

ESPON Database 2013 Phase II (2011-2014). Mutli Dimensional Database Desigh and Development (M4D). Final Report

Claude Grasland, Ronan Ysebaert, Isabelle Salmon, Nicolas Lambert, Timothée Giraud, Christine Zanin, Flora Hayat, Antoine Laporte, Martin Charlton, Paul Harris, et al.

▶ **To cite this version:**

Claude Grasland, Ronan Ysebaert, Isabelle Salmon, Nicolas Lambert, Timothée Giraud, et al.. ESPON Database 2013 Phase II (2011-2014). Mutli Dimensional Database Desigh and Development (M4D). Final Report. [Research Report] ESPON. 2015. hal-03591778

HAL Id: hal-03591778
https://hal.science/hal-03591778

Submitted on 28 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

ESPON Database 2013 Phase II (2011-2014)

Multi Dimensional Database Design and Development (M4D)

Final Report
January 2015



EUROPEAN UNION
Part-financed by the European Regional Development Fund
INVESTING IN YOUR FUTURE

This report presents the final results of a Scientific Platform Project conducted within the framework of the ESPON 2013 Programme, partly financed by the European Regional Development Fund.

The partnership behind the ESPON Programme consists of the EU Commission and the Member States of the EU28, plus Iceland, Liechtenstein, Norway and Switzerland. Each partner is represented in the ESPON Monitoring Committee.

This report does not necessarily reflect the opinion of the members of the Monitoring Committee.

Information on the ESPON Programme and 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 basic report exists only in an electronic version.

© The ESPON Monitoring Committee and the partners of the M4D project mentioned, January 2015.

Printing, reproduction or quotation is authorised provided the source is acknowledged and a copy is forwarded to the ESPON Coordination Unit in Luxembourg.

List of contributors to the M4D Project

UMS RIATE – University Paris Diderot (FR)*
Claude Grasland
Ronan Ysebaert
Isabelle Salmon
Nicolas Lambert
Timothée Giraud
Christine Zanin
Flora Hayat
Antoine Laporte

National Centre for Geocomputation - National University of Ireland (IE)
Martin Charlton
Paul Harris
Alberto Caimo
Conor Cahalane
Lars Pforte

LIG STeamer – University Joseph Fourier (FR)
Jérôme Gensel
Benoit Le Rubrus
Camille Bernard
Anton Telechev
Marlène Villanova-Oliver
Laurent Poulenard
Cassia Trojan Dos Santos

Universitat Autonoma de Barcelona (ES)
Roger Milego
Maria-José Ramos
César Martínez
Francisco Domingues

TIGRIS – Universitatea Alexandru Ioan Cuza (RO)
Alexandru Rusu
Alexandru Banica
Daniel Tudora
Octavian Groza

UMR Géographie-Cités - CNRS (FR)
Anne Bretagnolle
Antonin Pavard
Hélène Mathian
Marianne Guerois

* Coordinator of the project

Table of content

Introduction 6

Part 1 – Integrating Heterogeneous Geographical Data 7

1.1 Regional challenge 9
1.1.1. Core database strategy – theoretical background 11
1.1.2 Creation of ESPON Time-series – First “manual” experiments within M4D 12
1.1.3 Time-series creation 13
1.1.4 Potential use of core database indicators: Functional indicators 18
1.1.5 Discussions around the implementation of the Core Database Strategy 20

1.2. Urban Object challenge 21
1.2.1 Harmonised databases of European cities 23
1.2.2 Feasibility of FUA construction 28

1.3. Grid data as a method for the integration of urban and regional data 30
1.3.1 Background and introduction 32
1.3.2 Methodology overview 32
1.3.3 Results 34

1.4. Local data 36
1.4.1 The alternative geometry 38
1.4.2 VIGO and the LAU2 frame 42
1.4.3 Lessons learned from the implementation of alternative geometry and VIGO 45
1.4.4 Other work done 45

1.5. European Neighbourhood 46
1.5.1 European neighbourhood regions 48
1.5.2 European neighbouring cities 53

Part 2 - The ESPON Database Portal 55

2.1. ESPON Database 59
2.1.1 ESPON Data and Metadata Specifications 60
2.1.2 ESPON Nomenclatures Support 62
2.1.3 Database Tools and Software 64

2.2. The Web Application 66
2.2.1 Tracking and Checking Tools 67
2.2.2 Search Interfaces 70

2.3. A lively Database Portal: networking and updates 75
2.3.1 Support to ESPON TPGs 76
2.3.2 Support to ESPON Tools Projects 76
2.3.3 Application and Database Delivery 77
2.3.4 External networking, ESPON Coordination Unit support and dissemination 78

Part 3 – The M4D Experience 79

3.1 Lessons from the past 80

3.2. Recommendations for the future 83

Conclusion 87

References 88

Table of figures

Figure 1 - Tree representation of the data model 13
Figure 2 - Results of a 'walk' down the tree structure for Austria, visiting each region in turn 16
Figure 3 - Change of NUTS regions in Netherlands 17
Figure 4 - Missing data in Austria (total population, according to Eurostat), left 18
Figure 5 - Result of the data estimation and coherence analysis (right) 18
Figure 6 - Functional indicators based on M4D core indicators 19
Figure 7 – Urban delineations existing for Wien 25
Figure 8 - Typology of age structure by city at European level and class profiles 27
Figure 9 - FUA based on isochrones (Navstreet model) compared to FUA based on commuters: the case of Paris and Barcelona 29
Figure 10 - Example of unique combination of units within a grid cell 33
Figure 11 - General schema of data processing for OLAP integration 33
Figure 12 - Example of output from the Web tool without map and with map 35
Figure 13- Two examples of alternative territorial divisions: *Zones d'emploi* (France, left), and *Raumordnungsregionen* (Germany, right) 38
Figure 14 - Cartographic output for cases a1 and b2 42
Figure 15 - Illustration of the relation between the VIGO (selected services) and the LAU2 frame in France 44
Figure 16 - Data collection, inter-census estimates and harmonisation for Belgrad and Moskva regions 50
Figure 17 - Regional neighbourhood mapkit adapted to cartographic purpose (left) and layer adapted for GIS calculation (e.g. land cover calculation etc., right) 50
Figure 18 - Absolute variation of population 1990-2010 at regional level in the European Neighbourhood in 2D (top of the page) and 3D (bottom of the page) 51
Figure 19 - Modelised networks of cities in the neighbouring countries 54
Figure 20 - Dataset information element structure 60
Figure 21 - Spatial Binding element structure 61
Figure 22 - Indicator Identification element structure 61
Figure 23 - Source Reference element structure 62
Figure 24 - Main elements that compose the description of a nomenclature 63
Figure 25 - Overview of the territorial changes 63
Figure 26 - Overview of the ESPON Database software infrastructure 64
Figure 27 - Three possibilities for delivering the data 67
Figure 28 - ESPON Database Portal - The Key Indicators datasets tracking tool main/overview page 68
Figure 29 - The data quality check [M4D, 2012 (6)] and its related technical report [M4D, 2012 (4)] .. 68
Figure 30 - ESPON Database Portal - Case Study metadata editor, Step 1 (Dataset Information) 69
Figure 31 - ESPON Database Portal – Key indicator search User Interface 71
Figure 32 - ESPON Database Portal – Case Study search page 72
Figure 33 - ESPON Database Portal - Case Study Metadata page 73
Figure 34 - Overview page of the available resources 74
Figure 35 - Upload page: documents and templates 76
Figure 36 - Timeline of the database and/or application deliveries until June 2014 77
Figure 37 - A four circles strategy for promoting ESPON Core indicators? 84

Table 1 – Description of core indicators gathered by the M4D Project 11
Table 2 - Summary of the situation regarding the Core Database Strategy 20
Table 3 - Existing urban nomenclatures at European scale, main characteristics and potential use(s) in a policy context 24
Table 4 - Dimensions and measures of each ESPON OLAP Cube 34
Table 5 - Several methods tested to create the FUNC4 level 40
Table 6 - Four possible outcomes of the regionalization technique 40
Table 7 - Description of the possible relations between the VIGO and the LAU2 43
Table 8 - SNUTS levels identified for the 23 neighbouring countries of the ESPON Area 49
Table 9 - Three examples heterogeneous urban definitions in the European Neighbourhood: Belarus, Morocco and Israel. 53

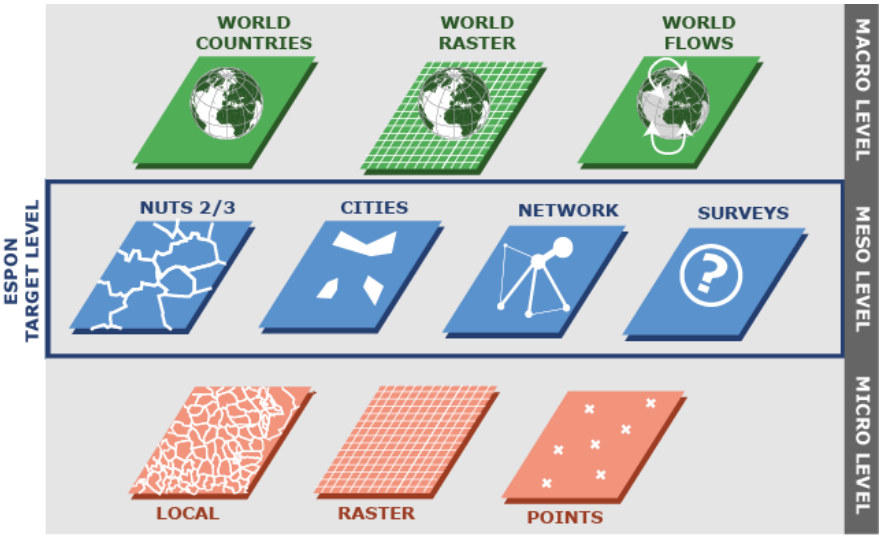
Introduction

At the end of the 2007-2013 Programming period, the final composition of the ESPON database could be described as an interesting reflection of the new trends of the European Union's political agenda in the fields of territorial cohesion, regional policy and urban development.

During the first period of activity of the ESPON Programme (1999-2006), the database only included regional data at NUTS2 and NUTS3 levels, with very short time series (1995-1999). There were very few attempts to cover areas located outside the territory of the European Union, Switzerland and Norway. Some urban object, environmental grid, local scale (LAU2) or global scale (state of the world) data was collected during this period 1999-2006 by some of the ESPON projects. However this data was collected in an experimental way and did not include any strategies for harmonisation, storage and future updates. With regards to the transfer of data between ESPON I and ESPON II, only a small part of the data that had been collected (mainly basic regional indicators at NUTS2 level) could actually be transferred and integration was difficult because of the change in NUTS delineation in 2003 and 2006…

During the second period of activity of the ESPON Programme (2007-2013), a new strategy was developed in order to (i) increase the diversity of geographical objects stored in the database, (ii) enlarge the time series, (iii) improve the possibilities of widening the scope of the analysis to various scales (e.g. from local to global), (iv) offer better quality data and better access to information through more detailed and efficient metadata, (v) propose a coherent and user-friendly web interface to retrieve ESPON data. Have we been successful? It is difficult to say. Indeed, it will only really be possible to evaluate the work achieved in 2007-2013 when the new ESPON Programming period has started.

Part 1 – Integrating Heterogeneous Geographical Data



Stakeholders interested in "territorial cohesion" always want more data whatever the period of time. "Data ! Data ! Data !" has been are current mantra of ESPON meetings from the very beginning of the project. The - researchers and policymakers of the current programme will remember (probably with some emotion), the victory that was the culmination of the SPESP programme (1997-1999). The first map with complete coverage of EU15 and candidate countries at a regional level. Maps of EU31 at NUTS2 level, which we consider "basic" today, were a major scientific and political achievement in 2000. Making it possible, for the first time, to evaluate the consequences of EU enlargement in terms of economic inequality, cohesion policy, CAP, TEN. SPESP and the first ESPON Programme were pioneers in the field, because they took the risk to harmonise heterogeneous data and estimate missing values in datasets that at the time, had not yet been produced by official European statistical organisations. Subsequently, some of the ideas proposed by SPESP and the first ESPON Programme were adopted by these organisations and are now regularly distributed by official European statistical organisations. In this context, what are the challenges that were explored by the ESPON M4D Project?

- **Challenge 1: Enlargement of time series of regional data in past and future.**

NUTS2 and NUTS3 data remains a major challenge for the future of territorial cohesion but not in terms of the amount of indicators available. What is vital is the creation of long term time series of basic data which would make it possible to anticipate the future trends and to derive substantial numbers of indicators through the combination of regional and accessibility data. This is a crucial challenge and so we propose a pragmatic and efficient solution called the "Core Database Strategy". This solution has been implemented within the M4D project: Firstly, we have delivered complete time-series for the ESPON Area at regional level. Secondly, we have developed a methodological framework to transform these Core Indicators into Functional Indicators, using a travel-time matrix.

- **Challenge 2: Launching solid foundations for urban data collection for both competitiveness and cohesion policy.**

Urban data has become more and more important in strategic terms for the European Union. The growing importance of competitiveness in the strategy developed by the European Union has indicated that perhaps the metropolitan area rather than region is the right geographical object to be used for the choice of places where investments have to be concentrated in order to compete

successfully with global challengers. At the same time, the cohesion policy insists also on the crucial role of “secondary”, “intermediate” or “medium” cities as a crucial spatial anchor for the resistance of local societies to demographic, economic or social crisis. These dual challenges (competitiveness and cohesion) imply in our opinion, that there should be different definitions of cities. One definition focusing on European metropolitan areas engaged in Global competition and another one relating more to synergies between urban and rural areas and access to services of general interest. We have therefore focused our work on the creation/harmonisation of two different urban databases.

- **Challenge 3: Crossing new thematic and scalar dimensions through grid data.**

Regions and cities are not the only useful geographical objects relating to the objective of territorial cohesion. Several of the environmental questions and even the technological questions asked by ESPON stakeholders concern phenomena that are not constrained by regional or urban limits. For example, the analysis of water management is better described by water-basins than by regional territorial units. The potential risk caused by a potentially dangerous chemical plant operating under Seveso directives is certainly not limited to the immediate vicinity of the district where it is located but should be described by a surface model resulting in a continuous map. We have handled these multiple challenges using a powerful process called the OLAP Cube which makes it possible to dis-aggregate data of various types and re-aggregate, in different geometries adapted to the needs of each stakeholder.

- **Challenge 4: Enlarging scales of analysis toward more local and global contexts.**

The process of globalization has created – according to many authors – the frequent development of “scalar shortcuts” which means that local communities are often directly connected to the global society or market, with less and less mediation of intermediate levels (cities, regions, state). This theory of “Glocalization” [Robertson, 1995] is not accepted by all researchers who instead are speaking of “inter-territoriality” or “Multiscalar governance”, suggesting that all the scales of analysis are linked and it does still matter. Whatever the conclusion of this theoretical debate, it appears clearly that ESPON should in future develop more and more the collection of data at the local level (LAU2, LAU1) but also at the global level (World) with a particular focus on the eastern and southern neighbourhoods, from the Sahara to the Urals. We have answered this dual demand by having complementary strategies of data collection at the opposite local and global scales. At the local scale, we have explored solutions to harmonise the delineation of local administrative data, and also to create innovative measures of accessibility to services and more generally Very Important Geographical Objects (VIGO). At the upper scale, we have concentrated our efforts on the extension of regional databases and geometries towards east and south, producing delineation “Similar to NUTS” in order to help the project ESPON ITAN, in charge of the “integrated territorial analysis of the Neighbourhoods”.

Each of these four crucial challenges were carefully explored by the research teams of the ESPON Database (2009-2011) and ESPON M4D (2011-2014) projects, on the basis of needs expressed by other ESPON Projects, ESPON CU and ESPON MC. The solutions proposed are pragmatic in the sense that they have always been tested on real datasets associated to precise political targets, either by the Database project itself or by other ESPON TPGs. The final result is therefore not only a set of data but also a collection of technical reports explaining what has been attempted with or without success. We hope that this accumulation of knowledge will be useful in general, for the success of the future ESPON III Programme, and in particular for our successors in charge of the ESPON Database after 2014.

Note from the authors: all URLs mentioned in this report were last accessed on 2015-20-01. We cannot guarantee that all links will continue to work after the closure of the project.

1.1 Regional challenge

Policy implications

The collection of regional data is a crucial topic for the development of EU policies in different fields. The lack of data has been pointed by the Barca's report as a major weakness of policies developed until now: "*Second, a methodological failure has prevented dealing appropriately with the issue of incomplete information. It is not only the case that the Commission has less information than Member States and Regions on what should be done to achieve results (as in the traditional paradigm of asymmetric information). It is also generally the case that, at the moment of drawing Programmes, Member States and Regions themselves have only a very partial understanding of what their targets should be and how to achieve them. The "degree of completeness" varies across policy fields, being quite high, say for major infrastructure – the managing authority presumably has a clear idea of which projects will be implemented during the programme period – but very low in other fields, like innovation*[1]". The development of place based policies which are recommended by Barca's report implies firstly the creation of a coherent database at regional level in order to support the new objective of territorial cohesion: "*Regions might need external support to find their own paths of sustainable development, with particular attention paid to those regions lagging behind*"[2].

Overall M4D strategy

NUTS2 and NUTS3 data remains a major challenge for the future of territorial cohesion but not in terms of the amount of indicators available. What is vital is the creation of long term time series of basic data making it possible to anticipate the future trends and to derive substantial numbers of indicators through the combination of regional and accessibility data. The description in the overall strategy proposes a pragmatic and efficient solution called the "Core Database Strategy" (part 1.1.1). This Strategy has initially been implemented by providing full time-series with manual estimation methods (part 1.1.2), which we have attempted to automate. (part 1.1.3). Based on this crucial information, it opens the door to the creation of innovative indicators, which may be useful for the EU Cohesion Policy, such as the functional indicators framework (part 1.1.4).

Key figures

- **10** time-series indicators in the DB at NUTS 2006 version with **manual estimation methods.**
- **100 %** degree of completeness of M4D time-series for the ESPON Area.
- Time-period of **20** years covered by M4D time-series at regional level.

Policy relevance of the related developments

The regional data included in the ESPON Database appears particularly relevant for the support of the political objectives mentioned above and offers three specific advantages:

- Wide range of topics, not limited to classical economic indicators of performance in order to fulfil the objectives of smart, sustainable and inclusive growth.
- Development of long term times series in the past, in order to propose more accurate projections toward the future.
- High quality of metadata indicators, making possible to understand in detail the content of each indicator and to update them in the future.

By doing that, the ESPON program is clearly in line with the expectations formulated in the alinea 51 p.9 of the Territorial Agenda : "*We suggest that the ESPON programme should take into account the priorities and challenges of the TA2020 in its research activity […] the current status, role and outputs of the ESPON programme should be adapted to the future period in agreement with the European Commission to better serve European policy-making related to territorial development and cohesion*"[3].

[1] Barca F., 2009, An agenda for a reformed cohesion policy (http://www.eu-territorial-agenda.eu/Related%20Documents/report_barca_v0605.pdf)

[2] European Union, 2011, Territorial Agenda of the European Union 2020 – Toward an Inclusive, Smart and Sustainable Europe of Diverse Regions. Final version approved in may 2011 under Hungarian Presidency.

[3] Territorial Agenda 2020, al. 51, p.9



Further readings - Related M4D deliveries and presentations

Presentations

- Functional indicators computation (*ESPON meeting on tools presentation, May 2013)*
- Transformation of regional indicators with functional neighbourhood *(ESPON Seminar and Eurostat presentations, June 2012*).

Technical reports - documentation

- **Dictionary of spatial units**: available in the "Help" section of the ESPON Database Portal at http://database.espon.eu/db2/dico.
- **The Core Database Strategy – A new paradigm for data collection at regional level**, available in the “Resource section” of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=32
- **The Core Database Strategy - A specialized subpart of the ESPON database dedicated to a long term development of monitoring tools** *(M4D Newsletter n° 1, June 2012)),* available in the Home Page of the ESPON Database Portal at http://database.espon.eu
- **Towards an overall data collection strategy (and database?) at the scale of the ESPON Programme?** (delivered to the ESPON Coordination Unit for internal use).
- **Time-series analysis**, available in the “Resource section” of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=32
- **A strategy for elaboration and update of coherent time series of hierarchical territorial units**, available in the “Resource section” of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=32

Datasets (http://database.espon.eu/db2/search)

- M4D time-series in the NUTS 2006 division
- M4D time-series in the NUTS 2010 division *(Datasets, June 2014)*

1.1.1. Core database strategy – theoretical background

The Core Database Strategy (CDS) is a proposal made by the M4D project for the ESPON Scientific Platform in order to monitor selected indicators through the creation of a specific part inside the ESPON database. The idea is to deliver a label "Core data" to a very limited number of indicators that are most relevant for EU Cohesion Policy[4] and should be regularly updated in the future for the purpose of territorial monitoring but also cartographic applications like Web-mapping tool, HyperAtlas, etc. Basically, the CDS is based on five bullet points:

Keep only count variables: In concrete terms, it means that a variable like GDP per capita is initially excluded from the core database as it is not a count variable but the ratio of two count variables (total population and total amount of GDP (in $ or pps). As a consequence, the CDS focuses only on a limited number of count variables that are the most relevant for the EU Cohesion Policy. In that order, the CDS may become an efficient strategy for the monitoring of regions and cities within the ESPON Area. It focuses precisely on the indicators that can be regularly updated in past and future. Moreover, the fact that the CDS only stores a limited number of count variables allows regular updates.

- ✓ The M4D project has built this Core Database for a set of basic indicators, which can be queried in the Search Interface portal (part 2.2). Starting from Eurostat and national statistical institutes data sources, and using missing estimation methods[5], 10 core indicators have been gathered yearly, from the NUTS0 to the NUTS2 or NUTS3 levels in the 2006 division (Table 1).

Indicator	Temporal extent	NUTS version	NUTS level
Total population	1990-2011	2006	NUTS0, NUTS1, NUTS2, NUTS3
Population by 5-years age-class	2000-2009	2006	NUTS0, NUTS1, NUTS2
Births	2000-2010	2006	NUTS0, NUTS1, NUTS2, NUTS3
Deaths	2000-2010	2006	NUTS0, NUTS1, NUTS2, NUTS3
Total land area	1990-2011	2006	NUTS0, NUTS1, NUTS2, NUTS3
GDP in Euros	1999-2008	2006	NUTS0, NUTS1, NUTS2, NUTS3
GDP in PPS	1999-2008	2006	NUTS0, NUTS1, NUTS2, NUTS3
Active population	1999-2010	2006	NUTS0, NUTS1, NUTS2
Employed population	1999-2010	2006	NUTS0, NUTS1, NUTS2
Unemployed population	1999-2010	2006	NUTS0, NUTS1, NUTS2

Table 1 – Description of core indicators gathered by the M4D Project

Store formula of indicators of interest derived from count variables (e.g. a dependency ratio is a ratio between "young" and "old" divided by "adult". There is therefore virtual data that is computed on the fly when a user requests it.

- ✓ This opportunity has not been directly explored and implemented within the M4D Project, due to a lack of available resources but seems to be promising. Among other things, it could avoid the necessity of storing hundreds of indicators that would actually be easily calculated automatically (e.g. ratios).

Use the statistical properties of count data to check the consistency of time-series. Count variables are in limited number which reduces the task of estimation of their time series. Moreover, they are more robust to statistical tests because it is not necessary to introduce weighting criteria as it is the case for ratio. They offer many opportunities of estimation in the ESTI framework and are likely to be checked in many ways (e.g. comparison of sums in lines and columns in the case of age structure). Once time series of count variable are completed, all the derived indicator can be immediately computed using their formula.

Develop automatic procedure to exchange count variables between geometries of various types: For example, thanks to the ESPON OLAP Cube, the GDP data at NUTS 3 level will be firstly

[4] The Core Data created within the M4D Project covers approximatively 30% of the regional maps of the 6th Cohesion Report on Economic, Social and Territorial cohesion. To fit with the indicators displayed in this report, data on education and professional learning should be considered more in depth. The M4D project has not focused its work of data collection on these indicators for one major reason: the data source derive from surveys at European level. It is consequently difficult to obtain long time-series at regional level without significant resources and skills on this topic. It could be a specific domain of interest for the ESPON Programme in the future.

[5] A complete listing of these estimation methods can be found in: Grasland C., Ysebaert R., Giraud T., 2012; the core database strategy – a new paradigm for data collection at regional level, ESPON M4D Technical Report.

estimated at grid level (using an ancillary variable like JRC 2001 or GEOSTAT 2006 population grids) and then re-aggregated to other geometries like water basin, FUA, etc.

- ✓ In that respect, section 1.3 describes how the M4D core indicators have been disaggregated in a 1 km grid and re-aggregated in the ESPON OLAP Cube, in order to provide basic data for several geographical dimensions (urban, regional, bio-geographical objects) (cf part 1.3)

Propose innovative procedures of multi-representation and multi-level analysis of indicators for territorial monitoring and political decision: Whatever the geometry of interest (cities, regions…) is, an isolated indicator can be transformed into a family of variants targeted to answer to specific political questions (comparison to the EU average, territorial discontinuities…).

1.1.2 Creation of ESPON Time-series – First "manual" experiments within M4D

The ESPON time series have at least two challenging characteristics. First, they are relatively short; typically a series may be no more than 20 years in length, and many are far shorter. This poses challenges for estimating level, trend, and any seasonality in the data. The second characteristic is that they are often part of a spatial hierarchy (NUTS0-1-2-3). Thus each series has a longitudinal component, represented by the variation across the time domain, and a cross-sectional component, represented by variation within and between the various levels in the NUTS hierarchy. This means that there are some constraints: the total of units at NUTS levels must add up to the value of the parent unit at NUTS level s-1.

In the past, ESPON produced interesting materials for overcoming the problem raised by missing values. The ESTI framework in the Data Navigator 2 project [ESPON 3.2, 2006] is one of the milestones to consider in that topic. The ESTI framework proposed a couple of solutions, based on four dimensions:

- **The Spatial dimension (E):** Estimates the missing values thanks to values known at an upper, lower or at the same hierarchical level.
- **The Source dimension (S):** Alternative sources of information can be used when the main source does not provide the targeted information.
- **The Time dimension (T):** Uses information provided in time before or after the value to be estimated.
- **The Thematic dimension (I):** The idea is to replace the missing value by a known value coming from an alternative indicator, controlled by a correlation factor.

This methodological background allows a lot of combinations for data estimation. In that order, the ESPON M4D project proposed a catalogue of solutions that are well documented and help the experts to choose the best one in each particular situation [M4D, 2011 (1)].

The M4D experience (M4D core indicators) shows that among all the existing possibilities for estimating missing values, one method is especially adapted to build consistent hierarchical time-series [M4D, 2013 (2)] . This specific data model combines the temporal dimension, taking into account the statistical properties of the NUTS hierarchy (Figure 1):

- **Count variable is estimated only for the top level of hierarchy:** for example, taking the example of active population, the population of Bulgaria in 1999 is estimated in number of active variables. Whatever the count variable chosen (active, unemployed, GDP, population etc.), this information is generally available at State level for large time-series.
- **All other territorial units are described by a frequency of the upper level unit:** for example the unit BG3 is described by the evolution of its share of BG and the unit BG31 as a share of BG3.
- **Consistency of hierarchy is imposed.** When the sum of the frequency of child units from the same parent is different from 1.00, the value is adjusted. For example, the sum of BG31+BG32+BG33+BG34 in 2003 is equal to 1411.9 when BG3 is declared as 1412.0. This is not an error but simply a question of rounded value. Nevertheless it is corrected.

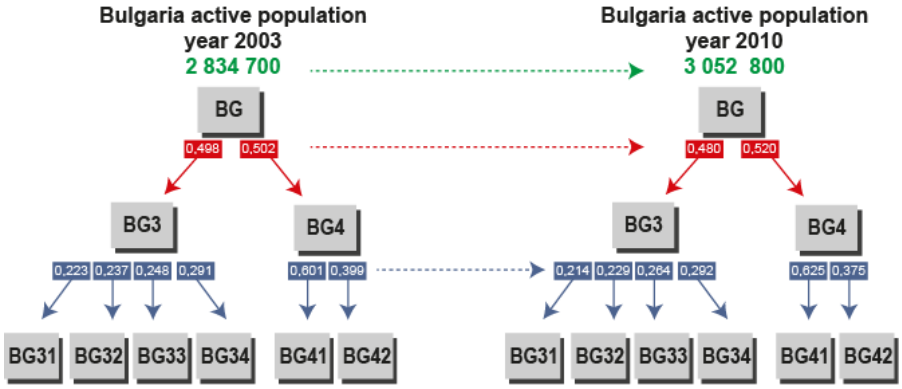


Figure 1 - Tree representation of the data model

The tree structure makes the way we propose to solve the problem of homogenization and estimation of missing values clearer by dividing a big problem into more manageable sections, according to René Descartes' method: "The second [principle] is to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution". Basically, the problem that we have to solve is reduced to a combination of vertical and horizontal analysis of the trees.

- **Vertical analysis** will allow us to simultaneously check for logical errors (when the sum of the frequency of child units from the same parent is always equal to 1) and to estimate some missing values (definition of the value of a parent by the sum of child or estimation of one missing child value by difference between the parent and the other children…).
- **Horizontal analysis** will enable us to estimate missing values using the method of time series analysis and also to check for time outlier and provide margin of errors. But the important point is that these estimations are made for small groups of time-series that are typically the frequency of all the child of the same parent. This frequency is related to internal redistribution and not to external or general trends that are only taken into account for the estimation of the raw count at the top level of the tree.

During the first steps of the M4D project, the work of data estimation was a "step by step" and manual process, done manually (by a "click" in an excel sheet). However, by reducing the problem into smaller sections, it appears easier to propose an automatic procedure of data checking and data estimation that verify objectives and can be implemented in a computer programme (part 1.1.3).

1.1.3 Time-series creation

1.1.3.1 Introduction

A time series is a collection of observations made sequentially in time (Chatfield, 1989). Time series are frequently encountered in (i) economics [Beveridge annual wheat price series], (ii) physical sciences [monthly average air temperature] (iii) marketing [monthly product sales] (iv) demography [annual population estimates] and (v) process control [weights of manufactured product sampled hourly] (vi) communication [binary series are common]. While many series are usually measured at regular intervals (e.g.: year, month, week, day, hour, minute), there are series which occur irregularly, for example, major railway disasters, which are known as point processes. Time series analysis is concerned with (i) description of the main properties of the series, (ii) explanation of the relationship between two series taken at the same time [monthly atmospheric temperature readings, monthly measurements of the North Atlantic Oscillation] and (iii) prediction of (usually) future values.

Time series description can take several forms, but is intended to reveal the underlying structure of the series. This structure can include several components (Shumway and Stoffer, 2010):

- Trend: an increase or decrease in the value of the series over time
- Seasonality: a regular pattern of high and low values related to calendar time
- Long term cycles: periodicity not related to seasonality
- Outliers: values which are unusually high or low in comparison with the rest of the data
- Abrupt changes: changes to the variation in the series or level
- Variance: this may be constant over time or increase/decrease

Estimation activity to date

Section 1.1.2 describes the work by the RIATE team to create methods for estimating missing data in the series for the population 1990-2011. This is described in detail in a separate technical document. The metadata reveals that 4 sources of data from EUROSTAT were used, as well as additional data from 27 national statistical agencies. The metadata then reveals that 52 additional estimation strategies were undertaken to estimate the missing data and ensure is spatial coherence. Their method was enormously time intensive and error prone as is was carried out manually in Excel. Our analysis of their estimation approaches led us to develop the Bayesian methods described below in which the estimation is largely automated. Details of our analysis are in a separate technical document[6] [M4D, 2013 (2)].

1.1.3.2 Characteristics of ESPON Time Series

The socio-economic time series which appear to be commonly encountered in the ESPON Programme are annual counts or ratios for a restricted time period (1990-2013)[7]. This implies that the longest series are less than 25 years, and some are far shorter. Missing data in the middle of a series shorten it still further. Hyndman has reported the effects of attempting to fit models to short series (Hyndman, 2014). In 95 short economic series about one third were random walks (they had no structure).

The series are also presented for the spatial units in the NUTS classification, The NUTS system is a hierarchical system for dividing up the economic territory of the EU for the purposes of the collection of regional statistics, socio-economic analyses and framing EU regional policies. We are concerned mainly with the first 4 levels of the NUTS hierarchy. Tables of the NUTS regions and their codes, and correspondence tables for region adjustments are available from EUROSTAT. The input and outputs from the ESPON database are in the form of rectangular files, as Excel spreadsheets – each line of data is preceded by its NUTS code. Spreadsheets, and the EUROSTAT tables present the NUTS codes of every unit in a single column (for example demo_r_gind3), which gives the impression of flattening the hierarchy, The key to the strategy for handling the time series is to think of the cross-sectional structure at each time period, implied by the NUTS codes, explicitly as a tree. We shall return to this later.

The NUTS0 series shows a diverse set of trajectories. Most show a gentle upward or downward trend. There are a number of interesting potential anomalies in some of the series. Bulgaria shows a downward trend, with an apparent discontinuity between 2001 and 2002. The Czech Republic falls, towards 2002 then rises rapidly – with a discontinuity between 2001 and 2002. Germany's rises to 2004 and then starts to fall. The series for Croatia will require some initial exploration before any modelling commences, and data for 2001 appears to be missing. Hungary is missing 2000. Italy's series is stable and then shows growth after 2002. FYRoM has one obvious anomaly in 1993, and possibly another 2002-3. Malta may have an anomaly in 2000-1. Poland's series shows a rapid fall in 1999-2000. Those for Sweden, Slovenia and Slovakia are similarly uneven, and Turkey's series appears to rise linearly to 2000. Clearly care will be needed when estimating any missing data.

1.1.3.3 Source of data

Many of the tables required can be obtained from EUROSTAT. However, there can be a lot of missing data. Depending on the particular series, the data may be complete to NUTS 3 back to 2000. For earlier years, data may only be available to NUTS2 level. This requires recourse to other sources, for which the most authoritative would be those from the National Statistical Offices (for example: Turkstat, Croatian Bureau of Statistics, Statistics Norway, and Statistische Ämter des Bundes und der Länder). There are other sources, such as the OECD, and the EUROSTAT NewCronos database which may be of assistance in completing the time series.

[6] It is important that the work of ESPON reaches as wide an audience as possible. Academic outputs include a paper presented at the Geographical Information Systems Research UK (GISRUK 2014, http://www.gla.ac.uk/conferences/gisruk/) meeting in Glasgow, and a paper to be presented at the Geographical Information Science (GIScience 2014, http://www.giscience.org/) meeting in Vienna.

[7] Commission Regulation 11/2008 specifies the required time series starting years for NUTS levels in 17 statistical domains. The domain Demography has a start year for 1990 for both NUTS2 and NUTS3.

1.1.3.4 Missing data scenarios

Populating the tables to NUTS 3 level over the period 1990-2013 using (a) Eurostat population data and (b) data from the National Statistical Offices reveals some recurrent patterns with regard to missing data. The question arises as to whether there can be some consistent method of ensuring (a) that the missing data can be estimated and (b) whether the spatial coherence of the estimated values can be ensured. The solution is not to treat each missing value as a separate case; this would arise if it were desired to use a spreadsheet application such as Excel to make the estimations. The solution we sought requires programming so that initial estimates of the missing data in each series are made, using the other values in the series as the 'known' evidence. Following the initial estimates, a second stage of adjustment is applied, to ensure the top-down spatial coherence of each series; thus the value for a NUTS1 region will equal the sum of the values for the NUTS 2 regions which are contained within it.

1.1.3.5 Strategy for completing the series

To estimate data for a time series we need to start with a model. There are many such models in the time series analysis literature which may be applied. However, typically to obtain reliable estimates of the various model parameters requires series of several dozen of observations across time. We do not have this luxury with the ESPON series. What we do have is a set of complete and partially complete series, and we also have a requirement for ensuring the estimated values for the lower levels in the NUTS hierarchy sum to the values for their containing regions in the higher levels. This means that our estimation will have three main components:

(a) a method of making initial estimates for each series based on the available longitudinal data;
(b) a method of ensuring the cross-sectional spatial coherence of the estimates in each time period
(c) a means of representing the spatial hierarchy of the NUTS regions in each country.

(a) Experimentation of the existing ESPON series suggests that some relatively simple models will yield reasonable results for the estimation component. Each series can be modelled with either a linear or quadratic trend, with an autocorrelated error term:

$$P_t = a_0 + a_1 t + \varepsilon_t$$

$$P_t = a_0 + a_1 t + a_2 t + \varepsilon_t$$

where the error term is $\varepsilon_0 \sim N(0,\tau)$ and $\varepsilon_t \sim N(\rho\varepsilon_{t-1},\tau)$ with $t > 0$ and $|\rho|<1$.. The parameters a_0 a_1, and a_2 are to be estimated. Both models can handle missing values – points for which no data is available – as well as provide forecasts, backcasts and interpolation of missing data in the middle of a series. The generic term *prediction* covers these three eventualities.
Without going into technicalities of the estimation, based on Hierarchical Bayesian Models, the approach has been implemented using Markov Chain Monte Carlo (MCMC) techniques, using the R language. This is consistent with the approach adopted earlier in the creation of the Outlier Detection software. The Database spreadsheets can be read directly into the R system, and the estimation functions applied to impute the missing data.

(b) The second component requires a set of constraints to ensure that predictions for NUTS3 units sum to the appropriate value for their containing NUTS2 units, and that predictions for NUTS2 units sum to the appropriate value for the containing NUTS1 units as so on. We refer to this as hierarchical spatial coherence. If we have a NUTS2 region with population P, and we know the populations of three of its 5 units, a, b and c, but not d and e, then we can include P=a+b+c+d+e as a constraint in the Bayesian forecasting framework.

(c) We require one further component in operationalisation to ensure the spatial coherence. The files of data from Eurostat and the National Statistical agencies usually have the NUTS codes in a single column, and a flat file structure: the rows represent spatial units and the columns variables of interest. This effectively flattens the hierarchical structure of the NUTS units, and, if the calculations are restricted to the spreadsheet, renders the task of ensuring hierarchical coherence enormously difficult. An alternative is to use the conceptual representation of the NUTS units as belonging to some sort of tree structure – like a family tree – in which the higher level units are represented as the parents of lower level units (Figure 2). Trees are common data structures in Computer Science. This has been implemented in R – the coherence functions take a country at a time, and 'walk' down the tree in a consistent fashion. Any NUTS1 estimates are first constrained to

the NUTS0 values. Then each NUTS1 region is taken in turn, and any NUTS2 estimates are constrained to the NUTS1 totals. When the NUTS1 regions have been completed, the NUTS2 regions are dealt with in turn and their NUTS3 totals constrained. The output is an internally consistent set of estimates.

```
Tree Walk
NUTS0 root node is AT
 NUTS1 children of AT are AT1 AT2 AT3
  NUTS2 children of AT1 are AT11 AT12 AT13
   NUTS3 children of AT11 are AT111 AT112 AT113
   NUTS3 children of AT12 are AT121 AT122 AT123 AT124 AT125 AT126 AT127
   NUTS3 children of AT13 are AT130
  NUTS2 children of AT2 are AT21 AT22
   NUTS3 children of AT21 are AT211 AT212 AT213
   NUTS3 children of AT22 are AT221 AT222 AT223 AT224 AT225 AT226
  NUTS2 children of AT3 are AT31 AT32 AT33 AT34
   NUTS3 children of AT31 are AT311 AT312 AT313 AT314 AT315
   NUTS3 children of AT32 are AT321 AT322 AT323
   NUTS3 children of AT33 are AT331 AT332 AT333 AT334 AT335
   NUTS3 children of AT34 are AT341 AT342
```

Figure 2 - Results of a 'walk' down the tree structure for Austria, visiting each region in turn

1.1.3.6 Other issues

One further issue is that of ensuring temporal coherence for the changes that have taken place in the system of NUTS units over the years. Boundaries of zones may be shifted, zones may be merged, and zones may be split. Correspondence tables are available at EUROSTAT. Figure 3 shows the changes in the NUTS regions for the Netherlands. Again, the split/merge/shift operations can be incorporated into the forecasting process through additional constraints in the Bayesian framework.

Mergers are easily handled since the output is merely the sum of the values for the zones being merged. Splits require auxiliary data. The Geographical Information Systems practice new values which may be apportioned from the original source zone using the ratio of the areas of the source to the target zone. Other auxiliary data can be used, such as population or workforce, but this is application dependent. Boundary shifts will need to consider the ratio of the areas which has been shifted relative to their original zones.

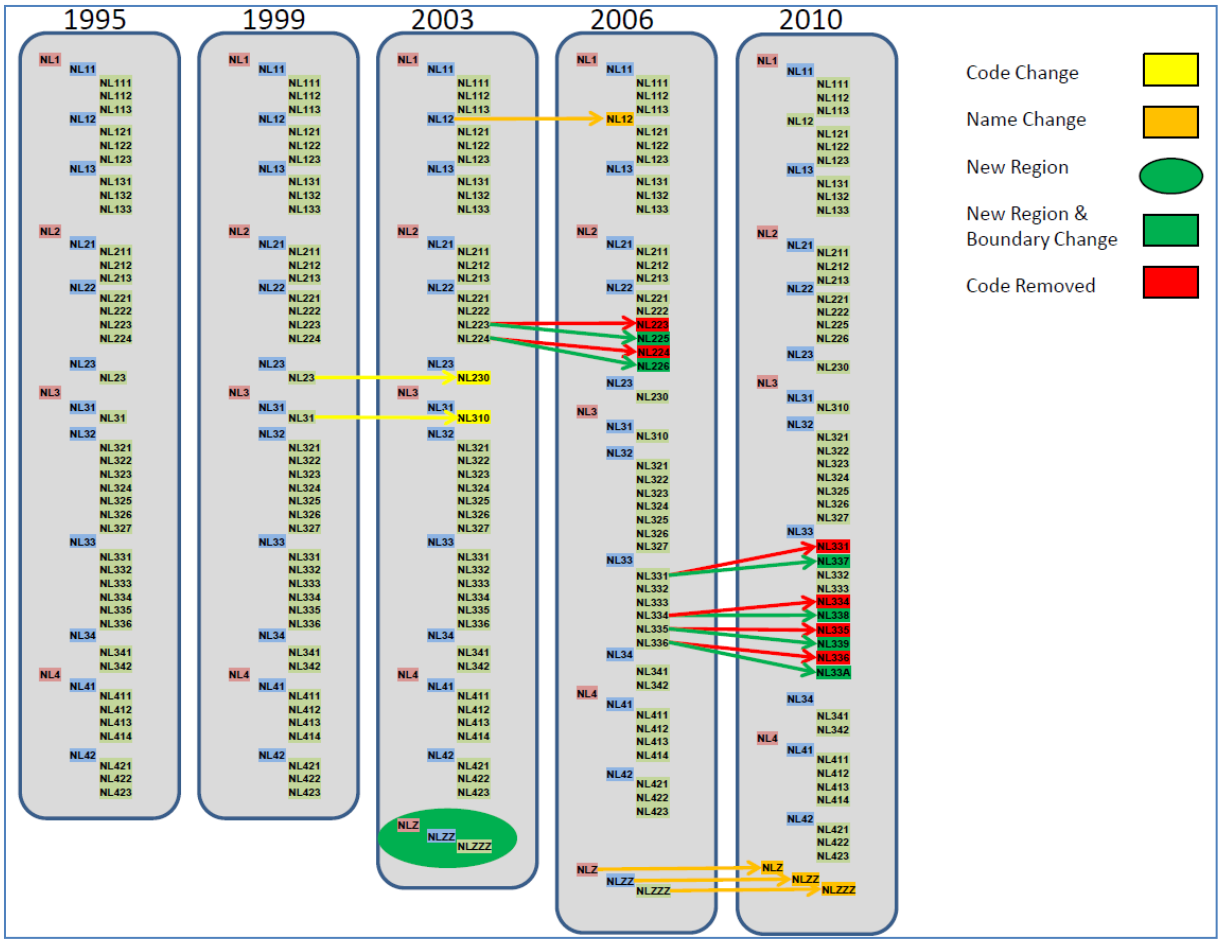


Figure 3 - Change of NUTS regions in Netherlands

1.1.3.7 Software

We have developed bespoke code in the R system which provides a means of both handling the Excel spreadsheets, the NUTS hierarchy, and the computations required for estimating missing data. Considerable experimentation was required to develop the models and implement them in a manner which works satisfactorily in practice. The approach has three stages:

a) Read the data from the spreadsheet and undertake whatever pre-processing required to organise the data for the estimation functions. The output includes, for each NUTS region, some identification information, and an indicator of the level in the NUTS hierarchy (0, 1, 2, 3), the time series (including the missing data) and an indicator of the parent NUTS region (for example: for NL421 this would be NL42). The rows in the output are indexed by the NUTS code

b) Apply the MCMC estimation to each series that has missing data to provide an initial estimation of the missing values. The output is a copy of the input, with initial estimates of the missing data. It should be noted that this stage is enormously computationally intensive. On a Dell T7400 workstation with a 3.16GHz processor, estimation for the example population data for the entire dataset took about 60 hours.

c) Apply the spatial coherence constraint. The output is a copy of the input, but with the values in the lower levels constrained to the values in the corresponding upper level (for example: estimates for NL421, NL422, and NL423 sum to that totals for NL42 in each year).

1.1.3.8 Estimation in practice

Figures 4 and 5 show (a) the missing data pattern for Austria and (b) the result of the data estimation and coherence analysis. Each row represents a NUTS region, and the ordering is by NUTS level, with the NUTS0 at the top of the image. Figure 4 shows that estimates are missing in their entirety for 1990 to 2001 at NUTS3. Data is present at NUTS0, NUTS1 and NUTS2 to enable reliable constraint. The colder colours (blue) represent low populations and the warmer colours (red) high populations;

missing data is shown as white. As has already been noted, the Markov Chain Monte Carlo method is computationally intensive, and the estimation of the missing data in the NUTS3 regions took just under 2 hours on a Dell T7400 workstation. The constraint is only required between the NUTS3 and their containing NUTS2 regions. As a result (Figure 5), the estimation process allows to provide estimates for missing values and also to smooth observed gaps in the time-series (usually due to statistical artefacts), which is very important to ensure the consistency of the backcast and/or the forecast.

It is a characteristic of the method that the more observations in a series that are missing, the shorter the time required to estimate them. However, it should be noted that the further away in the series from the observed data that forecasts are made, the greater the uncertainty in those forecasts. It has not been taken into account in the M4D estimations (backcast and interpolation mainly) but it should considered for projects interested by the creation of long-term forecasts (e.g. territorial scenarios for the future etc.). A subtle difference in the interpretation of forecasts using traditional methods and Bayesian methods means that we can be clear that the forecasted values are the most probable.

The Bayesian approach also allows to inclusion of covariates – they can be used to fine-tune the forecasts, and act as additional evidence in minimising the uncertainty in the forecasts.

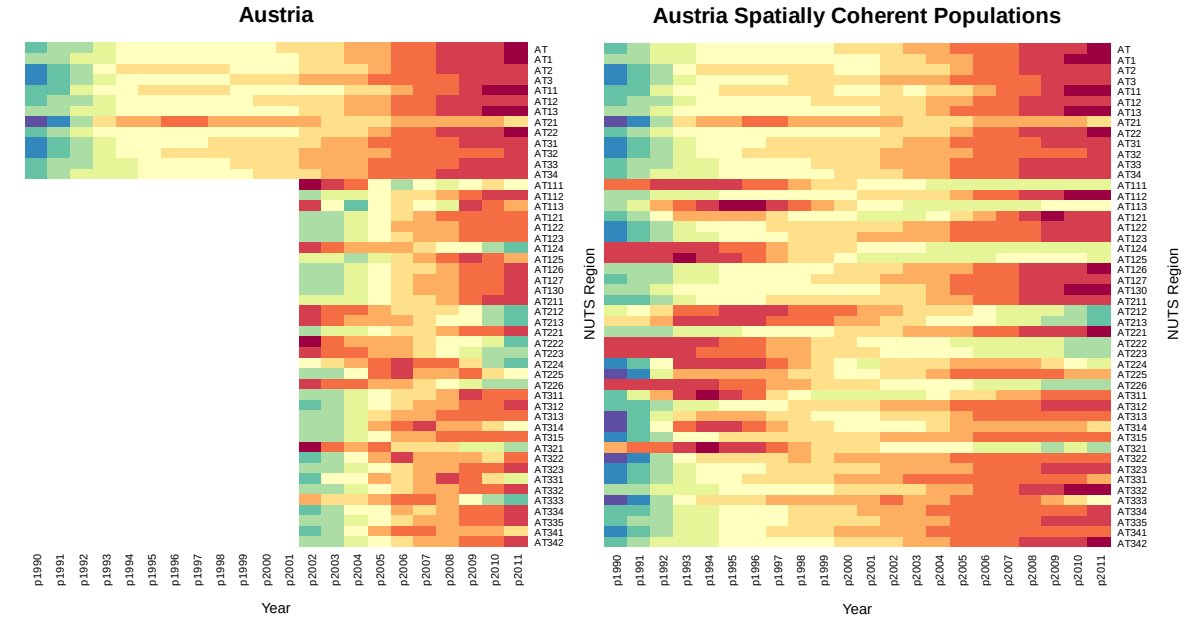


Figure 4 - Missing data in Austria (total population, according to Eurostat), left
Figure 5 - Result of the data estimation and coherence analysis (right)



It needs to be noted that the method is not a cure for all ills. One series we encountered had no observed data - no forecasts can be made using the MCMC approach. The further the estimation is away from the real data, the greater the uncertainty in the estimate. More complex underlying models require more observed data.

We do not need 5 separate methods of estimating missing data. The Bayesian method described in this section provides a coherent and consistent approach to the estimation of missing data in the ESPON time series.

1.1.4 Potential use of core database indicators: Functional indicators

As detailed in the core database strategy technical report [M4D, 2013 (2)], we do not need to gather too many indicators to create an innovative database. But it is important to very carefully select the more crucial ones and to put the maximum effort into their elaboration and verification (or data checking).

With a limited number of consistent core indicators, e.g. count data, available in time-series, which have been carefully statistically checked and using methods derived from spatial analysis (smoothing, discontinuities measures, multiscalar analysis), we can potentially create an infinity of indicators derived from this basic data. As proposed by the ESPON HyperAtlas, based on multiscalar analysis,

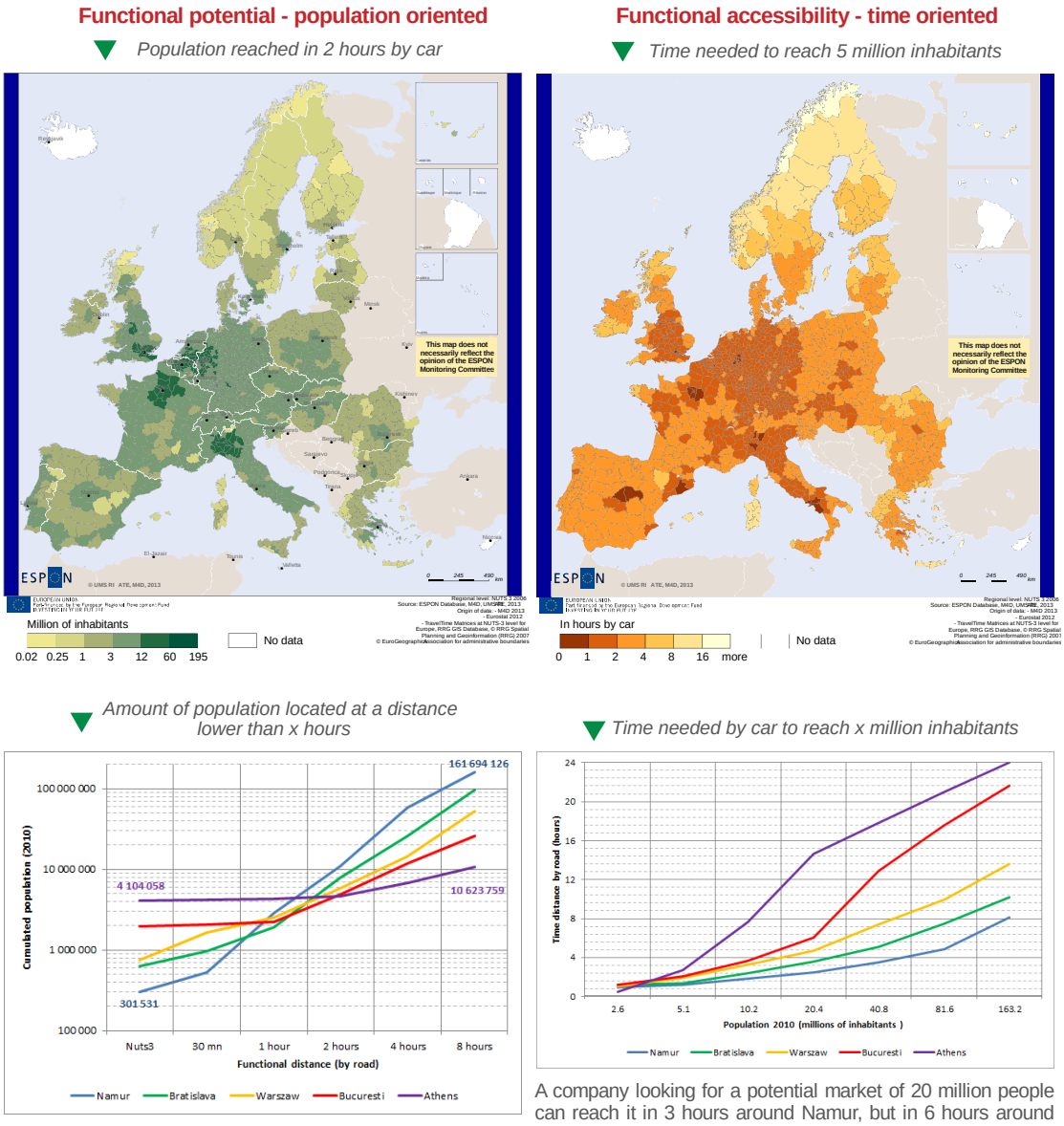
these methods are specifically adapted for proposing innovative views of the ESPON Space which may help the process of policy making [YSEBAERT et al, 2012 (2)].

These methods are also especially adapted to the EU policy context. A concrete example of the interest of the core database strategy can be given by the production of functional indicators (Figure 6). With a single dataset (time-series of population) and having a travel-time matrix, we can propose a high number of useful indicators for stakeholders and practitioners:

- What is the population potential accessible in a given time-distance around my city/ my region?
- What time-distance is needed to reach a given amount of population if we decide to implement a production or service in this particular place?

It is obvious that we can extend functional indicators to other M4D Core indicators (GDP, employment etc.). We can therefore answer more sophisticated problems, at the top of the political agenda:

- What are the areas of ESPON territory where the opening of borders could create complementarities for the labour market?
- What are the areas of the European Territory where positive and negative spillover are observed in terms of economic growth at local scale?



The potential population accessible by road from Namur is low at short distances but very high at long distances. It is the opposite for Athens.

A company looking for a potential market of 20 million people can reach it in 3 hours around Namur, but in 6 hours around Bucharest…

Figure 6 - Functional indicators based on M4D core indicators

1.1.5 Discussions around the implementation of the Core Database Strategy

Significant developments have been undertaken within the M4D project related to regional data and the Core Database Strategy. The left column in Table 2 reminds the main findings related to the Core Database Strategy (CDS), which are implemented (datasets, data model) and documented (technical reports).

Moreover, this work opens several promising perspectives. The right column in Table 2 proposes some development suggestions for the future ESPON Program. They are further described in the third part of this Final Report.

Elements of the CDS implemented within the M4D Project	Developments suggestions
✓ **Creation of count data time-series** with manual estimation methods for the indicators relying on the NUTS 2006 version. All the estimation methods are based on the conceptual work done for the Data Navigator 2 Project [ESPON 3.2, 2006]. Those methods are referenced in metadata associated to the delivered datasets. They are also described in a dedicated technical report [ESPON M4D, 2011(1)]. ✓ **The ESPON Metadata model proposes the field "Core" to tag Key Indicators**. The Search Interface uses this field to retrieve such Core Indicators delivered by the ESPON Projects [M4D, 2012 (1)]. Though the "How to Deliver my Data" documentation [M4D, 2012 (3)] invites TPGs to deliver Core data, only a few ESPON Projects have delivered such data at the end of their activities. ✓ The M4D experience on **time-series estimation** has been summarised in a dedicated technical report [M4D, 2013 (2)]. This report provides guidelines for computer scientists and statisticians for the automation of time-series estimation. ✓ Several technical reports suggest the use of the **Bayesian method** to estimate time-series. ✓ This procedure of estimation using the Bayesian method has been developed in a program written in **R language**. ✓ Developments around **Functional Indicator** prove that using Core Indicators and spatial analysis methods can dynamically provide indicators that are policy relevant.	✓ **Enlarge the set of Core Indicators to cover more themes** (education, income etc.) [ESPON M4D, 2013 (7)]. ✓ **Automation of time-series estimation**. The R solution proposed by M4D can be further developed. ✓ **Smooth the data in the time** in order to avoid the Modifiable Temporal Unit Problem (Grasland, ESPON Seminar in Aalborg, 2012). ✓ Store **"formula of interest**" to automatically compute indicators of interest for the EU Cohesion Policy: for instance, the GDP per capita should be automatically calculated using two core indicators: GDP (in euros or pps) and the total population (inhabitants). ✓ **Develop visualisation tools for time-series**. The solutions described in the ESPON Cartographic Language project [ECL, 2013] should be considered. ✓ **Make existing tools use and connect the Core Database** (HyperAtlas, Functional Indicator Tool). ✓ **Exploit the methods** developed within the ESPON Programme (convergence analysis, discontinuity analysis etc.). ✓ Make the ESPON Core Indicators accessible through Geographic **Web Services** for a better interoperability of the ESPON Database with other geographic systems.

Table 2 - Summary of the situation regarding the Core Database Strategy

1.2. Urban Object challenge

Policy implications

The elaboration of urban data is requested by at least two political objectives mentioned in the Territorial Agenda and further developed in the Leipzig Charter. First of all, the **development of a multilevel governance** which should not be limited to the administrative hierarchy EU>State> Region but should also involve local actors and urban areas which are able to develop linkage and networks not bounded by classical borders: "*We call states, regions, cities (including small and medium sized towns), other territories and sectorial policies at all relevant levels to contribute to common European territorial priorities*[8]". Second, cities (especially the major ones) are crucial nodes of organization which can support the objective to **promote polycentric and balanced territorial development**: "*We stress that polycentric and balanced territorial development of the EU is key element of achieving territorial cohesion. Where the most developed cities and regions within Europe cooperate as parts of polycentric pattern they add value and act as centres contributing to the development of their wider region*[9]". It is nevertheless clear that the structuration of territories could not be achieved only by a focus on major metropolitan area and also support the urban-rural partnership for development: "*In rural areas **small and medium - sized towns play a crucial role**; therefore it is important to improve the accessibility of urban centres from related rural territories to ensure the necessary availability of job opportunities and services of general interest. Metropolitan regions should also be aware that they have responsibility for the development of their wider surroundings*[10]".

Overall M4D strategy

Urban data has become more and more important in strategic terms for the European Union. The growing importance of competitiveness in the strategy developed by the European Union has indicated that perhaps the metropolitan area rather than region is the right geographical object to be used for the choice of places where investments have to be concentrated in order to compete successfully with global challenges. At the same time, the cohesion policy insists also on the crucial role of "secondary", "intermediate" or "medium" cities as a crucial spatial anchor for the resistance of local societies to demographic, economic or social crisis. These dual challenges (competitiveness and cohesion) imply in our opinion different definitions of cities, depending on the purpose of the analysis.

Key figures

- The UMZ database includes **4304** morphological objects larger than 10 000 inhabitants.
- **1** dictionary of correspondence between LAU2 and UMZ created.
- **1** alternative method of FUA construction has been explored in **12** EU cities to overcome the miss of commuting data at EU level.

The delivery of the new LUZ proposed by the DG REGIO is a first important step in that way. But we consider that it needs to be extended to other urban nomenclatures to cover all the themes of the EU policy agenda. However, few urban nomenclatures (associated with data) exist currently. And some of them needs to be regularly updated (FUA). From our point of view, it implies firstly to think about how to build urban nomenclatures (with comparable urban objects in term of methodology) and feeding them by indicators. It has been done with the Urban Morphological Zones (UMZ) Nomenclature (part 1.2.1). But it implies also to take care of the update of functional urban nomenclatures based on accessibility, taking into account missing harmonised information on commuting at European level. It has been explored with a feasibility study of FUA construction (part 1.2.2).

Policy relevance of the related developments

Considering the importance of the political objectives mentioned above, particular attention has been paid by the ESPON database project to the various possible definitions of urban area and the research of methods of harmonization of national data. If the data of the Urban Audit can offer some

[8] Territorial Agenda 2020, al. 10, p. 4
[9] Territorial Agenda 2020, al. 25, p. 6
[10] Territorial Agenda 2020, al. 29, p.7

comparative data on metropolitan area, they are not able to fulfil all the political objectives, especially when it comes to the role of small and medium city size. It is the reason why it has been proposed to develop also an urban definition derived from UMZ delineation and more generally to check the possibility of **Taylor-made definition of cities** adapted to each specific political problem.



Further readings - Related M4D deliveries and presentations

Presentations

- Urban delineations and databases in Europe (*ESPON Seminar presentation, June 2012)*

Technical reports - documentation

- **Dictionary of spatial units**: available in the "Help" section of the ESPON Database Portal at http://database.espon.eu/db2/dico?idDoc=66.
- **M4D Newsletter** : “The ESPON Database includes indicators related to cities” // “Large Urban Zones Harmonised” *(M4D Newsletter n°3, June 2013),* available in the Home Page of the ESPON Database Portal at http://database.espon.eu
- **UMZ: a database now operational for urban studies (M4D improvements),** available in the “Resource” section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=31
- **Functional areas, feasibility study**, available in the “Resource” section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=31

Datasets (http://database.espon.eu/db2/search)

- 10 indicators on Urban Morphological Zones (education, age pyramid, dwellings, population classification).
- 2 indicators on Morphological Urban Areas (basic indicators)
- 8 indicators on Functional Urban Areas (basic indicators)
- Geometries, nomenclatures and some additional indicators can be found in the Resource Part of the ESPON Database Portal: http://database.espon.eu/db2/resource?idCat=43

1.2.1 Harmonised databases of European cities

City is a very complex object that cannot be reduced to a sole approach and definition. In the thematic work packages of M4D linked to cities, we have focused on the importance of uploading several urban databases to the ESPON Database and not just one, which could be considered as “official”. Three different urban databases have been uploaded: (Morphological Urban Areas, Functional Urban Areas defined by the IGEAT in Brussels, Urban Morphological Areas of the EEA). One of the major issues for European statistics is maintaining two different types of urban databases, one defining cities as morphological objects (based on density or continuous built-up areas) and one defining cities as functional objects (based on commuters). We have also insisted on the fact that working at an international scale (Europe and/or neighbouring countries) means we have to define cities using harmonized criteria and not as a collection of national definitions, as it was done for example in Urban Audit 2004 for the Larger Urban Zones.

Considering the functional urban areas (FUA), there is also the possibility for using the New Harmonised LUZ, provided by the consortium Eurostat/DG Regio/Urban Audit/OECD (2012). However, the LUZ generally relate only to the largest cities and not to small and medium ones. For instance, there are 1645 new LUZ, which means that cities with lower than 50 000 inhabitants are not represented by this database, even though they raise crucial issues in terms of urban planning and knowledge on territorial dynamics. This is the reason why we have also considered the Urban Morphological Zones with great interest. There are 4304 UMZ larger than 10 000 inhabitants, and some of them lie inside regions lacking in any new LUZ but needing European support.

UMZ is the only database freely available, defined at the European scale with harmonised criteria and including all cities larger than 10 000 inhabitants. It is updated regularly by the EEA and the automation of the methods that have been constructed by ESPON M4D for giving them a name render the updating process very easy and quick (see section 1 of the technical report “Urban Morphological Zones”, [M4D, 2013 (3)]). Instead of using different computing languages and environments, the procedure now only uses Python and the Arc Gis tool. Furthermore, we have simplified this procedure by creating a graphical interface that allows for choosing geometric layers, parameters, directory files etc. In particular, the choice of a specific version of LAU (for example LAU2 2001 or LAU2 2006) may bring very different resulting names for countries that have dramatically reformed their administrative boundaries, such as Poland.

The indicators available with the three uploaded databases (MUA, FUA and UMZ) are interesting (population and surface) but too poor for a deep exploration of European urban structures and dynamics. This is the reason why we have investigated methods for populating urban databases with LAU2 indicators. The methods which are described for UMZ (see [M4D, 2013 (3)], section 2) are generic and can easily be transferred to other databases. The UMZ delineations are not based on a collection of LAU2 but on a collection of pixels, as they are created from CORINE Land cover. In order to facilitate the use of socio-economic indicators based on LAU2, we have worked on a generic methodology to create a dictionary of correspondence between UMZ and LAU2.

As displayed in the case of Wien (Figure 7, page 25), the comparison of the results obtained from those databases for the Wien area reveals huge differences:

- For the total area, 3375 % between the UMZ (400 km²) and the FUA (13500 km²)
- For the total population, 165 % between the MUA (1,675 million inhabitants) and the FUA (2,773 million inhabitants).

Table 3 summarises the existing urban nomenclatures at European scale, their main characteristics and their potential use for the future of ESPON, and more generally at the level of the European Union.

Urban Nomenclature	Data producer	Characteristics	Elementary unit	Basis of the delineation /methodology	Potential policy use / perspective
UMZ	➢ EEA (UMZ layer) ➢ ESPON – M4D (naming UMZ + populating with LAU2 indicators)	4304 units (UMZ above 10 000 inhabitants) covering the EU28	Collection of pixels : Corine Land Cover (EEA) + intersection with the LAU2 layer	1. An Urban Morphological Zone can be defined as "A set of urban areas laying less than 200m apart. Those urban areas are defined from land cover classes contributing to the urban tissue and function. 2. Intersecting the geometries of UMZ and LAU2 and computing, for each resulting area, the population and surface. A total of 23257 intersections have been considered for the UMZ larger than 10 000 inhabitants. 3. Giving a name to the UMZ 4. Populating UMZ by LAU2 data (dictionary of correspondence between LAU2 & UMZ).	UMZ layer regularly updated by EEA. Support to policies in favour of small and medium size cities. Building block for analysis of settlement system at local level. Linkage between economic and environmental data.
MUA	ESPON – IGEAT	1988 units, covering EU28 + Switzerland + Norway. Croatia not included.	LAU2	MUA = LAU2 contig. with at least 650 inh./km² and at least 20 000 inh. aggregated.	Historical interest. Not to be maintained in the future because very similar to LUZ 2012
FUA	ESPON - IGEAT	1553 units, covering EU28 + Switzerland + Norway. Croatia not included.	LAU2, then MUA	1. Urban Core : MUA 2. Part of commuters = Pop. of commuters (EU) going to the Urban Core / Economically active pop. (EU) * 100) ≥ 10% 3. Max part of commuters from EU to Urban Core; If equality, check at the most populous Urban core. + manuel changes 4. If the perimeter of functional area is fragmented, the limit corresponds to the largest perimeter. If fragmentation is caused by geographic element (Island, Fjord, ...), keeping all.	Update the FUA nomenclature (cf Feasibility study realized by M4D). Probably important to maintain as an alternative to LUZ2012, especially if accessibility can be used as a relevant alternative to commuters data.
LUZ 2004	DG REGIO	367 units covering EU28 + Switzerland + Norway	Depending of the country	Larger Urban Zones are based on various national definitions (functional areas, planning regions, local administrative units, etc.). Only a good knowledge of specifications allows identifying bias resulting from the national heterogeneity of these definitions	Historical interest. Not to be maintained in the future because too heterogeneous in terms of definition
LUZ 2012	DG REGIO	1645 covering EU28 + Switzerland + Norway + Iceland	LAU2	**The urban core (city)** lies on two criteria: a minimal density threshold and a minimal population threshold. It is defined through three main steps. 1. High-density population area is first identified by selecting grid cells over 1500 inh./km² and by aggregating contiguous high-density cells. 2. Only the clusters larger than 50 000 inhabitants are retained as urban cores. 3. Those cores are adjusted to LAU2 delineations: all the LAU2 with at least 50% of their population inside the urban core are retained and form the urban core. The commuting zone around this urban core is identified by using commuting patterns: it is composed by all the LAU2 where more than 15% of their employed resident population work in the urban core. Then, all the LAU2 which are enclosed within this functional area are included and all the non-contiguous LAU2 are dropped.	Official harmonized nomenclature, to be considered as reference for studies on metropolization. But not sufficient for analysis of small and medium cities and to be benchmarked with FUA.

Table 3 - Existing urban nomenclatures at European scale, main characteristics and potential use(s) in a policy context

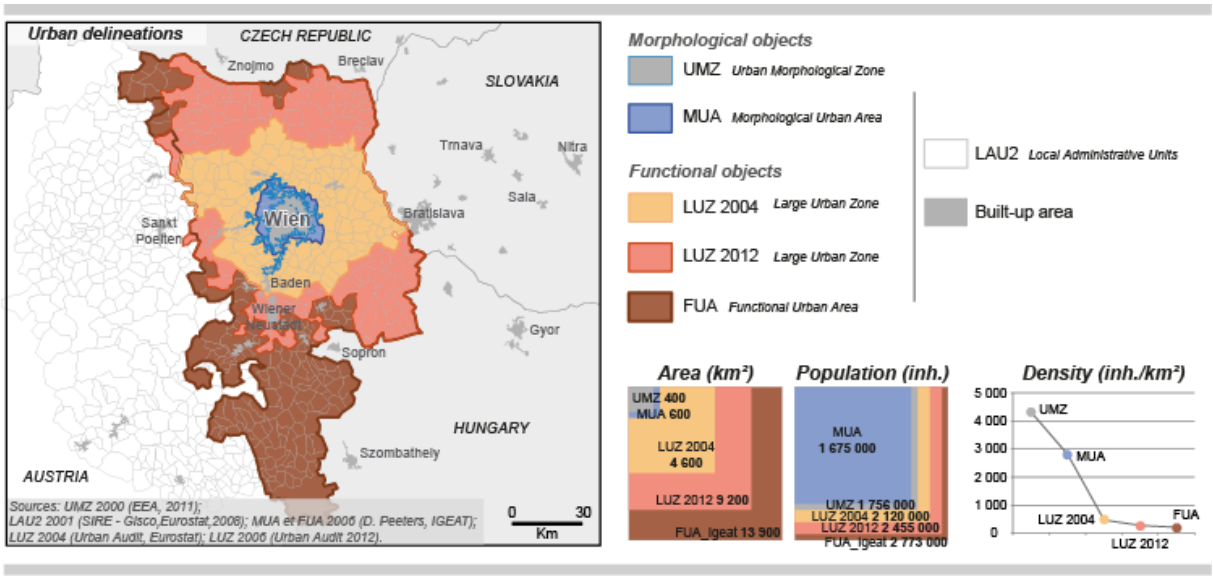
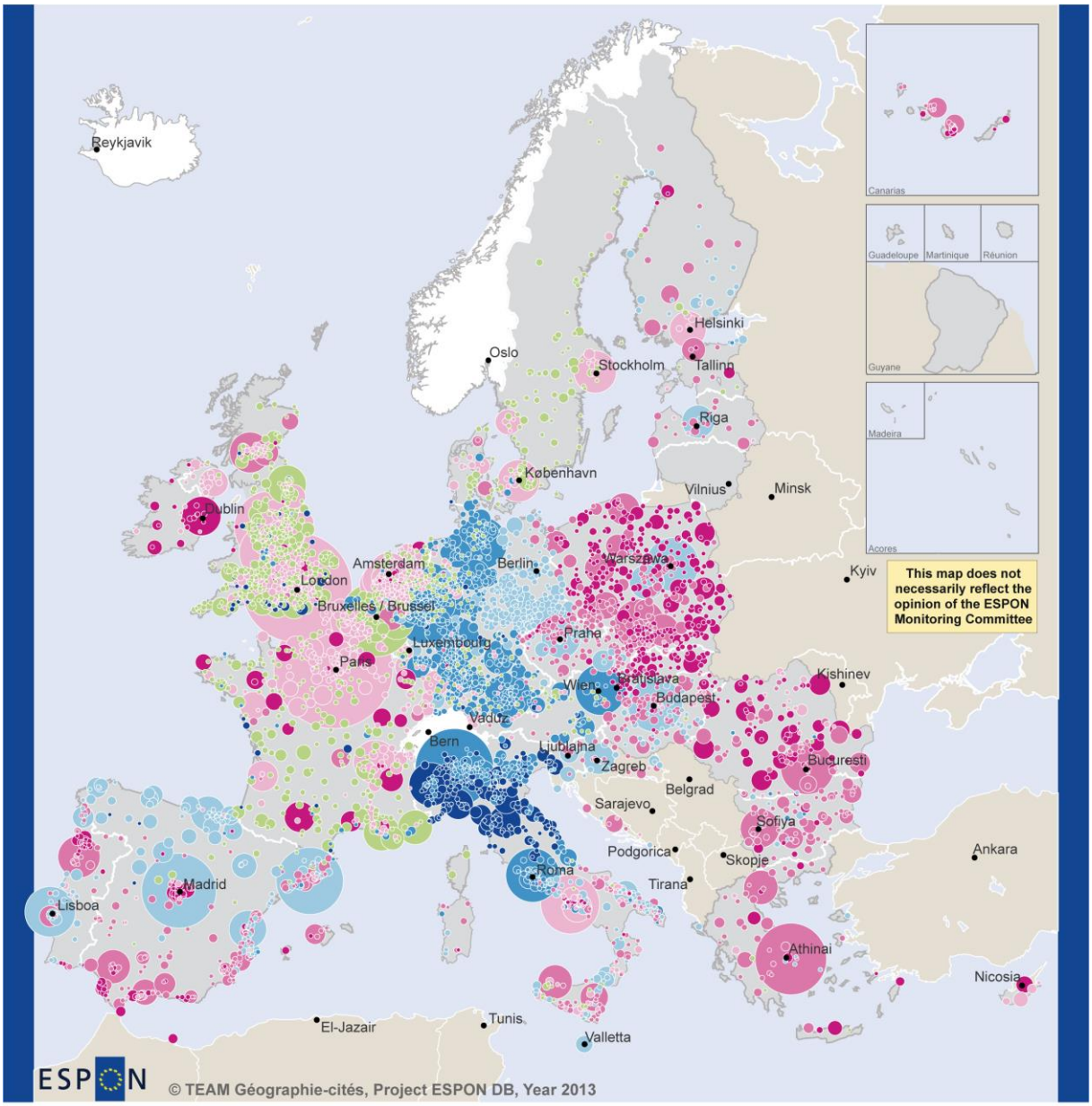


Figure 7 – Urban delineations existing for Wien

The dictionary of correspondence between UMZ and LAU2 enables the population of this database with local data, as it gives the composition of the UMZ in terms of LAU2. The intensity of each link LAU2-UMZ has been characterized by the share of the LAU2's population that is included in the UMZ (calculated with the JRC population density grid). Populating UMZ with social, economic or demographic indicators presumes however that we have robust and complete LAU2 databases. This is not the case with the current version of SIRE (year 2000). We have analysed in depth the different indicators that are proposed in SIRE and retained only the age structure (see the Technical Report).

The results (Figure 8) are very promising and show that there is an urgent need for improving the SIRE database in order to explore other dimensions of European cities. We have made an initial typology of the European cities according to the age structure considering all the UMZ larger than 10 000 inhabitants. The cluster analysis has been created with 7 classes. Cities marked in blue and green are characterised by an ageing population, particularly those marked in dark blue. At the opposite end of the spectrum, cities marked in pink are characterized by young populations, particularly those shown in dark pink. The different profiles are represented by age pyramid schemas: for each pyramid, we can see the European profile (average) on the left and the specific class profile on the right. To make it easier to read the pyramids, over-representations of some age classes in each pyramid have been underlined in bold. The map highlights a very clear opposition between three types of regions, ageing ones (Germany, Austria, northern Italy & Spain), intermediate ones (UK, France, Belgium, Netherlands, northern Europe) and young ones (Central & Eastern Europe, southern Italy & Spain, Greece, Ireland). But when realised at a more regional level (see in the Technical Report for central Europe), the typology also highlights city size effects. The largest cities (especially capital cities) tend to be populated by the oldest inhabitants (on average) and the smallest cities, the youngest. These spatial structures are very important to consider when analysing the current demographic trends (for example, questions of shrinking cities or shrinking regions, but also out-migration of young people). This is not visible when the age classes are mapped by NUTS2 or NUTS3 nor when mapped at LAU2 level. It raises the importance of having spatial nomenclatures of urban objects and coherent LAU2 databases for populating them.



EUROPEAN UNION
Part-financed by the European Regional Development Fund
INVESTING IN YOUR FUTURE

Regional level: NUTS 0
Source: ESPON DB, year 2013
Origin of data: Joint Research the European Environment Agency (UMZ 2000), SIRE database 2008.
© EuroGeographics Association for administrative boundaries

Type of demographic ageing:

(Classification elaborated for UMZ with 10 000 inhabitants and more. Seven indicators of ageing were used: 0 to 14 years, 15 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 to 74 years and 75 and more)

- 4 — 216 UMZ
- 2 — 766 UMZ
- 3 — 475 UMZ
- 1 — 846 UMZ
- 5 — 658 UMZ
- 6 — 722 UMZ
- 7 — 526 UMZ

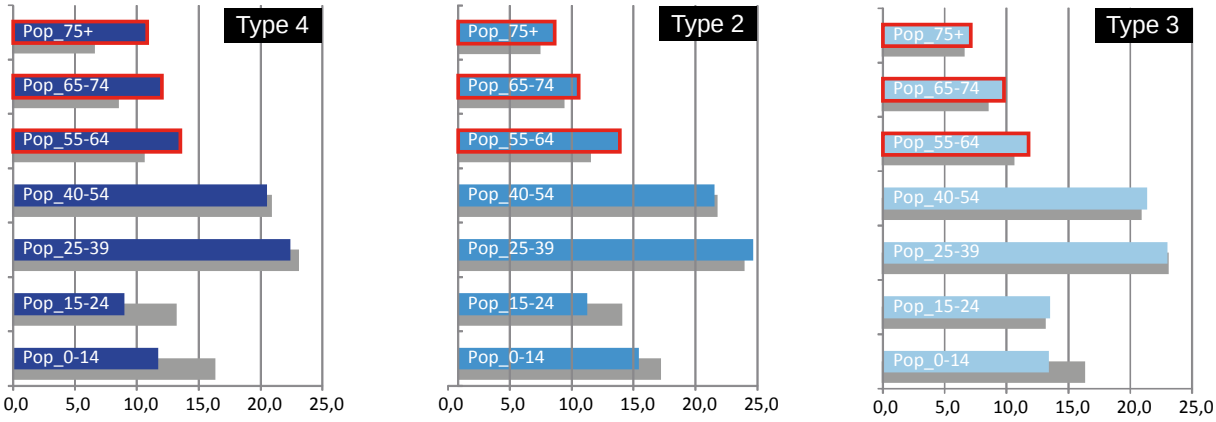
Total population:

- 9 700 000
- 2 300 000
- 150 000

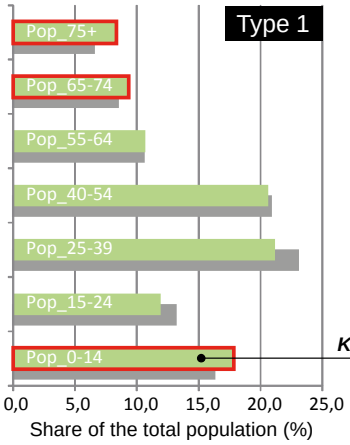
No UMZ named

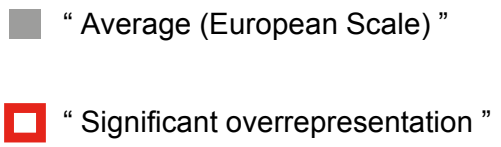
Out of Espon space

AGEING profiles



INTERMEDIATE profile





Key of lecture

The UMZ displayed in light green are included in the "Intermediate profile". As regard to the average of the overall European UMZ, these urban units are characterised by a significant overrepresentation of the population aged below 15 years, of the population aged from 65 to 74 years and of the population aged above 75 years.
In the average of this "Intermediate profile", the population below 15 years represents 17,9 % of the total population, while it represents 16,4 % for the overall European UMZ.

YOUNG profiles

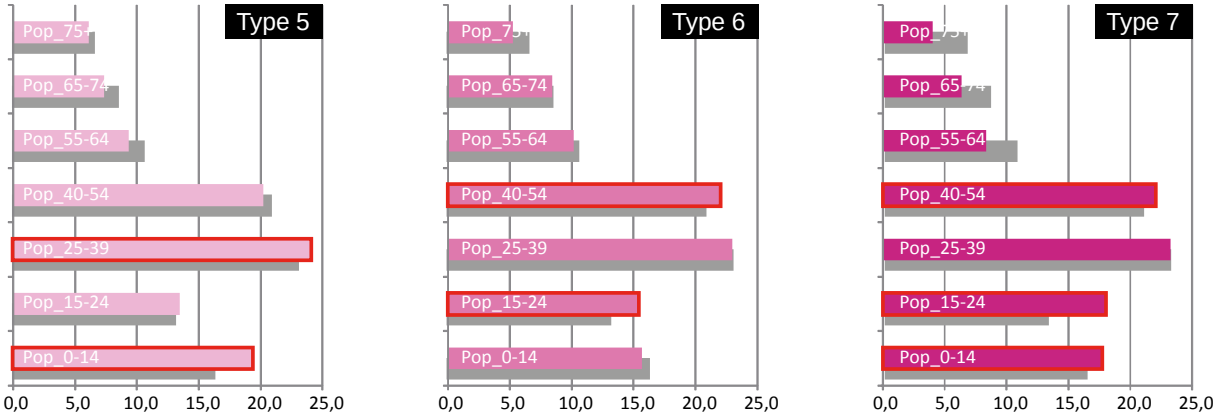


Figure 8 - Typology of age structure by city at European level and class profiles

1.2.2 Feasibility of FUA construction

The issue of international harmonisation is fundamental for every European urban database, but it is particularly complex when considering objects such as Functional Urban Areas (FUA). Over the last few years, significant progress has been achieved concerning the harmonisation of FUA databases defined from commuting flows. Indeed, compared to the first attempts that either concerned the very top of urban hierarchy or mixed functional criteria with administrative ones, FUA from IGEAT (version 2000, revised in 2012) and now LUZ from EC-OECD (2012) provide statistical homogeneity for an important set of cities, through a top-down process. However, major questions still remain about the comparability of these urban objects, in relation to the sources and to the choice of parameters. We still do not know if it is possible and relevant to implement the same parameters in each country, taking into account the variability of sources, the heterogeneity of the settlement contexts and the differences of LAU2 sizes. The aim of ESPON M4D was to shed light on this question by 1) reviewing the remaining obstacles to harmonisation for FUA based on commuting flows, 2) assessing the feasibility of an alternative method based on accessibility isochrones and speed parameters, and 3) questioning the complementarity between both methods. To that end, we constructed and compared two types of functional areas for 12 large cities that reflect different case studies in terms of urban structure (monocentric/polycentric pattern), regional settlements and urban sprawl dynamics (see [M4D, 2014 (1)]).

Concerning the issues raised by the construction of FUA based on commuting flows, we first focused on data availability issues (in particular commuting flows and employment data). In particular, we explained why the handling of SIRE database is still hampered by different types of inconsistencies and we identified the countries where SIRE data gave a good approximation of commuting flows between LAU2. The heterogeneity of sources has been further addressed by recalling the dramatic influence of the spatial resolution of LAU2 on the size of the polarized area. Lastly, we emphasized more specifically the sensitivity of FUA perimeters to the choice of the main parameter (proportion of each LAU2 active population working into the FUA core) and we compared different perimeters obtained for 12 European cities.

In order to further analyse the issue of harmonisation for FUA, we have tested the feasibility of an alternative method based on a time-definition of cities. Starting from the Zahavi law that underlines the huge stability of the time-budget spent by commuters (around one hour) [ZAHAVI, 1974], the method consists in constructing accessibility isochrones around urban centres. If recent works have already relied on such a concept for delineating European cities, the originality of the M4D approach lies on the modeling of traffic congestion and of speed parameters, for a significant set of cities. Different steps of such a modeling have been explored:

- 3 methods have been tested in order to select a central point for defining the centre of the city.
- The combined use of a road graph network and of a raster diffusion method has proved to be the most suitable method in order to measure the shortest time travel to the centre. We also assessed the relevance of the Euro Regional Map (ERM) for collecting the networks structure at a metropolitan scale, by comparisons with other national databases.
- The estimation of speed parameters has been approached through different models, following degree gradation in the complexity of the method. The simplest one considers congestion as a discrete function of the distance to the centre and consists in implementing a peak-hour index inside the continuous built-up areas (UMZ), defined by experts for each of the main European cities (Tom-Tom congestion index). A second original method considers congestion as a continuous function of the distance to the centre and relies on the construction of a congestion gradient around the city centre, thanks to route calculation websites. A third method draws the accessibility directly map from the Navstreet data about congestion. The results obtained through this last method were used as a reference situation for 3 cities (Paris, Berlin, Barcelona), in order to evaluate the previous models. Given the very high cost of the Navstreet database, such a validation of a congestion model and ways to improve it is of critical importance.

One of the first results is related to the modelisation of traffic congestion. We precised the importance of taking into account congestion speed instead of free-flow speed, as well as we could assess the accuracy of the congestion gradient model. In Paris, Berlin as in Barcelona, the distance that can be reached within 1 hour is about 70 to 80 km with the free-flow speed and peak-hour index model, 45 km with the congestion gradient model and 30 km with the Navstreet model. While the gradient

congestion model appears to be more relevant than the UMZ peak-hour index model, it could be improved as a better approach to Navstreet estimations, for instance by taking into account the way traffic slows down at intersections. Secondly, a first overview of how the “commuters-FUA” and the “isochrones-FUA” match each other has been given, especially for the case of Paris and Barcelona (Navstreet model) (Figure 9). It has shown that the “commuters-FUA” perimeters either fitted with the 1h15 isochrone (Barcelona) or were larger than the 1h30 isochrone (Paris).

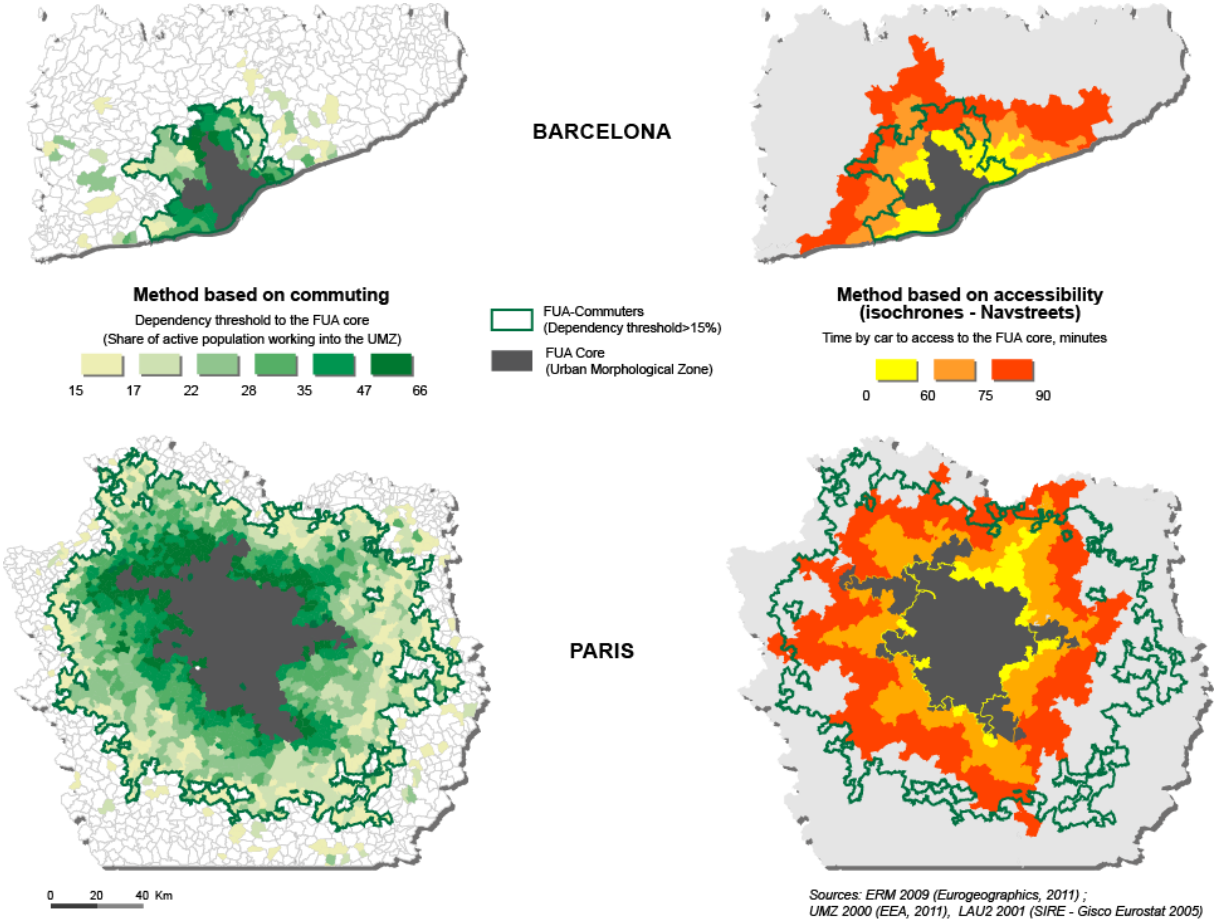


Figure 9 - FUA based on isochrones (Navstreet model) compared to FUA based on commuters: the case of Paris and Barcelona

Comparing a set of FUA based on commuters and FUAs based on accessibility through generic methods had never been realised until now and provides rich and new perspectives for the construction of European harmonised databases. In future, these first exploratory results deserve to be looked at in more depth, in order to study the differences between both perimeters on a larger set of cities. This would allow more systematic testing of the 2 or 3 main hypotheses about the main factors underlying these differences (either the size of the spatial units, or the value of travel time, between countries or between large or smaller cities).

1.3. Grid data as a method for the integration of urban and regional data

Policy implications

Grid data are not only a technical way to store geographical information. It is also the symbol of a political ambition of data harmonization at Eureopan level. The heterogeneity of the definition of regions and cities in the different Member States makes indeed very difficult to compare spatial distribution (Modifiable Area Unit Problem), especially when different periods of time has to be compared (changing delineation of territorial units). Grid data appears from this point of view **like a major step toward EU building from statistical as well as political point of view**. Grid data are also important when it comes to sustainable development perspective: "*Natural and cultural heritage are parts of territorial capital and identity*"[11]. The problem is that a lot of natural phenomena have to be observed in specific territorial divisions (Water Basin, Natura 2000, Green Corridor, mountain areas, Coastal zones, Maritime space…). All this different types of zoning can be easily derived from grid data, but not from NUTS or LAU units.

Overall M4D strategy

Regions and cities are not the only useful geographical objects relating to the objective of territorial cohesion. Several environmental questions and even technological questions asked by stakeholders concern phenomena that are not constrained by regional or urban limits. We propose to answer to these multiple challenges by using a powerful process called the OLAP Cube which makes it possible to dis-aggregate data of various types and re-aggregate in different geometries adapted to the needs of each stakeholder. We will discuss background and methodological process in 1.3.1 and 1.3.2 and we will describe the content of the several OLAP cubes available through the ESPON Database portal in 1.3.3.

Key figures

The 3 ESPON OLAP Cubes include:

- **11** geographical dimensions (NUTS, UMZ, LAU2…)
- **9** thematic dimensions (Corine Land Cover, Globcover…)
- **10** measures (GDP, population…)

Which means there are **huge opportunities for data interoperability.**

Policy relevance of the related developments

Thanks to the development of the family of OLAP cubes at different scales and for different topics, the ESPON program is now able to develop a database at grid level that can be further re-aggregated at different levels or in different zoning and for different time-periods. In that order, ESPON follows the different initiatives existing at European level: *Population density grid 2000 disaggregated with Corine land Cover by the Joint Research Center (2009)* [12] and the ongoing GEOSTAT project, launched at beginning of 2010 by Eurostat with the European Forum of GeoStatistics, "*to promote grid-based statistics and more generally to work towards the integration of statistical and geospatial information in a common infrastructure for the EU"*[13].

[11] Territorial Agenda 2020, al. 23, p.5

[12] http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2#tab-metadata

[13] http://ec.europa.eu/eurostat/web/gisco/geostat-project



Further readings - Related M4D deliveries and presentations

Presentations

- Monitoring phenomena: how to manage data over time and by different geometries*, ESPON Seminar presentation, December 2011.*

Technical reports - documentation

- The new ESPON OLAP Cube includes a population time-series 1990-2011 (*Newsletter n°2, December 2012*), available in the Home Page of the ESPON Database Portal at http://database.espon.eu
- **Disaggregation of socioeconomic data into a regular grid and combination with other types of data**, available in the "Resource" section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=31

Datasets (http://database.espon.eu/db2/resource?idCat=21)

- ESPON OLAP Cube v6, 1 km² resolution and its associated Visualisation tool.
- ESPON Urban OLAP Cube v1, 100 m² resolution
- ESPON Neighbourhood OLAP Cube v1, 1km² resolution
- Disaggregation of time-series population data (1990-2011) into a regular grid, available in the "Resource" section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=45

1.3.1 Background and introduction

With ESPON M4D and the previous ESPON 2013 Database, we faced the challenge of **integrating data** which are thematically, spatially and/or temporary **heterogeneous**, e.g. time series of socioeconomic data with land cover. The harmonisation of datasets in a fixed spatial division (NUTS3) solution, selected by several ESPON 2006 projects, presented some doubts and not always convincing results. Another strategy for data integration was needed.

The Modifiable Areas Unit Problem (MAUP) study [Grasland et al, 2006] highlighted the integration of heterogeneous databases as one of the most promising applications of **gridding methods** for ESPON. Two potential fields of application were pointed out: **time harmonisation** of changing territorial units and **thematic harmonisation** and combination of heterogeneous spatial sources.

The analysis of different experiences of data integration worldwide [Deichmann et al, 2001] ; [Gallego, 2001]; [Nordhaus, 2006], carried out in the first ESPON DB project, led to the conclusion that the best way to downscale socioeconomic data and make them comparable with other kinds of data, is using a **regular grid structure**, in which each cell takes a figure of the indicator or variable.

This work has continued under the ESPON M4D project, developing a complete methodology for data integration based on the **grid** and **OLAP** technologies, plus a **webtool** to make database queries easier. The main **outcomes**, which are detailed in this section, have been:

1. ESPON OLAP Cube, containing data for the ESPON space.
2. ESPON Urban OLAP Cube, containing data for urban areas (Urban Audit agglomerations).
3. ESPON Neighbourhood OLAP Cube, containing data for the ESPON space and its neighbourhood.
4. ESPON OLAP Cube Viewer webtool.

1.3.2 Methodology overview

In order to combine socioeconomic data with other types of datasets within a single database, 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], is used as the **common integrator layer**. This means, in the end, that each single grid cell must store a single figure for each variable, either numeric or alphanumeric. Actually, depending on the nature of the dataset or variable, we distinguish:

- **Geographic dimensions**: they define the geographical unit each cell belongs to (NUTS region, LUZ, UMZ, biogeographical unit, etc.)
- **Thematic dimensions**: they define a physical characteristic of a grid cell, such as land cover type.
- **Measures**: they are numeric variables which can be aggregated by any combination of the data dimensions available in the system. They are surface units (e.g. hectares) but also socioeconomic figures (population, unemployment, GDP...).

As for **geographic dimensions**, an overlay between the dimension and the 1 km^2 ERG is undertaken, in order to give each grid cell a single feature code (e.g. a NUTS3 code, a LUZ code, etc.). Aiming at avoiding coastal discrepancies, some dimensions, such as NUTS regions, are previously submitted to spatial extension processes generating proximity layers.

Taking into account that some **thematic layers** have a higher resolution than 1 km^2 (e.g. Corine Land Cover is available at 100 m^2 resolution), the combination of such information is carried out at **100 m^2**. In this way, it is possible to store, for instance, the different land cover classes and their surfaces for each grid cell. The grid cell index (unique identifier for each grid cell) might appear repeated in the resulting output table, as many times as different land cover classes exist within the square kilometre. Geographic dimensions will have their code repeated as well (Figure 10).

Index	NUTS99	NUTS03	NUTS06	NUTS10	FUA	LUZ04	LUZ12	MUA	UMZ00	CLC90	CLC00	CLC06	HA
26421951	1178	1202	1004	1141	745	267	535	1013	117593	111	111	111	58
26421951	1178	1202	1004	1141	745	267	535	1013	117593	112	112	112	42

UNIQUE COMBINATION OF UNITS

26421951-1178-1202-1004-1141-745-267-535-1013-117593-**111-111-111**
26421951-1178-1202-1004-1141-745-267-535-1013-117593-**112-112-112**

Figure 10 - Example of unique combination of units within a grid cell

When it comes to **socioeconomic data**, typically stored by NUTS3, the main challenge is how to disaggregate such figures and get a single value for each cell. In order to do so, population has been used as an **ancillary variable**. With the release of the **GEOSTAT 2006 population grid**, high quality 2006 population figures are available at 1 km^2 resolution. That valuable information is used to redistribute population figures at NUTS3 level for the time series 1990-2011 (compiled by ESPON M4D) proportionally to the distribution of population in 2006. When doing so, figures by each grid cell are a proxy of the actual value, but they compensate each other when grid cells are aggregated back to show the result of a query, which should always be done by a bigger reporting unit (NUTS3, LUZ, UMZ...). The rest of socioeconomic variables, such as GDP/capita or unemployment are redistributed in the same way, weighting the figures by population in each grid cell. In this way, those figures can be aggregated back in a query, despite the different NUTS breakdowns or geometries involved (urban delimitations, administrative levels, etc.).

Whenever a grid cell falls in the boundary of different NUTS figures, those figures are disaggregated proportionally to the share of area of each administrative unit within the grid cell. In Figure 11, a general schema of data preparation for integration within an OLAP database is shown.

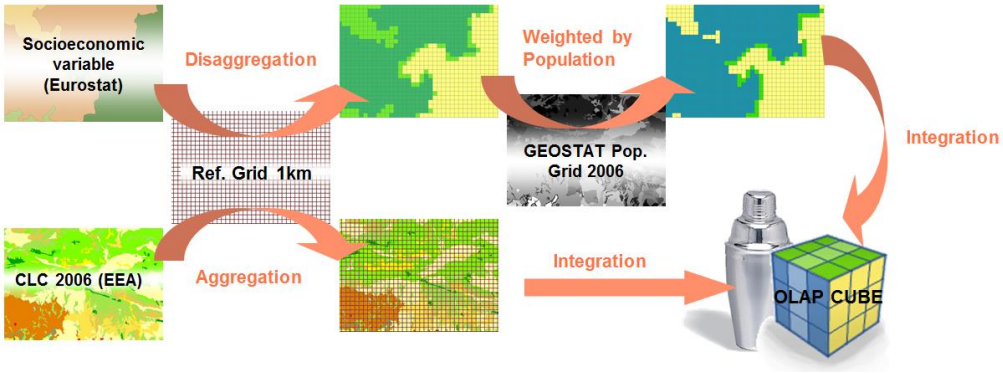


Figure 11 - General schema of data processing for OLAP integration

Once the data dimensions and measures have been overlaid with the ERG and each grid cell has a single figure for each variable (or more, in the case of Corine Land Cover), then it is possible to join all the information in a unique **OLAP database**.

OLAP (Online Analytical Processing) is a computer-based technique to answer multi-dimensional analytical queries swiftly. OLAP tools enable users to analyse multidimensional data interactively from multiple perspectives. There was a rise in the use of OLAP in the late 90s, but it was mainly used for business data in the search for business intelligence. The application of such technique to social and geographical information is rather innovative and provides added value to the integration of data by means of a grid.

The OLAP database (or **Cube**, so-called for its multidimensional nature) is created by a procedure which is fully detailed in the Technical Report, and culminates with a .CUB file as output, which can be queried using **Microsoft Excel** or an online connection to a remote server. A **User Manual** has also been produced to allow users to connect to and query the OLAP database.

1.3.3 Results

1.3.3.1 OLAP Cubes

Several OLAP Cubes have been produced throughout the project. The main one, updated regularly, covers the entire **ESPON space** at 1 km^2 and contains basic socioeconomic information, land cover information and different administrative breakdowns. An **Urban** OLAP Cube has been produced at **100 m^2** for the Urban Audit agglomerations. Finally a **Neighbourhood** Cube has been produced in order to extend the analysis beyond the ESPON space. Table 4 summarises the contents of each of the three OLAP Cubes. The methodological and technical details of each one are included in the Technical Report [M4D, 2014 (3)].

	Resolution	Geographic dimensions	Thematic dimensions	Measures
ESPON OLAP Cube (v6)	1 km^2	NUTS 1999, NUTS 2003, NUTS 2006, NUTS 2010, LUZ, FUA, MUA, UMZ	CLC00 Hierarchical, CLC06 Hierarchical, CLC90 Hierarchical, Land Cover Flows 1990-2000, Land Cover Flows 2000-2006, Land Cover Flows 2000-2006	Area in hectares, GDP 2000, GDP 2003, GDP 2006, GDP 2009, Population 2000, Population 2003, Population 2006, Population 2009
ESPON Urban OLAP Cube (v1)	100 m^2	LAU2 2008, LUZ 2004, LUZ 2012, FUA, MUA, UMZ 2000	Urban Atlas 2006	Area in hectares, Population 2006
ESPON Neighbourhood OLAP Cube (v1)*	1 km^2	SNUTS 2012	GLobcover 2004, Globcover 2009	Population 2004, Population 2009, GDP 2004, GDP 2009, Unemployment 2012

Table 4 - Dimensions and measures of each ESPON OLAP Cube

* we acknowledge the work of data collection and disaggregation processes, carried out by UNEP/GRID, as subcontractor of the UAB

1.3.3.2 OLAP Cube webtool

The **web-based** OLAP cube viewer was created to make the exploitation of the OLAP Cubes easier, providing a **user-friendly** platform to query them and get maps, charts and tabular information in one step.

The webtool has been steadily improved since it was created, and most recent efforts have been put in adapting it to the ESPON **layout** and the ESPON **Mapping guides**, a task that will be finalised for the final delivery.

At this stage, the tool allows the user to **select** what **output** they require from the query. For example they can choose to hide the map or the table.

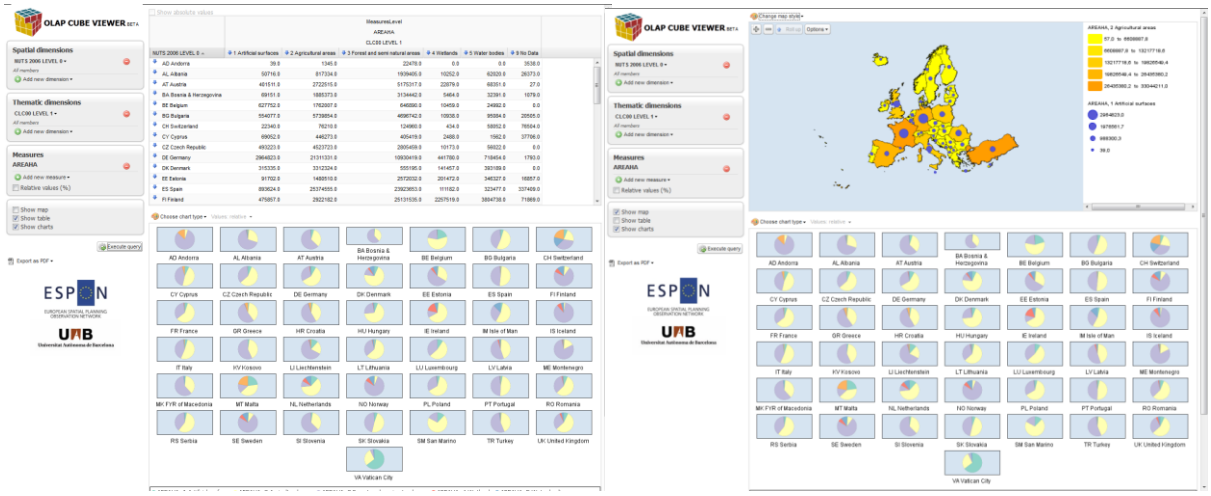


Figure 12 - Example of output from the Web tool without map and with map

All the technical details of the Web tool is included in the Technical Report [M4D, 2014 (3)].

1.3.3.3 Discussion

The methodology and results implemented under this task of the ESPON M4D project respond to a need for **data integration** within the ESPON context, useful for analysing territorial development and cohesion in Europe. They might also be helpful in the framework of **territorialisation** and **place-based** approaches of different sectorial policies, as they allow integrating and analysing spatial data together with socioeconomic figures. Although it has not been tested for policy-making purposes, the OLAP approach presented in this paper might be useful for **spatial planning** and **spatial monitoring** at different scales.

Furthermore, this methodology overcomes several classical problems, like working with different **NUTS breakdowns** or the combination of thematic with administrative and socioeconomic data.

Nevertheless, it should be stated that the purpose of this method it is not getting detailed disaggregated datasets, but using **disaggregation as an intermediate process** to combine data, and being able to query them to get sound results quickly that might help to analyse trends or facts regarding territorial and social development. Some **degree of error** is introduced when disaggregating data. Disaggregation and reaggregation at NUTS3 level might introduce a deviation of about 2-5% with regard to original figures. In this regard, the better disaggregated datasets are built up in Europe, or the better disaggregation techniques can be applied, the better will be the OLAP database produced. That is why we are closely following the developments of the **European Forum for Geostatistics** (EFGS) and the GEOSTAT initiative.

1.4. Local data

Policy implications

During a long time, the local scale was not the priority of the political agenda of the EU, which focused more on regional development (since the origins) and metropolitan area (since the development of the Lisbon Strategy and further development of Europe 2020). Local challenges has clearly been pushed by the development of the demographic crisis ("*Territories facing severe depopulation should have long-term solutions to maintain their economic activity by enhancing job creation, attractive living conditions and public services for inhabitants and businesses*[14]") and, furthermore, the economic crisis: "*Accelerating globalization and growing vulnerability to external shocks have been experienced by **regional and local** communities*"[15] The Committee of regions and the European Parliament has played a major role in the evolution of political priorities in favour of the development of studies at local scale because the average values of indicator at regional or urban level can hide the existence of internal discontinuities and pockets of poverty, deprivation and low accessibility to services of general interest. The political importance of local scale is now admitted in the framework of the concept of multilevel governance approach which "*aims to unleash territorial potential through development strategies based on **local and regional** knowledge of needs and building on the specific assets and factors which contributes to competitiveness of places*[16]".

Overall M4D strategy

Key figures

- **4** regionalisation techniques have been explored for proposing alternative geometries.
- **11** case-studies from ESPON P1 & P2 projects are available, mainly at local scale

It appears clear that ESPON should focus more and more on developing in the future the collection of data at local level (LAU2, LAU1). At this scale, the M4D project has explored solutions to harmonise the delineation of local administrative data and overcome the MAUP effect at local scale (part 1.4.1). Innovative measures of accessibility to services and more generally Very Important Geographical Object have been experimented also (VIGO, part 1.4.2). The VIGO framework could be a solution for overcoming the problem of the lack of data at local scale in a European context. The elaboration of indicators based on VIGO highlights how this scientific matter should be handled. The aim of this exploratory work was above all conceptual and methodological to explain the interest of the approach for exploring local data in the future of ESPON.
Another key element regarding local data has consisted by the metadata harmonisation of the heterogeneous case-studies delivered by ESPON projects. The workflow related to this part included challenging and time demanding components, starting with the data check for a large number of indicators and datasets and finishing with the full integration of the data into the ESPON Database Portal (part 1.4.3).

Policy relevance of the related developments

The development of case studies by ESPON 2013 has clearly been a relevant answer to this political demand, and the storage of ESPON "Case-study" data has been a first important step in the development of local scale in ESPON database. But an archipelago of local case-studies is not aan adapted answer to the political demand for new evidences on local trends for the whole ESPON territory. The experience realized on LAU2 data has demonstrated that it was not only possible to collect statistical data at this level but also to produce new indicators based on accessibility to Very Important Geographical Object like cities, airport, universities. More, we have elaborate solutions for the automatic design of harmonized functional areas, more adapted to the observation of the effect of demographic and economic crisis than classical administrative units. This experience will be without any doubt developed in the future ESPON 2020 program, as answer to a growing political demand.

[14] Territorial Agenda 2020, al. 28, p.6
[15] Territorial Agenda 2020, al. 11, p. 4
[16] Territorial Agenda 2020, al. 11, p. 4



Further readings - Related M4D deliveries and presentations

Technical reports / Tools

- **Local data in M4D : LAU2 and VIGO**, available in the Resource part of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=31
- **A GIS tool for regionalization** (ask to the ESPON Coordination Unit to access this tool)

Datasets (available in the "Resource" section of the ESPON database portal: http://database.espon.eu/db2/resource?idCat=21)

- Location Quotient for Land Use at LAU2 level
- Regional demographic centrality index
- Shift and share analysis of the demographic dynamics between 2006 and 2011 (ask to the ESPON Coordination Unit to access this dataset)
- Distance to the nearest highway entries at LAU2 level (ask to the ESPON Coordination Unit to access this dataset)
- ESPON Case-studies interface, available in the "Search" section of the ESPON Database Portal at http://database.espon.eu/db2/case-study-search

1.4.1 The alternative geometry

1.4.1.1 Theoretical background and relevance of the concept at the European scale

The ESPON M4D argues that ESPON should proceed as the OECD and build an original hierarchy of territorial units. This new hierarchy would not replace the current official definition of Eurostat (NUTS or LAU) that is still relevant when works are produced for subjects related to EU regulations like structural funds. But this new hierarchy could be used for strategic planning and more generally studies on territorial agenda and cohesion.

In concrete terms, we insist on the interest to create a coherent hierarchy of functional divisions with 5 levels (FUNC1 to FUNC5) where we use the best compromise between conceptual coherence and minimization of cost for collection of data [M4D, 2011 (2)]. Several attempts exist in national context to create such intermediate territorial divisions. In Germany, the BBSR has proposed an intermediate territorial division between the NUTS2 and the NUTS3 (ROR) especially adapted for cross-border comparisons (Figure 13, on the right) [BBSR, 2014]. This territorial division is already officially introduced in European context as German OECD TL3 level and could be an appropriate candidate for the FUNC3 level (meso-regional level).

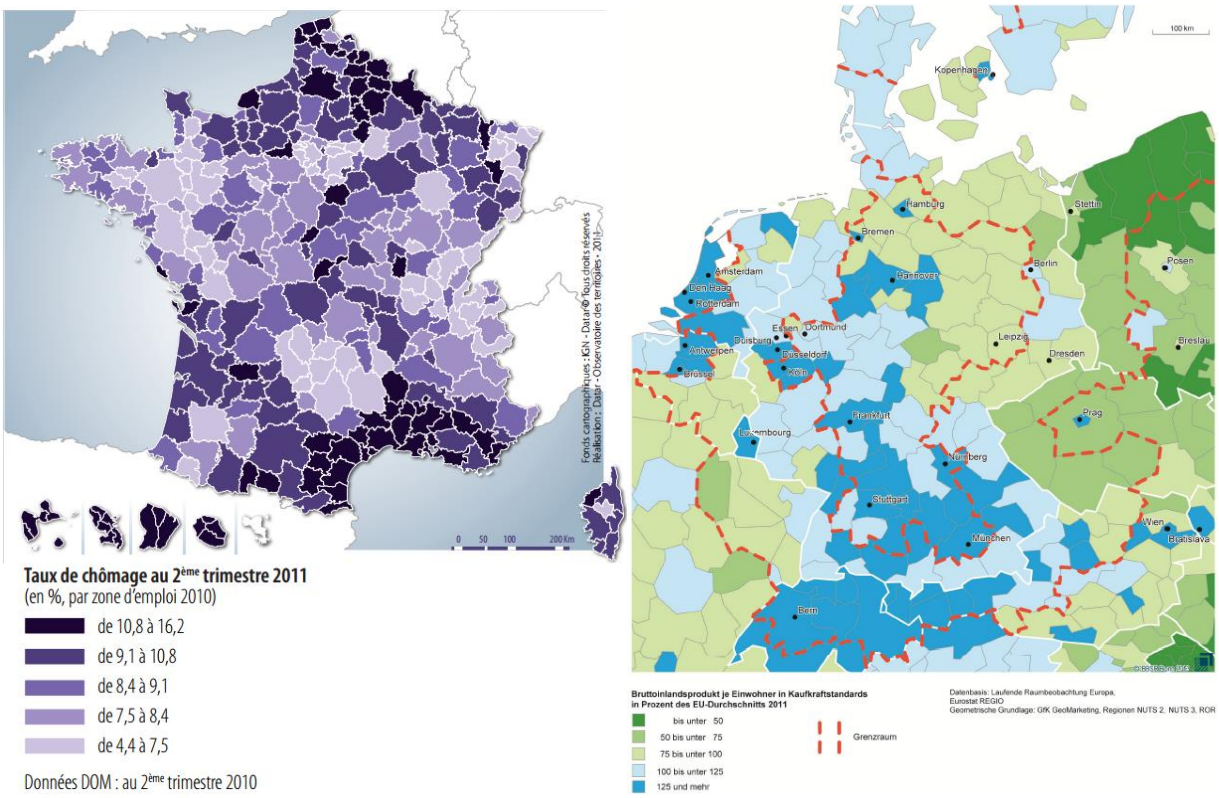


Figure 13- Two examples of alternative territorial divisions: *Zones d'emploi* (France, left), and *Raumordnungsregionen* (Germany, right)[17]

The micro-regional level (FUNC4) is in our opinion the most complicated but the most important for the ESPON program as a whole. It can be defined in different ways but is related to concepts of daily relations for labour, education, or access to services. In other words, it sticks to different concepts like "labour market area" or "life basin", which exist more or less in all countries (Figure 13, on the left [DATAR, 2011]). From our point of view, a mixture of LAU1 and NUTS3 is the adapted territorial level if we are looking for a pragmatic approximation. In the Local Data Working Package of the M4D Project, we focused the analysis on the creation of this "FUNC4" level. We present some proposals to create this micro-regional level by proposing a common methodology for all the ESPON Area.

[17] On the left: unemployment rate in 2011. Territorial division : *bassin d'emploi* [DATAR, 2011].
On the right: gross domestic product per inhabitant in pps in percent of the EU average 2011. Territorial division for Germany: Raumordnungsregionen [BBSR, 2014]

The elaboration of an alternative geometry for the ESPON area is basically a classical problem of regionalisation). As it was specified in many occasions, this alternative is scientifically oriented and it does not intend to replace any administrative frame, at LAU1 level. From a scientific point of view, it serves to a multiple purpose:

- in multi-scalar analysis an intermediate layer is missing for some countries (the LAU1 level), making the analysis impossible on a cross-border situation.
- the elaboration of this alternative geometry is able to eliminate the MAUP when translating the spatial analysis from one scale to another
- it minimizes the number of spatial units involved in some studies, conserving the general results and cartographic outputs. Our objective intersects some work already done on this topic (Study on comparable market areas, Research Report)

The solutions we propose are derived from recent studies that emphasize the potential utility of the regionalization for decision-makers and policy designers (districting and re-districting problems, gerrymandering, planning regions), such as the Study on Comparable Labour Market Areas launched by the European Commission [COOMBES M. et al., 2012], which explores "*the possibility of a consistent statistical classification of the whole EU territory, defined on a functional basis"*.

1.4.1.2 Problems set within the M4D Project

Our intention in the "alternative geometry framework" is to propose an **harmonized methodology** for aggregating LAU2 units in larger regions (smaller than the NUTS3) and building the milestones of this homogeneous FUNC4 nomenclature. Based on an objective function that minimizes or maximizes the dissimilarity between statistical regions, several consequences must be clarified:

1. How to define a similarity/dissimilarity index between the LAU2?
2. How to choose the optimal number of new regions (pseudo LAU1)?
3. How to introduce territorial constraints, so that the new regions respect the NUTS3 limits?
4. How to control the quality of the model?

Methodology and solutions

Each of the four problems implied a pragmatic solution, having in mind the idea of working with one software package, eliminating the extra-time needed for the conversion of files.

1. The similarity/dissimilarity index can be implemented for one or more than 1 variable, using a classical technique. The implementation in a GIS is described in the Technical Report [M4D, 2014 (2)].
2. Multiple solutions can be tested. In fact the number is a sub-problem of the objective function. For example, one will intend to maximize the coverage of the candidate centres in a specific impedance (similarity index) or maximize the attendance of these centres, considering that the similarity index is weighted with a distance decay function. More details will be provided in the Technical Report.
3. In a common approach, one will use a similarity/dissimilarity matrix of LAU2 that ignores the territorial belonging. However, a simple constraint can be introduced to purify the matrix from not needed values (*if not in the same NUTSx, eliminate the dissimilarity between i and j*)
4. The control is provided by calculating a territorial autocorrelation coefficient (TAC).

TAC = 1- (Average of internal dissimilarities between i and j/ Average of external dissimilarities between i and j) where :
internal: i and j belong to the same new region (pseudo LAU1)
external: i and j don't belong to the same new region (pseudo LAU1)

A class profile for each new region can also be added. Taking into account that the output of the method involves a large number of classes, this profile is optional. In the case of Poland, 100 pseudo-LAU1 were created and for each new unit a class profile based on the standard deviations of the basic indicators was conserved.

Based on the literature, in the elaboration of an alternative geometry based on the LAU2 layer, we have tested several methods to delineate the pseudo-LAU1 regions. Table 5 summarises the results derived from these methodological experiments. A larger description of these methods and the comparison with other strategies can be found in a dedicated technical report [M4D, 2014 (2)].

Possible solution	1. Selection of aggregation center (expert opinion)	2. P-median solution	3. Hierarchical clustering with spatial constraints	4. Pseudo K-mean clustering with spatial constraints.
Description	Selection of aggregation centers (expert opinion) that will eventually be used to coagulate the pseudo-LAU1.	1/ Selection of a given number of LAU2 centers regularly placed that will be aggregated in pseudo LAU1. 2/ Using a p-median algorithm to aggregate LAU2 that interrogated an origin-destination distance matrix between the LAU2 from selected countries calculated in the road network.	This method weights the similarity matrix with a spatial matrix (distance, contiguity, k-number of neighbours etc.) and it implements the classical algorithm by merging the individuals in respect to a statistical criterion (minimum linkage, complete linkage, Ward's criterion etc.).	Combination between the network analysis techniques and the exploration of a similarity/dissimilarity matrix between the LAU2. It represents a methodological mixure between solutions 2 and 3.
M4D result	Attempts in Romania, quickly abandonned	Tested on a large number of countries. It provided a first version of the alternative geometry for Eastern Europe	Implemented in RedCap[18] for the North-East region of Romania.	Implemented for the North-East region of Romania and Poland.
M4D Expertise	• Engage the expertise of a specialist. • Lacks of scientific consistency.	• Need a network of reference in the attempt to identify the candidate centers. For some countries of the ESPON area it is a problem. • The solution is based on the identification of the most "central" candidate points in a given region, making this methodology dependent on the morphology of the network. • From a computer point of view, can be implemented relatively quickly.	• Very robust from a scientific point of view (most cited methodology in the litterature). • It provides generally regions with unequal area / heterogeneity of resulting clusters. • Implementation in GIS unfriendly due to iterative process for merging the spatial units. • Not reliable for large dataset such as French LAU2.	• Solution proposed by our approach. • Need time to compute the results. We suggest to create alternative geometries country by country.

Table 5 - Several methods tested to create the FUNC4 level

Implementation, results and discussions

The benchmark of these several methods guides us to focus the work on the technique based on the theoretical combination between the P-median problems and the K-mean clustering techniques (solution 4 in Table 6). Having in mind the fact that we need to aggregate the LAU2 units in larger units and that the LAU2 are contiguous, the similarity matrix can be weighted with a binary spatial matrix (1 if i and j are neighbours, 0 otherwise). This weighted matrix of dissimilarity can be transformed in a spatial graph and this graph can be analysed using network techniques of exploration. Aggregating the LAU2 is now reduced to a problem of location-allocation in a graph, for n facilities (number of desired regions), respecting a minimum impedance (which is derived from the similarity index) between the LAU2 and the candidate centres.

	Territorial constraints (a)	No territorial constraints (b)
Similarity (1) (maximizes the similarity between the LAU2)	It creates homogeneous pseudo LAU1 regions, not overlapping the NUTSx limits.	It creates homogeneous pseudo LAU1 regions, overlapping the NUTSx limits.
Dissimilarity (2) (maximizes the dissimilarity between the LAU2)	It creates heterogeneous pseudo LAU1 regions, not overlapping the NUTSx limits.	It creates heterogeneous pseudo LAU1 regions, overlapping the NUTSx limits.

Table 6 - Four possible outcomes of the regionalization technique

[18] www.spatialdatamining.org

The case **a1** has as output homogeneous pseudo LAU1 regions that respect the NUTS3 borders. For the best results, this solution should be applied when:

- the data is spatially autocorrelated and when the NUTS3 limits make sense (some geographical distributions are independent to the territorial belonging, such as the land use)
- the data is territorially imbricate and the pseudo LAU1 layer is an intermediate territorial reference of analysis.

The case **b2** refers to pseudo LAU1 regions that involve some polarization from a centre or spatial concentrations (density of population, for example).
The validation tests (territorial autocorrelation coefficient) indicate which of the fours solutions is optimal for the regionalization/zoning aim. As the possible outcomes might be time consuming, a GIS tool is also provided as deliverable and all the intermediate steps is largely described in the technical report [M4D, 2014 (2)].

Steps of implementation:

1. Define the variables needed to create the similarity/dissimilarity matrix.
2. If an option is made for contiguity between the LAU2, build a binary spatially weight matrix. Or, choose an alternative spatial weight policy (Inverse Distance Weighted, for example).
3. If an option is made for territorial belonging, eliminate the unnecessary similarity/dissimilarity values.
4. Transform the spatially weighted similarity/dissimilarity matrix in a graph.
5. Define the number of needed regions (for example, aggregate the LAU2 in 84 pseudo LAU1)
6. Implement the location-allocation model on this graph, choosing the P-median solution.
7. Evaluate the result by a territorial autocorrelation coefficient.

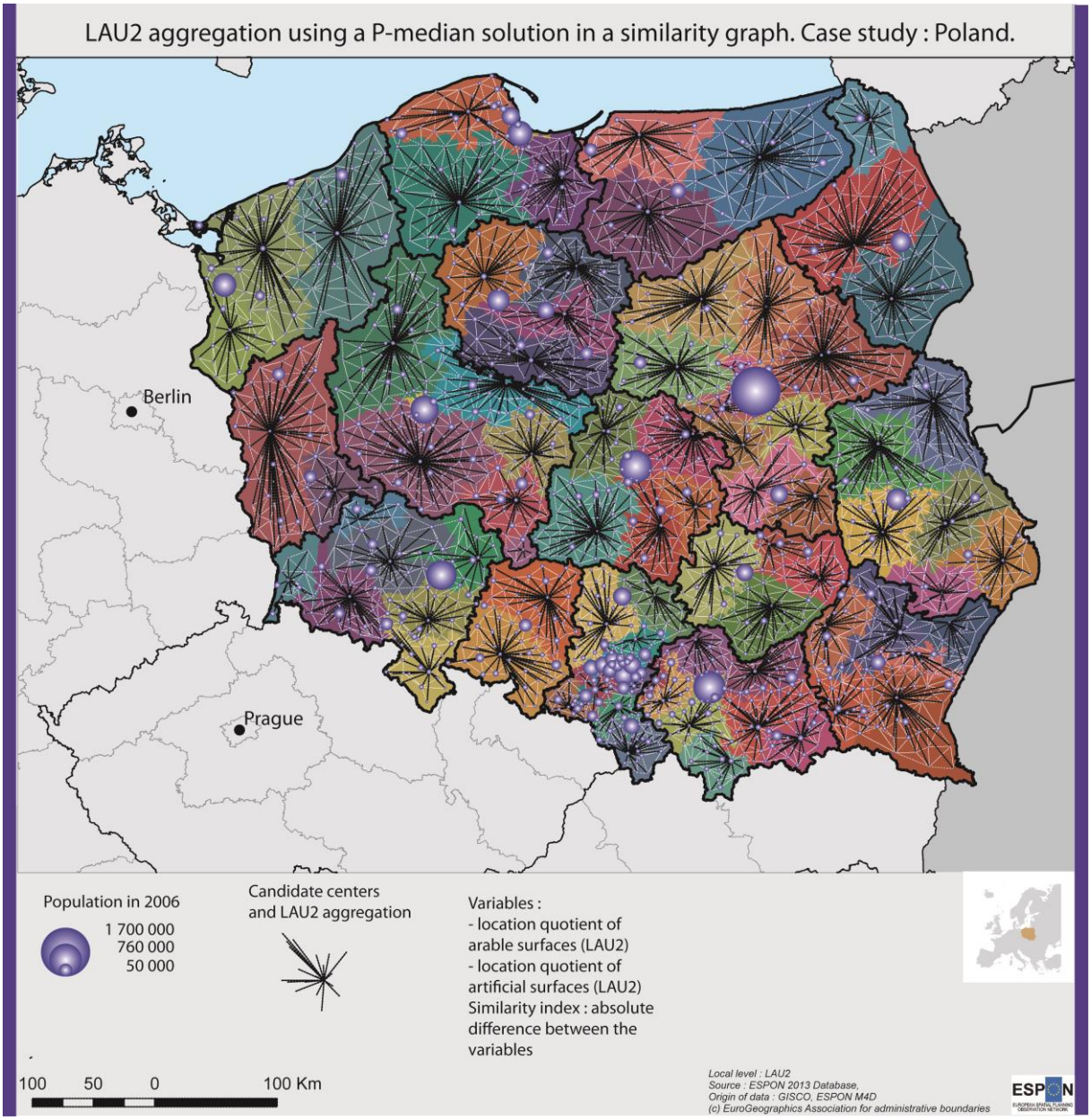


Figure 14 - Cartographic output for cases a1 and b2
(see Table 6) pseudo LAU1 in Poland with a dissimilarity index based on land-use (location quotients calculated for arable and artificial surfaces, CLC 2006 Seamless vector, http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version).

1.4.2 VIGO and the LAU2 frame

1.4.2.1 Theoretical background

The very important geographical objects (VIGO) can be defined as elements of territorial endowment that are essential in order to evaluate the territorial cohesion at local scale and to promote territorial competitiveness. Ranging from infrastructural objects (ports, airports, motorway slip roads etc.) to services of general interest or environmental objects (natural protected areas), analysing the relation between these VIGO and the LAU2 frame is a promising way to create innovative indicators at local scale. The VIGO framework offers also interesting perspectives to overcome the problem of data availability / harmonisation observed at LAU2 scale in Europe.[19]

[19] The last version of the SIRE database produced by Eurostat (2008) offers data at LAU2 scale around 2000.

As the VIGO is a in-house concept of the ESPON M4D project, it will be quite difficult to execute a literature review on this topic. For some countries, the problems related to the VIGO can be partially assimilated to more specific studies based on the territorial equipment; it should be particularly the case of France, where both DATAR and INSEE (the French NSI) provide analysis[20] and statistics[21] on this topic. The analysis of the relation between the VIGO and the LAU2 frame in the ESPON space is a matter of compromise between the territorial extension of the studied area and the availability of data.

1.4.2.2 Problem set

If for a statistician a sample of VIGO (e.g. 99 % of the universities present in a country) should be enough to evaluate them, in geography and spatial analysis one missing value will create unreliable cartographic results. Updating the data is another problem. For example, in Eastern Poland, only one airport was registered in 2012 (Rzeszow). In 2013, the airport of Lublin started to function, making the maps obsolete. All these aspects make the study of VIGO more complex than the construction of simple distance based relation with the LAU2.

The issues we face when dealing with VIGO are the set of issues we want to solve. This problem can be summarized by a description of the spatial relation between any VIGO (i) and any LAU2 (j), in a given study area. Optionally, this description should take into account the mass of i and eventually the mass of j.

	The VIGO is a LAU2	**The VIGO is not a LAU2. It is a geographical object, such as port or airport**
The VIGO has a mass. The Study area covers the entire ESPON Area.	Few datasets available from official sources. Reliability of data is uncertain and interfering with the updating problems. Could be a challenge for ESPON in the future.	
The VIGO has a mass and does not cover the entire ESPON Area	A classical model of potential of interaction* can be implemented. For example : available urban population in a gaussian span of n km. **Case 1**	The relation between the LAU2 and the natural protected areas (environmental topic) was analysed for selected countries in Eastern Europe. **Case 3**
The VIGO does not have a mass and covers the entire ESPON Area	Not relevant. If the VIGO is a discrete point spatial pattern, it can be summurized by a LAU2 frame and reduced to the Case 1.	Huff model. A probabilistic model that describes the potential interaction between j and i. **Case 4** Average distance between any j and k nearest i. **Case 5**
The VIGO does not have a mass and does not cover the entire ESPON Area	The case study of France. A description of the relation between the LAU2 and the distribution of commercial services. **Case 2**	

Table 7 - Description of the possible relations between the VIGO and the LAU2

* The model is implemented using road network distances (time, km) rather than the Euclidean ones.
** By mass we define any variables able to describe the VIGO (sq. km. of protected areas, number of passengers, number of inhabitants etc.)

1.4.2.3 Methodology and solutions

The major methodological problems to be solved are related to the quantity of information needed to be treated for each analysis. The limits of the analysis are often linked to the software architecture and ignoring them demand fragmented and iterated procedures. Basically, the methodological challenges are reduced to the translation of the geographical models (Huff, potential of interaction, average distance to the k nearest VIGO etc.) in steps of implementation in GIS. It might look a trivial problem, but it is not, especially when we need to weight the VIGO with a mass variable. As in the case of the alternative geometry, tools were designed to accelerate the model's implementation. These tools are described in a technical report, together with the methodological approach.

1.4.2.4 Implementation, results and discussions

The implementation of the methodology we propose is easier to understand with case studies. These case studies are derived from Table 7:

[20] http://www.datar.gouv.fr/observatoire-des-territoires/fr/acc-s-par-indicateurs/coh-sion-territoriale?ind=681
[21] http://www.insee.fr/en/methodes/default.asp?page=sources/ope-adm-bpe.htm

Case 1: The VIGO is an LAU2, has a mass (demographic mass in 2011) and does not cover all the ESPON space. It is a case study on selected countries in Eastern Europe. The definition of the urban LAU2 respects the national definitions. The values of the model can be optionally constrained by NUTSx borders.

Case 2: Using distances and the cumulated population by VIGO in the vicinity, we can evaluate their territorial impact more easily. A case study provides relevant information on the relationship between the day to day VIGO (commercial services) and LAU2, at different scales (Figure 15): following this methodology, the map below highlights remote LAU2 (displayed in blue) and central LAU2 (displayed in green) regarding these typical commercial services.

Case 3: The environmental dimension of the relation between the VIGO and the LAU2 was intercepted by this case study. The problem is even more interesting when we take into account the fact that some of the European natural areas are overlapping the LAU2 and that these VIGO need a double spatial approach - as polygons and as points (centroids).

Case 4: The Huff model was implemented in a GIS, in order to evaluate the potential of interaction between a VIGO (airports) and the LAU2 from the ESPON Space. The results were impossible to calculate for Greece and Bulgaria, due to topological errors or missing transport network. The model and the associated tool can be easily implemented to other VIGO analysis.

Case 5: In this case study, the relationship between VIGO (highway entries) and the LAU2 is analysed. The connectability to the major transportation corridors is key information in the evaluation of the territorial competitiveness. It also allows us to compare the LAU2 in different territorial contexts.

The different methodological approaches and the outputs, together with other case studies, is largely described in the technical report [M4D, 2014 (2)]. Illustrations and detailed descriptions is also provided.

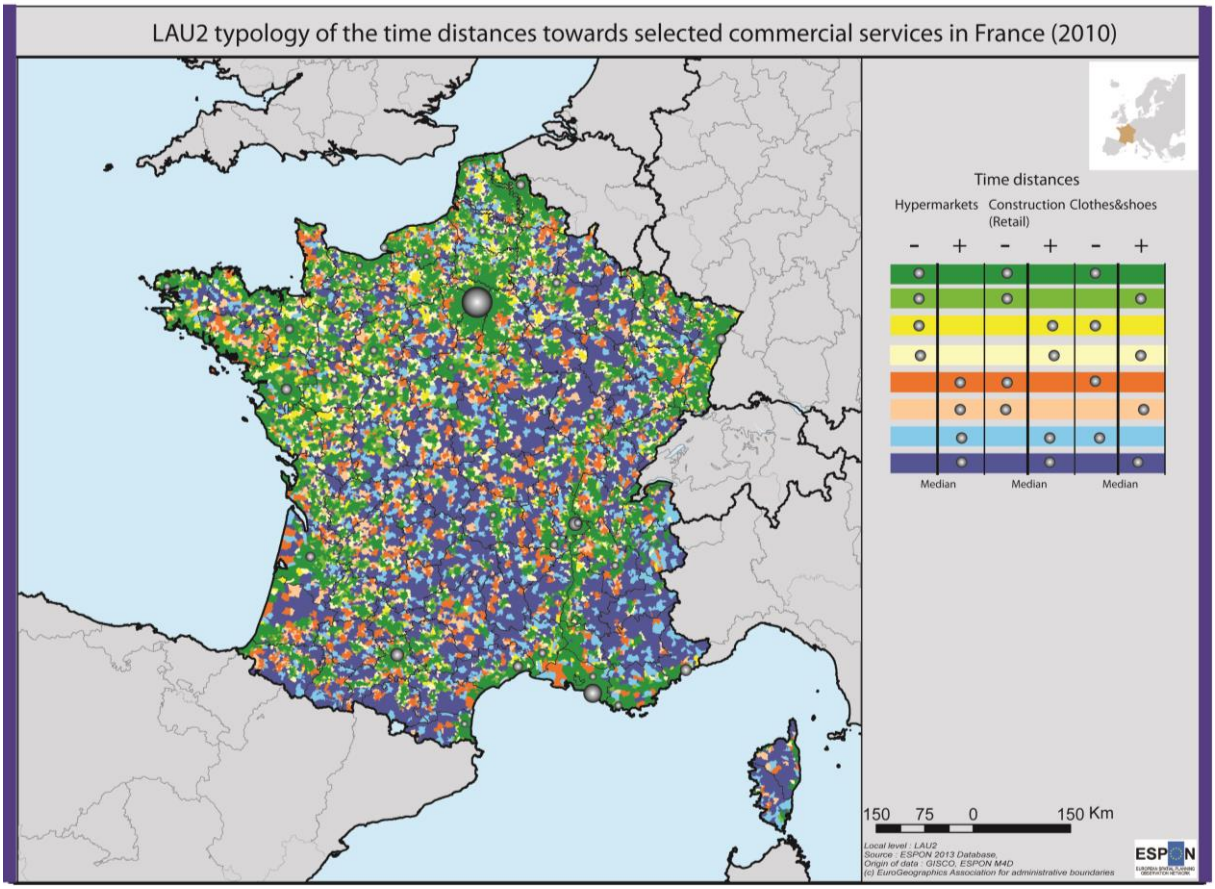


Figure 15 - Illustration of the relation between the VIGO (selected services) and the LAU2 frame in France

1.4.3 Lessons learned from the implementation of alternative geometry and VIGO

From a scientific point of view, the analysis realised within the M4D project show that the methodological background exists and is operational for building alternative nomenclatures and relevant measures of accessibility related to VIGO. It opens very promising perspectives for the ESPON Programme and territorial analysis at European scale in general: analysis of daily relations for labour, education, access to services, comparison of cross-border areas at local scale etc.

However, the problem still relates to data availability. It is obvious that these analyses are dependant of three major components:

- **Availability of data at LAU2 level** to create alternative geometries based on relevant indicators (aggregation of LAU2 territorial units);
- **Collection of VIGO with precise geographical location and a common definition for the all ESPON Area** adapted to the needs of the EU Cohesion Policy (hospitals, universities, airports for instance).
- **Availability of time-distance matrixes at a local scale** for measuring effective accessibility to VIGO.

Lessons learned from meeting with Eurostat show that for a large set of reasons, it is difficult to obtain in the short term this information directly from Eurostat. From our point of view, it means that it opens promising perspectives for the ESPON Program to offer useful and innovative material for the European statistical community. The results derived from the study launched by the DG REGIO: *Population data collection for European local administrative units from 1960 onwards* [GLOERSEN, LÜER, 2013], shows that it is possible. This kind of initiative should be extended in the future.

1.4.4 Other work done

The metadata harmonisation of the ESPON Priority 2 case studies is a task of major importance in the working flow related to the local data. Eleven projects from the Priority 1 and 2 that provided local data were checked, harmonised and integrated. This task involved a consistent time budget, if one will consider the large number of indicators and datasets that needed harmonisation. The main challenge in implementing this activity is related to the high degree of heterogeneity in the datasets, the geometries and the indicators, heterogeneity that complicates a task that is already repetitive and time consuming.

1.5. European Neighbourhood

Policy implications

The extension of ESPON database to eastern and southern neighbourhoods is probably not at the top of the political agenda of the European Union, at least when it comes to the Cohesion Policy which is, by definition, focused on the reduction of internal disparities. But policymakers are aware of the fact that the European Union cannot be an island of prosperity surrounded by closed borders. The ESPON ITAN Project reminds in its Final Report [ESPON ITAN, 2014], that "*the European Neighbourhood Policy launched in 2007 brings a general framework for cooperation, security and common development and provides a transversal financial instrument (ENPI) for a large set of actions. The Europe 2020 Strategy gives also a role to enlargement and regional integration with neighbouring countries. The Cohesion Policy promotes cross-border and transnational cooperation and macro-regions including neighbouring countries since it encourages economic activity and territorial development across internal and external EU borders. The "Fifth report on Economic, Social and Territorial Cohesion" (2010) highlights the need for peripheral EU regions to enhance transport infrastructures, cross-border links and communication".*

Finally the Territorial Agenda shows also that the European neighbourhood is not ignored. It is namely mentioned that "*cohesion at external borders [of the EU] is crucial as disparities and differences in legal, social and political systems have important consequences, especially in terms of migration and trade*[22]" or that "*attention shall be paid to areas along external borders of EU*[23]".

Overall strategy

Therefore, the emergence of the neighbourhoods issue in the EU political framework is a well-known occurrence of the last years: European Neighbourhood Policy, cross border and transnational cooperation promoted within the EU Cohesion Policy, EU2020 Strategy giving a role to enlargement and regional integration with neighbouring countries etc. But creating synergies between the EU and its neighbourhoods implies that we also need to have tools which enable us to measure potential disparities or complementarities between these geographical areas, which do not currently exist in a harmonised way. In that respect, we have concentrated our efforts on the extension of regional databases and geometries towards east and south, producing delineation, nomenclature and basic data "Similar to NUTS" (part 1.5.1). The extension of urban databases to the neighbourhood is another striking element in that context. The part 1.5.2 highlights the huge heterogeneity of urban definitions across the EU neighbouring countries. Despite the large amount of data collected for neighbouring cities, it underlines the urgent need for launching a process of data/definition harmonisation in the future.

Key figures

- **23** neighbouring countries covered by the SNUTS nomenclature
- **5** mapkits available for displaying SNUTS units on maps.
- **3** types of urban definitions existing in neighbouring countries

Policy relevance of the related developments

The M4D project has not directly collected statistical data on the regions of the European neighbourhood (this task was dedicated to the ESPON ITAN project). However, the M4D project has provided a technical support for cartography, data collection and data organisation. One important task consisted in the creation of a territorial hierarchy for non-UE countries: SNUTS (similar to NUTS) and monitoring the territorial changes of this nomenclature through time. More, the M4D project has started to explore the possibility to collect urban data for the European neighbourhood and also designed a first prototype of a OLAP cube for the European neighbourhood (cf part 1.3). ESPON 2020 program will be therefore ready to further develop this topic if the political priority for cooperation with the European neighborhood increases in the future.

[22] Territorial Agenda 2020, al. 17, p.4
[23] Territorial Agenda 2020, al. 31, p.7



Further readings - Related M4D deliveries and presentations

Presentations

- Territorial Factors for Global Competitiveness and Economic Growth in the ESPON Area and its neighbourhood (*ESPON Seminar presentation, December 2012).*
- Demographic change in the ESPON Area and its neighbourhood *(ESPON Seminar presentation, December 2012).*

Technical reports - documentation

- **European Neighbouring cities**, available in the "Resource" section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=32
- **Territorial units in Russia, Ukraine, Belarus and Moldova and NUTS classification**, available in the "Resource" section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=32
- **RIATE contribution for the Paphos Report**, delivered to the ESPON Coordination Unit for internal use.

Datasets - Mapkits

- **European neighbourhood regional data** (nomenclature, basic data and geometries)**,** available in the “Resource” section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=44
- **European neighbouring cities** (point location and basic data)**,** available in the “Resource” section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=44
- **European neighbourhood mapkit**, available in “Map Kits” section of the ESPON Database Portal (restricted access) at http://database.espon.eu/db2/mapkits
- **ESPON Neighbourhood OLAP Cube v1**, 1km² resolution, available in the “Resource” section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=21

1.5.1 European neighbourhood regions

The aim of this challenge consisted of enlarging the data collection at regional level to the Eastern and Southern neighbourhoods in order to make possible comparisons with European regions. In fact no harmonised regional database was available for Europe and its neighbourhoods at the beginning of the M4D project. In strong collaboration with the ESPON ITAN Project, this task has consisted in defining a nomenclature comparable to the NUTS classification (SNUTS), creating seamless geometries with the ESPON Area for displaying results on maps and proposing a methodology for gathering harmonised time-series on the considered study area.

1.5.1.1 SNUTS Nomenclature definition

Following the principles of the NUTS classification[24], a nomenclature «Similar to the NUTS units» (SNUTS) has been elaborated for 23 neighbouring countries. Table 8 summarizes the SNUTS levels identified by countries. In all cases, level SNUTS0 corresponds to the country level.

It is important to mention that the SNUTS3 level has not been systematically assigned: in Russia for instance, no administrative division exists between the *rayon* (populated enough to be considered as a SNUTS3) and the *oblast* (populated too highly to be considered as a SNUTS3).

Country	SNUTS Level	Administrative name	Number of territorial units	Average population 2010 (million inh.)
Albania	1-2	Country	1	3,204
	3	Prefectures	12	0,267
Algeria	1-2	Does not exist officially	5	7,093
	3	Wilaya	48	0,739
Belarus	1	Does not exist officially	2	4,798
	2	Oblast	7	1,371
Bosnia and Herzegovina	1	Country	1	3,760
	2	Autonomous region/district	3	1,253
	3	Canton	18	0,209
Egypt	1	Region	4	20,280
	2	Governorate	27	3,004
Faroe Islands	1-2-3	Country	1	0,049
Georgia	1-2	Country	1	4,352
	3	Mkhare/republic/city	12	0,360
Greenland	1-2-3	Country	1	0,057
Israel	1	Country	1	7,418
	2	District	6	1,188
	3	Sub-district	16	0,360
Jordan	1	Country	1	6,187
	2	Region	3	2,062
	3	Governorate	12	0,516
Kosovo	1-2	Country	1	1,776
	3	District	7	0,254
Lebanon	1-2	Country	1	4,228
	3	Governorate	6	0,704
Libya	1	Country	1	6,355
	2	Does not exist officially	3	2,118
	3	Shabiyat	22	0,289
Mauritania	1	Country	1	3,460
	2	Region	13	0,266
	3	Province	44	0,079
Moldova	1-2	Country	1	3,573
	3	Republic	2	1,786
Morocco	1	Does not exist officially	3	10,650
	2	Region	16	1,997
	3	Province	61	0,524

[24] http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/principles_characteristics

Country	SNUTS Level	Administrative name	Number of territorial units	Average population 2010 (million inh.)
Palestinian Territory	1	Country	1	4,039
	2	Region	2	2,020
	3	Governorate	16	0,252
Russian Federation	1	Okrug	7	20,423
	2	Oblast/kray/republic/city	83	0,323
Serbia	1	Statistical region	2	4,040
	2	Statistical region	4	2,020
	3	District	25	0,323
Syria	1	Country	1	20,410
	2	Governorate	14	1,458
	3	District	61	0,335
Tunisia	1	Does not exist officially	3	3,494
	2	Planning regions	6	1,747
	3	Governorate	24	0,437
Northern Cyprus	1-2-3	Country	1	0,285
Ukraine	1	Does not exist officially	4	11,362
	2	Oblast/republic/misto	27	1,683

Table 8 - SNUTS levels identified for the 23 neighbouring countries of the ESPON Area

For some countries (Ukraine, Algeria, Morocco, Libya, Belarus), a SNUTS level has sometimes been aggregated from official administrative divisions for statistical purposes. It generally refers to simple geographical characteristics (North region, South region etc.), with no administrative existence in the respective neighbouring countries. This first proposal has been validated by national experts coming from the ESPON ITAN Project but must be checked further by policy experts.

This is a first version of the SNUTS nomenclature. It reflects the territorial structure of neighbouring countries observed in 2012 (year of the data and geometries collection / creation). Since the collection of the data displayed above, the European neighbourhood has experienced significant geopolitical events (Arabic revolutions, Ukrainian crisis), which impacts the territorial belonging (Crimea) or hierarchy (territorial reform in Tunisia) of some territorial units available in this version of the SNUTS nomenclature.

1.5.1.2 Methodological proposal for gathering time-series in the European neighbourhood

After having created the SNUTS nomenclature, a population time-series has been gathered for the entire ESPON neighbourhood at regional level (1979-2010). This work has consisted of a three steps process (Figure 16):

1. **Gather the raw data** "as it is" from the National Statistical Institutes. Main data sources were from the National Statistical Offices Websites, national censuses and *Statistical Yearbooks* publications. All the sources have been carefully mentioned in the ESPON Metadata template.
2. No neighbouring country delivers complete time-series. Generally, one national census is organised every 10 years. It means that to obtain yearly time-series, the first step consisted in making **inter-census estimates**. These estimates have been produced following a linear trend.
3. **Calibration to the National Population Prospect Database (United Nations)**. All the regional values coming from the National Statistical Institutes are described by a frequency relating to the value of the country level (sum equal to 1) and multiplied by the value provided by the United Nations. Consequently, the sum of the count variable corresponds to the one delivered by the UN.

At the end, the M4D project has produced two tables. The first one with "official" national data only (including a high number of missing values); and a second one with estimated values adjusted to the United Nations Population Prospect database. To download these datasets, open the "European neighbourhood regional data" in the ESPON Database Portal at the URL displayed in footnote[25].

[25] http://database.espon.eu/db2/resource?idCat=44

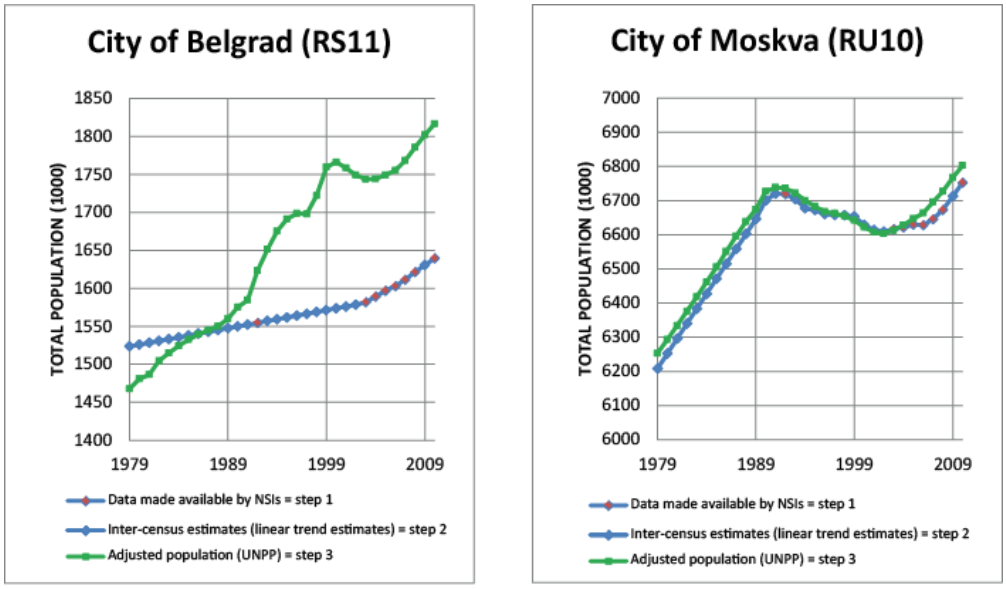


Figure 16 - Data collection, inter-census estimates and harmonisation for Belgrad and Moskva regions

1.5.1.3 Two mapkits to display European neighbourhood data on maps

Three main sources were used to generate the ESPON Map Kit on the Neighbouring regions:

- The Global Administrative Area (GADM) database for the GIS layer. This resource provides seamless geometries for all the countries of the World at different administrative levels. However, some of the shapefiles available in the GADM are outdated, due to territorial reforms not taken into account yet.
- National Country Profiles (usually downloaded from the National Statistical Institutes websites) in order to identify the levels 0, 1 & 2 of the SNUTS nomenclature, and, if possible, the level 3.
- External sources (Wikipedia, City Population, GeoHive…) when incoherencies where found between the GADM shapefiles and the National Country Profiles.

At the end, the M4D Project has delivered the following files to the ITAN Coordination team (Figure 17):

- GADM geometries with SNUTS codes, with precise boundaries (adapted for GIS calculations).
- M4D geometries with SNUTS codes, with generalized boundaries (adapted for mapping purposes).

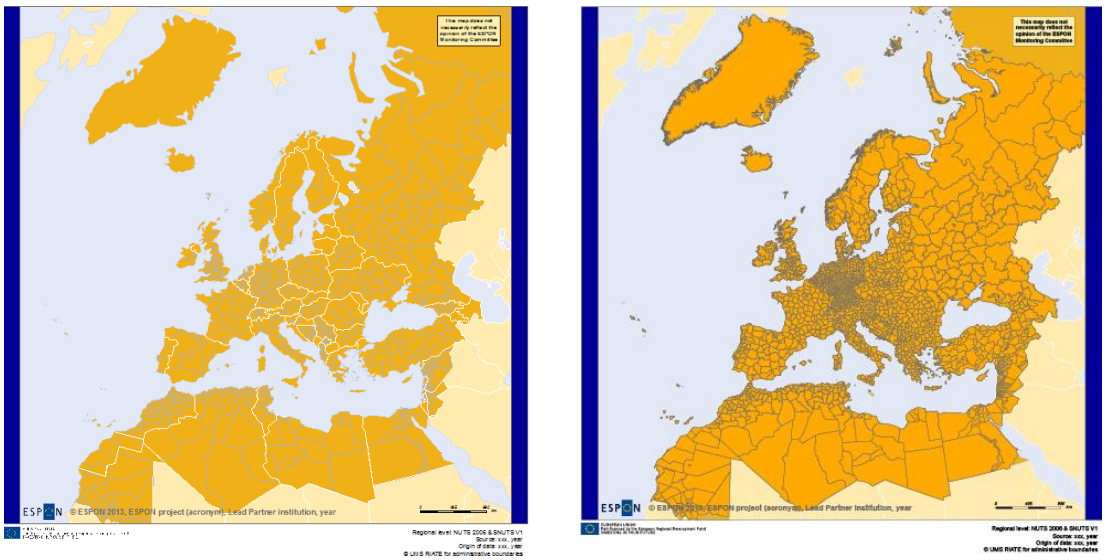


Figure 17 - Regional neighbourhood mapkit adapted to cartographic purpose (left) and layer adapted for GIS calculation (e.g. land cover calculation etc., right)

This methodological framework produces innovative pictures of the European territory and its neighbourhoods. It gives indeed new findings to feed the policy debate on the EU Neighbouring policy (Figure 18).

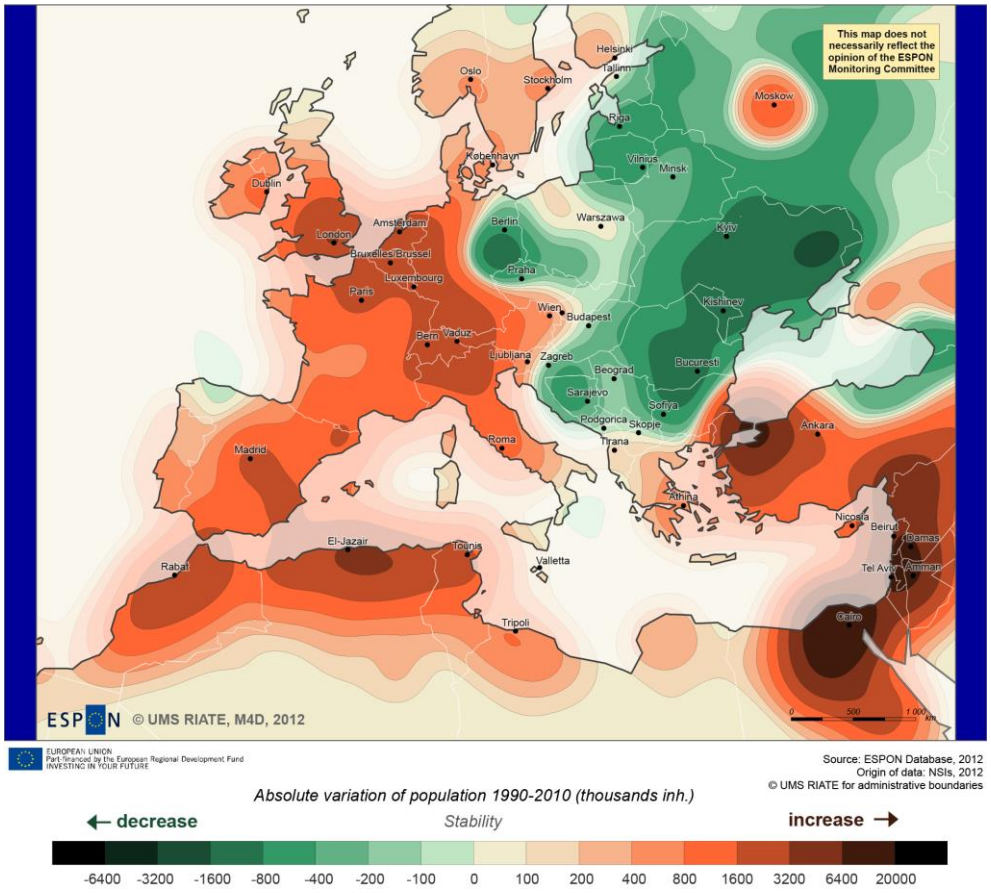
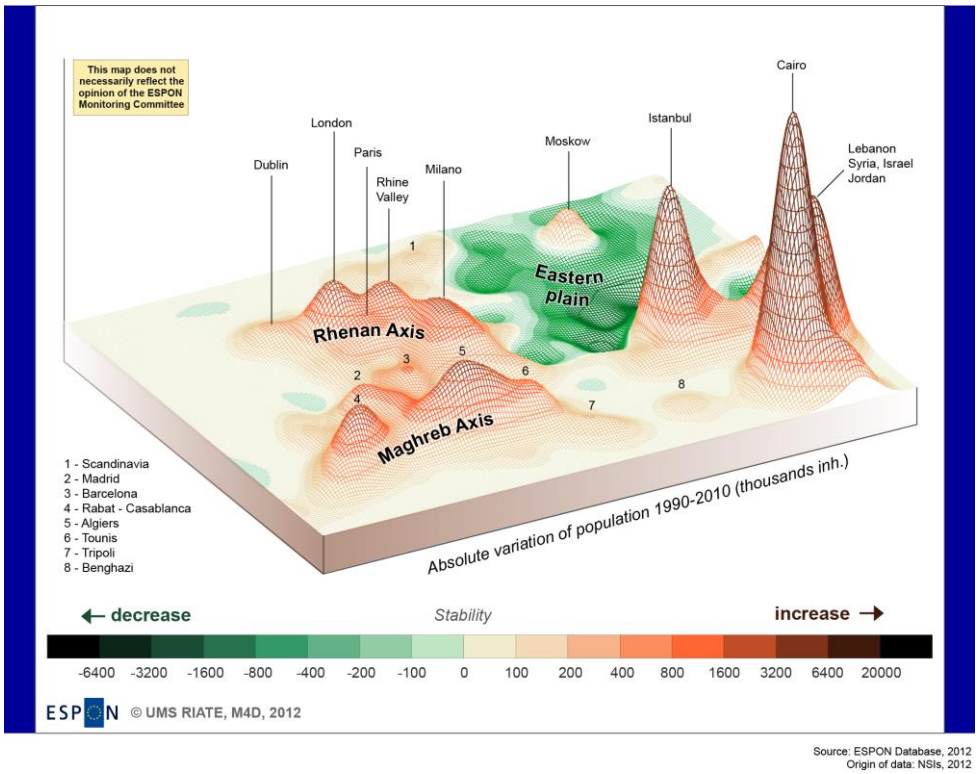


Figure 18 - Absolute variation of population 1990-2010 at regional level in the European Neighbourhood in 2D (top of the page) and 3D (bottom of the page)



The map of absolute variation of population during the last 20 years (1990-2010) reveals the intensity of demographic variation in terms of gains or loss of inhabitants in a cross border perspective (as variations are smoothed in a neighbourhood of 160 km). For a correct interpretation of the map, it is important to keep in mind that the importance of variation is firstly related to the initial amount of population in 1990. The peripheral areas with low density of population (Sahara, Eastern Plains of Russia, Northern Scandinavia) are not likely to produce important variations in absolute terms but they can experience strong relative variations, not visible on this map. That being said, it is possible to identify three major changes of the demographic structure in this study area:

 North-Western Europe remains characterised by important peaks of population growth, following the classical axis of the “Blue Banana”, from London to Milan, through Paris, Brussels, Amsterdam or Zürich. In the periphery of this main axis, we can also observe smaller peaks corresponding to growing metropolitan area of Ireland (Dublin), Spain (Madrid, Barcelona) or Scandinavia (Copenhagen-Malmö, Oslo, Stockholm and Helsinki). The absolute importance of these peaks is created by the importance of the initial concentration of population but remains limited in relative terms.

 Maghreb coastal area presents an equivalent axis of population growth from Morocco to Tunisia through Algeria. As in the previous case, the size of the peaks is related to the initial distribution of population and it is not surprising to observe maximum growth in the central part of the Algerian Coast around the capital of Algiers. Because of its polycentric urban structure, Morocco presents a peak of lower absolute intensity but covering a larger area. Tunisia appears as a continuum of the growth area of Algeria but Libya is identified by an isolated peak of growth around the capital Tripoli. The fact that the growth pole of Northern Africa appears equivalent to the one of Western Europe is remarkable, considering the fact that the initial population was much lower. In relative terms, the intensity of growth is several time higher which means that the pressure for education and employment is very strong.

 Turkey and-eastern Mediterranean Coast offer the most impressive growth peaks of the area, with three maxima observed around the economic capital of Turkey (Istanbul), around the low delta of the Nile and finally in the cross-border area of Israel, Lebanon, Syria and Jordan. The third peak reveals the potential of cross-border cooperation between countries that are currently subject to war and political conflict. In case of peaceful solutions to the Israel-Palestinian conflicts or the Syrian civil war, this part of the neighbourhood would be a major area in terms of demographic and economic cooperation. Cyprus offers from this point of view a remarkable location, exactly between the three growth poles of the eastern Mediterranean area. And the recent discovery of an off-shore oil resource in the area confirms its strategic interest for the EU. Concerning Turkey, the fact that the maximum growth pole appears in Istanbul and Ankara is also of high importance for EU, because it suggest a high potential of cross-border cooperation with Greece, Bulgaria or Romania.

 Central and Eastern Europe appears finally as the most challenging area from a demographic point of view, as it is has been characterised by very strong peaks of population decrease during the last 20 years, and this trend will probably go on in the future. The most severe population loss has been observed in eastern Ukraine and eastern Germany, with large decline in Romania, Bulgaria, Serbia and central part of Russian Plain as well, with the exception of the Moscow region. A kind of extension of the area of declining population is observed in the Caucasian area and the north eastern part of Turkey. Considered from EU point of view, this situation is certainly a major challenge because it has a lot of consequence in terms of cross border cooperation. For example, it is clear that the attraction of highly qualified migrants from Russia or Ukraine is not a sustainable strategy at a moment when these countries are facing a major shortage of young population. It is also complicated for EU firms to proceed to long-term invest in area where the decrease of population will necessarily limit the absolute economic growth. Last but not least, we can really suspect that the major driving force of the future is the complementarity between declining Eastern Europe and growing SE Mediterranean area. As long as Turkey is not member of EU, we can say that EU has really no direct influence on what will be the future of this area.

1.5.2 European neighbouring cities

One of the aims of the ESPON M4D project was also to collect urban data in a systematic way in 18 countries (for the detailed list, please see the Technical Report "European neighbourhood cities", [M4D, 2013 (4)], first by requesting official urban perimeters in Census Boards, secondly by gathering urban definitions built by researchers.

Concerning official data on urban population, we gathered results for each country except Libya, Iraq, Lebanon and Syria, and for different dates ranging from 2001 to 2013. The documentation collected in the different national census boards has been analysed in depth. A common "syntax" has been used for the description of the specifications of the available information (administrative divisions, urban definition, data availability, references) and each country is described by a single sheet following the general syntax of the Technical Report [M4D, 2013 (4)].

A typology of the definitions has then been constructed. The analysis of these definitions brings to light a huge heterogeneity: three types of urban definitions are currently used, the most frequent being the administrative one (8 cases on 14), followed by the multi-criteria approach (administrative and statistical criteria, 4 cases) and by the statistical definition (used in Israel and Jordan). The urban level is very heterogeneous and extends from 40% (Egypt) to 91% (Israel). It is however very difficult to comment on these results, regarding the diversity of the national definitions, as displayed in Table 9.

Administrative definition: Belarus	Multi-criteria Approach: Morocco	Statistical definition: Israel
There are two types of urban objects: **Cities (113 units in 2008).** • 15 cities (town of oblast subordinance), with a population of 50 000 and more. These are large administrative, economic and cultural centers. These have a good development of the production and the social infrastructures; • 98 Gorad (town of raion subordinance), with a population of 6 000 and more. These have industrial companies, social and cultural establishments, and good prospects for development and population growth. **Urban Settlements (94 units in 2008)** according to the law of the Republic of Belarus (n°154-3) of the 5 May 1998, divided in three categories: • Urban settlements (гарадскія пасёлкі) with population over 2 000, industrial enterprises and developed residential infrastructure (87 units in 2008); • Resort settlements (курортныя пасёлкі, resort towns) with population of at least 2000, sanatoriums, resorts or other health recuperation establishments, and developed residential infrastructure (1 unit in 2008). • Worker settlements (рабочыя пасёлкі): with population of at least 500, servicing industrial enterprises, construction sites, railroad stations, electric stations, or other industrial objects. (6 units in 2008).	• **Administrative approach**: the "Communes urbaines" are defined by decree ("décret n°2-92-651 du 17 août 1992"), 199 units in 2004); • **Statistical approach**: An urban center must meet at least four of the seven conditions: ✓ Have an electricity network; ✓ Have a network of drinking water; ✓ Have sewerage network (sewage evacuation); ✓ Have a hospital or clinic; ✓ Have a high school; ✓ Have a court; ✓ Have a proportion of the non-agricultural workforce of over 50%. And this center has to contain a population of at least 1500 inhabitants.	There are two definitions: • **Urban localities** (233 units in 2009) include all localities with 2,000 or more inhabitants and are classified by size. • **Metropolitan area**: for two areas (Tel Aviv area, and the Haifa area): A large number of local authorities (municipalities and local authorities), as well as regional councils and rural localities that are adjacent to one another and constitute one functional entity that integrates economic, social, and cultural relations within the boundaries of the metropolitan area. ✓ **Core**: The territory of the primary city serving as the focus of activity for the population of the metropolitan area. ✓ **Inner Ring**: The area containing the localities surrounding the core. ✓ **Middle ring**: The area containing the localities surrounding the Inner ✓ **Ring** (in the Tel Aviv and Be'er Sheva metropolitan areas). ✓ **Outer Ring**: The area containing the localities surrounding the Middle Ring. ✓ Sections (Northern, Eastern, Southern); subdivisions of the metropolitan rings.

Table 9 - Three examples heterogeneous urban definitions in the European Neighbourhood: Belarus, Morocco and Israel.

Many indicators have been collected (urban population, and sometimes urban surfaces, shapes files, centroids…) and have been enriched with associated metadata.

With regards to harmonised databases built by researchers, we worked with Menapolis (e-geopolis project) which provides morpho-statistic delineations (based on continuous built-up areas) of cities from 5 countries from the 18 under study. The comparison of the official and harmonised data shows high discrepancies of up to 30% in the level of urbanisation (Egypt, Tunisia) and/or in the number of cities (Israel, Jordan, Morocco). This major result underlines the urgent need for a harmonised urban database in the neighbouring countries.

Figure 19 outlines the interest from the international perspective, despite the uncertainty concerning official data. Networks of cities (e.g. in this case, several cities located in a restricted area) do not stop at the political frontiers of the European Union, and draw important social and economic links at international level. It is well known with the cross-border functional urban areas, for daily migrations of workers, but it is also the case with more distant cities, even if the dataflow is sometimes difficult to gather. Considering the proximity of cities allows to see the potential for such networks and so, we have modelized them by drawing proximity links. The method was developed by Céline Rozenblat in 1995 [ROZENBLAT, 1995] and consists of drawing a link on the map for each pair of cities located less than 25km apart. This distance is considered as small enough to allow the development of strong relationships between the cities. The results are striking. Different potential clusters appear clearly in Russia, Ukraine, Turkey and Maghreb.

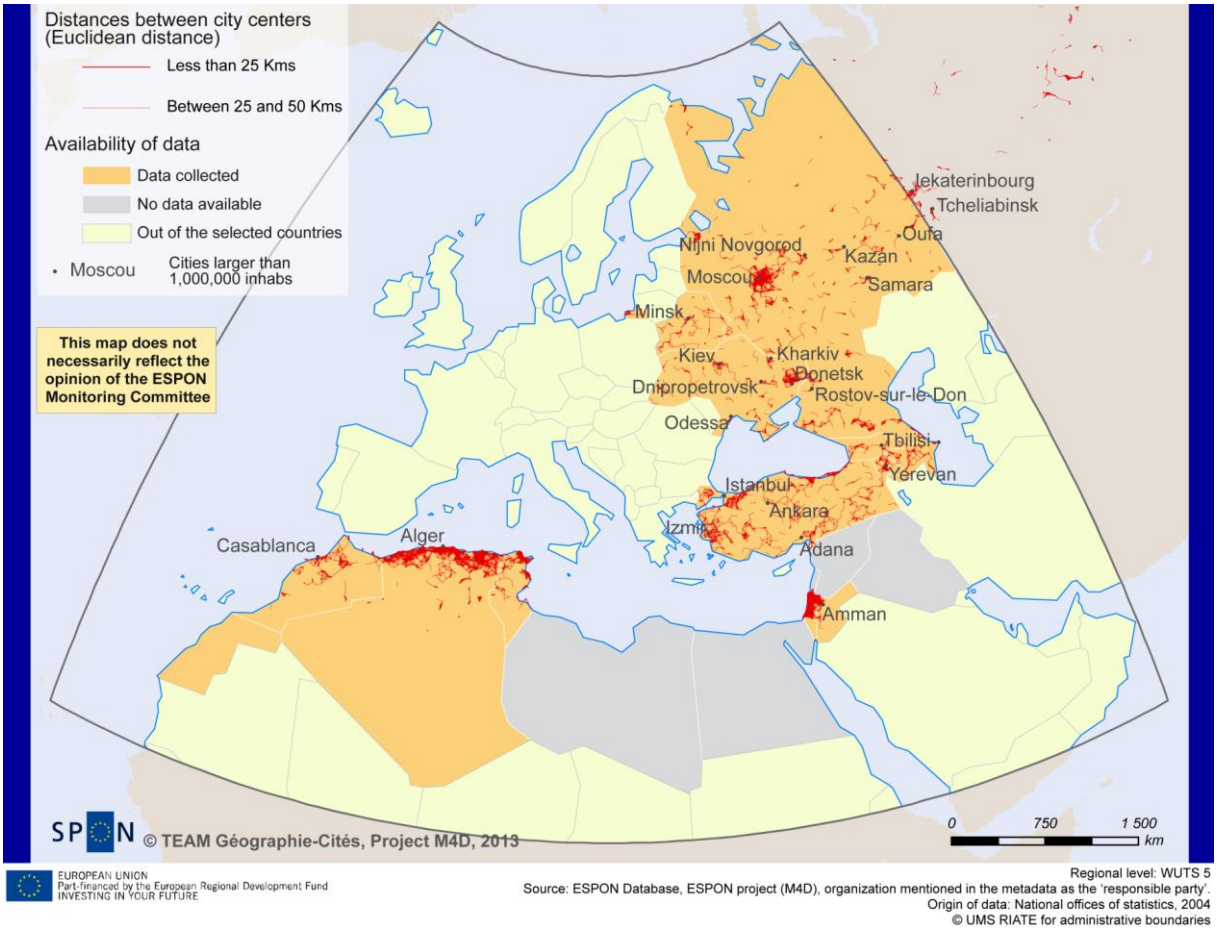
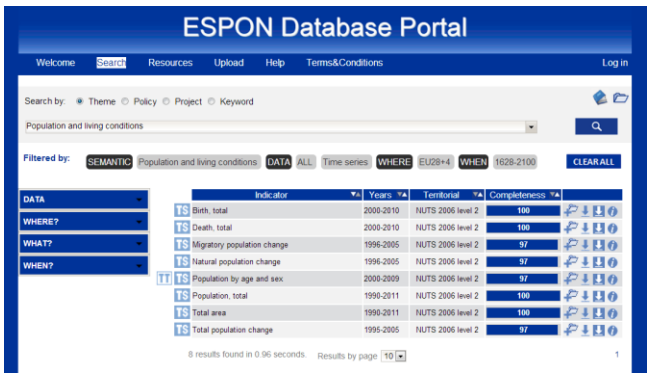
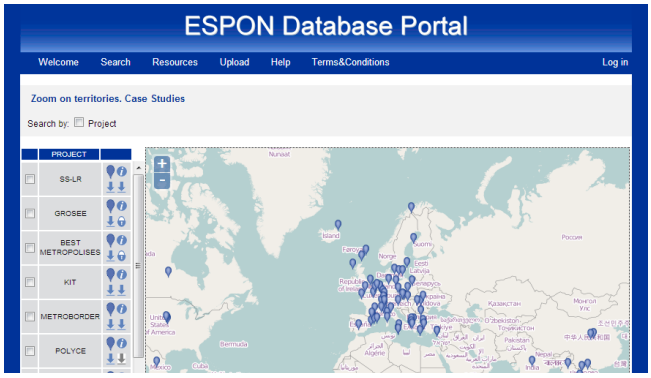


Figure 19 - Modelised networks of cities in the neighbouring countries

Part 2 - The ESPON Database Portal





Introduction

Since 2011, taking over the preceding ESPON 2013 Database Project, partners of the Project have worked together to improve the so-called ESPON Database in two main directions: data quality and data access. Thereafter, the term of data should be understood as a generic noun encompassing the term of “data” itself, but as well those of “datasets” and “metadata”. Moreover, in the context of the ESPON Database, most of the time “data” refer to “values of indicator” (i.e. statistical measures).

- **Data Quality**

Regarding data quality, a strong effort has been made towards metadata. Now, not only is ESPON metadata (describing ESPON data) compliant with the INSPIRE Directive and compatible with the ISO 19115 standard, but new metadata specifications have been designed to address the minimal and common set of information required for describing each dataset provided by other ESPON TPG Project or imported from some external organisation (for instance, Eurostat). Such specifications result in a pre-formatted Excel© file, that has to be filled out and compulsorily provided with its associated dataset file containing indicator values. However, supplying standardized metadata is just a first step towards data quality as a whole. Assessing and asserting a certain level of quality relies also on data (and metadata) checking. Our efforts have also focused on this point.

- **Data Access**

Regarding data access, the ESPON Database Portal constitutes the web point of entry towards data and metadata delivered by the ESPON Database. But, the ESPON Database Portal is much more than a dynamic web site. It is the web interface (actually a set of web interfaces) that is used by the three principal actors of the ESPON data workflow: data producers (ESPON TPG Projects), ESPON CU, and end-users (policy makers, European citizens). But, most of all, it is also the tip of the iceberg which supports the ESPON data workflow: from the delivery to the search and retrieval of ESPON data and metadata, without forgetting their quality control.

The ESPON Database Portal gives access to three different kinds of data (and metadata): key indicators, case study data, and background data. The latter – background data – are datasets resulting from applied research, targeted analysis, or in scientific platforms by ESPON TPG Projects or external European data providers (DG REGIO, Eurostat etc.). They are considered as useful and auxiliary data and as such are stored in the ESPON Database into zipped files. Case studies data correspond to data, which do not cover the entire ESPON area. They refer to local data produced or used for specific studies concerning some European regions, or to data described in a nomenclature that is not supported by the database, or even to data about non-ESPON areas and non-ESPON neighbourhood areas. In order to better understand the results, conclusions, or recommendations of such case studies led by ESPON TPG Projects, access to such data (and metadata) is paramount.

Key indicators are ESPON data. They are data (and metadata) entirely stored and managed by the ESPON database.

The first and most innovative development of the ESPON Database Portal is the data upload. Indeed, the main challenges that the project faced were firstly managing, standardising and coordinating a wealth of diverse data, ensuring data quality and traceability [YSEBAERT R., SALMON I., LE RUBRUS B., BERNARD C., 2014]. This functionality concerns data producers, namely ESPON TPG Projects. While background dataset are simply uploaded in the ESPON Database as archived files, the validity of supplied maps and metadata is checked for case study datasets. In order to upload key indicators, a semi-automatic checking process has been designed and implemented. Last but not least, an application, called the tracking tool, enable verification of the five steps – namely syntactic check, semantic check, outlier detection, ESPON CU approval, and final integration – of the key indicators upload process. This tracking tool and the underlying checking procedures are a key asset for establishing the quality of ESPON data held by the ESPON Database, and probably one of its most innovative features. This checking process alternates between automatic and manual actions. Thus, algorithms perform syntactic checks by parsing the pre-formatted data and metadata files, while the outlier detection phase relies on the execution of a set of statistical aspatial and spatial analysis methods. Experts' intervention is required for the semantic check and approval. Lastly, a script is launched in order to integrate the checked and approved data and the metadata sets into the ESPON Database.

- **A database schema adapted to ESPON needs**

The second innovative feature of the ESPON Database Portal lies at the very heart of it: the database and its schema (the "database schema" refers here to the organization of data as a blueprint of how the database is constructed and divided into database tables). Designing and implementing a database dedicated to the storage and the querying of key indicators was a very challenging task that we began to address during the former ESPON 2013 Database Project. Besides the handling of a standardized metadata schema, the ESPON Database had to take into account the intrinsic dynamics of the European territorial divisions: different kinds of nomenclatures (NUTS being the principal) whose components – territorial units or geographical objects – may change in names or contours, or may appear, or may disappear through time quite frequently… The schema of the ESPON Database has been designed for supporting such a flexibility and dynamicity (new version of a nomenclature for instance). Thus, each census indicator value stored in the ESPON Database is linked with the definition of the indicator, the way it has been calculated, the territorial unit to which it refers at, its census period, the version of the considered nomenclature, some information about the producer (for instance, an ESPON TPG Project), etc. All this information contributes greatly to the quality of the ESPON data. Again, some procedures that support the scalability and the evolution of the ESPON Database content are provided. These procedures are detailed in section 2.1.3 entitled "Database Tools and Software". For instance, through the ESPON Database Portal, a new nomenclature or a new version of an existing one can be defined, referenced and stored, which ensures some sustainability to the ESPON Database.

- **Several Interfaces**

The ESPON Database Portal is above all an interface for accessing and querying the ESPON data. While background datasets can be found through the Resources Menu, the Search Menu allows users to query respectively the key indicators and case study datasets. The query interface dedicated to the key indicators fully exploits, not only the main dimensions of such data but also the associated metadata. Thus, two levels of interrogation are provided. At the first level, key indicators can be searched by providing a theme, a policy, the name of an ESPON Project, or more broadly by keywords. Then, search can be even more refined by specifying what kind of data is expected (Data item), which area and level of territorial division is to be considered (Where? Item), which project, theme, or policy is targeted (What? item), and over which period in time (When? item). Such flexibility in the formulation reveals both the powerful expressiveness of the query and the high level of detail that can be reached when exploring the key indicators database and in the delivery of the results. Such a granular query process would not have been possible without the richness of the metadata schema and without the high level of disaggregation (from one dataset to one value of indicator for one territorial unit at one given period for one version of a nomenclature) performed by the ESPON Database schema. Results of such a query are built by assembling values into a new dataset that can be then extracted (downloaded) in the shape of an Excel© file.

Special effort was made on the visual aspect of the key indicators query interface in order to make it user-friendly, easy and fairly intuitive to use. Thus, the basket allows users to advance step by step while researching datasets. As well, for any set of indicator values returned as a result, the end-user is given the opportunity to consult the associated metadata page before deciding either to store this result temporarily in her/his basket, or to download either an Excel© file generated on the fly that contains the indicator data and metadata, or an Excel© file that contains the full dataset to which the selected indicator belongs. These options enhance the flexibility of the key indicators query interface.

The case studies interface is based on a visual query paradigm. The user is first invited to select one case study produced by an ESPON TPG in a list and then the corresponding studied areas are shown on a cartographic component (map). Again, the user is given access to a metadata page, and can download the archived file containing the data of the selected case study, or the geometries file of the selected case study made available for mapping purposes.

A large part of the ESPON Database portal is dedicated to resources. They relate to statistical production within a European context and are hopefully useful for all people interested by such information. These resources have been produced either by Database 1 and the M4D projects themselves (technical reports, documentation, datasets) or gathered from other European Institutions (DG REGIO, EEA etc.). The organisation of these resources in a thematic tree aims to provide this information in an intuitive way.

- **A lively database**

All the information included in the ESPON Database Portal is the benefit of intensive networking. Firstly with the other ESPON Projects, in order to support them with the different possibilities of data integration and to help them to upload their data at the end of their activities. Secondly, with the European producers of statistical information to share with the ESPON Community the most relevant information for their studies (LAU2 shapefiles, SIRE database etc.)

The design and implementation of ESPON Database and, consequently, the ESPON Database Portal is an evolving and continuing process. Since May 2011, every six months, a delivery of both the ESPON Database and the ESPON Database Portal is made, deployed and installed on the ESPON Website server. Both versions integrate upgraded, improved, or new datasets and functionalities. Even more frequently, new contents of the database or some application upgrades are deployed. In total, more than twenty deliveries or upgrades have been performed. It should be noted that, by the end 2014, the ESPON Database Portal would be in its final version, only the content of the ESPON database would need to be upgraded in the future. As aforementioned, the quality of the ESPON Database content relies heavily on the semi-automatic checking process monitored through the tracking tool. Although such a checking process can be deactivated, preserving and sustaining such a quality in data service is essential.

Regarding the management of both the ESPON Database Portal and the ESPON Database contents, key indicators datasets and case study datasets can be created, modified, or deleted by registered users. Also, the “News” list and users accounts can be created, modified, or deleted through the Administration back-office section of the ESPON Database Portal accessible to registered users. For the sake of completeness, integrating a new nomenclature into the database is possible, however automatically providing a new territorial division (e.g. a supplementary criterion) in the key indicators query interface would require some modification to the code of the application.

- **Implementation of Web Services: a priority for the future of the ESPON Database Portal**

As a conclusion, the ESPON Database and its sets of interfaces available through the web application called the ESPON Database Portal, forms a Spatial Data Infrastructure (SDI) that handles the ESPON data workflow, supplied on the one hand by the ESPON TPG Projects and searchable on the other hand by end-users with various profiles (policy makers, spatial planning technicians or experts, citizens, etc.). This SDI is INSPIRE compliant in relation to the edition of metadata that follows the ISO 19115 standard. On the subject of data accessibility through Open Geospatial Consortium (OGC[26]) compliant geographic web services (such as WMS, WFS, CSW, etc.), it has been demonstrated through a study led in 2013 [UNEP 2013], that key indicators stored in the database could be made searchable and extractible by means of the OGC web services technology, through an automatic

[26] Open Geospatial Consortium Web site: http://www.opengeospatial.org (last visit : 2015-01-09)

procedure. This task should be considered as a first priority in the evolution of the ESPON Database for it will allow ESPON data to be universally readable not only by individuals through the ESPON Database Portal, but also and above all, by interoperable computer systems (machine readable) over the Internet.

- **Outline of the second part of the report**

The second part of this Final Report and its associated documentation (cf further readings in the first page of each sub-sections) describe in detail the elements related to the development of the Web Application. The section 2.1 describes the data and metadata model and its associated software modules for the generation and feed of the whole ESPON Database. The section 2.2 is more user-oriented. First, this section describes how to upload data into the ESPON Database Portal, thanks to the ESPON Tracking Tool (ESPON TPG oriented). Then, it describes the main query functionalities of the ESPON Database Portal: Key Indicators Search Interface, Case-Study Interface and Resources (end-user oriented). The section 2.3 describes the networking activities that have taken place within the M4D Project. It presents the interactions with ESPON TPGs, ESPON Tools projects, infeurope and external institutions like Eurostat. Actually, this section aims at highlighting the management of interactions with the different stakeholders, which is a corner stone of such a data infrastructure.

2.1. ESPON Database

Overall strategy

Efficiently responding to the needs of the ESPON Database users, and in particular to the multi-dimensional criteria research from the ESPON Database Portal, required a consequent study and review of the existing tools delivered at the end of the ESPON Database Phase 1. The first year of the M4D project was devoted to focusing on the underlying input data workflow processes, and to redesigning the schema of the database. These two preliminary tasks were necessary and indispensable in order to support, on the one hand, a higher quality data (content of the database), on the other hand, the attractiveness, easy and efficient retrieval of the data thanks to the output search request interface. Therefore, the most prominent achievements of the first year were the delivery of the ESPON Data and Metadata Specifications and ESPON Nomenclature support documents. These specifications documents are the theoretical foundations for the model of the ESPON Database.

Key figures

- **4** NUTS versions available in the database (1999, 2003, 2006, 2010).
- **3** urban nomenclatures available in the database.
- **5** main elements composing an ESPON Dataset: dataset, indicator and source metadata, spatial binding, data.

Based on these specifications and on the redesign of the database schema, the M4D team has implemented flexible software modules for the generation and feed of the whole ESPON Database in three steps: create schema, insert nomenclatures with the module named SUNI (*Statistical Units Nomenclature Integration*), and insert datasets with the module named DatasetIntegrator (Figure 26). The implementation of these software modules is essential to ensure the sustainability of the ESPON Database infrastructure.

The following sections summarizes the key points of the specification documents and the associated software, then propose a status regarding a study about the access to data and metadata of the ESPON Database by the means of Web Services.



Further readings - Related M4D deliveries and presentations

Presentations

- Videos on ESPON Database Portal *(ESPON Seminar presentation, June 2013)*
- Monitoring and benchmarking the European territory *(ESPON Seminar presentation, December 2013)*

Public documents

- **ESPON Data and Metadata Specifications**: available in the "Help" section of the ESPON Database Portal at http://database.espon.eu/db2/metaspecifs
- **ESPON Nomenclatures Support**: available in the "Help" section of the ESPON Database Portal at http://database.espon.eu/db2/spatialspecifs
- **Dictionary of Spatial Units**: available in the "Upload" section of the ESPON Database Portal (restricted access to registered users) at http://database.espon.eu/db2/dico.

Tools and relative technical documents

- ESPON Database and ESPON Database Portal delivery
- ESPON Database and Database Portal Install Instructions
- Database schema SQL scripts, SUNI.jar, DatasetIntegrator.jar

2.1.1 ESPON Data and Metadata Specifications

Scarcity of data documentation within the ESPON Programme Phase 1 had been seen as a key impediment to the building and use of the ESPON database. Difficulties stemmed from uncertainties about legal constraints, sources, units, etc. Building a rich database would have been useless without this effort to describe precisely the data that had been gathered and integrated into the database. Creating and organizing metadata is therefore an additional, important, and transversal challenge for the ESPON community. Indeed, to be useful for ESPON projects and other end-users, data must always be accompanied by metadata, including information about their quality and sources. It is also particularly important that the metadata is created in a manner that is consistent with International (ISO) and European (**INSPIRE**) Standards, to ensure the long term sustainability of the database, and to make it compatible with other national and international database initiatives.

The **ESPON Data and Metadata Specifications** document is a specification of the metadata model intended to be applied by the ESPON Projects providing data for the ESPON 2013 Database. Firstly, it describes the generic conceptual model of the ESPON Metadata (the Abstract Metadata Model). Secondly, it presents the implementation of the abstract model using the international standards (ISO-19115 and INSPIRE Directive). Finally, it explains the implementation of the abstract model in a tabular file format (Excel files), the most commonly used format during Phase 1, and Phase 2.

To ensure correct data processing and integration into the ESPON 2013 Database, all data providers participating in the ESPON projects must respect the specifications carefully. In the context of the statistical results produced by different ESPON projects, data is represented by sets of indicator values that have spatial, temporal and source bindings. The indicators values represent the data in the strictest meaning of the term. Each indicator value is linked to a statistical unit; it is valid for a particular period or instant of time and must refer to its source of production. These bindings are bridges to the metadata, where the spatial, temporal, ontological and source contexts are detailed.

The ESPON Metadata is developed according to the requirements of the INSPIRE Directive and the heterogeneity of the data coming from different ESPON projects. It represents a complex system of pieces of information grouped into several hierarchical elements that have distinct semantic roles. The following figures aim at providing the reader with an overview of the main elements that compose the expected set of metadata in a Key Indicator dataset. The four main elements of metadata are:

1. The **Dataset Information** element groups general metadata properties describing the entire dataset. It is the main element of the metadata model and the root of the metadata elements hierarchy.

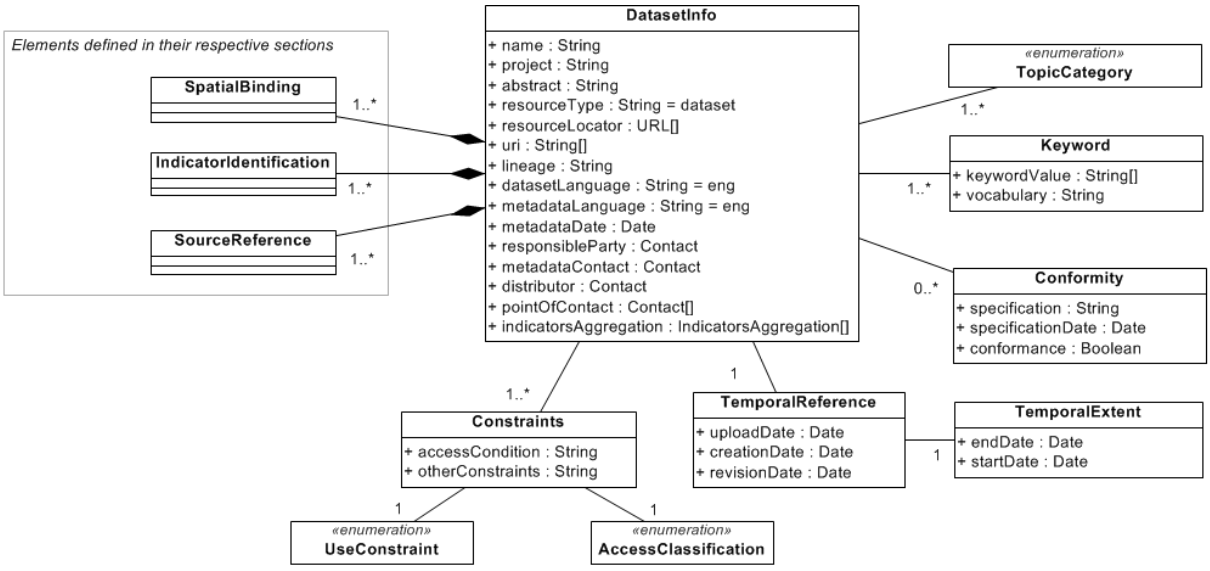


Figure 20 - Dataset information element structure

2. The **Spatial Binding** element describes the nomenclature(s) of statistical/spatial units used by the dataset and their geographical extent.

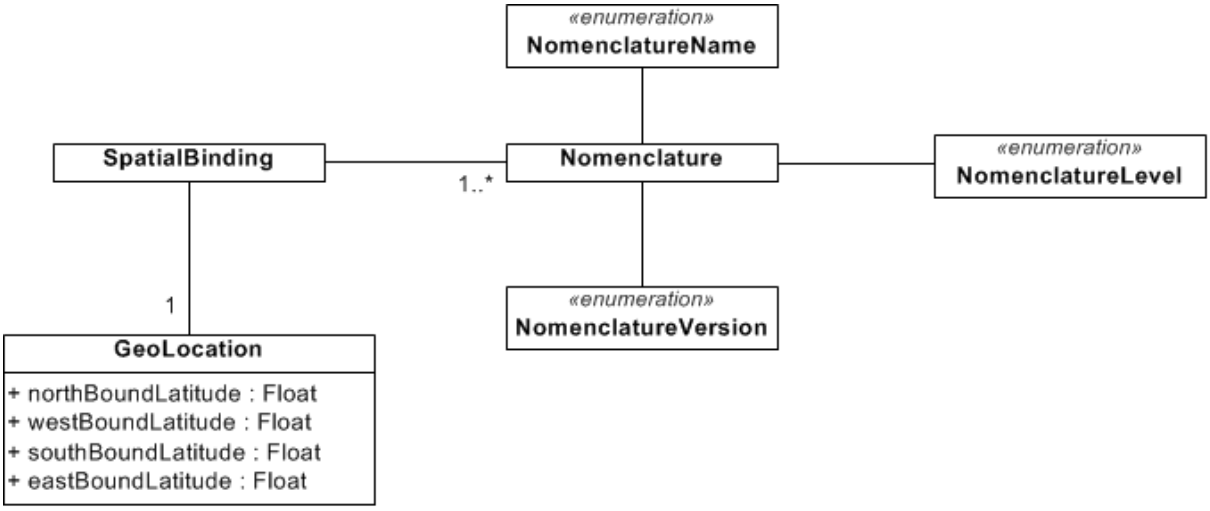


Figure 21 - Spatial Binding element structure

3. The **Indicator Identification** element provides metadata on the indicators existing in the dataset.

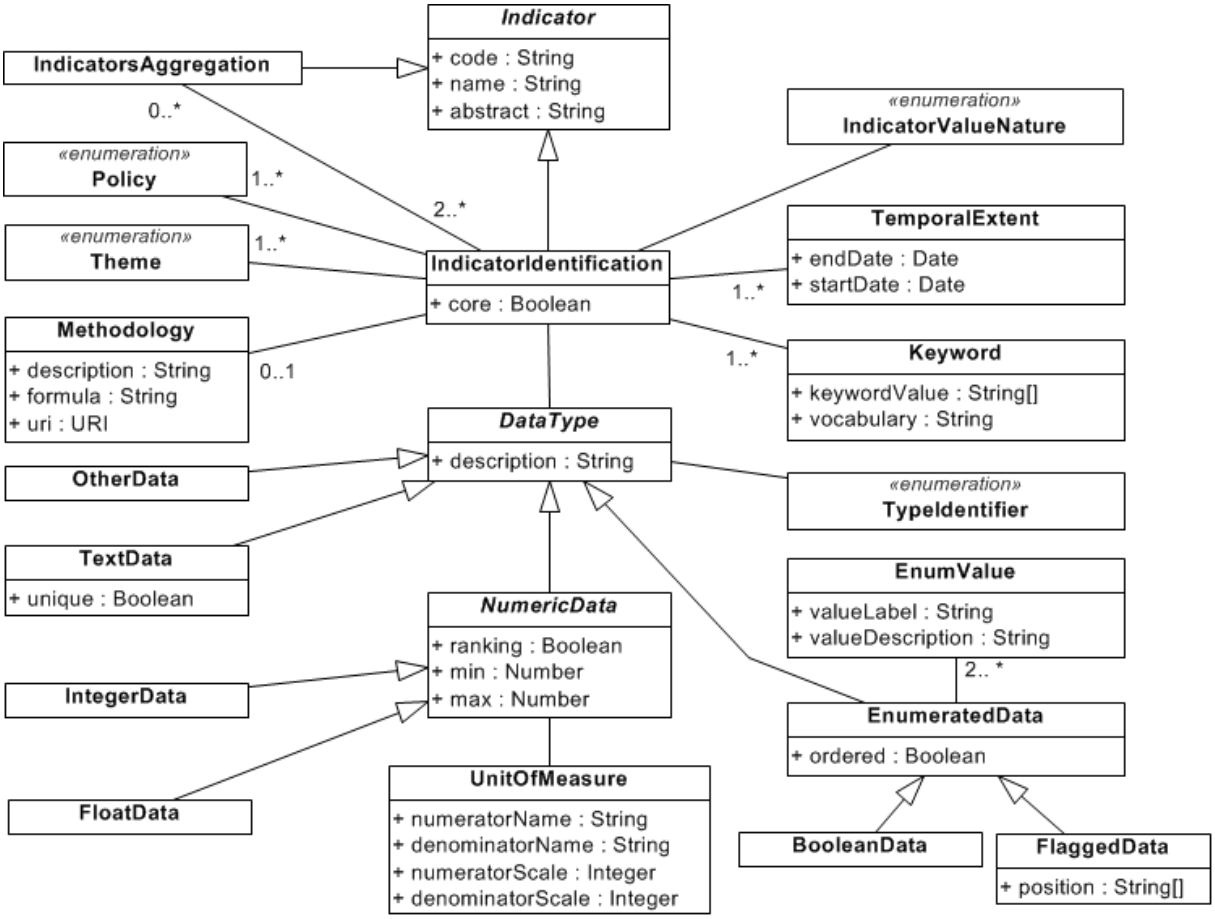


Figure 22 - Indicator Identification element structure

4. The **Source** Reference element gives details about the origin(s) of the indicator values in the dataset.

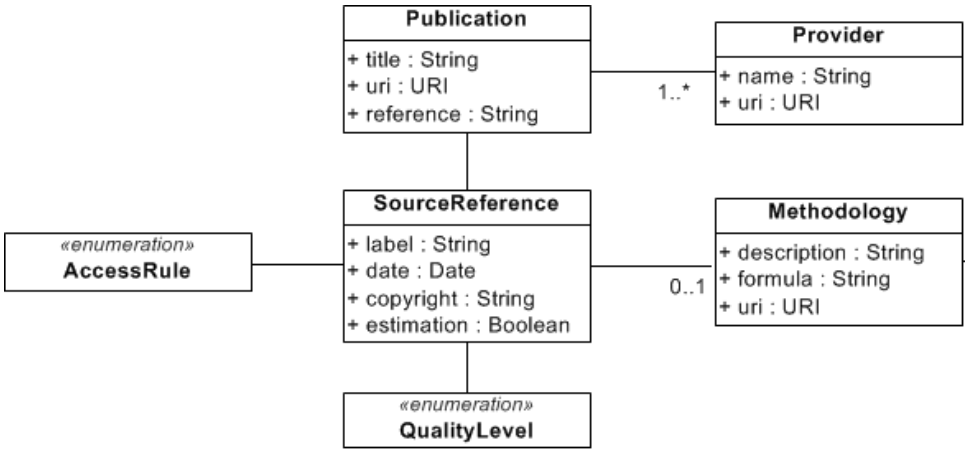


Figure 23 - Source Reference element structure

Whenever possible, the elements of the Metadata model follow the requirements of the INSPIRE Directive. However, some elements refer to ESPON-specific data structures that are not addressed by the INSPIRE Directive. The **ESPON Data and Metadata Specifications** document establishes the links between the elements of the ESPON Metadata model and the INSPIRE Directive. It describes all the elements of the metadata model and their properties. For further information, please consult [M4D, 2012 (1)].

2.1.2 ESPON Nomenclatures Support

The TPG deliver datasets containing indicators that rely on statistical units. Those statistical units belong to a consistent set, which must be clearly identified as a nomenclature. Thus, a nomenclature lists all the statistical units (name, code, geometry) of a territory at a given time.

During the ESPON 2013 Database Phase 2 project (aka M4D - Multi Dimensional Database Design and Development), we aimed at extending the support of different geographical objects. Besides the existing library of NUTS units of all public official versions (1999, 2003, 2006, 2010), the database now supports the libraries covering the nomenclatures UMZ (Urban Morphological Zones), MUA (Morphological Urban Area) and FUA (Functional Urban Areas).

Several datasets, already created during the first phase of ESPON Database project, use statistical units of these nomenclatures, but do not provide information about these units. Before the integration of such datasets, the database needs to know some details about the statistical units used, their hierarchy and relationship with other nomenclatures. This information, collected into the ESPON database, allows afterwards integrating any number of new datasets relying on supported statistical units libraries.

To integrate a library of statistical units into the database, the automatic module created for this purpose (SUNI.jar) reads the data from a nomenclature definition file. Once the nomenclature definition validated, the module inserts the nomenclature information into the ESPON database.

The **ESPON Nomenclatures Support** (available from the Help Menu of the portal [M4D 2012 (2)] is another indispensable reference document new to Phase 2 of the project. This document precisely describes how to define a new nomenclature based on spatial units. It describes the structure of the package and the layout to use in order to prepare a new nomenclature (or a new version of an existing nomenclature) for the integration into the ESPON Database. Indeed, several datasets, already created during the first phase of ESPON Database project, use statistical units of different nomenclatures (usually NUTS), but do not provide information about these units. The delivered TPGs' datasets must now rely on a supported nomenclature available in the database. The TPG can now focus on his/her dataset, but can still propose a new nomenclature definition for particular datasets needs. A nomenclature is integrated into the database when several Key Indicator datasets are likely to rely on it.

The schema shown in Figure 24 illustrates the main elements that compose the description of a nomenclature.

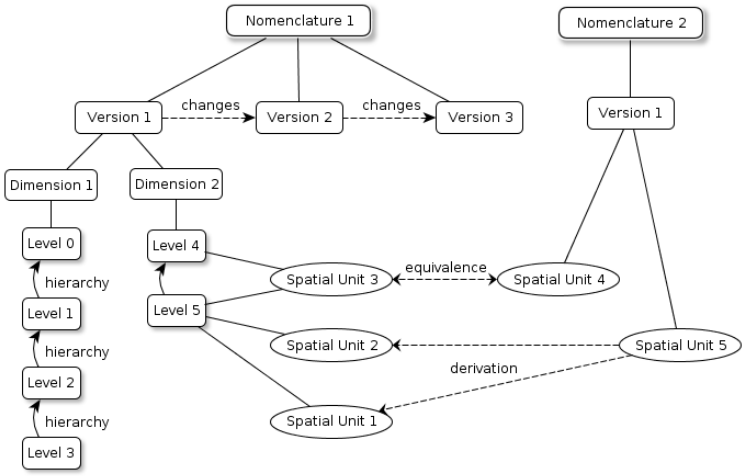


Figure 24 - Main elements that compose the description of a nomenclature

- (1) One version of nomenclature may be composed of several dimensions. Example: the UNEP classification distinguishes the "geographical" and the "political" dimensions [UNEP, 2012].
- (2) One dimension may be composed of several levels that define a hierarchy. Example: NUTS levels 0, 1, 2 and 3.
- (3) The description of the nomenclatures also aims to define relationships and bridges between units of different versions, equivalences (for example, the unit code LU in Nomenclature N1 is equivalent to the unit code LUX in Nomenclature N2), and derivations (for example, the spatial unit Ui defined in the Nomenclature N1 is an aggregation of the spatial units Uj and Uk in the Nomenclature N2).

One nomenclature may be declined in several versions (example: NUTS revision 1999, NUTS revision 2003, etc). The nomenclature description aims at tracking the changes between these versions, i.e., describing the evolution of statistical units. Based on work by [PLUMEJEAUD C., 2011] and adapted for the ESPON needs, the typology of statistical units evolution includes three main type of changes: (1) existential changes (new unit, end of unit), (2) non-territorial changes (code change, name change) and (3) territorial changes, the 6 main cases are illustrated in Figure 25.

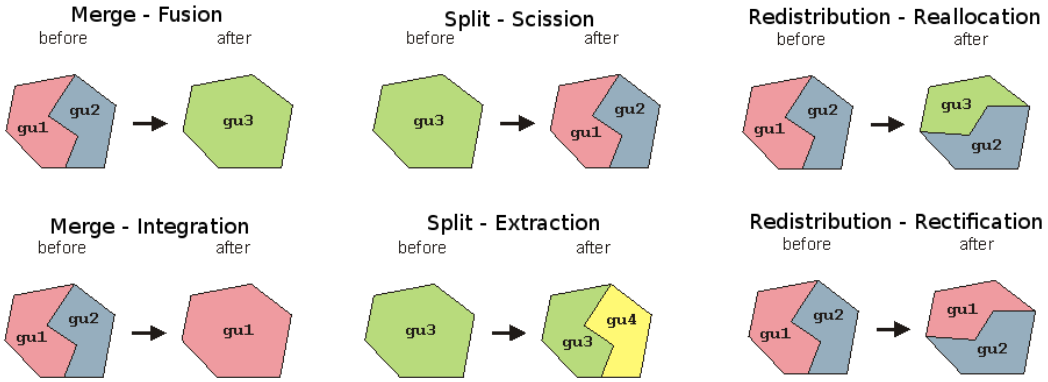


Figure 25 - Overview of the territorial changes

To summarise, the ESPON Nomenclature document mainly aims to respond to the need to extend the support of different geographical objects in the database, and then, to integrate Key Indicator datasets based on various geographical objects. Libraries covering NUTS2/3 and urban objects UMZ (Urban Morphological Zones), MUA (Morphological Urban Areas) and FUA (Functional Urban Areas) have been added to the existing library of ESPON NUTS units (all public official versions 1999, 2003, 2006, 2010) have been added in 2013-2014. Thanks to the nomenclature specifications, any new nomenclature respecting the defined constraints may therefore be available in the ESPON Database: LUZ (Larger Urban Zones), SNUTS (*Similar* to NUTS) or WUTS (World Unified Territorial System) may be built and integrated in the future. Hence, the Nomenclature Support specifies the expectations regarding the creation and the integration of a new nomenclature, which is a quite complicated task. But to help TPGs to create their Key Indicator Datasets, the M4D Team has completed the set of documentation with the document entitled **Dictionary of Spatial Units** (available online in the Help section of the ESPON Database Portal [M4D, 2013 (1)]). This document proposes a list of the available nomenclatures integrated (or to be integrated) into the ESPON Database, and guidelines on how to correctly reference the used nomenclature in the delivered datasets.

2.1.3 Database Tools and Software

Application Programming Interfaces (APIs) have been implemented to automatize and manage the creation and content of the ESPON Database. These tools aim at insuring the portability, reproducibility, modularity and full control over the content of the database. Figure 26 shows how they are involved in the ESPON Database workflow.

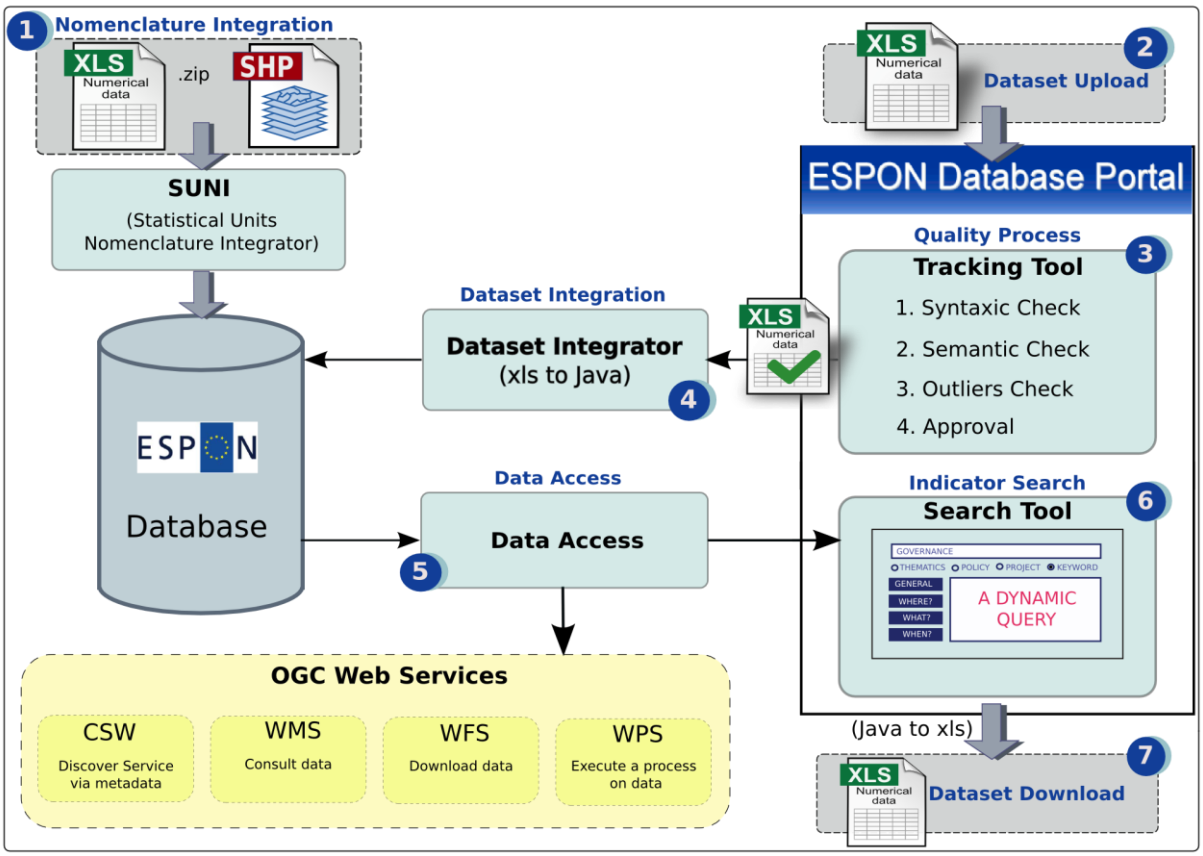


Figure 26 - Overview of the ESPON Database software infrastructure

The *SUNI* (Statistical Units Nomenclature Integrator) component integrates territorial nomenclatures in a relational database based on the data model described above (see box 1, Figure 26). The TPGs deliver their dataset thanks to the *ESPON Database Portal* upload section (see box 2). Then the *Tracking Tool* component enables the checking and improvement of the dataset quality (see box 3). Afterwards, the *Dataset Integrator* component integrates the dataset into the database (see box 4) by converting all its elements (XLS format) into operable Java objects. The *Data Access* component provides search capabilities to access data in database (see box 5). This component is used by the *ESPON Database Portal* web application to access and download data (see box 6-7). It can also be involved in the creation of OGC web services (Open Geospatial Consortium) like WFS (Web Feature Service) and WMS (Web Map Service) in order to access data in an interoperable way (see yellow box, Figure 26).

The workflow infrastructure is ensured by a systematic process to deliver the database. Throughout the M4D project, the deliveries of new versions of the database have been composed of:

- the “espondb2” script command, which executes a set of SQL (Structure Query Language) scripts to create the structure of the database, composed of more than 100 tables dispatched in 8 thematic schemas. This set of SQL scripts also inserts the initial constants and enumerations expected by the metadata and nomenclatures specifications;
- the set of nomenclatures packages (zip archives) to be integrated by the the **SUNI** module;
- the set of validated Key Indicator (xls files) and Case Study (xml files) datasets to be integrated by the **DatasetIntegrator** module;
- a document entitled **ESPON Database Install Instructions** describing how to use these tools, including also a full description of the database tables and overview diagrams of the structure;

- a document entitled **delivery** describing precisely the content of the current delivery and the changes occurred since the previous one.

This set of modules and documents enables the installation and population of the ESPON Database at any moment and on any standard workstation. M4D provides administrators with three high level scripts that are enough to reproduce the full system in three steps: (1) create the database, (2) insert nomenclatures and (3) insert datasets. Though each command is customizable for optional needs, the default behavior inserts all available nomenclatures and validated datasets. This easy process has been validated on multiple Operating Systems and environments (Linux Debian, Ubuntu, CentOS, Mac OS X), including the ESPON Production environment hosted on infeurope, and ESPON Tools teams and partners (RIMAP, ETMS, UNEP Grid Geneva).

2.2. The Web Application

Overall strategy

The ESPON M4D Multi-Dimensional Database Design and Development Project entails integrating, verifying and presenting the territorial data produced by ESPON Applied Research Projects.

The main challenges that the project faced were managing, standardising and coordinating a wealth of diverse data, ensuring data quality, traceability and creating search tools for territorial data.

To meet these challenges, implementing specialised methods and tools has provided solutions.

Key figures

- **136** Key Indicators Datasets
- **11** Case Study Datasets
- **807** indicators
- **12** themes
- **33** policy categories
- **130** Resources

It is clear that centralising the data management and processing helps a wider audience to access data. Therefore, within the M4D project, the Web application called "ESPON Database Portal" has been developed. The aim of this portal is to centralise the management of the ESPON data flow by seamlessly coordinating the different actors.

First of all, the portal allows projects to build their datasets so they are consistent with the expected format. For example, the documents entitled "How to deliver my data?" [M4D, 2012 (3)] and "ESPON 2013 Database Dictionary of Spatial Units" [M4D, 2013 (1)] provide the TPGs with useful information on how to create their datasets.

Then, from the portal, TPGs can upload their datasets via a specific tool named "tracking tool" that coordinates stakeholders at different stages of the integration process. This online tool enables stakeholders to comment, correct and ensure the quality of the dataset.

Finally, once the dataset integrated, the portal allows users to search and download datasets or indicators through adapted search interfaces. Depending on his or her profile (policy maker or scientist), the users can quickly find the data they are looking for.



Further readings - Related M4D deliveries and presentations

Presentations

- **ESPON Database** *(ESPON Seminar presentation, December 2011)*
- **A new interface and data model for the ESPON Database** *(ESPON Seminar presentation, June 2012)*
- **The Input Process of the ESPON M4D Database** *(ESPON Seminar presentation, June 2012)*
- **Monitoring and benchmarking the European territory : The M4D contribution: Time-series, Urban Data and Case Studies** *(ESPON Seminar presentation, December 2013)*
- **The Usefulness of the ESPON Scientific Platform for Policymakers. ESPON Database : Case Studies Interface** *(ESPON Seminar presentation, June 2014)*

Public documents

- **User manual**: available in the “Help” section of the ESPON Database Portal at http://database.espon.eu/db2/usermanual
- **Data quality check: methods and procedure**: available in the “Resource” section of the ESPON Database Portal at http://database.espon.eu/db2/resource?idCat=31

Tools and relative technical documents

- **Administrator manual**

2.2.1 Tracking and Checking Tools

One major challenge of the M4D Project has been to collect the heterogeneous data and indicators produced by the ESPON TPGs. This was done through a dedicated Web interface. TPGs choose to provide datasets in at least one of the three types established:

- **Key indicators** dataset: it covers the entire ESPON Study Area and references one of the available nomenclatures (NUTS, FUAs, MUAs, UMZ);
- **Case-study** dataset: it does not necessarily cover the ESPON Study Area, and/or its nomenclature is not available in the ESPON Database yet;
- **Background data**: any data produced by the project, whatever their format.

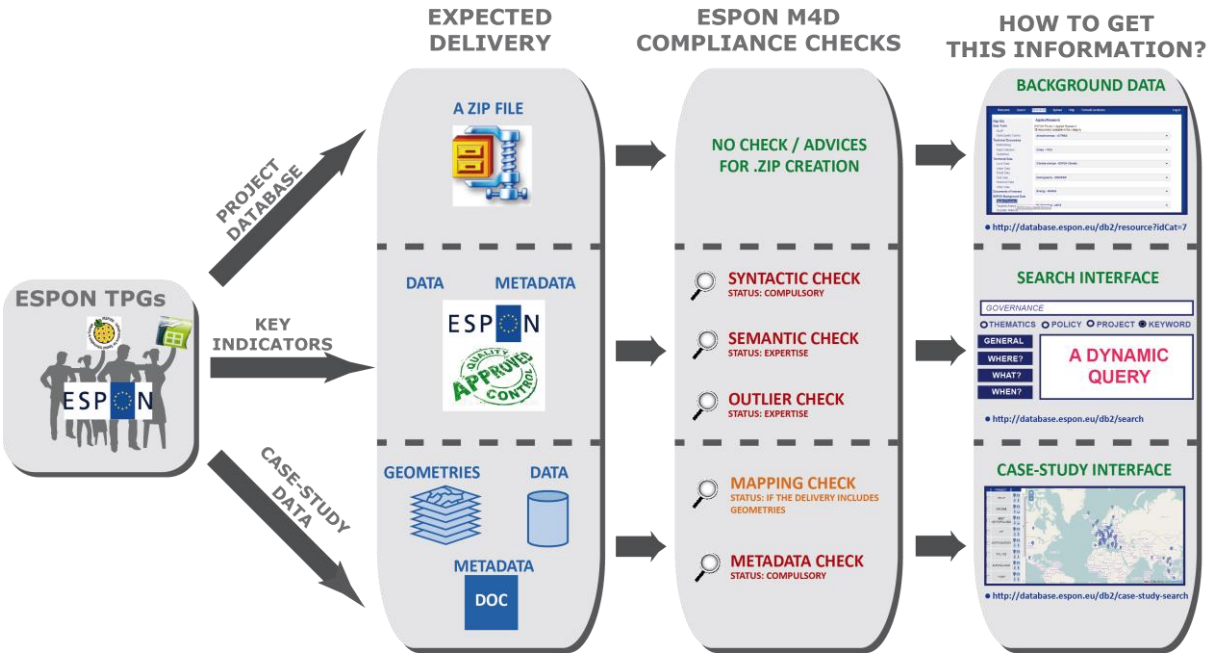


Figure 27 - Three possibilities for delivering the data

As shown in Figure 27, the dataflow depends on the nature of the delivery. To ensure compliance and quality of datasets integrated in the ESPON database, a monitoring tool has been developed under the M4D project: the tracking tool. This tool meets several objectives:

- monitor the datasets compliance with the expected format;
- ensure the consistency between the description of the indicators and data;
- involve stakeholders during the integration process: data providers (TPG), database administrator (M4D) and the ESPON CU;
- visualise in real time the integration process status.

Access to this tool in the dedicated menu of the ESPON Database Portal is restricted to registered users: TPGs, database administrator (M4D) and the ESPON CU. The tool manages different kinds of integration process, depending on the dataset type. The following sub sections respectively describe the dataflow for Key Indicators, Case Study and Background Data datasets. Depending on the type of dataset the TPG decides to deliver, it uploads its dataset on the appropriate section of the portal.

2.2.1.1 Key Indicators Dataset

1. After the upload of a Key Indicator dataset, the TPG is redirected to a dashboard page shown in Figure 28, which displays all the datasets being checked before integration into database.The Key Indicator dataset integration process consists of five steps: **Syntactic check**: the TPG uploads its dataset. An automatic procedure is triggered to ensure among other things, that all mandatory fields are filled.
2. **Semantic check**: after a notification email, one of the M4D Project Partners verifies the consistency of the delivered metadata, in particular the free text fields. This analysis results in a report that the TPG is invited to consult. Then, the TPG can decide to submit (or not) a new version of its dataset in order to improve its semantic quality.

3. **Outlier check**: a notification is sent to a statistician of the M4D NCG Partner to search for exceptional values. A report is generated and uploaded to the portal (Figure 29). Again, the TPG is invited to consult the report in order to decide the next step (abandon, resubmit or continue).
4. **ESPON CU Approval**: the ESPON CU consults all the reports and decides if the dataset can be integrated into the database.
5. **Integration**: the dataset is integrated into the database. Its indicators can be accessed through the search page of the ESPON Database Portal.

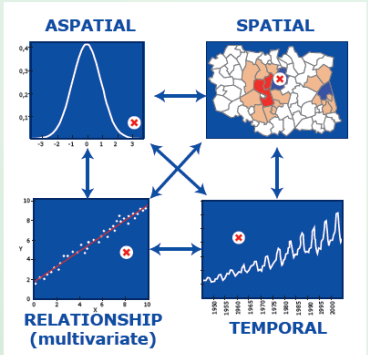


Figure 28 - ESPON Database Portal - The Key Indicators datasets tracking tool main/overview page

What happens to your data during the data flow process known as the "**data quality check**"? And why, after you've carefully put the data files together, do we need a data quality check?

As well as being part of the evidence you may have assembled for your project, data is quite interesting in its own regard. All data has context – you may have collected some of your own data, or you may have used someone else's data. And, if the latter, that data may have been originally collected for an entirely different purpose. Furthermore, **data collation, data analysis and data interpretation is a subjective process**. We are researchers ourselves, and we're very aware that when you live very close to your data, you can sometimes overlook a minor problem.

So, after we've checked the metadata for correctness and completeness, and that the data matches its metadata (syntactic and semantic checks), **we carry out some checks on the data values themselves**. Some of these checks can be simple. For example: an indicator whose metadata tells us it's a percentage should have values between 0 and 100; counts should be positive integers. If the metadata states that the values of a typology are 1, 2, 3 and 4, then there should be no other values. Are any data values unexpectedly missing?



We carry out other more complex checks: are any values for an indicator extraordinary high or low? We examine indicators singly and, where appropriate, in groups. As many projects use spatial units in the NUTS hierarchy, we check whether data values are unusual compared with their neighbours.

The ESPON metadata properties help us to proceed the appropriate tests: some indicators are ratios, some are counts, some are typologies, and some form time series. **Actually, checks depends on the statistical property of the indicator.**

When we find an unusual or extraordinary value, we flag it. If a NUTS region has many unusual values, then we'll seek assurance that these values are correct – often they are (major cities can be very different from surrounding settlements).

The output from the quality check is a report[2] which we share with you as a data supplier. When we all agree that the dataset is as correct as we can make it, it is finally loaded and made available into the ESPON Database.

[2] This report will be available under the Online Tracking Tool (cf previous section)

Figure 29 - The data quality check [M4D, 2012 (6)] and its related technical report [M4D, 2012 (4)]

2.2.1.2 Case Study Datasets

The ESPON M4D defines a “Case Study” as a dataset that does not necessary cover the entire ESPON Area (EU28+4). Various data can be delivered as a Case Study dataset:

- local data for a region or a group of regions (e.g. Greater Manchester at LAU2 level, Ile-de-France at employment basin level etc.)
- data described in a nomenclature not supported by the ESPON database (airports, water basins etc.)
- non ESPON Area and non ESPON Neighbourhood data (e.g. data on American, Brazilian or Japanese regions).

Following the same principles used with Key Indicators dataset type, the life cycle of a Case Study dataset consists of the following steps:

1. **Upload file(s)**: the TPG uploads its Case Study (including data and geometries) to the tracking tool;
2. **Metadata registration**: the M4D Tigris Partner edits metadata through the online editor (Figure 30);
3. **Dataset Check and Integration**: once completed and validated, the Case Study is publicly available and can be queried through a dedicated page of the portal.

Once a TPG has submitted a case study dataset in the upload section of the portal, the team in charge of the online metadata edition is automatically notified by email. The online metadata editor takes into account the specificities of the Case Study dataset and, to some extent, the fields expected by the INSPIRE directive. The case study metadata editor allows semantic experts to edit and register the datasets metadata in the ESPON Database. This semantic expertise aims at enriching and harmonizing the dataset metadata. Moreover, the metadata are arranged for the visualization tool. This metadata edition step makes metadata compliant with the expected metadata model and improves the quality of the case study dataset. The Case Study metadata editor is split into 6 successive parts: the dataset information (name, project, abstract, etc.), the constraints applied to the data and geometries, the dataset contacts (responsible contact, metadata contact, point of contact), the spatial binding of the dataset, the indicator identifications, the source references. The Figure 30 illustrates the first step.



Figure 30 - ESPON Database Portal - Case Study metadata editor, Step 1 (Dataset Information)

Once the metadata is validated and published, the case study is available from the dedicated search page shown in Figure 32.

2.2.1.3 Background Data Datasets

Following the same principles used for the tracking of Key Indicators and Case Study datasets, the integration process of a Background dataset consists of three steps:

(i) **Upload file**, the TPG uploads its Background dataset to the tracking tool (must be a zip file)
(ii) **Dataset Check**, after a notification email, one of the M4D Project Partners verifies the structure and consistency of the zip file content.
(iii) **Integration**, once validated, the Background data dataset can be access through the "Resources > ESPON Background Data" section of the ESPON Database Portal.

2.2.2 Search Interfaces

The portal allows users to search datasets through adapted search interfaces. They have been designed to provide the users with the convenient data representation for each type of dataset.

2.2.2.1 Key Indicators



The ESPON Database portal offers the possibility to search among hundreds of indicators produced by ESPON Projects.

The search query UI allows policy makers or scientists to quickly find indicators for various areas (EU28 + Switzerland, Liechtenstein, Iceland and Norway) at various geographical levels (NUTS or city levels).

The aim of this section is to explain the way to find indicators contained in the multidimensional ESPON database, considering both scientist and policy maker profiles.

1 Search by theme, policy, project or keyword

The first step consists in selecting a semantic search criterion. It may be a theme (population and living conditions, education, governance, territorial structure etc.), an EU Cohesion Policy Document (EU2020, Territorial Agenda...), an ESPON project or a keyword. The user choices are assisted by keyword autocompletions and dropdown lists.

2 Launch the query

Taking into account the search criteria, the query is performed by clicking on the "launch search" button ( ). All the chosen query parameters are displayed in the *filters toolbar*. At any moment, the user can choose to refine her/his query by adding or deleting some filters.

By clicking on the *clear* button, all the filter parameters are refreshed.

3 Filter the results

Advanced options allow the user to further filter the search results. The indicators of the ESPON Database are heterogeneous. They are available at various geographical levels, time period. In order to comply both with policy makers and scientists needs, various filter options are available:

- Filter *Data* to query indicators by their statistical properties (typology, count data etc.) or by the dataset type they belong: thematic table or time-series.
- Filter *Where* to select specific study areas (EU28, ESPON Area, a given country). Please note that the completeness calculation depends on the study area chosen. It is also possible to choose a specific territorial level (NUTS, UMZ, FUA or MUA).
- Filter *What* to add a semantic criteria. For instance, if a user has selected a thematic criterion, he may add a policy keyword or a project name criterion.
- Filter *When* to filter the results for a given year or time-period.

4 Search result output

The search results return all the indicators that match the search criteria, displayed in a table with:

- the indicator name together with the *dataset type of the indicator* (TT for thematic table, and TS for time-series);
- *the year of reference*;
- *the territorial level*;
- *the completeness*, calculated as regard to the chosen study area.

By default, results are alphabetically sorted by the indicator name. They may also be sorted by the year of reference (oldest or latest temporal extent available), by territorial level or by the value of the completeness.

5 Download options

Four download options are available:

- *Metadata* - Open the metadata page of the indicator. It gives information related to the data provider, the methodology behind the indicator, the data sources etc. This metadata page can be exported in a .pdf file.
- *Download the indicator* - Download the indicator data and metadata (.xls file). If the number of values of the indicator exceeds 3000 (time-series, contingency table etc.), the time of download might be long. In this case, the arrow is grayed ( ) and the user is invited to download the full dataset.
- *Download the full dataset* - Download the original dataset to which the indicator belongs to (.xls file).
- *Basket* - Put the selected indicators into a .zip file.

6 Additional options

The overview option - icon - displays key figures (number of indicators, number of themes etc.) and the alphabetical list of all the integrated indicators. This list can be exported to a .pdf file.

What is a thematic table and a time-series?

A thematic table TT is considered as a set of indicators which are semantically linked. It means that it is impossible to download them separately. In practical terms, it relates to two kind of figures: contingency tables (the sum of the columns makes sense, e.g. age structure, employment by branch etc.) or indicators linked together (e.g. a typology and its related raw data, indicators included in the 5th Cohesion Report etc.).

A time-series TS is considered as an indicator available at different temporal extents (5 years at least) and sharing the principles of the core database strategy, e.g.:

- Most of missing values have been estimated.
- Each time-series has been statistically checked (outlier check).
- It is possible to easily update the indicator in the future.
- The statistical types of time-series indicators is necessarily absolute stocks or absolute flows.

Figure 31 - ESPON Database Portal – Key indicator search User Interface

2.2.2.2 Case study Search

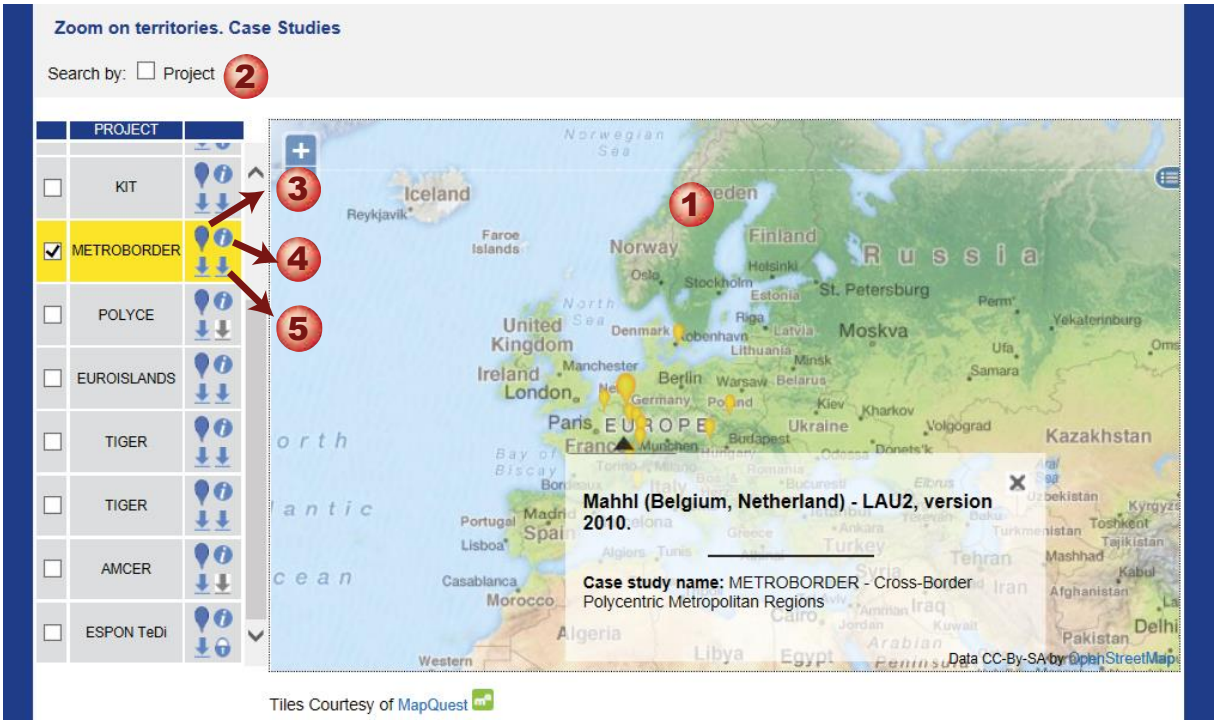


Figure 32 - ESPON Database Portal – Case Study search page

The Case Study search page allows to quickly identify the list of Case Studies recorded in the database, and their spatial binding in a world map. A typical scenario to search and download Case Study datasets is the following:

1 Search from the map

The world map offers an overview of Case Study locations. This map allows to discover data through their spatial binding, since the ESPON projects may provide data out of the European territory. The table and map are interactively binded: clicking a pin on the map highlights the matching Case Study row in the table, and vice-versa.

2 Filter by project

The list of Case Studies can be filtered by the name of the project that provided the dataset. Selecting a project name in the dropdown list causes a filter on the data table as on map pins: only case studies of the project and their study areas are displayed on the UI.

3 Visualise a Case Study spatial binding

Selecting a table row (i.e. one Case Study) shows on the map the study areas corresponding to the selected Case Study. Therefore, it is possible to visualize the spatial coverage of one case study.

4 Consult Case Study metadata

As shown in the figure on the right, the metadata page of the Case Study displays information related to the data provider, the methodology for the list of indicators, the data sources, etc. It can be obtained by clicking on the icon.

5 Download Case Study dataset

The dataset and its associated geometries (if available) can be downloaded from the search UI table by clicking on icon or from the metadata page.

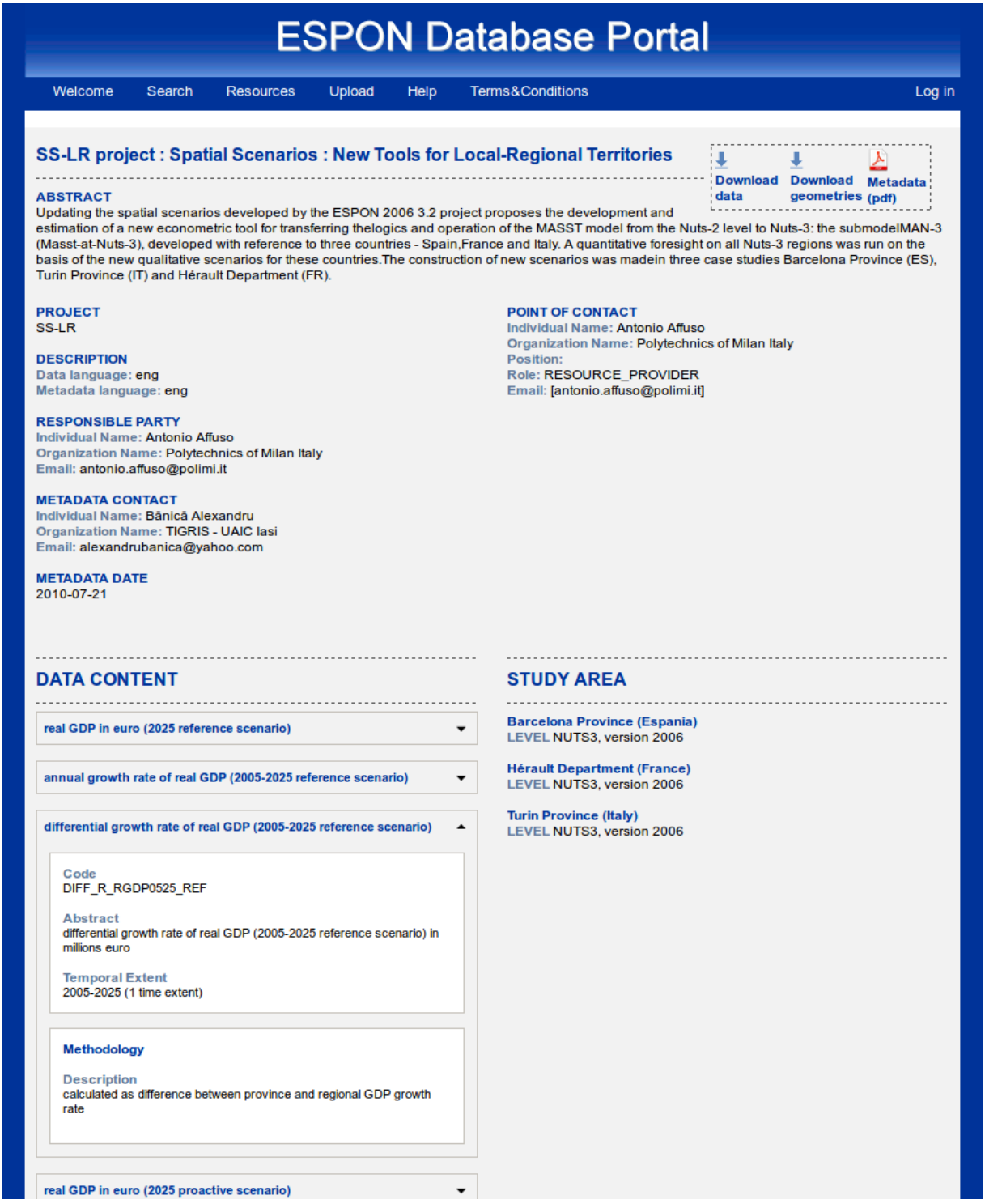


Figure 33 - ESPON Database Portal - Case Study Metadata page

2.2.2.3 Resources

The Resources part of the ESPON Database Portal aims to coherently structure and share all the useful information related to data availability, creation, transformation and representation in an European context. It is structured in 6 sub-categories (Figure 34).

- **The Map-Kits** section provides the necessary material to create maps for several study areas (ESPON, Euromed, World etc.). It includes georeferenced geometries and map templates, which follow the corporate identity of the ESPON Programme. These Map Kits are available in several formats (Q-GIS, ArcGIS, Philcarto) and their access is restricted to logged persons.
- **The Data tools** section includes programmes (in a R language) and applications (OLAP Cubes) useful for analysing or transforming territorial data.
- **The Technical documents** section contains a set of standalone booklets called technical reports. Technical Reports focus on methodological questions related to data and metadata that are regularly asked in ESPON projects and try to summarise collective knowledge.
- **The Territorial data** section is structured in several categories (local data, urban data, World data, grid data, historical data, other data), taking into account the different types of data that it is possible to download from the resource part of the ESPON Database Portal.
- **The Documents of interest** section includes reports and datasets of reference which can be especially useful for people interested by data production in an European context. At the moment, the documents available in this section have been provided by Eurostat and European DGs.
- **The Background data** section relates to data provided by ESPON Projects, whatever their format. In fact, in some cases, it is especially important to keep the coherence of a database produced within a scientific project since it includes generally coherent data, metadata, documentation and examples which cannot be restituted entirely by the ESPON Search interfaces.

The files available in this sub-part of the database have been mainly provided by the ESPON Database 1 and M4D projects partners. Ideally, the resource section of the Database Portal will be further expanded with information from other ESPON projects.

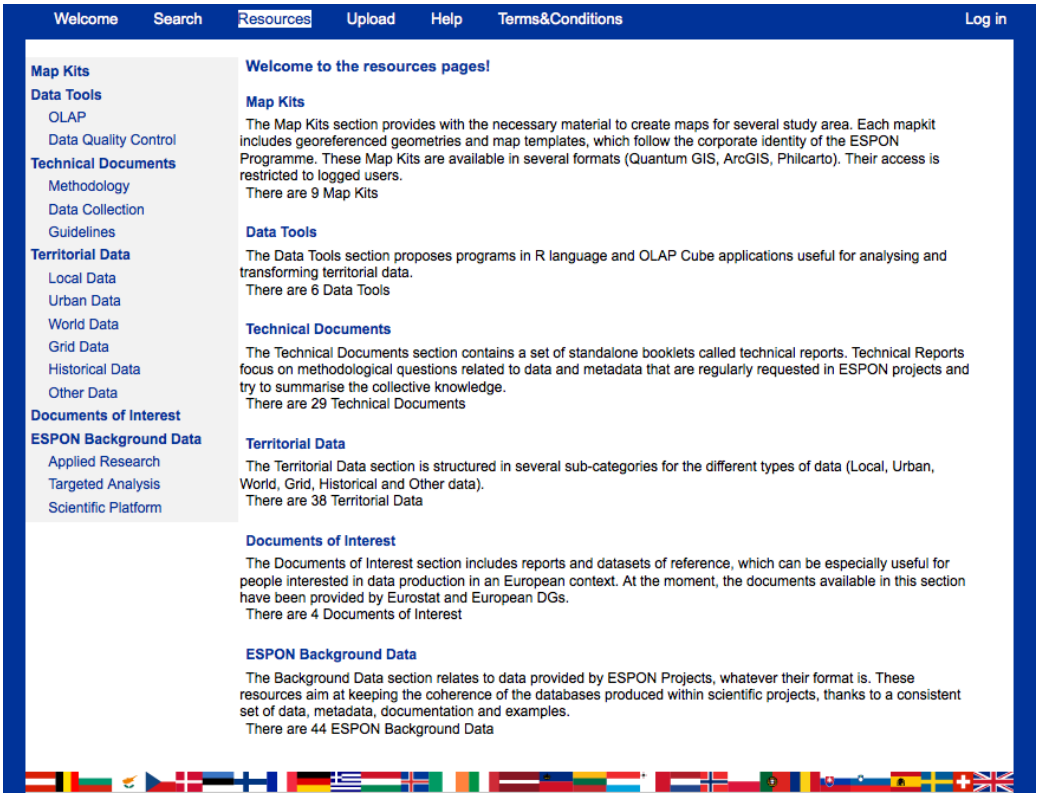


Figure 34 - Overview page of the available resources

2.3. A lively ESPON Database Portal : networking and updates

Overall strategy

In relation to networking and support of ESPON projects and ESPON CU, several activities have been addressed in order to ensure (1) an efficient data and information flow, (2) efficient updating of the database and of the application.

During the M4D Kick-Off meeting in May 2011, this task (1) had initially been considered as part of the "*Working Package C.1 - Platform*". However, the ESPON Database Portal was quickly seen (by December 2011) as a central online location to provide the ongoing ESPON Projects with guidance, recommendations and dataset submission. Therefore, to facilitate networking, "Platform" tasks were centralised by creating extra functionalities in the ESPON Database Portal: e.g. availability of useful documents and resources, "Tracking", and management of registered users.

Key figures

- **24** application deliveries and…
- **18** database deliveries from 2011 and 2014.
- More than **300** emails have been sent to **ESPON TPGs and the external world** and more than **400** to the **ESPON Coordination Unit** in the lifetime of the project by *manager@espondb.eu*

The following sections describe the main actions:

- Support the ESPON TPGs.
- Support the ESPON Tools.
- Deliver updated versions of the database and application to the ESPON website service provider (**InfEurope**).
- External networking.



Further readings - Related M4D deliveries and presentations

Presentations

- **The input process of the ESPON M4D Database** (*ESPON Seminar presentation, June 2012*)
- **How to deliver my data?** Preparation of a dataset – upload – tracking – integration, (*ESPON Seminar presentation, December 2012*)

M4D Involvement in ESPON Publications (data & map creation)

- **Second ESPON Synthesis Report**, Territorial insight: where to focus what type of investments (June 2013)
- **Third ESPON Synthesis Report** (October 2014)
- **Europe's neighbourhood from a territorial perspective,** ESPON Seminar Report (February 2013)
- **ESPON Atlas – Territorial dimensions of the Europe 2020 Strategy** (June 2013)

Technical Reports - documentation

- **How to deliver my data?** available in the "Upload" section of the ESPON Database Portal at http://database.espon.eu/db2/upload (registration needed)
- **Templates in a .xls format** to help TPGs to feed metadata, available in the "Upload" section of the ESPON Database at http://database.espon.eu/db2/upload (registration needed).
- **Frequent Asked Questions (FAQ),** available in the "Help" section of the ESPON Database Portal at http://database.espon.eu/db2/help

2.3.1 Support to ESPON TPGs

Several documents and resources are available from the ESPON Database Portal, providing the ESPON TPGs with guidelines and help in order to deliver their datasets:

- First of all, the document entitled "**How To Deliver My Data?**" [M4D, 2012 (3)] targets the TPGs and proposes an entry point as an overview of the expected roadmap for the delivery of a dataset, depending on its type.
- Additionally, for each dataset type, a template is available from the portal (see Figure 35).

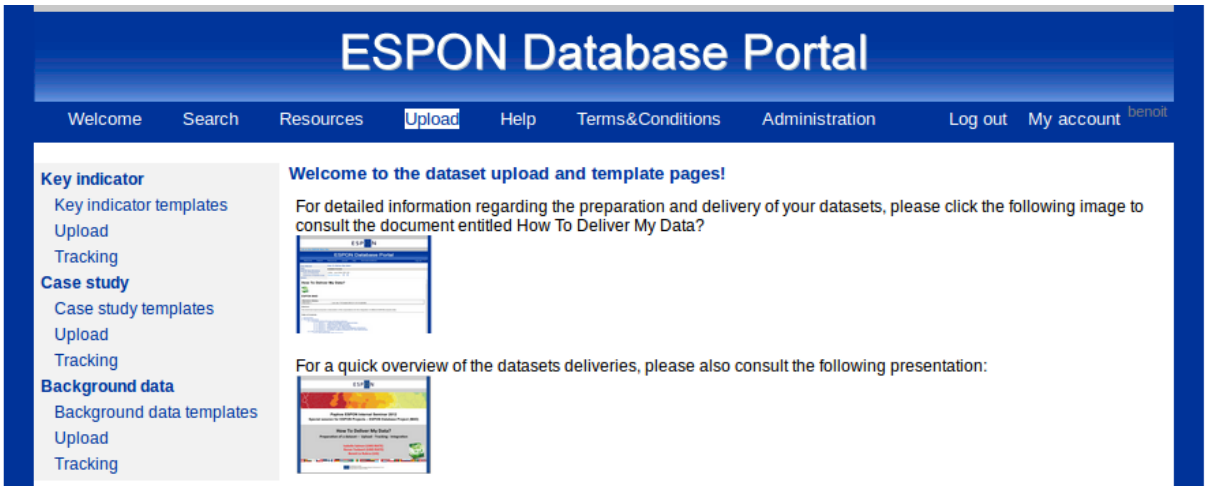


Figure 35 - Upload page: documents and templates

- For the delivery of Key Indicators Datasets, the TPG is invited to follow the expectations described in the following specifications and help documents (further described in Section II.1): **ESPON Data and Metadata Specifications**, **ESPON Nomenclatures Support**, **Dictionary of Spatial Units**, **FAQ** (*Frequently Asked Questions*).

Besides these documents and resources, any ESPON Project has been assigned to a dedicated M4D partner, who has been in charge of the "*semantic check*" during the Key Indicators dataset delivery process (see the "*Tracking Tool*" section in Section 2.2). The tracking tool has automated the management of a complex flow of information. At each stage of a dataset integration process (syntactic, semantic, outlier checks, ESPON CU approval), the tool assigns responsibilities to the relevant actors and allows a more efficient coordination. Moreover, all reports are centralised. Thus, the TPG only focuses on the quality and consistency of his/her dataset. Consequently, thanks to the tracking tool, the project management has been facilitated and the tool has provided to the ESPON Coordination Unit with a synthetic dashboard of the ongoing integration process.

2.3.2 Support to ESPON Tools Projects

In charge of the ESPON Database design, the M4D partners have had a particular role of support and coordination with other ESPON Tools projects, specifically to aid with access to the database. This coordination included email exchanges and regular delivery packages, which were composed, as described in Section 2.1, of the necessary tools, software and relevant versions of the documentation: mainly, the document "**ESPON Database Install Instructions**" which describes the structure of the database, the document "**delivery**", and summarises the main changes to be considered. This cooperation with ESPON Tools projects has meant that the LIG STeamer M4D partner has supported ESPON Online Mapping Tool (RIMAP) and ESPON Territorial Monitoring System (ETMS) projects.

Since almost all new versions of the portal application come with an updated version of the database, RIMAP and ETMS have received such "database delivery" updates at the same time as the delivery of the online official version to the ESPON Website service provider, infeurope. This "official" delivery topic is further described in the following section.

2.3.3 Application and Database Delivery

Besides the delivery of the database to ESPON Tools for the development purposes, described above, the LIG STeamer M4D partner has been in charge of a regular and continuous update of the official online version of the ESPON Database Portal Web application. For this activity, the LIG STeamer M4D partner has supported the ESPON Website service provider infeurope, providing them with the necessary material (software and documentations) for the setup of the environment.

Usually, the delivery of a new version of the portal came with an updated version of the database structure. Every delivery of the database and of the application has consisted in automatically creating a distribution folder. Externalising a convenient installation procedure to the service provider infeurope led M4D to set up an usual practice regarding the content of the distribution folders, which are composed of the following list of sub-directories and files:

- Directory **delivery**: a document (in HTML and PDF formats) whose objectives are threefold:
  - describing precisely the full content (files and directories) of the delivered distribution folder (e.g. a user-friendly readable manifest);
  - describing the purpose and the implemented functionalities in this new version;
  - proposing a convenient summary of all main changes and a history of the previous deliveries (the table of contents of this section is shown in figure 36).
- Directory **ESPON_Database_Portal_Install_Instructions**: a document (in HTML and PDF formats) describing how to install the database and the application. Compared to the **ESPON_Database_Install_Instructions** document, this one provides the infeurope administrators with additional information regarding the setup of the architecture, the deployment of the application and the configuration of the connection between the application and the database.
- Directory **database**: SQL scripts and API software tools to build and populate the database from scratch (see Section II.1 for further details about these tools).
- Directory **nomenclatures**: list of nomenclatures archives files to be integrated into the database.
- Directory **validated**: list of original Key Indicator and Case Study validated datasets files to be integrated into the database.
- Directory **webapp_resources**: list of new and modified resources (changes made since the previous delivery, available for download from the "Resources" section of the portal);
- File **e2.war**: the ESPON Database Portal Web Application archive file ready to be deployed in the infeurope application server.

For each delivery, a distribution folder has been uploaded to the File Transfer Protocol (FTP) server hosted at infeurope. Then, a notification email is sent to ESPON CU, with the **delivery** document attached. At a glance, the Figure 36 provides the list of the 29 deliveries (in January 2015) that have been performed since the beginning of the M4D project.

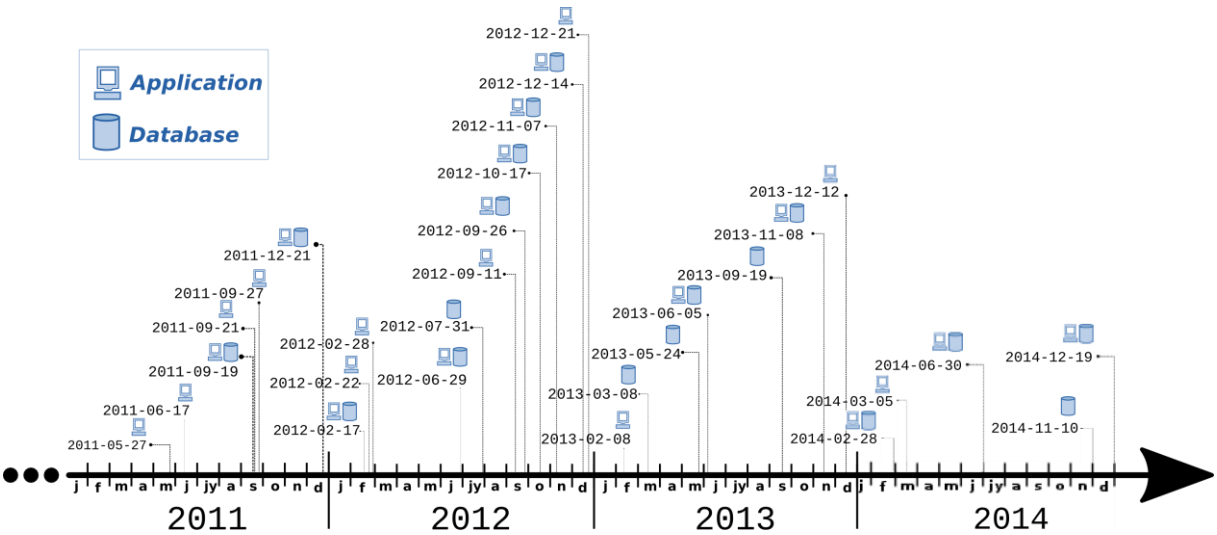


Figure 36 - Timeline of the database and/or application deliveries until June 2014

2.3.4 External networking, ESPON Coordination Unit support and dissemination

Important networking activities were undertaken during the lifetime of the project. In summary this was predominantly external networking with other European institutions; support of ESPON Coordination Unit for map and statistical creation; and the dissemination of M4D findings in several research conferences.

Regarding external networking, the M4D consortium attended a meeting with GISCO in March 2012 in order to disseminate information about the ESPON OLAP Cube and the adaptation of the INSPIRE directive to socio-economic territorial data. Also, the M4D teams attended a meeting with Eurostat in June 2013 to present the main discoveries of the M4D Project.

The M4D project participated in major ESPON Publications: ESPON Synthesis reports 2 and 3, ESPON Atlas and the Paphos Seminar Report. The project has also created posters for ESPON events and participated in the data collection / estimation for some of the ESPON Maps of the Month.

Last but not least, major findings of the project have been disseminated in academic colloquium and scientific papers. There is a list below of the relevant papers (in French and English). This highlights the fact that the M4D findings also link GIS analysis and computer sciences within the scientific community.

Bernard C., Le Rubrus B., Villanova-Oliver M., Gensel J., Ysebaert R., Salmon I., 2014, STedi : une infrastructure logicielle pour renforcer la qualité des données territoriales statistiques, Conference SAGEO 2014 (Spatial Analysis and GEOmatics), Grenoble, France.

Charlton M., Brunsdon C., Cahalane C., Pforte L., 2014, Estimating missing data in hierarchical space-time series with a short temporal extent, colloquium GIS Research UK, 15-16 April 2014, Glasgow.

Harris P., Brunsdon C., Charlton M., 2012, Multivariate spatial outlier detection: a comparison of techniques. geoENV 2012, Valencia, Spain.

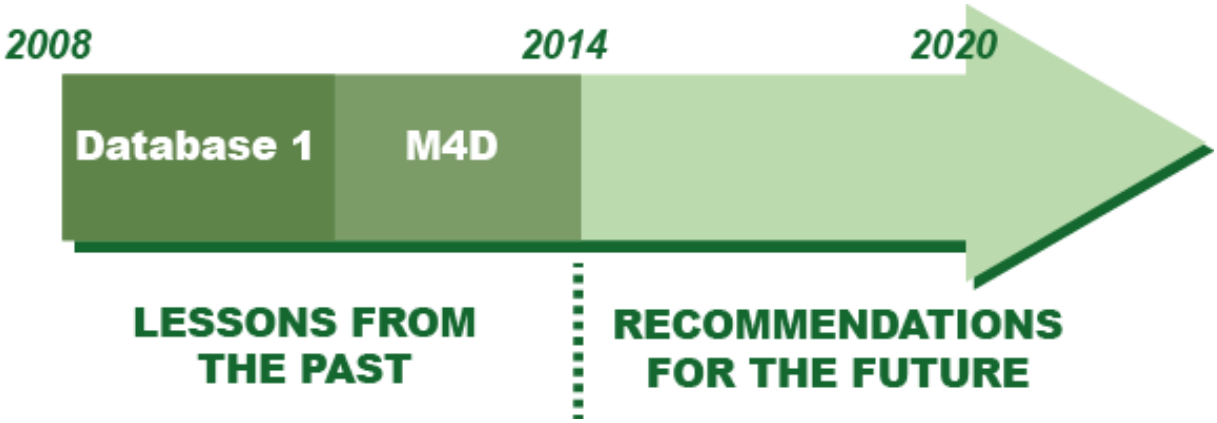
Harris P., Brunsdon C., Charlton M., 2011, Multivariate spatial outlier detection using geographically weighted principal components analysis. 7th International Symposium on Spatial Data Quality, Coimbra, Portugal.

Harris P., Brunsdon C., Charlton M., Juggins S., Clarke A., 2014, Multivariate spatial outlier detection using robust geographically weighted methods. Mathematical Geosciences 46(1) 1-31.

Harris P., Clarke A., Juggins S., Brunsdon C., Charlton M., 2014, Enhancements to a geographically weighted principal components analysis in the context of an application to an environmental data set. Accepted for publication in Geographical Analysis.

Ysebaert R., Salmon I., Le Rubrus B., Bernard C., 2014, Recueil, traçabilité et restitution des données territoriales du programme ESPON, seminar « Fronts et frontières des sciences du territoire », 27-28 March 2014, Paris, France.

Part 3 – The M4D Experience



A self-evaluation of the collective results of M4D project has been launched by the coordinator (UP7D UMS RIATE) in partnership with all associated partner teams (UJF LIG, UAB, NCG, UIC Tigris, CNRS Géographie-cités) and discussed during the last TPG meeting held in Greece, in June 2014. The initial proposal was to realize a self-evaluation on a SWOT basis, eventually we did prefer to use a simpler way to present the lessons from the past and proposals for the future.

3.1. Lessons from the past…

Strengths

- **The ESPON database is extensible and evolutionary within 4 dimensions (Space, Time, Thematic and Scales):** the acronym of the M4D project (**M**ulti **D**imensional **D**atabase **D**esign and **D**evelopment) has a clearly defined main aim of enlarging data collection within the second phase of the ESPON programme (2008-2013). This main aim has been successful thanks to all the M4D partners who collect several type of data:
  - **Space**: The M4D Project has provided the methodological background to extend the data collection of territorial indicators to the European Neighbourhood. The ITAN Project has successfully taken this material and has extended the data collection to other socio-economic indicators for this study area.
  - **Time**: Time-series have been gathered for a set of basic indicators at regional level (Core Database Strategy).
  - **Thematic**: It is now possible to cross environmental (land use) and socio-economic data thanks to the OLAP Cubes implemented at several scales (new geographical objects have been explored (urban data and European neighbourhood).
  - **Scales:** An innovative urban nomenclature (UMZ) has been created. The implementation of a dictionary of correspondence between LAU2 and UMZ allows to characterise for than 4 000 urban morphological objects by socio-economic indicators. On top of that, the VIGO framework has provided innovative proposals related to the accessibility to point of interest, such as airports, roads etc. at a local scale.

- **An operational data flow including a compulsory upload of the associated metadata to ensure a high quality of data**. The ESPON TPGs data was processed from the very beginning (upload) to the very end (query and download). The quality of data was also improved thanks to the checking/tracking tool. We are convinced that quality is more important than quantity in terms of data collection or data delivery. We have therefore (1) constrained the ESPON TPGs to provide a complete and specified description for all the figures and indicators to be stored into the ESPON database and (2) obliged the ESPON database users to download simultaneously data with the associated metadata. The ESPON metadata model is not only fully INSPIRE compliant, but also proposes an extension of the INSPIRE framework to include social, demographic and economic data.

- **A wide collection of technical reports, newsletters and support documents have been created/collected to help users understand the concepts of the ESPON Database Portal.** Through each query interface, and more specifically, through the Help Menu, the ESPON Database Portal provides access to documentation such as User Manual, FAQ, Dictionary of Spatial Units, Data and Metadata Formats, Nomenclatures Support… These documents are of great importance as they will enable users to take advantage of all functionalities of the ESPON Database Portal. Moreover, the technical reports address methodological issues of high interest for European spatial planning themes such as definition of cities, estimation of missing values in time-series, outlier check, creation of local indicators of accessibility, functional transformation of classical indicators by linking this information with accessibility, etc.

- **Innovative proposals to extend applied Research on European Territorial Cohesion Policy at new scales in terms of space and time.** Our response to the request of policy makers and researchers to have “more data” was not to accumulate more and more heterogeneous data at regional level (NUTS2, NUTS3) for a short period of time (2006-present), but rather to focus on how to easily extend the data collection of a limited number of core indicators in terms of space, time and scales. (1) Extension of time series in the core database: apart from being able to store the lineage of any territorial unit, care and work have been brought to the expansion of long-term series of core indicators which makes the ESPON Database very useful and trusted resource. (2) Exchange of data between regional and urban delineation through LAU2 or GRID: through the OLAP cube on the one hand, and the analysis of urban nomenclature on the other, we have suggested practical and promising solutions.

- **A multidisciplinary consortium sharing its various skills and expertise to achieve a complex and a project that works.** The implementation of the ESPON Database Project required knowledge and expertise from various different fields: computer science, geography, statistics, GIS, management, etc. For example, the combination of computer scientists to create the complex ESPON Database infrastructure and subject experts to provide smart inputs to the multidimensional ESPON Database (core indicators, grid indicators, innovative urban nomenclatures). Or even a close cooperation between partners to propose a "user-friendly" web interface. The M4D TPG has always done its utmost to take into account the ongoing changes through the lifetime of the project. Eventually, thanks to the diversity and the complementarity of the teams involved in the M4D project, the multidimensional tool is fully operational.



Weaknesses

- **The opportunity to extend the dimensions of data collection was not further developed** Data delivered by the ESPON TPGs are usually at a regional level: NUTS2 and – less frequently – NUTS3. Alternatives to the heterogeneous NUTS delineation could have been explored in more depth within the ESPON Programme (cf part 1.4.1 and the technical report on FUNC). This framework also follows the research study launched by the European Commission (Study on comparable Labour Market Areas). Concerning the urban data, it was nearly the same: the M4D project had launched an ambitious approach of various functional and morphological delineation of cities comparison and creation, but too much pressure (at the political level) to reduce the number of urban objects stored in the ESPON database and to focus on the one considered to be the most important according to the (changing) political agenda.

- **Some tasks related to the ESPON Database Portal could have been explored/automated more.** Among other tasks that could have been explored more, we had to deal with ongoing specifications and then to be very flexible and less ambitious. For example, (1) the .xls dataset template file is not easy to use. A solution could have been to create an online metadata editor to ease the production of the ESPON metadata. The use of autocomplete fields, for example, could ease metadata production by providing the project with examples of metadata expected. (2) The workflow could have been automated further, in particular the outlier checks. Moreover, we could have further explored the opportunity to reduce the number of manual tasks. (3) The focus on the Excel file has shown its limits: exported files are not very convenient for the needs of users like GIS experts, statisticians etc.

- **Misunderstanding the objectives of the ESPON Database Portal.** The main difficulty was not to answer to scientists and policymakers' needs to access data. Very few policy makers and very few scientists are directly interested in accessing the data stored into the ESPON Database. The majority of users (scientists or policy makers) are firstly interested in accessing the data through graphics, diagrams, maps, time plot, etc. If data is downloaded in a tabular format, it is usually after having seen an interesting graphic or map but not before. In other words, the real effort should have been put on the connections between ESPON M4D and other priority 3 projects delivering such tools, not on the internal interface of the ESPON database project, which should be used only by experts or specialist researchers. The Query Interface that we have created is a very powerful tool to extract data. As a consequence, it requires practice for non-expert users. But thanks to the other ESPON tools, users can access the data through maps and graphics.

- **Management of both the archive data and the core data.** A final difficulty encountered by the ESPON M4D project was related to the dual objectives of (1) storing data gathered by the ESPON TPGs and (2) updating or extending core data of major interest. Both objectives were of important to ESPON, undoubtedly, but it was certainly a mistake trying to integrate them to the same database with the same interface for data research. (1) Archive data is the data produced by ESPON TPGs displaying their scientific results. When reading such reports, access to data AND metadata can help to understand or verify a map or a graphic or understand the methodology used, for instance to reproduce the research done by ESPON TPGs. The M4D project has done this by proposing two specific Graphical User Interfaces to access to data produced by projects for the ESPON area and Candidate Countries but also for case studies (zoom-in or zoom-out). As described in the next page, (2) Core data is considered as a crucial building block to help researchers in ESPON and should be established in long-term time series

and at different scales. It can be combined with accessibility matrix, transformed into grid data, re-aggregated in new geometries, etc. The majority of the ESPON tools need such core data as first inputs as well as the TPGs related to scenarios and foresights. The ESPON M4D project has made the best efforts to ease the availability of this core data. Nevertheless, we consider that mixing Archive data and Core data has introduced a lot of confusion for the end users. This explains why the ESPON M4D was late in the delivery of the core data, because the majority of time was allocated to store and check archive data.

3.2. Recommendations for the future

Opportunities

- **To carefully study the advantages/disadvantages to store archive and core data in the same database infrastructure**. On one hand, it seems interesting to provide a unique access to all data produced/gathered/updated by ESPON and to maintain the current situation of ESPON database and portal in the future. On the other hand, disadvantages of the current situation could be solved by the division of ESPON 2013 database into two separate parts with more specialized targets. Whatever the choice, such a decision has to be made before the beginning of the ESPON 2020 activity period. We have described below the possible opportunities that dividing the ESPON database into two parts might create.

- **Database 1: The "treasury" of the ESPON projects**: this database 1 could use the current interface with minor changes. Indeed, its main objective would be to insure the storage of best indicators produced by ESPON projects in the past and future with high quality control. The tracking tool will obviously be kept in order to ensure high level of quality but the search interface would be focused on research "by project" rather than "by policy options" or by "subject/theme". The idea would be to link this database with the final reports of the project and with integrated tools like "Online Map finder" or "ESPON Atlas". We believe it would be appreciated by people interested by a report to be able to visualize quickly the map associated to this report and to download related data. The search interface could combine data covering all ESPON space as well as case studies. In relation to the case studies, it could also be interesting to visualize on the map the place of case studies not associated to data but to a part of a written report.

- **Database 2: A multidimensional dynamic core database:** the content of this second database should be only composed of a limited number of existing indicators available in long-term time series (population by age and sex, active by branch, GDP, land cover, etc.). The data involved in this database would be "neutral" in terms of project affiliation because they should be a common basis, a reference to be used by all other projects. They should not necessarily be "official" data (i.e. figure exactly conformed to Eurostat or National Statistical Institutes) but they could also be homogeneous time-series harmonised by ESPON. In other words, some of this data would be "ESPON stamped" with guarantee of quality and perfect documentation concerning the methods of estimation and harmonisation. The difference with a classical database is related to the fact that the majority of indicators offered by this database will not be stored but likely to be computed dynamically according to four circles (Figure 37):
  - ✓ **First circle (1) – Count data**: Download and display a limited number of time-series in a user-friendly way, from the NUTS0 to the NUTS2/NUTS3 levels and hopefully other geographical objects.
  - ✓ **Second circle (2) – Indicators for the EU Cohesion Policy**: Propose classical "simple" indicators of the EU Cohesion Policy (the calculations should be made automatically). Generally, these indicators are obtained by dividing two count data together (GDP per inhabitant). In other terms, when a time-series is updated, all the derived indicators are recalculated automatically.
  - ✓ **Third circle (3) – Dynamic indicators**: Use the ESPON knowledge (production or methods) to propose dynamics/evolution indicators for territorial cohesion using time-series data properties: evolution indicators, convergence indexes, updated ESPON Typologies (when adapted). Here again, it should be calculated automatically.
  - ✓ **Fourth Circle (4) – Spatial indicators**: Combine previous indicators with spatial information like geographical distance, time distance, location in the same state, distance to airport etc. and derive automatically new indexes from this combination. A simple example of that is the three deviations produced by the HyperAtlas (global, national, local). But a lot of other indicators belong to this type like the family of indicators based on functional distance (potential of population or GDP available in 1 hour by road, minimum distance necessary to

reach 10 million inhabitants), etc. The Functional Indicator Tool has been implemented following this approach[27].

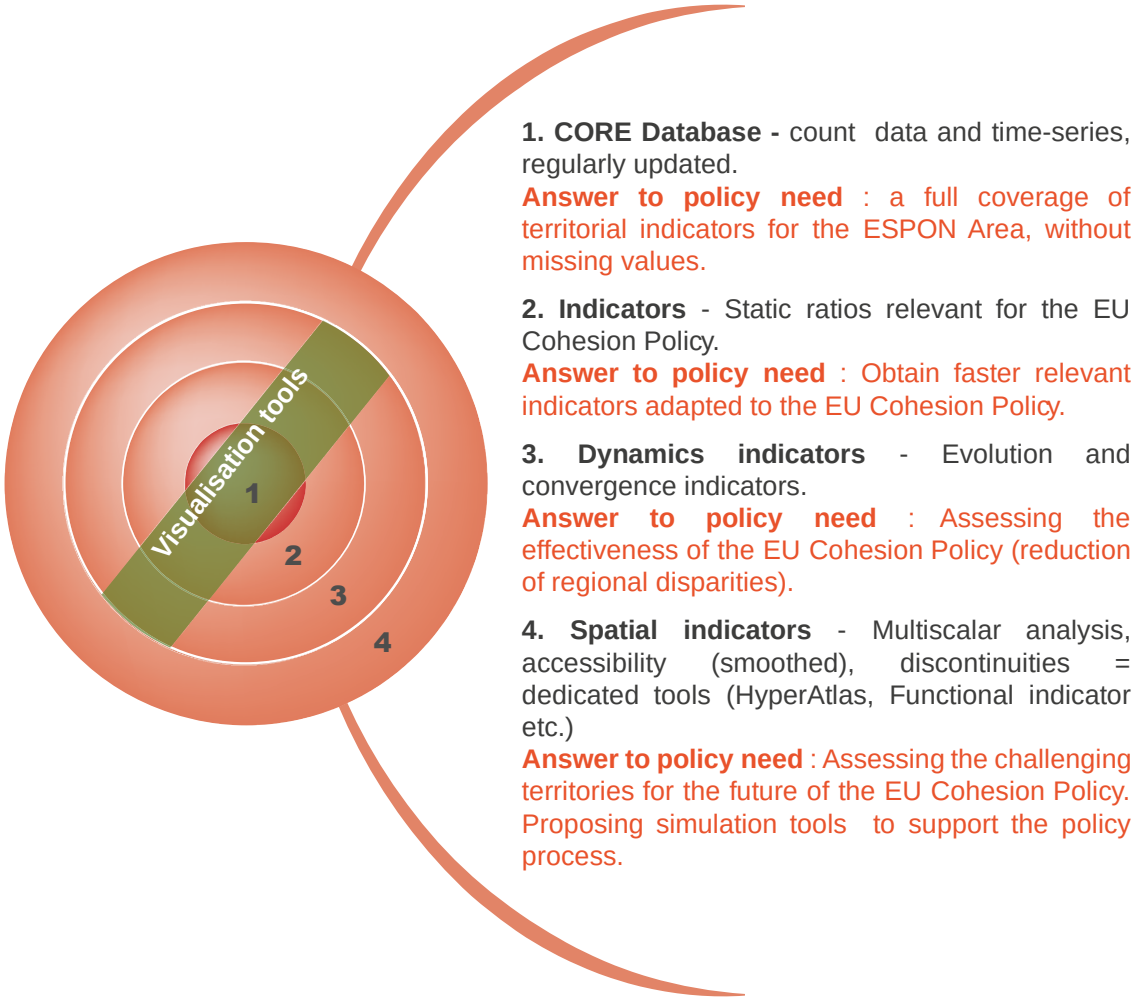


Figure 37 - A four circles strategy for promoting ESPON Core indicators?

As displayed on Figure 37, we consider that each "circle" of this strategy could be able to answer to specific policy needs related to data collection/production in the broad context of the EU Cohesion Policy: geographical coverage of the statistical information (1), need for a territorial information regularly updated (2), territorial assessment of the EU Cohesion Policy through convergence analysis (3) and dedicated tools highlighting "challenging territories" (4).

- **Connections between ESPON Core Data and web application**. To go further, the first technical steps consist in proposing the metadata available in the Catalogue Service for the Web (CSW) according to the ISO 19115 standard; extending the list of metadata to ESPON specific fields (Policy codes, ESPON Projects, Units of measure, for example); automatically enriching the catalogue with new Web Feature Service (WFS) layers for each triplet [indicator / nomenclature level / temporal extent] found while integrating a Key Indicators dataset into the database.

  Once this change made, two major opportunities would be possible**. Both the OGC Web Services (1) and Open Data initiative (2) aim at improving the visibility of the ESPON data and the interoperability of the ESPON Database through standards**. The OGC web services focus on the communication with other geographic systems. The Semantic Web technologies link data over the web, allow referring and accessing ESPON data on the World Wide Web.

  Those two opportunities are complementary: (1) The first one concerns the set of OGC Web Services that would make it possible to index, access, map and process any data stored in the

[27] http://fit.espon.eu/

ESPON Database. A proof of concept has shown the potential of such a technological approach. Interoperability will be increased. This means that not only human would be able to access and exploit the ESPON Database's content, but also computers, that is to say, programme and also other Spatial Data Infrastructure around the world, contributing a lot to the dissemination of ESPON's works.

(2) the second one takes place in the Open Data initiative. It could be possible to transform some or all the content of the ESPON Database so that data become accessible in the shape of RDF graphs, somewhere in what is called the Linked Open Data (LOD) Cloud that forms now an ever-growing and huge amount of data, a concrete part of the Semantic Web. Again, interoperability and dissemination are targeted. By making ESPON data be part of the LOD Cloud, by simply Linked Open ESPON Data to other LOD sets, they will be accessible and ready for use using Semantic Web technologies, which will allow referring and accessing ESPON data through URIs on the World Wide Web. According to the W3C, "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries".

- **The ESPON Programme as a leader in the exploration of new frontiers for data collection.** The three previous opportunities are related to innovation that could be limited to classical geographical objects like regional data or urban audit objects. But we are convinced that ESPON could obtain a comparative advantage by the exploration of new frontiers of data collection. From this point of view, the ESPON Programme has currently an advance in the interoperability, thanks to the different OLAP Cube creation and the technical reports on urban database building. A lot of opportunity will appear in the next period 2014-2020, thanks to the development of coherent and complete collection of data at LAU2 level. Thanks to the existence of a network of ESPON contact point, ESPON could take the leadership in the elaboration of an harmonised definition of labour market area or daily activity basin covering all European territory. Instead of being a "follower" of normative definition produced by Eurostat, ESPON should propose innovative approaches and compete with OECD in the field of innovative territorial approaches.



Few difficulties to take into account for the future

- **What about the strategy on the ESPON Database in 2014-2020:** Based on our experience in the two phases of the ESPON Database projects from 2008 to 2013, we would firstly underline that it is a real threat to modify the strategy related to the ESPON Database every 1 or 2 years according to the pressure of circumstances. It is certainly important to take into account the feedbacks of users (policy makers and researchers) but only at the end of a 3-years contract. It is fair to admit that in the past period (2007-2013), it was difficult to have a clear view on the possible database strategy. But for the next period (2014-2020) we strongly suggest to be much more precise in the expected task. Doing that implies probably to split the "Database Project" into 3 or 4 small and medium projects much more targeted and realised on a shorter period of time. The danger with this solution would be the risk of lack of coordination between the different small and medium projects working on database. To be really efficient, ESPON CU should be executive coordinator of the ESPON database and engage directly on its own resource a team of people specialized in GIS, computer Science, Spatial Analysis, … This "ESPON database team" would define the strategy for all the period 2014-2020 and launch then a limited number of small and medium projects related to specialized tasks (e.g. harmonisation of time series, development of web services, outlier check, local data collection, grid data integration through OLAP cube).

- **Maintenance of the ESPON Database?** One of the obvious threats would be that the ESPON Database would no longer be maintained and used after the end of the ESPON Programme phase 2, which would undermine the efforts made by the ESPON M4D Project to confer a high level of sustainability to the ESPON Database. Another threat would be to integrate datasets, without being checked as the tracking tool allows to, which would degrade a lot the level of quality reached in the ESPON Database. It is very clear that this question is related to the choice of the overall strategy of ESPON 2020 concerning database. We are convinced that the ESPON CU could easily maintain and use the tools developed during the period 2009-2013, making therefore easier the start of the new ESPON projects in 2015. But we are also convinced that many changes or improvement should be made quickly, according to the strategy adopted.

- **HOW TO TAKE INTO ACCOUNT the political demand related to the use of the ESPON Database and its interface**: Policy makers are without any doubt the final users of the ESPON database portal. The choice of priority in terms of data collection should therefore be in line with the priority of policy makers (and try to anticipate this demand or create it). But being the ***final user*** does not mean that policymakers are ***direct users*** of the ESPON database. Nobody can imagine the Minister in charge of spatial planning from a Member State trying to download an excel file of GDP per capita of EU regions… What this Minister really would like may be a map of GDP per capita, a simulation of regions eligible to convergence policy, a diagram of evolution of GDP per capita since the beginning of the crisis of 2008, a policy brief explaining if regional inequalities has increased, a research note concerning the spillover/backwash effects etc. The important point is therefore not to improve the interface of ESPON database with political keywords (these political keywords are always changing…) but rather to ensure a smooth and efficient connection between ESPON database and dedicated tools such as “City Bench”, “HyperAtlas”, “Monitoring tools”, etc.

Conclusion

Under the ESPON programme and the M4D project, the metadata model and the process of updating the database has been set up to make it intelligible and accessible. In this sense, this database is a tool that gives easy access to indicators on the cohesion policy of the European Union. These indicators are precisely documented so that their methodology can be replicated and updated or discussed or questioned.

ESPON database becomes a strategic part of the ESPON programme, not only searchable from the query interface, but also by other analysis (HyperAtlas) and visualization tools (ESPON Online Mapping Tool). Ultimately, access to metadata Web services (Catalog Service for the Web - CSW) and data (Web Feature Service - WFS) will release the contents of the base ESPON other spatial data infrastructures, to third party applications and other operating and performance indicators and metadata initiatives.

If the heterogeneity of the content of the database is well aware of the richness of the ESPON programme, it also highlights the difficulty in interpreting and synthesizing the content for users unfamiliar with this kind of production. This underlines the need for a strategic long-term vision on the content and organization of the database.

The experience of the M4D projects led to many advances in organizing, sharing and dissemination of complex data. For these advances do not go unheeded, the issue of sustainability and recovery of database-wide coordination of the ESPON programme is now a question that must be asked. The absence of permanent central structure has a short-term impact on the quality monitoring data. Currently, this process is semi-automatic, since most quality checks are performed by human operators. What will happen at the end of the project without those resources?

An optimal solution would be based on a strong central technical structure (such as data center) at the scale of the ESPON programme. It would therefore be necessary to provide such technical structure of human resources to manage the content of such database and define the issues and long-term goals.

References

REGIONAL DATA / TIME-SERIES CREATION

ANDRIENKO N., ANDRIENKO G., GATALSKI P., 2001, *Exploring change in census time series with interactive dynamic maps and graphics*. Computational Statistics, vol. 16, no. 3, pp. 417–433.

BEN REBAH M., PEETERS D., PLUMEJEAUD C., YSEBAERT R., 2011, *Modeling territorial changes and time series database process, empirical approach and applications*, ESPON, Technical Report, 32 p.

BOX, G.E.P and G.M. JENKINS, 1970, *Time Series Analysis, Forecasting, and Control*, HoldenDay, Oakland, C.

BOX, G.E.P and G.M. JENKINS and G.D. REINSEL, 1993, *Time Series Analysis, Forecasting, and Control*, Third Edition, Prentice Hall, Englewood Cliffs, NJ.

CHARLTON M., HARRIS P., 2011, *Spatial analysis for quality control*, ESPON Databse 2013, Technical Report, 81 p

CHARLTON M., BRUNDSON C., CAHALANE C., PFORTE L., 2014, *Estimating missing data in hierarchical space-time series with a short temporal extent*, colloquium GIS Research UK, 15-16 April 2014, Glasgow.

CHATFIELD, C., 1996, *The Analysis of Time Series: An Introduction*, Fifth Edition, Chapman & Hall, Boca Raton, F.

DEICHMANN U., BALK D., YETMAN G., 2001, *Transforming Population Data for Interdisciplinary Usages: From census to grid*, NASA Socioeconomic Data and Applications Center (SEDAC), 19 p.

ESPON 3.2., 2006, *Part 1 – Handbook for data collection*, Data Navigator 2, Final Report, 153 p.

ESPON 3.4.3, 2006, *The modifiable areas unit problem*, Final report, 254 p.

ECL, 2013, *ESPON Cartographic Language, options and recommendations related to further development of an ESPON Cartographic Language,* 80 p.

FULLER, W.A., 1996, *Introduction to Statistical Time Series*, Second Edition, John Wiley & Sons, New York.

GAYNOR, P.E. and R.C. KIRKPATRICK, *Introduction to Time-Series Modeling and Forecasting in Business and Economics*, McGraw-Hill, New York, 1994.

GRASLAND C., MATHIAN H., VINCENT J.M., 2000, *Multiscalar analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences*, Statistical Journal of the United Nations Economic Commission for Europe, IOS Press, Volume 17 (2), 157-188

LEVY M.-L., OMOLUABI E., 1992, *Le recensement du Nigéria, Population et Sociétés*, n°272, 4 p.

MADELIN M., GRASLAND C., MATHIAN H., SANDERS L., VINCENT J-M., 2009, *Das 'MAUP' : Modifiable Areal Unit – Problem oder Fortschritt ?*, Informationen zur Raumentwicklung, n°10/11, 645-660.

MILEGO R., RAMOS M.J., 2011, *Disaggregation of socio-economic data into a regular grid and combination with other types of data*, ESPON, Technical Report, 37 p.

PLUMEJEAUD C., 2011, *Modélisation et représentation de données spatio-temporelles évolutives pour la cartographie interactive*, Thèse de doctorat, 310 p.

PLUMEJEAUD C., VINCENT J.M., GRALAND C., BIMONTE S., MATHIAN H., GUELTON S., SHUMWAY, R.H. and STOFFER D.S, 2000, *Time Series Analysis and Its Applications, SpringerVerlag*, New York.

WEST, M. and J. HARRISON J., 1997, *Baysian Forecasting and Dynamic Models*, Second Edition, Springer-Verlag, New York.

YSEBAERT R., LAMBERT N., GRASLAND C., LE RUBRUS B., VILLANOVA-OLIVER M., GENSEL J., PLUMEJEAUD C., 2012, *HyperAtlas, un outil au service du débat politique, application à la politique de Cohésion de l'Union Européenne* in : Fonder les sciences du territoire, Bekouche et al., Karthala, 293 p. (an English version will be soon available).

URBAN OBJECTS

BRETAGNOLLE A., GUEROIS M., MATHIAN H. PAULUS F., VACCHIANI-MARCUZZO C. et al. 2011. *Harmonie-cités : bases de données harmonisées sur la dynamique et les compétences des villes en réseau selon les régions du monde*. Rapport final de l'ANR Corpus et Outils de la Recherche en SHS, 54 p. En ligne.

BRETAGNOLLE A., GUEROIS M., MATHIAN H. et al., 2011, *Naming U.M.Z.: a database now operational for urban studies*, Technical report, ESPON Data Base 2013, 39 p.

BRETAGNOLLE A., GUEROIS M., MATHIAN H., PAVARD A., 2014*, Functional Urban Areas in Europe. A feasability study through the comparison of two models: commuting flows and accessibility isochrones*, Technical report, ESPON Database 2013.

DJIKSTRA L., POELMAN H., 2012, *Cities in Europe*, The new OECD-EC, European Commission, Regional Focus 1/2012, 16 p.

GUEROIS M., BRETAGNOLLE A., GIRAUD T., MATHIAN H., 2012, *A new database for the cities of Europe? Urban Morphological Zones (CLC2000) confronted to three national databases of urban agglomerations (Denmark, France, Sweden),* Environment and Planning B, vol. 39 (3), pp 439-458.

MATHIAN H., SANDERS L., 2006, *Scientific approach of the MAUP?,* in : Grasland C., Madelin M. (eds.), The modifiable areal unit problem, Final Report of ESPON 3.4.3, 1-33.

ROZENBLAT C.,1995. *Tissu d'un semis de villes européennes*, Mappemonde, n°4, 22-27.

ZAHAVI Y., 1974, *Traveltime budgets and mobility in urban areas*. Report prepared for the U.S Department of transportation.

GRIDS

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

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

GALLEGO J., *A downscaled population density map of the EU from Commune Data and Land Cover information*, JRC-ISPRA, 10 p.

ESPON 3.4.3, 2006, *The modifiable areas unit problem*, Final report, 254 p.

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

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

ZAUCHA J. and SWIATEK D., 2013, *Place-based territorially sensitive and integrated approach. A NTCCP report*, Ministry of Regional Development, Wspólna 2/4, 00-926 Warsaw.http://www.stfk.no/Documents/Nering/Kysten%20er%20klar/Report_place-based_approach_29_03_2013.pdf

LOCAL DATA

BBSR, 2014, Raumbeobachtung Deutschlands und seiner Nachbarregionen, Workshop, 09 September 2014, Berlin.

COOMBES M., CASADO-DIAZ J.M., MARTINEZ-BERNABEU L., CARAUSU, F., 2012, *Study on comparable Labour Market Areas* (Final research report), Framework contract 6001. 2008.001 - 2009.065

DATAR, 2011, *Dynamiques, interdépendance et cohésion des territoires*, Rapport de l'Observatoire des Territoires 2011, 219 p.

DECROLY J.M., GRASLAND C., 1993, *Frontières, systèmes politiques et fécondité en Europe*, *Espace Population Sociétés, 2,* pp. 135-152

GLOERSEN E., LÜER C., 2013, *Population data collection for European local administrative units from 1960 onwards*, Final Report, Spatial Foresight, Contract No 2012.CE.16.BAT.054 (DG REGIO), 17 p.

GUO, D. and WANG H., 201, *Automatic Region Building for Spatial Analysis*, Transactions in GIS, 15(s1), pp. 29-45.

GUO, D. (2008), *Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP)*, International Journal of Geographical Information Science, 22(7), pp. 801-823.

GUO, D. (2009), *Greedy optimization for contiguity constrained hierarchical clustering*, The 4th International Workshop on Spatial and Spatiotemporal Data Mining, IEEE International Conference on Data Mining, pp. 591-596.

HAGGETT, P., ARNOLD, E., 1965, *Locational Analysis in Human Geography*, St. Martin's Press London. 337 p.

SANDERS, L., 2001, *Modèles en analyse spatiale*, Hermès-Lavoisier, Paris, 333 p.

EUROPEAN NEIGHBOURHOOD

EUROPEAN COMMISSION, 2008, *Green Paper on Territorial Cohesion. Turning territorial diversity into strength*, 12 p.

ESPON ITAN, 2014, Draft Final Report, Main report, 52 p.

ROBERTSON, R., 1995, *Glocalization: Time-Space and Homogeneity-Heterogeneity*, pp. 25-44 in M. Featherstone, S. Lash, and R. Robertson (eds.), Global Modernities, London: Sage.

NSIs WEBSITE, 2012-2014, National Statistical institutes Websites.

ESPON DATABASE PORTAL

PLUMEJEAUD C., 2011, *Modélisation et représentation de données spatio-temporelles évolutives pour la cartographie interactive*, Thèse de doctorat, 310 p.

UNEP, 2012, *UNEP Environmental Data Explorer. Search - Map - Graph – Download*, http://geodata.grid.unep.ch/extras/geosubregions.php.

UNEP/Grid Geneva, LIG Steamer, June 2013, *Access to ESPON Database by Third party applications, an expertise on the potentialities of web services for accessing the ESPON metadata and data*.

YSEBAERT R., SALMON I., LE RUBRUS B., BERNARD C., *Recueil, traçabilité et restitution des données territoriales du programme ESPON*, Colloquium « Fronts et frontières des sciences du territoire », 27-28 March 2014, Paris.

M4D DOCUMENTATION [online] http://database.espon.eu (last visit: January 2015)

M4D, 2012 (1), *ESPON Data and Metadata Specifications*

M4D, 2012 (2), *ESPON Nomenclatures Support*

M4D, 2012 (3), *How to Deliver my Data?*

M4D, 2013 (1), *Dictionary of Spatial Units*

M4D PUBLIC TECHNICAL REPORTS [online] http://database.espon.eu/db2/resource (last visit: January 2015)

M4D, 2011 (1), *The Core Database Strategy – A new paradigm for data collection at regional level*, 108 p.

M4D, 2011 (2), *Towards an ESPON hierarchy of functional regions*, 15 p.

M4D, 2012 (4), *Detecting and Handling Anomalous Data in M4D*, combined report, 38 p.

M4D, 2012 (7), *Time-series analysis*, 20 p.

M4D, 2013 (2), *A strategy for elaboration and update of coherent time series of hierarchical territorial units*, 45 p.

M4D, 2013 (3), *UMZ: a database now operational for urban studies (M4D improvements)*, 33 p.

M4D, 2013 (4), *European Neighbouring cities*, 31 p.

M4D, 2013 (6), *Territorial units in Russia, Ukraine, Belarus and Moldova and NUTS classification*, 40 p.

M4D, 2014 (1), *A feasibility study for the comparison of two models of FUA: commuting flows and accessibility isochrones*, 35 p.

M4D, 2014 (2), *Local data in M4D: LAU2 and Very Important Geographical Objects (VIGO); Delineating an alternative geometry at local scale,* 22 p.

M4D, 2014 (3), *OLAP technologies applied to the integration of geographic, thematic and socioeconomic data in the ESPON Context,* 54 p.

M4D TECHNICAL REPORTS, RESTRICTED ACCESS

M4D, 2013 (7), *Towards an overall data collection strategy (and database?) at the scale of the ESPON Program,* 18 p.

M4D NEWSLETTER [online] http://database.espon.eu (last visit: January 2015)

M4D, 2012 (5), *ESPON M4D Newsletter n°1*, June 2012, 6 p.

M4D, 2012 (6), *ESPON M4D Newsletter n°2*, December 2012, 6 p.

M4D, 2013 (5), *ESPON M4D Newsletter n°3*, June 2013, 6 p.

ISBN Number : 978-2-919777-91-4.