id, primary key
name	character varying(255)	Analytical unit shape file name
table_name	character varying(255)	Analytical unit name on the database
origine	character varying(255)	Origin of the analytical unit
type_of_region	character varying(255)	Reporting unit type

Analytical_units table (see Table A3.7 below) summarizes the information related to each functional region.

Table A3.7. Analytical units (i.e., functional regions) table

ID	NAME	TABLE_NAME	ORIGINE	TYPE_OF_REGION
3	CENSUS_UNIT_RG_01M_2011_TCOA16-NUTS2.shp	tcoa		Coasts
4	CENSUS_UNIT_RG_01M_2011_FUA.shp	fua		Functional Urban Areas
8	CENSUS_2011_ESP-FUA_v3_MSA_NUTS2.shp	msa		Coasts
6	CENSUS_UNIT_RG_01M_2011_Mtn_NUTS0-subTyp.shp	mtn		Mountains
5	CENSUS_UNIT_RG_01M_2011_Isl.shp	isl		Islands
7	CENSUS_UNIT_RG_01M_2011_Spa-SubTyp.shp	spa		Sparsely populated Areas
9	CENSUS_UNIT_RG_01M_2011_v3_GI-NUTS2-13.shp	gi		Green Infrastructure
2	CENSUS_UNIT_RG_01M_2011_BDA34_NUTS0border.shp	bda34		Border "large" (90 minutes)
1	CENSUS_UNIT_RG_01M_2011_BDA1_NUTS0border.shp	bda1		Border "narrow" (45 minutes)

Following we report the descriptive tables of all the functional regions that are link to the Analytical units table (see “Table name” field in Table A3.7 on Analytical units above):

bda1		
COLUMN NAME	**TYPE**	**DESCRIPTION**
gid	integer	Primary key
cntr	character varying(50)	Country code
bord1_nam	character varying(254)	
bda_type1	double precision	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

bda34		
COLUMN NAME	**TYPE**	**DESCRIPTION**
gid	integer	Primary key
cntr	character varying(50)	Country code

bda_all	double precision	
bord34_nam	character varying(254)	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

Fua		
COLUMN NAME	**TYPE**	**DESCRIPTION**
gid	integer	Primary key
cntr	character varying(50)	Country code
fua_code	character varying(254)	
fua_name	character varying(254)	
overlapping	double precision	
source	character varying(254)	
areai	double precision	
name	character varying(254)	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

gi		
COLUMN NAME	**TYPE**	**DESCRIPTION**
gid	integer	Primary key
cntr	character varying(2)	Country code
nuts2_2013	character varying(254)	
gi	double precision	
gi_area	double precision	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

lsl

COLUMN NAME	TYPE	DESCRIPTION
gid	integer	Primary key
cntr	character varying(50)	Country code
nuts_3_cen	character varying(50)	
isl_typo	double precision	
isl_numer	double precision	
isl_name	character varying(254)	
isl_admlev	character varying(254)	
shape_area	double precision	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

Msa		
COLUMN NAME	TYPE	DESCRIPTION
gid	integer	Primary key
cntr	character varying(254)	Country code
nuts2_2013	character varying(254)	
msa	double precision	
shape_area	double precision	code used for the disaggregation process
source	character varying(12)	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

Mtn		
COLUMN NAME	TYPE	DESCRIPTION
gid	integer	Primary key
cntr	character varying(50)	Country code
mtn_typo	double precision	

mt_massn	character varying(254)	
nuts0_mass	character varying(254)	
shape_area	double precision	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

Spa		
COLUMN NAME	**TYPE**	**DESCRIPTION**
gid	integer	Primary key
cntr	character varying(50)	Country code
spa_typo	double precision	
spa_subtyp	character varying(254)	
shape_area	double precision	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

Tcoa		
COLUMN NAME	**TYPE**	**DESCRIPTION**
gid	integer	Primary key
cntr	character varying(50)	Country code
nuts_2_cen	character varying(50)	
tcoa16	character varying(254)	
shape_area	double precision	
geom	geometry(MultiPolygon,3035)	Geometry column
objectid	integer	code used for the disaggregation process

For each functional region, a transitional geometry is used for the disaggregation and aggregation of each indicator (see Work flow section above).

As an example about the previous, we show tables referring to FUA functional region in Table A3.8.

Table A3.8 – Schema of the transitional geometry for each functional region

GRID	
COLUMN NAME	**DESCRIPTION**
id	serial number of each feature
geom geometry	geometry column
gridcode	code for each

NUTS3	
COLUMN NAME	**DESCRIPTION**
id	nuts3 id numerical code
Geom	geometry column
Cntr	country code
nuts3_2013	nuts3 alphanumeric code

WEIGHT SUM	
COLUMN NAME	**DESCRIPTION**
nuts3_pk	nuts3 alphanumeric code
weight	sum of weight for nuts3

TRANSITIONAL_GEOM_FUA_WEIGHT	
COLUMN NAME	**DESCRIPTION**
geom geometry	geometry column
sum	weight factor for each feature of transitional_geometry
id	same code from grid.id
objectid	objectid from each reporting unit
cntr_code	country code
nuts3_pk	nuts3 alphanumeric code
area_init	area before spatial intersection with reporting unit
area_fin	area after spatial intersection with reporting unit
weight	weight factor for each feature of transitional_geometry after spatial intersection with reporting unit
nuts3_weig	sum of weight for nuts3 (weight_sum.weight)

Annex 4: User requirements for the Data Analysis Toolbox

Disaggregation/ aggregation processing has been implemented in Jupyter Notebooks using Python programming language and PostGIS. For this reason, the user environment must have the following requisites to enable the processing with the Data Analysis Toolbox:

Python libraries

- SQLAlchemy==1.2.14
- Shapely==1.6.4.post2
- scipy==1.1.0
- rasterio==1.0.10
- rasterstats==0.13.0
- Rtree==0.8.3
- psycopg2==2.7.6.1
- pandas==0.23.4
- numpy==1.15.4
- ipykernel==5.1.0
- ipyleaflet==0.9.1
- ipython==7.1.1
- ipython-genutils==0.2.0
- ipywidgets==7.4.2
- jedi==0.13.1
- jupyter==1.0.0
- jupyter-client==5.2.3
- jupyter-console==6.0.0
- jupyter-core==4.4.0
- Fiona==1.8.2
- GDAL==2.2.2
- geopandas==0.4.0

Software version

- Python 3.6.7
- psql (PostgreSQL) 10.6
- "POSTGIS="2.4.3 r16312"
- PGSQL="100" GEOS="3.6.2-CAPI-1.10.2 4d2925d6" PROJ="Rel. 4.9.3, 15 August 2016" GDAL="GDAL 2.2.3, released 2017/11/20" LIBXML="2.9.4" LIBJSON="0.12.1" LIBPROTOBUF="1.2.1" TOPOLOGY RASTER"
- pgAdmin III 1.22.2
- QGIS 2.18.17

Required format of the outputs

The format of the outputs generated by the user using the Data Analysis Toolbox must be the following in order to allow their inclusion in the web tool. Firstly, output files must be a valid JSON file format. Secondly, all disaggregated-aggregated indicators must have the following structure:

COLUMN NAME	DESCRIPTION
objectid	Primary key
yyyy	Years of time series from 1990 to 2016
Cntr	Country code
fua_ind	Indicator code
parent_id	Functional region code
id	Unique values for the indicator

Annex 5: Weighting scheme: data sources and methods for the estimation of weights to refine the downscaling process guided by ESM/ GHS-BUILT and NACE ancillary datasets

Demographic, Education, ITS (class 1)**:** Batista & Poelman 2016

- **Category 1:** Land use classes for which population is assumed to be directly proportional to the amount of built-up detected in the ESM;
  - 'Urban fabric' classes
  - 'Agricultural areas'
- **Category 2**: Land use classes assumed to contain only residual amounts of resident population:
  - 'Industrial, commercial, public, military and private units'
  - 'Port areas'
  - 'Sports and leisure facilities'
- **Category 3**: Land use classes assumed to have no resident population:
  - All remaining classes.

Land use classes in category 3 are not allowed to contain any population. For classes from categories 1 the total built-up area, as derived from the ESM, is determined, whereas for category 2 only half of the built-up area was considered.

Employment, Society: Thousands hours worked – Eurostat (average values from 1995 to 2017):

Table A4.1 – Thousands hours worked, source: Eurostat. Average values from 1995 to 2016.

GEO// Thousands hours worked - Eurostat		
NACE_R2	**average 1995-2016**	**sector**
A - Agriculture, forestry and fishing	25806344.6	primary
B-E - Industry (except construction)	67371542.9	secondary
C - Manufacturing	60812078.6	secondary
F - Construction	28649036.4	secondary
G-I - Wholesale and retail trade, transport, accommodation and food service activities	92296112.7	tertiary
J - Information and communication	10094223.5	tertiary
K - Financial and insurance activities	10049700.7	tertiary
L - Real estate activities	3689584.74	tertiary
M_N - Professional, scientific and technical activities; administrative and support service activities	37288696.3	tertiary
O-Q - Public administration, defence, education, human health and social work activities	75427808.6	tertiary
R-U - Arts, entertainment and recreation; other service activities; activities of household and extra-territorial organizations and bodies	18773397.5	tertiary

Table A.4.2 – Weights for Class 2 (see Table 3.4) from thousands hours worked (Eurostat)

Sum by sector (Table A.4.1)	**CLR-R reclassification (Class 2)**		**weights**
25,806,345	1	primary	0.10421773
156,832,658	2	secondary	0.63336144
247,619,524	3	tertiary	1

Economy, R&D: World value added estimated from total production for each activity (OECD)

Table A4.3 - Value added estimated from total production for each activity in the European Union (OECD)

Activity	**EU**	**CLR-R reclassification (Class 2)**
AGRFORTFISH	1.40023152	1
CONSTR	3.0454509	2
FINANCEINS	3.64070828	3
INDUSENRG	11.2716172	2
INFCOMM	5.08656493	3
MFG	9.61474727	2
OTHSERVACT	2.36658918	3
PROSCISUPP	6.48331112	3
PUBADMINEDUSOC	9.94366763	3
REALEST	6.51278568	3
SERV	71.7648195	3
WHLEHTELTRANSP	10.6455421	3

Table A.4.4 – Weights for Class 2 (see Table 3.4) from value added (OECD)

Sum per class (Table A.4.3)	**CLR-R reclassification (Class 2)**	**weight**
1.4002315	1	0.0120249
23.931815	2	0.2055221
116.44399	3	1

Energy: Green House Gas emissions by aggregated sector in EU-28, 1990-2015 (in EEA 2017. Analysis of key trends and drivers in greenhouse gas emissions in the EU between 1990 and 2015. EEA Report No 8/2017. ISSN 1977-8449).

Figure A4.1 - Source: EEA 2017 (Figure 2.4 GHG emissions by aggregated sector in EU-28, 1990-2015)

GHG emissions by sources EU 28, 1990-2015 (kt CO_2 eq.)

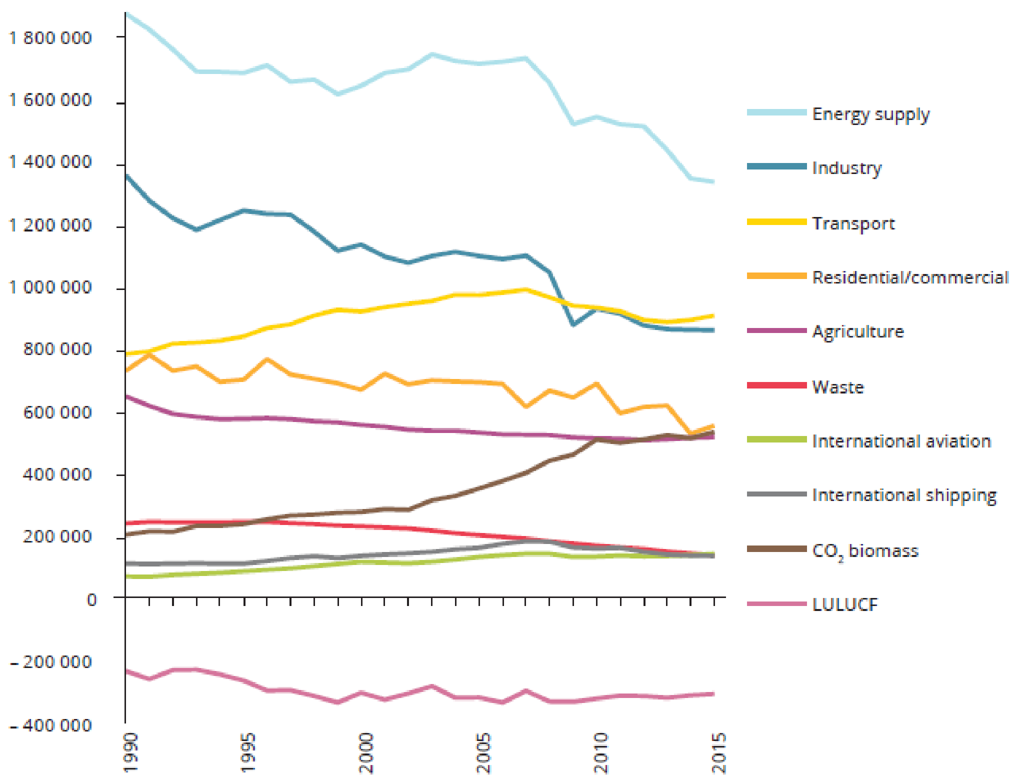


Table A4.5 - Green House Gas emissions by aggregated sector in EU-28, 1990-2015. Average values from 1990 to 2015. Source EEA Report No 8/2017, Fig. 2.4 (rough mean values for the period)

Data derived from Fig. A4.1.		**CLR-R reclassification (Class 3)**	**%**	**weight**
1800000	Energy supply	3	0.36842105	1
1200000	Industry	4	0.24561404	0.66666667
1000000	Transport	5		
700000	Residential/commercial	2	0.12280702	0.33333333
600000	Agricultural	1	0.10526316	0.28571429
300000	Biomass	3		
200000	Waste	4		
100000	Shipping	5		
100000	Aviation	5		

Annex 6: Analysis of cloud service providers

PROVIDER	Web	PLAN	PRESTACIONES					Prices (month)	Prices (hour)
			Numero de cores	RAM memoria (GB)	Disco Duro (GB)	Transfer (TB)	Add-ons		
clouding.io	https://clouding.io/		2	2 (/core)	285	2		38 €	
			2	2 (/core)	500	2		60 €	
			2	4 (/core)	285	2		43 €	
			2	4 (/core)	500	2		65 €	
			4	4 (/core)	500	2		80 €	
Digital Ocean	https://www.digitalocean.com/pricing/	App.	1	2	50	2		$10	0.015
		App.	4	8	160	5		$40	0.060
		App. Optimiz.	4	8	50	5		$80	0.119
		Storage			200			$20	
		Storage			400			$40	
Time4VPS	https://www.time4vps.eu/kvm-linux-vps/		4	8	160	16		18 €	
			6	16	320	32		34 €	
			8	32	640	64		66 €	
ovh	https://www.ovh.com/world/vps/vps-cloud-ram.xml		2	12	50			23 €	
			4	24	100			42 €	
Amazon	https://aws.amazon.com/ec2/pricing/on-demand/		4	16					0.1888
			8	32					0.3776
ikoula	https://www.ikoula.es/es/cloud-vps#cloud		2	4	50	ilimitado		20 €	
			4	8	50	ilimitado		40 €	
			8	15	50	ilimitado		80 €	
gigas	https://gigas.com/cloud-vps		4	4	100	400GB + Ilimitado		29 €	
			6	5	150	600GB + Ilimitado		49 €	
			8	6	200	1200GB + Ilimitado		69 €	
sys4net	https://sys4net.com/vps		2	4	75			45 €	
			3	6	100			60 €	
adw	https://www.adw.es/servidores-cloud.html		4	8	100			69 €	
Hostalia	https://www.hostalia.com/servidor-cloud		4	8	160			60 €	
ginernet	https://ginernet.com/es/servidores/vps/openvz/ssd/		3	4	50			30 €	
	https://ginernet.com/es/servidores/vps/openvz/nvme/		3	15	180			88 €	

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.