ESPON M4D
Multi-dimensional Data Design and Development

INCEPTION REPORT
REVISED VERSION

2nd February 2012

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

This inception report represents the first results of a research project conducted within the framework of the ESPON 2013 programme, partly financed through the INTERREG III ESPON 2013 programme.

The partnership behind the ESPON Programme consists of the EU Commission and the Member States of the EU25, plus Norway, Switzerland, Iceland and Liechtenstein. Each country and the Commission are 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 document produced by finalised and on-going ESPON projects.

ISBN number:
This basic report exists only in an electronic version.
Word version:

© The ESPON Monitoring Committee and the partners of the projects mentioned.

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

List of contributors to the inception report

UMS RIATE (FR)
Claude Grasland
Ronan Ysebaert
Isabelle Salmon
Nicolas Lambert
Timothée Giraud

NCG (IE)
Martin Charlton
Paul Harris
Paul Lewis

LIG-IMAG (FR)
Jérôme Gensel
Benoit Le Rubrus
Anton Telechev
Marlène Villanova-Oliver

UAB (ES)
Roger Milego
Maria-José Ramos

UMR Géographie-cités (FR)
Anne Bretagnolle
Hélène Mathian
Marianne Guérois
Antonin Pavard
Florent Le Néchet

TIGRIS (RO)
Octavian Groza
Alexandru Rusu
Daniel Tudora
Aurelian Nicolae Roman

TABLE OF CONTENT

1 Introduction: from ESPON Database to M4D Project

1.1 Context

It is very important to emphasize the continuity but also the differences between the two phases of the ESPON Database 2013 project. For simplicity reason, we propose to use the abbreviations *ESPON DB1* for first step and *ESPON M4D* for the second step.

1.1.1 Continuity

The figure 1 summarizes the improvements and innovations of the ESPON DB1. The ESPON DB1 was structured as a cube to show the different dimensions such as time, scale and themes. This is the main aim of the project.

- *Extension of time series* - 1995-2005 in the ESPON 2006 – in a long term database (1950-2050)
- *Extension of scales of analysis:* Starting from the NUTS divisions to other geographical dimensions, from the local scale to the global one, passing through various geographical objects (UMZ, grids, flows…)
- *Extension of thematic fields*: Making possible the integration of various topic of analysis, from the classical socio-economic field to pluri-approaches possibilities (crossing environmental dimensions to socio-economic indicators and so on).

Indeed ESPON DB1 project has built strong conceptual foundation (Technical reports, datasets, metadata profile, interface of data mining...) to develop innovative and operational database compliant with the INSPIRE directive. In general, ESPON M4D aims to improve the results done in the first phase of the project, following the information Cube strategy, in order to build an operational database. The general objectives for the ESPON M4D Project, defined in the project specification (version 23 August 2010, box 1) follows globally the recommendations developed by ESPON Database Phase 1 (Second Interim Report1), namely in point 2, 3 and 4 (e.g. Diffuse data, time-series and various geographical objects).

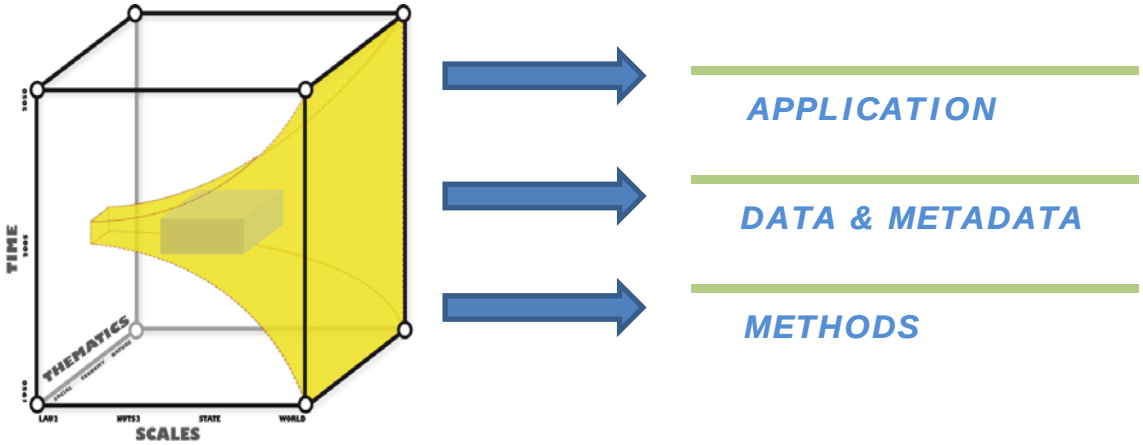


Figure 1 - The "information Cube": ESPON M4D goal

[1] http://www.espon.eu/main/Menu_Projects/Menu_ScientificPlatform/espondatabase2013.html

1.1.2 Differences

On the other hand, ESPON M4D cannot be (and will not be) a single follow-up of the previous ESPON DB1 for at less three different reasons:

- *The position in the lifecycle of the ESPON 2013 program as a whole is different*. ESPON DB1 took place at the very beginning of the program (2008-2011) and was focusing on foundations; ESPON M4D project take place in the middle and end of the ESPON 2013 cycle and should therefore focus much more on consolidation and achievement.
- *Networking with other ESPON projects has become a priority*. In ESPON DB1, the number of other ESPON projects was less important than now, but the networking with them appeared nevertheless to be a very time consuming task, especially in the final step of data delivery and check. The initial budget allocated to this task in ESPON DB1 was not sufficient and the distribution of networking activity between all project partners of ESPON DB1 had not been sufficiently strictly organized. The result was a huge additional amount of work granted for free by the coordination team (RIATE) without any financial counterpart. The ESPON M4D has taken into account this failure and allocated a more substantial part of budget to networking activities (20%) and introduce a specific work-package with clear targets and rules of division of work between M4D team in charge of the follow-up of other ESPON projects.
- *New political priorities are addressed to ESPON Program in general and therefore to ESPON database in particular*. Since the great economic crisis of 2008-2009, some obvious modifications has been introduced in the objectives of regional policy, that are visible in the 5th Cohesion Report, the strategy Europe 2020 and the revised version of the Territorial Agenda and the TSP adopted in June 2011 under Hungarian Presidency. The period 2011-2013 is characterized by a great uncertainty on the future orientation of European Budget in general and European Cohesion Policy in particular (including the future mission of ESPON after 2013). In this "grey" period 2011-2013, the statistical data will play a crucial role because they are a crucial input for political decision maker and they will be a central element for discussion and compromises to be found between member states. In this context, the large foundations of the ESPON database launched by ESPON DB1 in 2008-2011 are at the same time an opportunity and a threat. They are firstly an opportunity because they make possible to address a large number of political issues, from local to global scales, from smart to sustainable and inclusive growth, from cities to region or water basin objects … But they are also a threat because such a large scope is partly contradictory with the reactivity to request of short-term that will be addressed by political decision makers to ESPON in next months.
- *The in-between situation of ESPON database* as compared to official data producers on one hand (Eurostat) and expert international organization on the other hand (OECD, EEA) is more and more obvious. This point can be illustrated with the problem of the choice of relevant regional division (or delimitation of functional urban areas). On one hand, ESPON should use official EU definitions (like Eurostat NUTS2 and NUTS3) because it is the best way to provide useful inputs to political debate for future discussions on structural funds after 2013. On the other hand, ESPON should provide new evidences for long term strategic planning and feel free to use alternative divisions that are more suitable for the analysis of territorial dynamics and elaboration of long term time series in past and future. In this case, the good choice is the elaboration of alternative territorial divisions that describe better the social, economic or environmental trends, like OECD mixture of NUTS or EEA grids ad water basin. It is worth to remind that ESPON Program has been a pioneer in many aspects and that many of its most

recognized results are precisely based on non-official divisions, new geographical objects and combination of upper and lower scales of analysis[2].

- *The focus on metadata, quality of data and outlier analysis* was introduced in ESPON DB1 but will be more and more a priority in ESPON M4D because it appears as the most significant element of "corporate identity" that ESPON Program is likely to develop in order to find an original place in the landscape of European statistical producer. The ambition of the ESPON program is at the same time to provide analysis of high level from scientific and political points of view (the final report of priority 1 and priority 2 projects), but also to offer the possibility to end-user to criticize the results obtained and to explore in more details and to get more information of any result of particular interest. If this end-user is interested by an ESPON map on regional energy shortage, he/she should be able to download the datasets behind the map, explore the definition of the indicator, combine it with its own data, etc. If a regional political decision maker is surprised by the exceptional situation of its territory on an ESPON map, he/she should be able to check the exact origin of the figure behind the map, discover what was the original source or the model of estimation, what make this region exceptional as compared to the situation at different period, in neighbouring regions or for other criteria. We are convinced that this request that are not currently fulfilled by other European database (or only partly) are the best and the most original contribution that ESPON Program can offer.

1.2 Involving other ESPON Projects?

As a member of the Scientific Platform of ESPON, the M4D project will work in closer contact with other ESPON Projects and will try to take more and more into account their demands. As a result, some needs have been already identified following three main streams: Networking, Interface and Data collection.

1.2.1 Networking

Networking is an important component of the project. Whilst a central concern of the project is the creation of a comprehensive, accurate, reliable, and updated database, the users of the database require support for their use of the Database across a wide range of concerns. Networking will involve not only the creating of online materials for users of the Database but frequent visits to the Partners undertaking these research programme to determine user needs and aspirations with regard to the Database.

[2] We develop this strategic question in the general conclusion of this document with a SWOT analysis of external and internal role of ESPON M4D project.

1.2.2 Interface and computer application

Figure 2 – The current ESPON Database Interface

The integration of new geographical objects and new scales using previous methods developed in the phase 1 of the Database project, the main objective is to improve the interface, to make it more user-friendly and attractive and more intuitive. The idea is to have an efficient interface usable for heterogeneous users.

1.2.3 Data collection

Thematic work package consist to develop methodology and to integrate various geographic objects in the ESPON Database. This work package enhances the ESPON Database application through methodology and data and supports the implementation of the interface.

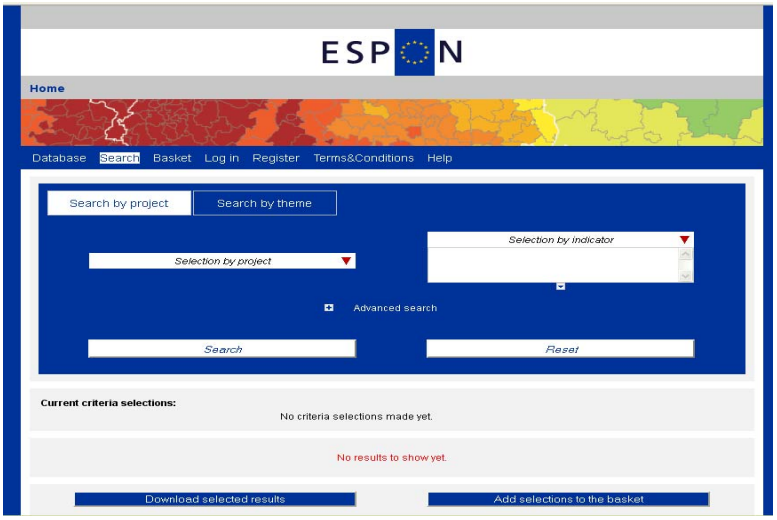


Figure 3 - ESPON Regional data, the current query interface

1.3 Operational consequences

On the basis of the changing context (see. 1.1) and the bottom up request expressed by the other ESPON projects (see. 1.2), the ESPON M4D has decided to focus on three priorities, taking into account what is the most important in the limits of the budget allocated.

(1) **Make an operational database**. Even if ESPON DB1 Project has provided strong foundations in term of data development, it is not only Technical Reports that are expected for integrating the different dimensions of the Information Cube. The ESPON Database interface will have to integrate recommendations and results given by Technical Reports. It is particularly the case for reconstitution of time-series, detection of outliers' challenges. We have therefore prepared a specific technical report called "**Core Database Strategy**" that has been delivered to ESPON CU in January 2012 in order to clarify our strategy and demonstrate that it is fully operational on the basis of various tests realized between October and December 2011.

(2) **Create a more efficient synergy in the field of data exchange within the ESPON Network.** The first phase of the ESPON Program (2007-2010) has launched few projects - 6 under priority 1[3] and 2 under priority 2[4] - as regard to the second phase of the program. Considering that ESPON Database Project has to provide guidelines and support to these projects, this task will become stronger. Added to this task which is internal to ESPON, a strategy for organizing the networking is also needed for exchanging with external organizations (Eurostat, OECD, INSPIRE…). We are currently negotiating a financial agreement with ESPON CU a solution in order to cover all ESPON projects and not only the one existing at the moment where ESPON M4D project was launched. The meeting with ESPON CU in Luxembourg the 16th January 2012 has lead us to agree in principle to a revision of our contract on this point. But the final agreement depends on the amount allocated by ESPON to this task (as regard to the number of project to be followed) and the acceptation of the distribution of work between TPG members. As long as ESPON CU proposal will be received, we will organized a meeting with our partners in order to conclude this point. During the meantime, we have also deeply improved and revised the content of the ESPON platform and reallocate the task, because of the problems encountered by the partner initially in charge of this task (departure of S. Fotheringham from NCG).

(3) **Take more benefit of other Priority 3 projects contributions to the ESPON platform**. The ESPON DB1 and ESPON M4D projects are certainly an important part of the ESPON platform but they should be completed by many other projects that are not currently achieved or just about to be launched. This problem is particularly obvious in the case of the selection of most interesting indicators for the monitoring of territories: ESPON DB1 has stored all indicators sent by other projects but this abundance is a problem for end user that are disappointed by the lack of hierarchy between very specialized data (e.g. residuals of variant of TIA models), very common data (population or GDP) and most original ESPON contributions (e.g. update of multimodal accessibility). It is clearly in relation with the TERCO project and the future MONITORING project that such hierarchy should be achieved. Another frustrating aspect of the current ESPON database is the lack

[3] RERISK, ESPON Climate, FOCI, DEMIFER, TIPTAP & EDORA
[4] SS-LR, TeDi

of connection with simple visualization tool for quick and easy access to data. As it is well illustrated by OECD regional database, it is necessary to complete the access to statistical tables by more cartographic or graphic oriented tools. It is not an easy task, precisely because such tools do not make sense for all variables of the database and implies a first selection of data to be visualized. HYPERATLAS (achieved in June 2011) offers a first possibility to do so, but do not cover all types of data. The new Priority 3 projects launched by ESPON in summer 2011 are therefore crucial and will help to solve this problem. Unfortunately, we have been informed by ESPON CU that for the second time, no application had been selected for the project "ESPON TERRITORIAL MONITORING" under priority 3 and that a third call will be launched a.s.a.p., but with no results before September 2012 at the very best. Considering the situation, ESPON M4D has decided to prepare as much as possible the work of this future project by introducing "political" tags and key-words in the database, as a complement to the thematic nomenclature.

2 Overview of the project

In order to fill the priorities defined above, we have organised the ESPON M4D project in four workpackages.

2.1 Organisation by working packages

The project is organised in 4 main Working Packages (A, B, C, D) further divided in Sub-activities (A.1, A.2, …) where precise tasks are identified and allocated to specific members of the project.

UMS RIATE ensures the coordination of the overall project and is responsible for the WP (B), (C) and (D). The Laboratoire d'Informatique de Grenoble ensures the coordination of the WP (A) and (C).

NCG, Géographie-cités, TIGRIS and UAB are responsible for one or several sub-activities.

- **Working Package A: Application (coordinated by LIG)**

The Workpackage A encompasses all the development activities that will be led in the Phase 2 of the ESPON M4D Project. Under the umbrella of this Workpackage are gathered four kinds of sub-activities. Computer Sub-Activity refers to the implementation of the next ESPON database application. It has been divided in five tasks: i) development of the core of the database; ii) development of interfaces of the ESPON M4D application; iii) development and integration of the OLAP cube; iv) integration of semantics and themes; and v) development of visualization tools on top of the OLAP cube. Statistics Sub-Activity groups the work that will be done on: i) harmonization of time series; ii) detection of outlier values; and iii) the evaluation of data quality. Map Kit tool Sub-Activity deals with the production of maps in conformance with the ESPON Corporate Identity. Finally, "Updates" Sub-Activity concerns updates of the ESPON database with valid data and metadata sets and updates of the ESPON M4D application on the ESPON Web site.

- **Working Package B: Thematic Group (coordinated by RIATE)**

This Working Package is dedicated to the consolidation of innovative solutions explored by previous ESPON DB1 for the production of new data, new maps, and new indicators … interesting for the ESPON program as a whole. ESPON DB1 had indeed proposed innovative solutions for the integration of data at different scales, based on different geometries, at different period of time, for different political use. The output was innovative maps or datasets related to manual indicated how to proceed in order to reproduce these methods in different situations. In ESPON M4D, we will build on this pioneer phase but avoid to open new areas of investigation except when it appears necessary. The most important task of the period 2011-2013 is indeed to build on the strong and wide foundation elaborated by ESPON DB1 and to develop the connections between the different types of data currently involved in the database and to make them available to a wider audience.

In more concrete terms, this WP can be considered as an expert group specialised in data collection, integration and harmonisation. In fact, it is impossible to envisage the building of an efficient application without the expertise of specialist of the different geographical

objects (cities, regions, grids, …) or the different thematic fields (environment, economics, social, transport, …) who: (1) are familiar to problems and solutions linked to the field of the data management; (2) have the knowledge to estimate what could be the most interesting achievement in term of data development for the final step of the Database Project. The Working Package B is divided in six sub-activities. Regions sub-activity refers to data collection of NUTS level in the ESPON Area (EU27+4). City Sub-Activity aims to define a suitable methodology for collecting and harmonising data linked to cities with different definitions. Grid Sub-Activity is dedicated to the thinking about the integration of raster information in the database. Local sub-activity focuses on the collection of local data in LAU levels. Neighbourhood Sub-Activity is linked to data collection in the Southern and Eastern Part of the European Union. This Sub-Activity is shared in three tasks: i) Regional data; ii) Urban data; iii) Environmental data. Surveys Sub-Activity aims to explore data collected through surveys and define in which conditions these data sources could be used or not.

- **Working Package C: Networking activities (coordinated by RIATE and LIG[5])**

Networking is an important component of the project. Whilst a central concern of the project is the creation of a comprehensive, accurate, reliable, and updated database, the users of the database require support for their use of the Database across a wide range of concerns. Such concerns include not only data, metadata, scale, analysis, indicator building, and data transfer issues, but also advice on the operation of the database itself and the associated tools for search, query, display, exploration, visualisation and download. Networking is achieved partly through the Scientific Platform, but also the provision of support for ESPON Priority 1, 2 and 3 projects and the development of mutual support between the project researchers and the Database itself. Networking will involve not only the creating of on-line materials for users of the Database but frequent visits to the Partners undertaking these research programme to determine user needs and aspirations with regard to the Database

- **Working Package D: Coordination (coordinated by RIATE)**

The Working Package D, relative to the coordination is not an obvious task since it ensures the fact that all other Working Packages (A, B, C) will work in synergy and efficiency. This task is coordinated by the lead partner and is divided in four sub-activities. The **Reporting** sub-activity aims to structure all the expected contributions of each partner (Intermediate reports, final report, and methodological notes). The **meetings** sub-activity organises all the elements linked to the organisation of meetings inside and outside the project. The **management** sub-activity is dedicated to the coordination of the information flow inside the project. The **support to the Coordination Unit** sub-activity is firstly related to all the common exchange of information. But also to more targeted support for selected request formulated by the ESPON Coordination Unit (production of posters, elaboration of original statistical tables or maps for an ESPON Synthesis Report…)

[5] Initially this task was under the full responsibility of NCG. But the responsibility is now shared with RIATE and LIG, because of the resignation of NCG's initial coordinator.

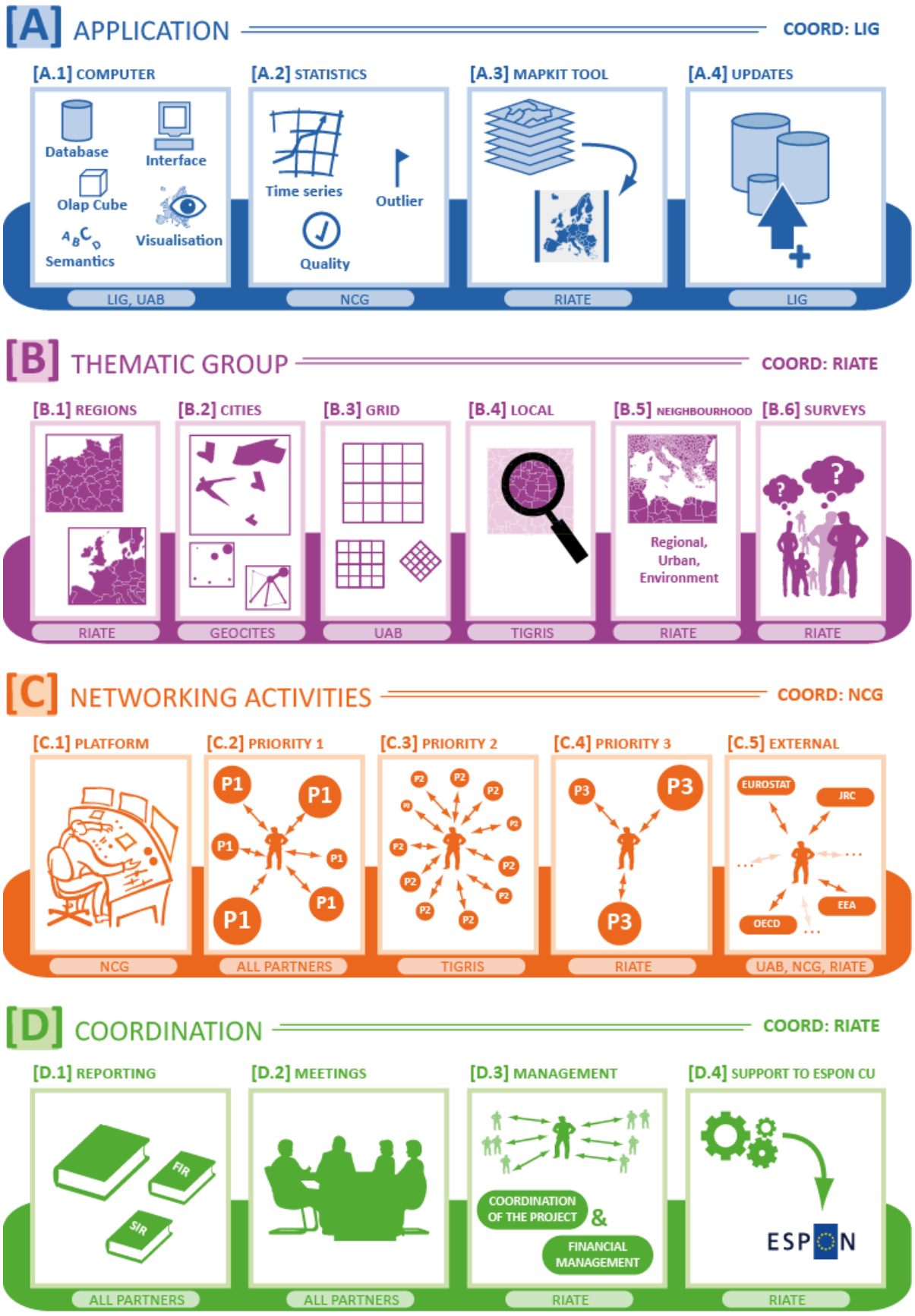


Figure 4 - Division of the project in activities and work-packages

2.2 Affectation of resources by WP, by time and by partners

The budget which is a relatively good proxy of the allocation of workforce by project partners[6] has been precisely detailed in this part. It takes into consideration a lot of factors: lifetime of the project, resources available and teams involved in the project, scientific goals of the project, feedbacks received at the ESPON seminar of Gödollö and Krakow, priorities defined in the terms of reference and the annex III to the Subsidy contract with ESPON Coordination Unit.

This part – deeply revised and improved after first comments made by ESPON CU on the draft version of Inception report - defines a clear strategic agenda, explaining what can be expected from the M4D Project and when. It defines also what cannot be done according to time and workforce available.

Firstly, we remind that it is not by chance or mistake that most of the resources are concentrated from the 31 July 2011 (Semester 2) to the 31 July 2013 (figure 2) due to the lack of resources. For most of the partners, resources are affected from Semester 2 because the M4D Project was still waiting for the signature of the Subsidy contract. It was impossible for some institutions to employ workforce on the M4D Project without this document[7]. The expenses made for the first reporting period are very low due to this fact. During the lifetime of the project, most of the resources are related to staff costs. To that respect, the workforce employed on the project will be done between semester 2 and semester 6. The amount allocated to the project was a problem from the very beginning of the project. We have alerted ESPON CU on that difficulty and the solution of a concentration of work from semester 2 to semester 6 was agreed.

Obviously the M4D project will do its best to maintain the DB during the lifetime of the project but without any innovation or hard work after semester 6.

We would like to remind ESPON that most of the time costs for permanent staff are VERY underestimated. Most of the permanent staff of the Lead partner RIATE but also at LIG and for others, is working on the project ESPON M4D without claiming any costs in refund. It is only because of the public status of RIATE that such indirect contribution to the ESPON Program is feasible … But it is obviously not a sustainable approach in the long term and we remind in the conclusion of this report that the ESPON Program has certainly underestimated the cost of the "technical part" of the Program, the Database project and should allocate more funds to Priority 3 projects in the future.

[6] Even if some differences exist between the two criteria. The cost of scientific labor force is indeed not the same in each partner country as well as VAT and percentage of funds taken or returned by legal authorities of research team.

[7] The subsidy contract was signed on the 27th June, 2011.

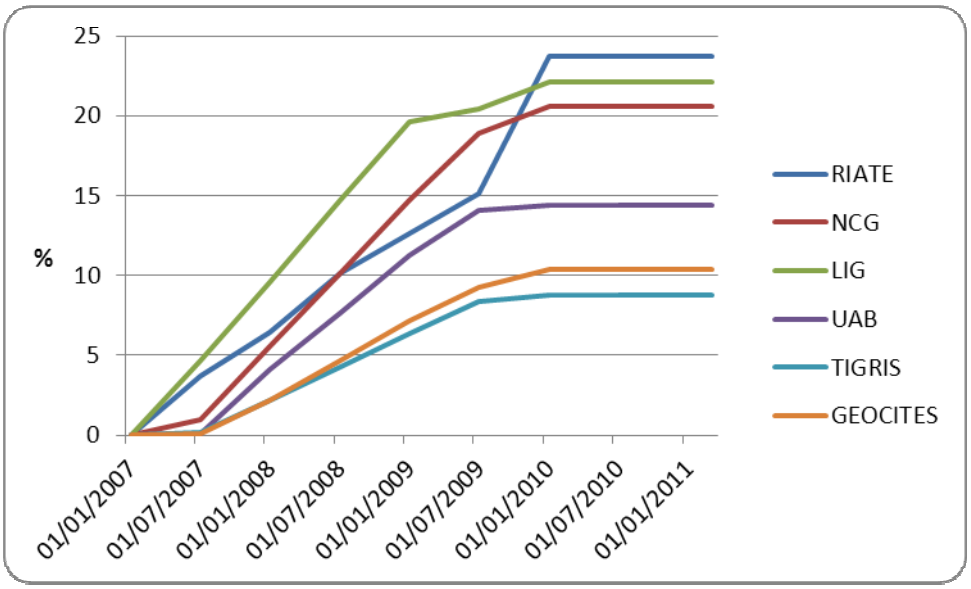


Figure 5 - Cumulative allocation of workforce by partners (%)

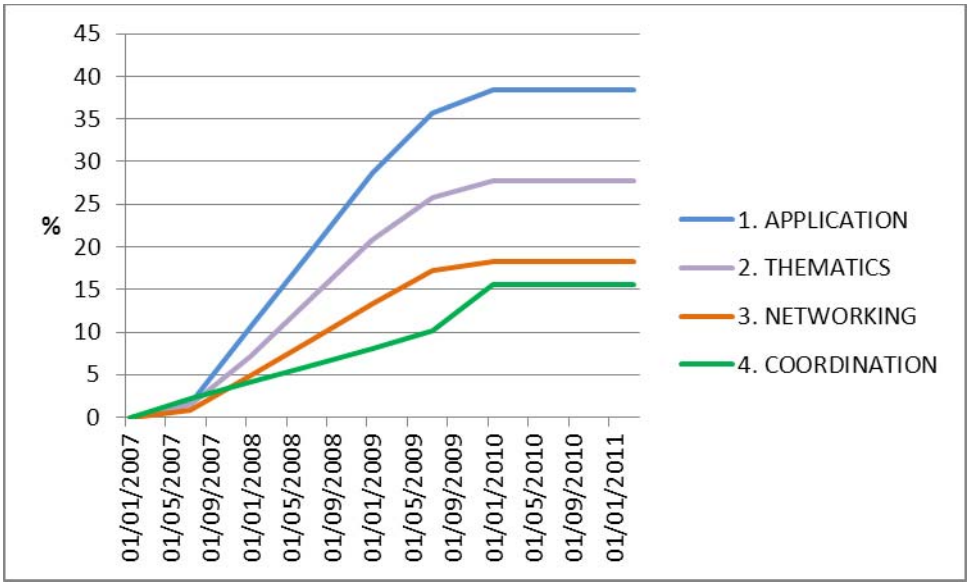


Figure 6 - Cumulative allocation of workforce by activity (%)

Considering the priorities of consolidation and quality of ESPON database (see. Section 1), most of the resources are affected to the WP Application (figure 6 & 7), with around 40% of the overall budget of the project. In this WP, the sub-activity A.1 (Computer development) takes a large part of the work (20 % of the overall budget). Statistics, which includes time series, automation of outlier detection and quality check of the database takes around 10 % of the global budget. The map kit tool and update sub-activities is a minor part of this WP as most part of the work has been done previously by ESPON DB1.

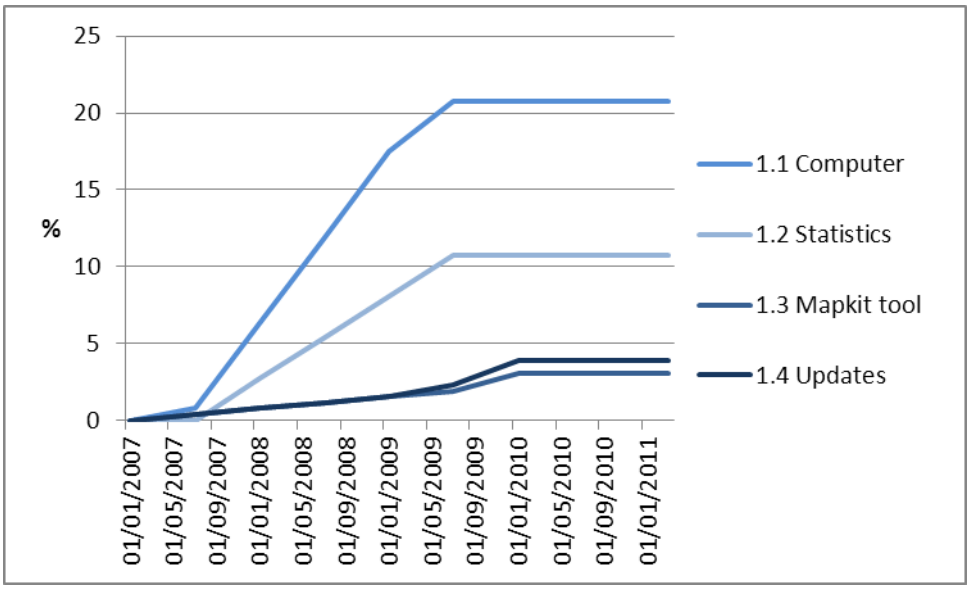


Figure 7 - Cumulative allocation of workforce by sub-activity of the application activity (%)

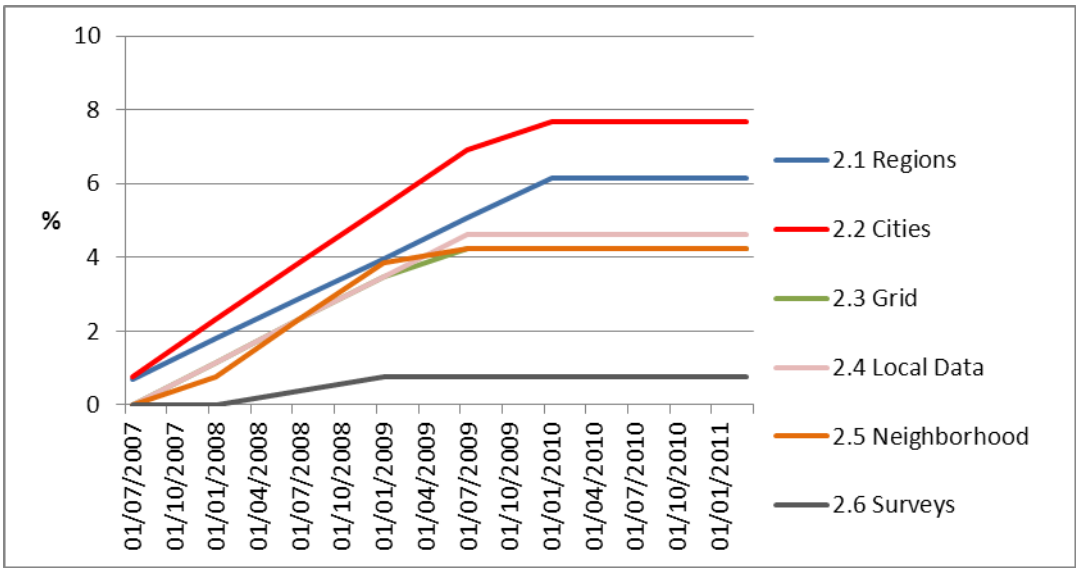


Figure 8 - Cumulative allocation of workforce by sub-activity of the thematic activity (%)

The thematic WPB represents the second most important budget allocation of the project (more than 25 % of the credits) since its aim is to provide innovative inputs in the database (figure 6). Following the requests of different types of users, the thematic WP (figure 8) aims to give a large importance to regions (WP B.1), cities (WP B.2) and local data (WP B.4) with more than 5 % of the overall budget of the project. Grid data, neighbourhood and surveys are also explored, but with less budget allocation. Following the feed-back received in Gödollö and Krakow meetings by other ESPON projects, we consider the possibility to allocate more work-force to the local data. More precisely, we propose to develop a more common strategy between the teams responsible from NUTS data (RIATE) and LAU data (TIGRIS). This common strategy is for example necessary in order to fulfil the objective to explore the interest of a new functional hierarchy of territorial divisions at level LAU1/NUTS3. But the exploration of new functional division is postponed

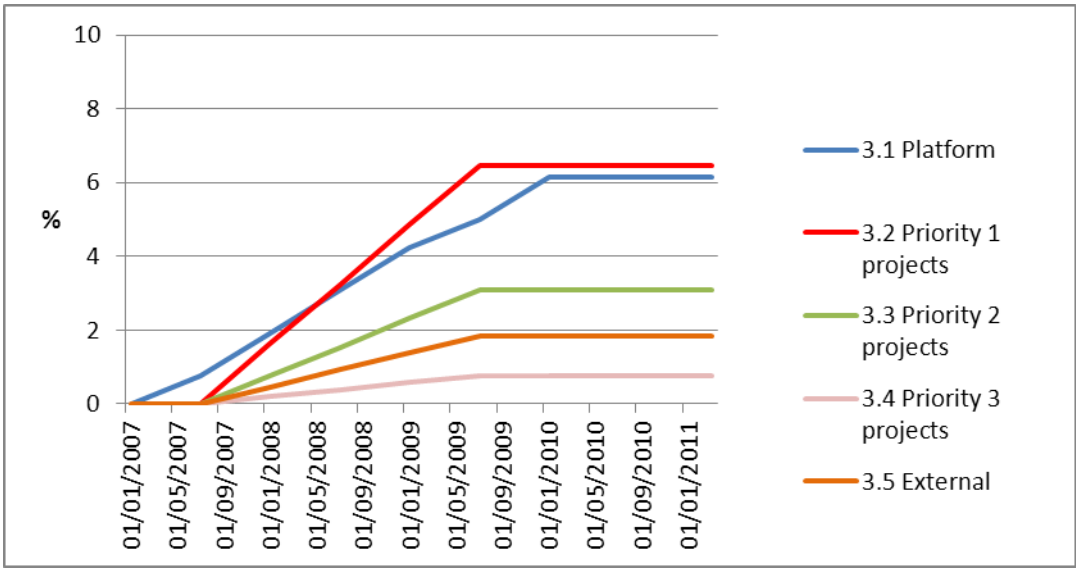


Figure 9 - Cumulative allocation of workforce by sub-activity of the networking activity (%)

Networking activities (WP C) has been taken as a special WP, considering the partial shortcomings observed in ESPON DB1. It represents for around 20 % of the overall budget of the project ESPON M4D. Importance has been given (figure 9) to the follow-up of the Priority 1 projects and a creation of a platform that can be used by the whole ESPON Community (more than 12 % of the resources). ESPON Projects under priorities 2 and 3 are also taken into account but with a less important allocation of funds since it represents less important volumes of data that covers not necessarily the whole ESPON Area. External networking with other institutions (like OECD) is also taken into account (around 2 % of the funds available).

15 % of the budget is affected to the coordination of the project (WP D) which is relatively limited according to the size and complexity of the ESPON M4D project (figure 10). Most of the budget is allocated to the management of the project (WP D.3) i.e. to the salary of non-permanent staff members during the period of activity of the project.

A significant part of the resources has also been dedicated to meetings (3,8% of the total), reporting (4,1 %) and support to ESPON Coordination Unit (1,8 %). The budget allocated to support ESPON CU is intent to answer to urgent requests. But for specific requests we advise to launch the tendering process.

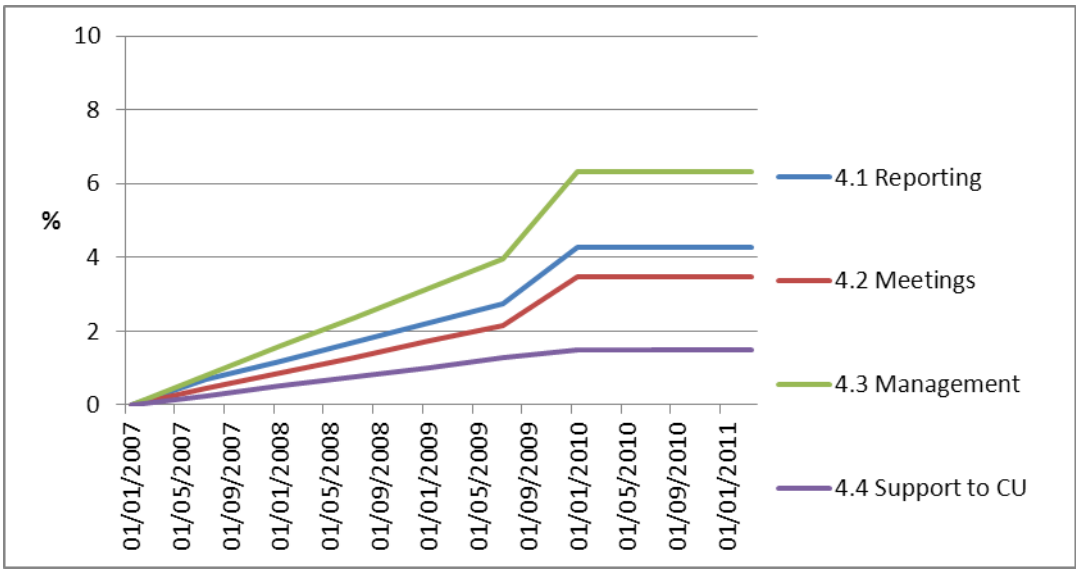
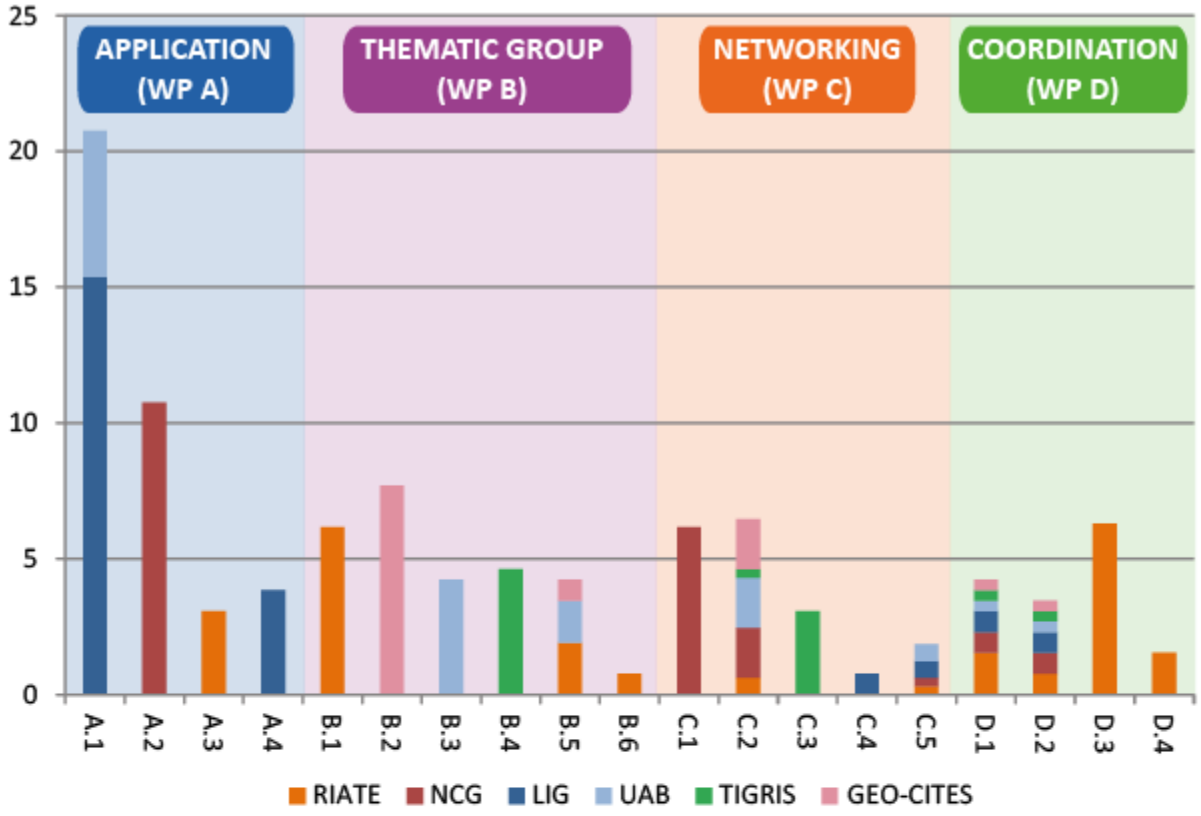


Figure 10 - Cumulative allocation of workforce by sub-activity of the coordination activity (%)

Each of the teams involved in the M4D Project has detailed activities planned in each Work Package in order to facilitate a sound evaluation of contributions of each of them in the project (figure 11). In this organisation, the lead partner of each Working Package will received a more important allocation of funds (respectively LIG, RIATE and NCG) considering the fact that these teams are responsible of these different parts of the work. This polycentric approach makes possible more efficient work, with for example extra-meeting between teams involved in each WP.



A.1 Computer
- A.1.1. Database
- A.1.2 Interface
- A.1.3 Semantics
- A.1.4 Olap Cube
- A.1.5 Visualisation

A.2 Statistics
- A.2.1 Time series
- A.2.2 Outlier
- A.2.3 Quality

A.3 Mapkit tool

A.4 Updates

B.1 Regions

B.2 Cities

B.3 Grid

B.4 Local Data

B.5 Neighborhood
- B.5.1 Regional Data
- B.5.2 Urban Data
- B.5.3 Environnement Data
- B.5.4 Balkans

B.6 Surveys

C.1 Platform

C.2 Priority 1 projects
- C.2.1 ATTREG
- C.2.2 SERVGE
- C.2.3 ARTS
- C.2.4 LAND USE
- C.2.5 TERCO
- C.2.6 TRACC
- C.2.7 SGTPDE
- C.2.8 GEOSPEC
- C.2.9 KIT
- C.2.10 TIGER
- C.2.11 SEAS

C.3 Priority 2 projects

C.4 Priority 3 projects

C.5 External
- C.5.1 JRC
- C.5.2 EEA
- C.5.3 OECD
- C.5.4 INSPIRE
- C.5.5 EUROSTAT

D COORDINATION
- D.1 Reporting
- D.2 Meetings
- D.3 Management
- D.4 Support to CU

Figure 11 - Involvement of M4D Team in the different sub-activities of the project on the overall period of the project (%)

2.3 Cross-activity organisation of work

Each WP activity (A,B,C,D) is carefully planned with an detailed timetable. Interim reports will propose summaries of main results.

Basically, WP A will receive the inputs from the WP B and try to include them into a computer application where expert knowledge can be reproduced as much as possible in a standardized and automatic way (cities, grids). WP A will also receive inputs from the WP C with new data coming from ESPON Projects, which will have to be checked by the outlier detection model (WP A.2) and integrated into the database (WP A.1 and A.4).

The same is true inside the WP B, where the administrative repartition of the work (B.1, B.2…) does not necessarily reflect the nature of the entire results which can be expected from the M4D Project. For instance it is also possible to re-draw the different sub-activities by geographical levels:

- "**macro level**": composed of the following items: world at state level, world with states grid, world as flows
- "**meso level**": composed of the nuts2-3, cities, neighbourhood items, surveys
- **"micro level"**: LAU data, continuous raster

As explained in the section 1, we have decided that ESPON M4D will not explore new geographical objects or scales but will focus on internal relations between the different one that has been proposed by ESPON DB1. Many potential innovative relationships exist between the different geographical objects and scales currently collected in ESPON database. We have displayed on the figure 12 six different possibilities of cross thematic involvement of teams, which will be extended to other possibilities considering the results obtained by the different partners:

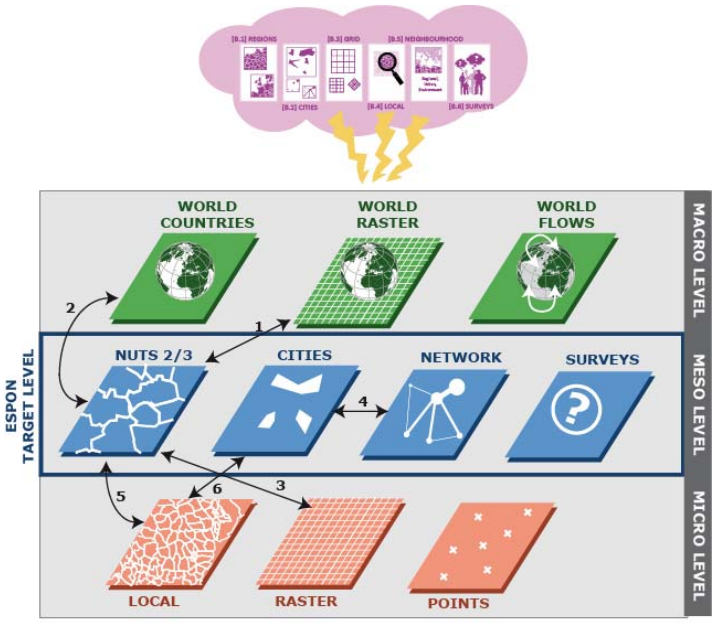


Figure 12 – WPB, Towards a cross sub-activity approach of organization of work
1: Estimating Corine Land Cover in ESPON Neighboring regions
2: Long time series at World level who could be considered for ESPON regional studies
3: Expertise for NUTS changes (through the JRC population grid for instance).
4: Functional distance between geographical objects
5: Explore the feasibility to create alternative territorial hierarchy (LAU/NUTS3)[8]
6: Check urban databases by local data available through National Statistical Institute.

[8] Not before the year 2013

3 The ESPON Database Strategy

This section summarizes the result of the first Project meeting held in Paris the 14-15 September where all Project Partners were represented as far as our project officers Sandra di Biaggio and Marjan Van Herwijnen from the ESPON coordination unit[9]. The main output of the meeting was the definition of a set of activities and deliveries to be engaged according to a precise agenda of realization.

3.1 Working Package A - Application

3.1.1 WP A1 – Computer

3.1.1.1 Database (LIG)

During the M4D project, the development of the core of the ESPON 2013 Database mainly consists in:

- Extending the schema of the database in order to store an enlarged collection of data, integrating new geographical objects. The ESPON database currently contains only NUTS data, whose levels range from 0 to 3. The updated ESPON 2013 database will store other types of geographical objects:
  - - Data collection for the Eastern and Southern European Neighborhood (SNUTS)
  - - Urban data: LUZ (Larger Urban Zones), and/or UMZ (Urban Morphological Zones), and/or FUA (Functional Urban Areas)
  - - Local data: LAU1, LAU2 (Local Area Unit)
  - - Raster data (Grid)

For each of these new geographical objects to be integrated, a selected set of indicators will be provided. Depending on the available data, all or part of the ESPON31 territory and possibly some parts of European Union neighborhood will be covered.

- Improving the integration process of data and metadata sets.
- Improving the compliance with European and international standards for spatial referencing and data storage, in particular with the INSPIRE recommendations. This task mainly consists in studying and improving the compatibility of the ESPON metadata extension of the ISO-19115 norm with the INSPIRE recommendations. The current status of the three formats is represented in figure 13 as three sets. Our goal is to enlarge the intersection of the three specifications.

In order to handle subsets (islets) of LAU1 and LAU2 units, a study will also be achieved in order to investigate the integration of local data, considering that such data come from different sources (in particular from ESPON projects under Priority 2) into the ESPON 2013 database, and may currently not cover the entire European Union territory[10].

[9] And also with reference to the conclusions of the special meeting between ESPON CU, RIATE and LIG, the 16th January 2012.

[10] Our contract does not imply the elaboration of an integrated LAU database which is obviously under the responsibility of Eurostat. But we have to take into account the fact that many data are delivered by Priority 2 projects at this level and also the growing interest of priority 1 projects for this local data. In our opinion, if the remaining program budget is sufficient, ESPON should launch a specific priority 3 project or contract on this local (LAU1/LAU2) or local–medium (LAU1/NUTS3) scales.

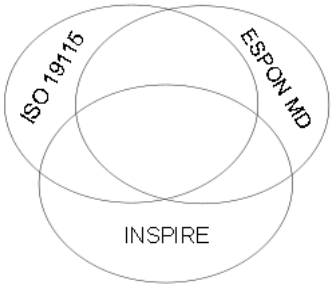


Figure 13 metadata formats as sets

The datasets integration flow, e.g. the update of the ESPON database contents is a major key point of the M4D project. It is summarized in figure 14:

- One first loop (1) involves the data provider, who is expected to provide a coherent set of data together with their associated metadata. The very first step will consist in performing automatic consistency checking (and when possible, corrections) in the uploaded data file, concerning for instance:
  - the general data file structure (expected column headers, rows order…);
  - the expected types of values for the indicators (for instance, a failure will occur when some decimal is encountered while an integer is expected);
  - the consistency of labels in the columns headers (indicators, data sources labels, time periods, etc.).
- Once the data file is correct, the Metadata Editor software will display a pre-filled metadata template form.
- Once the data provider fills correctly the metadata template (*i.e.* all required fields are filled), he will be authorized to definitely send the data and metadata file via the Upload interface to the Web server. The project manager (RIATE) is then notified that the next step can be started.
- A second loop (2) invites the corresponding partner of the ESPON M4D Project, to check further the syntactic content of the provided data/metadata. This task mainly consists in checking the text fields of the project files, for example:
  - Do the new indicators respect the naming conventions?
  - Are the names of new providers correctly spelled, considering the existing records in the database?
  - Are the new units of measure correctly spelled?

Once the content of the data/metadata set is considered syntactically valid, the NCG partner is then invited to execute some advanced statistical processes on the data/metadata set (outliers detection, for example). This loop (2) can lead to the fixing of the content of uploaded data/metadata set by the data provider. Once the content of the data/metadata set is considered as correct, the integration of the dataset into the database starts.

As the next step in the dataset integration flow, (3) in figure 14, we consider of the highest priorities the design and implementation of a tool, called *ESPON DB Admin*, which aims at automatizing this final step, in order to easily, integrate data/metadata sets into the ESPON Database. One major objective of this tool is to significantly increase the potential updates of the available content in the database (add/remove datasets functionalities). Moreover, this ESPON DB Admin tool will allow the management of

various glossaries (the list of providers, units of measures, etc.). Please also consult the "Platform" section (part 3.3.1) for further information about the complex dataset integration workflow.

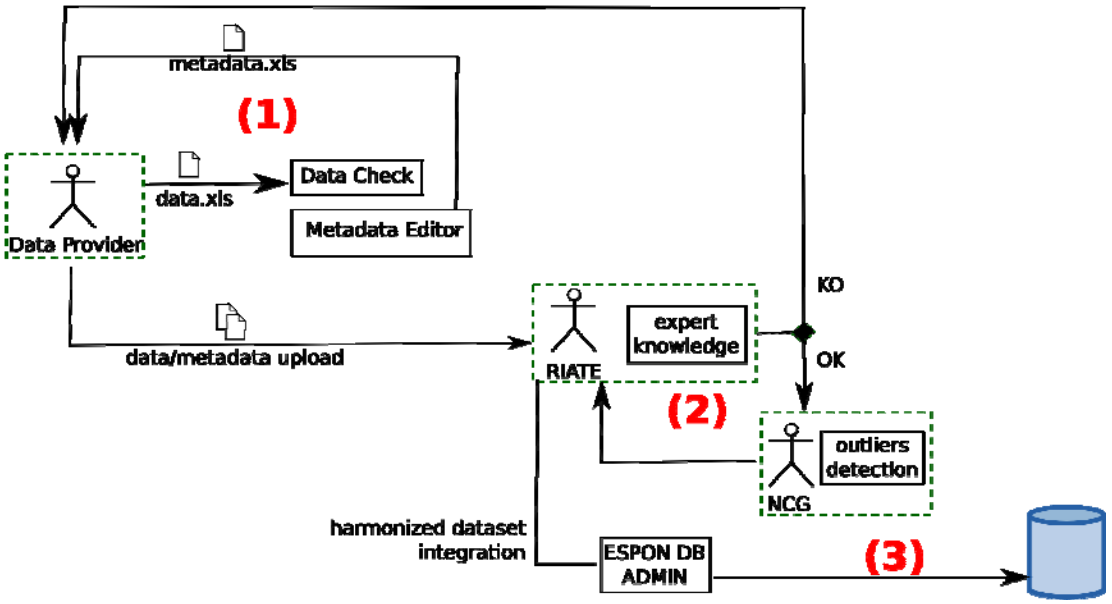


Figure 14 - Dataset integration flow

❖ ***Deliveries – timetable***

June 2012

Designed to be extensible to several kinds of geographical objects, the Metadata Editor and the Upload interface will allow by default the integration into the Database of datasets based on regional NUTS geographical objects, and urban datasets as a priority among the various new geographical objects to be integrated later.

The following tasks are planned for June 2012:

- Design and develop a Java API for metadata that will be possibly shared with any partner interested in developing tools for processing ESPON data, and will represent an interface ensuring data exchange and compatibility.
- Update the implementation of the Metadata Editor (and Data Check).
- Design and extend the database model for new geographical objects.
- Design and create some software for integrating and checking more efficiently data from all ESPON projects (ESPON DB Admin).

December 2012

- As soon as possible, the Metadata Editor and the Upload interface will be improved so that they take into account the upload of urban data, as well as the integration of Grid data.

June 2013

- Maintenance and minor evolutions.

3.1.1.2 Interface (LIG)

This work package subdivision concerns the twofold set of main functionalities available from the user interface of the ESPON 2013 database Web Application: the "Platform" and the "Query".

Please note that the previously named "Upload interface" set of functionalities, mainly targeted at ESPON Projects registered users, has been re-thought as a subset of the "Platform" set of functionalities, which is further described in 3.3.1. More precisely, M4D partners have agreed to design and to consider these "Platform" functionalities as a fully integrated part of the Web Application. Indeed, integrating the "Platform" functionalities into the Web Application will definitely increase the usability, browsing fluidity (Web pages consistent layout), the efficiency and the tracking of the datasets integration process.

Regarding the improvements of the Query and Download interface, one major objective to be addressed by the M4D Project concerns its user-friendliness, so that the ESPON database can be used by experts (researchers, scientists) as well as by non-expert users (policy-makers). The improvements will not be implemented under a "Computer Science point of view", but with the purpose of becoming useful in supporting policy formulation and implementation at different geographical levels. In practical terms, this means that inputs and concrete requests from policy makers, in particular from the ESPON MC (June 2011 MC Meeting in Budapest), will be fully considered for the further development and improvement of the interfaces of the ESPON Database.

Thus, the Query and Download interface will be improved in order to take into account the following key points:

- The improvement of the usability: the current Query and Download interface will be improved so that search and extraction in the next ESPON 2013 database becomes more user-friendly and adapted to the policy maker's profile.
- The integration of themes and keywords: the definition, integration, and selection of appropriate themes, sub-themes and potential keywords will improve the search possibilities.

- The study of:
  - The compliance of the ESPON 2013 Database with the INSPIRE recommendations for Network Services supporting View, Discovery and Download functions.
  - The access to the ESPON Database, not only through a Web query interface, but also via direct integration of the data into third party applications, including spatial data portals.

Based on the ESPON CU feedbacks following the MC Meeting in Budapest (June 2011), please note that an important set of improvements have already been brought, implemented and delivered in December 2011 (indicators overview page, query criterion, performance issues, for example). Moreover, in order to significantly improve the user-friendliness of the interface, the M4D partners have already thought of and proposed two new search paradigms to ESPON:

- During the Krakow Seminar ESPON Database Workshop, on the 29th of November 2011, Jérôme Gensel has presented a proposal of search interface based on the "Where What When" concept (respectively the spatial, the thematic and the temporal search criteria). The interface proposed visual prototype (figure 15) can be consulted in the presentation available on the ESPON Web Site [11].

[11] Jérôme Gensel. Workshop 1.4: ESPON Database. http://www.espon.eu/export/sites/default/Documents/Events/InternalSeminars/KracowNovember2011/26-ESPON_Database.ppt (last visit: 2012-01-25).

- During a meeting in Luxemburg (ESPON CU – UMS RIATE – LIG STeamer), on the 16th of January 2012, a visual prototype of the new ESPON DB query interface has been proposed (see figure 15)

Figure 15: Draft version of the future Query interface

As agreed during the meeting on the 16th of January 2012, the design and implementation of the future query interface will follow the following agenda:

- 25th January 2012: ESPON response regarding the agreement of the visual prototype of the new ESPON DB query interface;
- From 25th January to 8th February 2012: M4D partners will write specifications for this query interface. These specifications will detail the screens, the possible actions and various use cases;
- From 8th February to mid-March: ESPON will have to validate the specifications (and/or to suggest modifications);
- From mid-March to June 2012 ESPON Seminar in Aalborg, based on the agreed specifications document, this new version of the interface will be implemented.

Considering each forthcoming ESPON Seminar as an important milestone, the M4D LIG partner will deliver new official versions of the Web application for each of them (June 2012, November 2012, June 2013). Moreover, between the seminars, several prototypes

versions will be delivered in order to make new developed functionalities available as soon as possible.

For each delivery, a listing of the performed updates will be provided in order to clearly identify at a glance:

- what has been added to the database since the previous delivery (see also the section "Platform" relative functionalities: news and newsletter).
- the additional functionalities of the application between each delivered version.

Regarding the access to the database from third party applications, the following table 1 proposes a short SWOT analysis[12].

Criteria	Discussion
STRENGTH	The ESPON Database schema has been designed to allow flexibility and to support evolutions (integration of new geographical objects, for example). The revisited Metadata specifications (taking into account the INSPIRE directives) aim at improving the search and query possibilities. The revisited classifications of indicators (under construction) will allow multiple search possibilities.
WEAKNESS	The database has not been connected to any third application yet. For security issues, only the Query and Download interface of the Web Application can currently connect the database. Only one format (Excel files) is currently available as output.
OPPORTUNITY	Networking activities (EUROSTAT, EEA, OECD, UNEP, ESPON Online Mapping Tool project) may help at defining the needs in terms of the access to the content of the database from third party applications.
THREAT	The major risk regarding the design of the access to the database is to design a too much specific tool, that may respond to some requests and needs for the ESPON Online Mapping Tool, for example, but not to future unknown needs. The key idea of this threat is the difficulty to think of a generic access without knowing the needs and goals of these future accesses. The design and the implementation of a generic access to the ESPON Database is a very time-consuming task.

Table 1: SWOT analysis of the access to the ESPON 2013 Database

To sum up, this SWOT analysis highlights the difficulty to design a generic access for third party future applications

❖ ***Deliveries – timetable***

June 2012

- The delivery of a version of the Web Application including the functionalities which will be described in the specifications agreed in March 2012.

December 2012

[12] We propose in conclusion of this report a more general SWOT analysis of internal and external position of ESPON database, as regards to its specific role in ESPON strategy as a whole. That is the reason why the current SWOT is voluntary limited.

- The delivery of the application in December 2012 will mainly consist in taking into account the remarks and feedbacks about the previous June 2012 version.

- A technical study will be presented, based on the SWOT analysis shown in table 1 and conclusion, of the report, for querying the database from external third party applications via the Web Services technology.

June 2013

- An official Query and Download final application will be delivered, including all expected functionalities.
- The implementation of one of the solutions studied in the technical study delivered in December 2012 (Web Services for querying the database) will be provided in the form of WMS.

3.1.1.3 Semantics (LIG-RIATE)

During the Phase 1 of the ESPON 2013 DB Project, a thematic classification and structuring have been established in order to allow, on the one hand, the querying using keywords and themes, and, on the other hand, to easily describe data. The thematic structure resulting from Phase 1 will be associated with datasets based on new geographical objects. Also, the naming conventions proposed during the Phase 1 in order to identify indicators and to avoid duplicates will be adopted and implemented.
Thus, the thematic structure can be exploited:

- from the Upload interface (the Metadata Editor), as a guide for classifying indicators
- from the Query and Download interface, as a guide for selecting search criteria

After analysis of the situation and validation by ESPON CU in Luxembourg the 16th January 2012, it appears to us that an important revision of the thematic classification has to be launched in order to fulfill different objectives:

- **extraction of indicators by statistical type** (ratio, count, variation, typology…) in order to simplify connection with other applications like ESPON HyperAtlas (typically using ratio) or ESPON webmapping (producing different semiology for mapping counts variable or indicators).
- **Extraction of indicators by thematic type** (agriculture, population, economy…) in order to make more easy the connection with other databases following Inspire directive. Typically, we should combine a hierarchical thesaurus (specific to ESPON and adapted from results of ESPON DB1 and Interco) and a list of keywords (using standard like the ones of GMET).
- **Extraction of indicators by political challenges or objectives.** This last point is the most difficult as mentioned in introduction because it should be done in strong interaction with the Priority 3 MONITORING Project, unfortunately not launched until now and not forecast to start before the end of 2012.
  Taking this into account and as discussed with the ESPON Coordination Unit, this task will be based on the outputs provided by the INTERCO project (Annexes 2 and 3 of the INTERCO Final Report). In concrete terms, we propose to add a policy filter for querying the database. In that order, it will be possible to know what are the indicators contained in the ESPON database which relates to the 5th Cohesion Report; the Lisbon, the Gothenburg and the Sustainable Development Strategies; the Regions 2020 or the Europe 2020 documents. Also, it will be possible to query the database on the basis of the 26 top indicators selected by INTERCO Project.

But we want to precise that this policy labeling of indicators will be based **only on the INTERCO Final Report**. If updates are needed, it has to be coordinated by the ESPON Coordination Unit.

❖ ***Deliveries – timetable***

November 2011

- Definition of a new strategy for the codification, in order to avoid repeating the huge work of recoding existing data. This work is achieved with the present inception report that defines the strategy.

June 2012

- Construction of an extended thematic classification covering the three types of criteria defined here: statistical type, thematic type, political challenge[13]. The new functionalities expected from this classification will be presented at the Denmark meeting for validation by ESPON community.

December 2012 and June 2013

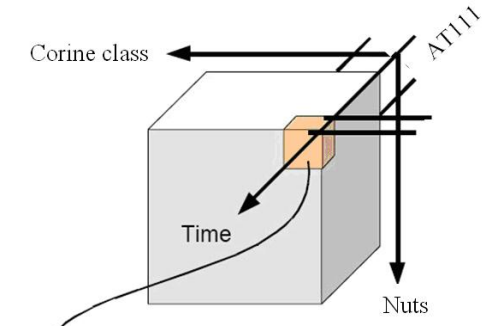
- Updates of the thematic classification in partnership with the other Priority 3 projects, in particular the Monitoring project.

3.1.1.4 OLAP Cube integration (UAB)

Analysis of the situation

The OLAP multidimensional data model will be integrated inside the ESPON database in order to make it available to users through a high user-friendly interface, providing a great level of usefulness.

The Online Analytical Processing (OLAP) technology is proposed as the integration framework for grid/raster data together with NUTS data. Databases customised for OLAP use a multidimensional data model borrowing aspects of navigational and hierarchical databases that make them faster than conventional relational databases.



Dimensions		Measures	
Nuts3 Code	Corine Class Level 1	Ha	Unemployment Total
AT111	Artificial surfaces	3795	697

Figure 16: ESPON OLAP Cube

[13] This is a very expensive task because of the amount of data still collected that we are now oblige to modify in order to introduce new metadata. A full time Post-Ph'D students has been engaged by RIATE to do this task that implies high skills in the political and thematic fields.

The environmental and socio-economic data will be redistributed based on the 1km European Reference Grid in order to harmonize both types of data. Once the variables have been distributed by 1km cell, either by aggregation or disaggregation, they can be compared to other variables or indicators on a cell-by-cell basis, and they can be integrated into an OLAP cube.

Different OLAP Cubes will be created according to the needs of the ESPON Community (e.g. OLAP Cube for Urban analysis purposes). It is foreseen to launch a survey amongst ESPON projects in order to catch user needs. In parallel, the different OLAP Cube updates will include the integration of the new basic indicators available in the Database.

In some cases, the big size of the data to be integrated into the OLAP Cube schema will suppose a challenge to be solved.

Objectives

To this regard, it will be needed to find innovative adaptations and solutions. The ESPON OLAP cubes will be integrated into the ESPON database platform in order to make them available to users. The integration could be made at three levels:

- Distribution of .CUB files: for advanced or technical users, the OLAP Cube files (*.cub) will be made accessible. This is what is currently delivered in every Cube update.
- Online connection: OLAP Cube files can be remotely queried by means of a connection to a server through MS Excel or ESRI ArcGIS software. This is technically possible. To be discussed in the First Interim Report and agreed with ESPON CU, as a server is needed.
- User-friendly web visualisation and query tool: an easiest way to query the OLAP Cubes and produce visual and useful results (tables, maps and charts) should be provided to the ESPON community. The next section shortly explains how this new tool should work and look like. In May 2012 there will be a first draft available. A demo could be ready for Denmark. The First Interim Report will include all the details about this tool.

❖ ***Deliveries – timetable***

November 2011

- Delivery of an updated ESPON OLAP Cube (v.4) and updated ESPON OLAP Cube Tutorial: ESPON Seminar in Cracow (November 2011)

3.1.1.5 User-friendly OLAP visualization tool (UAB)

Analysis of the situation

ESPON DB1 has developed an OLAP cube technology which is statisticially powerful but difficult to use for non-specialized users. We propose therefore, for a better use of GRID data produced by M4D, to build this client-server application by using the existing Spatial OLAP Technology and several open source tools: GeoMondrian, GeoTools and PostGIS on the server side, and OpenLayers and GeoExt on the client side.

It is important to keep in mind that this visualization tool does not intend to duplicate what is done by other priority 3 projects on web mapping or territorial monitoring. It is

only an internal functionality of ESPON database and not directly a tool for external diffusion of ESPON maps or political synthesis.

Objectives

A suitable web visualisation tool will be adapted and integrated into the ESPON database framework allowing the visualisation of grid and socio-economic ESPON OLAP data.

An open-source web application tool will be adapted and integrated into the ESPON Data Base framework to visualise, explore and analyse the data coming from the ESPON OLAP cubes.

Considering the different types of users accessing the data, we propose to test two specific environments: one for a basic user just for simple visualization, and a more advanced one with other capabilities, like the creation of interactive maps, graphs and tables.

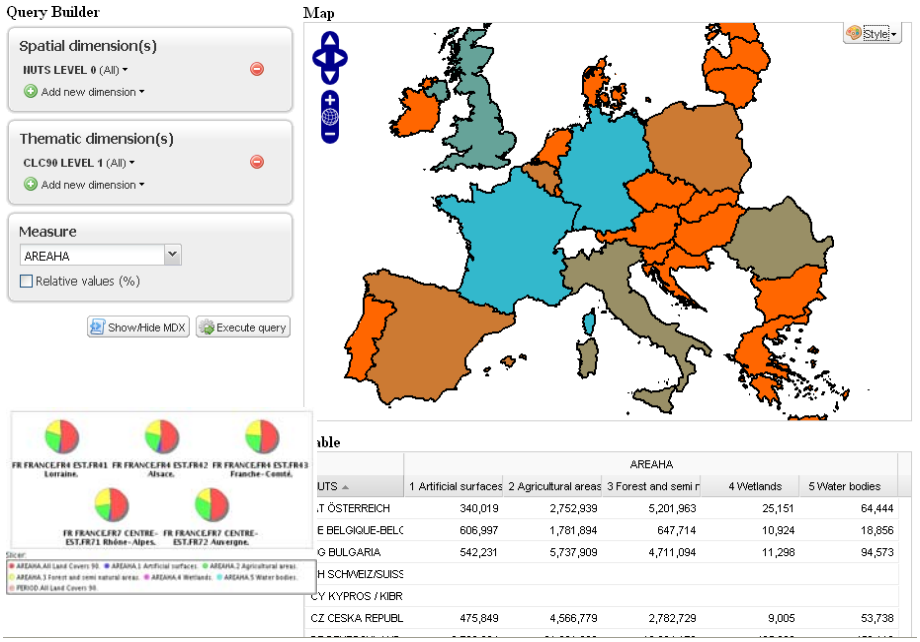


Figure 17: Draft view of the ESPON Visualization Webtool

The resulting application will be easily accessible through the ESPON Data Base platform or directly through the ESPON website if requested. Its main goals will include an intuitive user interface and the ability to perform customized analysis through interactive map, table and charts. It will be possible to export map and chart results to other formats, such as PDF. Contacts with other projects of the ESPON Platform (Priority 3) are forecast at Denmark meeting, in order to integrate this tool in a coherent strategy of diffusion of results (ESPON Mapping tool, ESPON monitoring…).

- ***Deliveries – Timetable***

November 2011

- Presentation/demo of the advancement of the work: ESPON Seminar in Krakow

June 2012

- Draft version of the ESPON OLAP Visualisation tool

December 2012

- Update of the ESPON OLAP Visualisation tool

June 2013

- Final version of the ESPON OLAP Visualisation tool

3.1.2 WP A2 Statistics

3.1.2.1 A.2.1Time-series (NCG)

Analysis of the situation

The nature of the times series datasets in the Database have been outlined in section A2.1. There are also techniques for imputing values for time and space-time series where there are missing observations, or an observation that has been flagged as exceptional requires correction. Time series for a single entity may have missing or exceptional data – the series can be modelled and the model used as a basis for imputing a different value, and providing an estimate of the reliability of the imputation. Space-time series with fine scale spatial and temporal granularity are not common – employment counts for aggregates of LAU1 or LAU2 units for example. Modelling the space-time structure of the data and examining any residuals for exceptional cases is one possibility here. However, the development of techniques at this level might well be carried out in conjunction with the Partner from Priority 1 2 or 3 who is intending to use such data so that the best technique for the application can be chosen. One-size-fits-all approaches are not necessarily the best for such applications. We may draw on external expertise to assist with the development of such applications.

The visualisation of space-time structures is also an important component in the exploration and analysis of time or space-time series. The development of suitable visualisation tools will draw on external expertise, again to ensure that the Database users have the best tools at their disposal.

Objectives

The NCG team will be concentrating on 3.1.2.2 and 3.1.2.3 up to December 2011. From January 2012 we shall extend the techniques to deal with time series data. Much of the analytical framework to deal with this will be in place, as a result of the work on outlier handling and data quality; time series data is likely to include data at different levels of measurement (see 3.1.2.2 Objectives below). However, development work will need to be undertaken to handle the temporal element. They will also have to be incorporated into the platform, and suitable protocols created for their use – however, these are likely to be along the lines of protocols which will be developed in 3.1.2.3 for interaction with the CU and the project provider.

This work will focus in priority on time series of absolute indicators (count) in order to support the Core Database Strategy. The application to relative indicators (ratio) will be develop later as it is based on different statistical methods.

- ***Deliveries – timetable***

June 2012

- Technical report ; Data quality methods for time series data of absolute indicators
- Check of data related to the Core Database Strategy

3.1.2.2 A.2.2 Outlier handling (NCG)

Analysis of the situation

In Phase I a series of techniques was developed for flagging exceptional values in a range of test datasets of the type likely to appear in the Database. Test data was also made available which exhibited a variety of exceptions for the Data Quality team to test their software against. We may define two main types of errors: errors of logic and statistic exceptions. Logical errors might include impossible value such as negative employment rates or household counts; they might also arise from typing or scanning errors during data collection/input (incorrect NUTS codes, for example), or from poor metadata (where missing value codes used by the data supplier have not been identified. Statistical outliers may include unusually high employment rates; these might be 'real' in the sense that an area does have genuinely unusual characteristics or they might arise from data input errors (transposition of numbers when typing, for example).

The errors of logic require a set of mechanical filters: percentage data must lie in the range 0% to 100%. Data for adjacent columns in the dataset can be testing or local or global duplication; data may be locally misaligned with the NUTS codes in a given column, or swapped between columns, the wrong NUTS regions may have been used (1996 for 2006 data).

In Phase I a series of methods was developed which considered indicators both in univariate and multivariate senses, as well as techniques which took into account the spatial structure of the areal units used for the data collection and recording. These include exploratory techniques such as boxplot display, Hawkins' test, examination of residuals from regressions (including OLS, locally weighted in the attributes and geographically weighted variants), principal components analysis, locally and geographically weighted principal components analysis. There is no single 'best' technique – indeed for categorical data, univariate tabulation or bivariate contingency analysis may reveal unusual values or pairings in the data.

Objectives

In Phase II we will extend these techniques to handle other data types. A common classification of data is by level of measurement: ratio, interval, ordinal, nominal. Ratio data is that which has a well-defined zero and for which the ratio of two values has meaning. Interval data does not have a well-defined zero; temperature is often suggested as an example: in the Fahrenheit scale the ratio of the freezing point of water to its boiling points 212/32=6.625 whereas in the centigrade scale it is 100/0=∞. Rank values are an example of ordinal data – adjacent data values do not express notion of the distance between the data objects, but there is an ordering. Nominal data consist of unordered categorical values (dog, cat, mouse, bear…).

That these data are available for spatial units adds further complexity. Land use classes are an example of nominal data, as are binary values (where 1/0 => yes/no). Checking may potentially extend little further than establishing that no logically impossible values exist in the data. Data for spatial units may be present as a count or a proportion. Count data values should be zero or positive integers; however, if the underlying at-risk population varies with the size of the spatial units, or its density varies, then treating the count data without reference to this will lead to misleading inferences about unusual values in the data.

- ***Deliveries – timetable***

June 2012

- Technical report : Outlier detection techniques for spatially normalized values (completion),
- Technical report : Detection techniques for spatially un-normalized data, interval, ordinal and nominal (completion).

3.1.2.3 A.2.3 Data quality (NCG)

Analysis of the situation

The Database must contain data that is sound and reliable. The various techniques that were developed in Phase I have been coded in the R language. This allows rapid development and improvement of techniques, and also rapid including in any test or production database structures. It is also important to decide what steps to take to ensure good data quality.

A working approach is to take a selection of detection techniques appropriate for the data concerned (so we do apply techniques for ratio level data to nominal level data, for example). An observation is flagged if is unusual on any of the technique that have been applied. This allows us to build up the weight of evidence that an observation is likely to be exceptional. We can also examine the components of the weight of evidence for any observation to determine whether it is aspatially, spatially, temporally, or structurally exceptional.

Issues in any approach to the detection of unusual data are concerned with the presence of false negatives (an exceptional value is missed) or false positives (a value is flagged as exceptional when it is not). Setting the appropriate cut-off levels may involve some simulation exercises on samples of test data from the Database or the application of corrections for multiple testing (such as the Benjamini-Yekutieli correction for spatially dependent tests).

Having identified outliers, we can then decide how to handle them. We can leave the raw data, together with flags in the metadata for experienced users. We can attempt the imputation of a 'more reasonable' value (and 'reasonable' depends on the type of outlier we have found); there are a range of suitable techniques – again discussion with data users may help to inform decisions on which are 'good'. We can consult one or more experts with suitably domain specific knowledge. We can also discuss these problems with the data users and arrive at solutions which are generically 'best' for their applications.

It has been observed that outlier detection can reveal results that are unexpectedly interesting from the point of view of any researcher – this may assist in improving the nature of the research through insight that may be missed had the detection strategy not been implemented. For the prediction of missing data, outlier detection methods may also be used (geographically weighted principal components can be used to impute values at locations where data has not been collected or is missing).

Objectives

The outlier detection and reporting methodologies will be incorporated into the Platform (C2: 8.2 and 8.2 in Figure 13a, p57)). Procedures will be put in place to allow the outlier testing to be run on each data set as it arrives on the Platform and to ensure that the results of the quality check are fed back to the CU and the project provider.

- ***Deliveries – timetable***

June 2012

- Web function ; Operationalisation of data quality check on Platform and reporting procedures

3.1.3 WP A3 Mapkit tool (RIATE)

Analysis of the situation

As a general rule, maps are used to display geospatial data and to enhance statistical data to understand phenomena. In ESPON Program, there is a need to produce a lot of maps. They have to be harmonized and comparable. This is the role of the map kit tool. It follows 3 main objectives:

- Ensuring compatibility with the ESPON 2013 database application
- Enhancing information with maps to understand phenomena and also disseminate scientific results.
- Ensuring harmonization of maps.

In coherence with these objectives, a map kit has been built in the previous EPSON DB project. Actually, it was composed by a set of 6 map kits according to different geographical levels (ESPON Area, ESPON Area + western Balkans + Turkey, Zoom-in, Zoom-out, global and local). In every case, these different elements are available in Arcgis format (mxd + shapfiles), Quantum GIS (a user friendly Open Source Geographic Information System licensed under the GNU General Public License), and Philcarto which is a free software for thematic cartography.

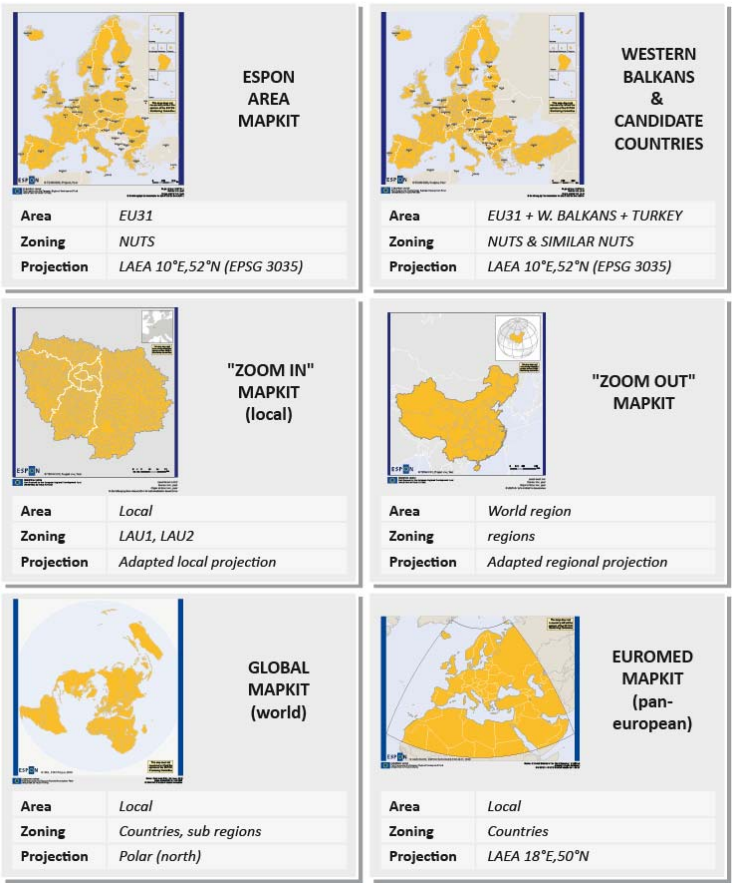


Figure 18 – Mapkits delivered during the ESPON DB1 Project

In addition, a technical report (the mapping guide) explaining how to use this map kit but also showing good and bad practices in Cartography was produced.

Objectives

In the M4D project, the approach is not to build entirely a new version of the mapkit produced in the phase I of the project, but to update it. Consequently, we will use the same projection (ETRS-LAEA system, EPSG 3035), same generalization of the delineations (Eurogeographics geometries) and same graphical aspect (colors, layouts). But, to fit with the developments of the content of the Database and also to the TPGs needs, improvements and updates of the mapkit will be done.

Fit more and more to TPGs needs: An important issue of the work is to fit the more possible to TPGs needs. For instance, the ESaTDOR project (European Seas and Territorial Development, Opportunities and Risks) which is working on Seas, need an updated mapkit covering specific areas: Arctic Ocean, Atlantic Ocean, Baltic Sea, Mediterranean Sea and North Sea. Consequently, a new mapkit with adapted delineations and projection will be added.Moreover, a new version of the Global MapKit will be probably done to include the updated version of WUTS delineation by TIGER project.

Improvement according to NUTS changes: One of the roles of the Mapkit is to ensure the possibility to display on a map current data with the official administrative delineations. But, NUTS delineations are changing all the time and the next change will arrive in January 2012. Consequently, one objective of the work is obviously to integrate the NUTS 2010 version in the mapkit ensuring interoperability with older delineations.

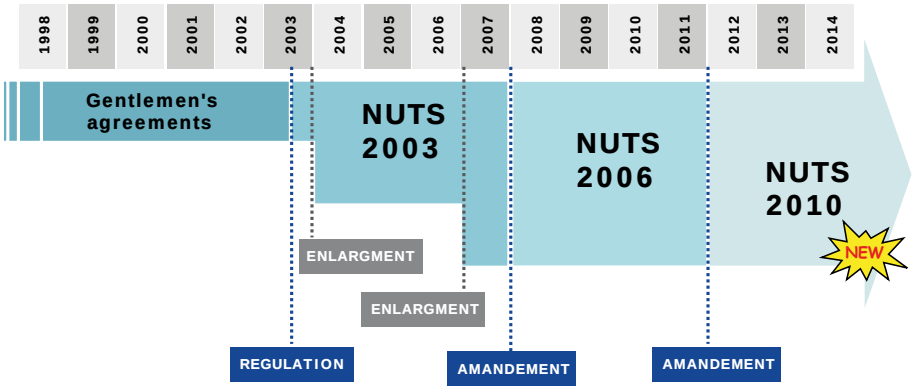


Figure 19 – Timetable of NUTS amendments

New mapkits, following the improvements of the database (neighbourhood, cities, grids, FUNCs):

One task consists also to follow improvements of the ESPON database and make available geometrical objects to all the EPSON projects. Consequently, the work done in the M4D project about neighbourhood will be, at the end, included in the MapKit. An adapted projection and an efficient level of geometrical generalization (adapted to cartographical issues) will be defined. As the same than for neighbourhood data, new MapKits will be created for fitting with non NUTS data (grids, cities, FUNCs). However, it is impossible to predict by advance the timetable of such deliveries considering that the development of such data depends not only on the M4D Project.

Interoperability: During the entire project, a background task will consists to try to enhance the more possible, the possibility to cross different kind of geographical objects (vector, raster, urban). To do that, several ideas will be explored. For instance, we will explore the possibility to store geometries in an online database (Postgis) which is accessible directly from a GIS.

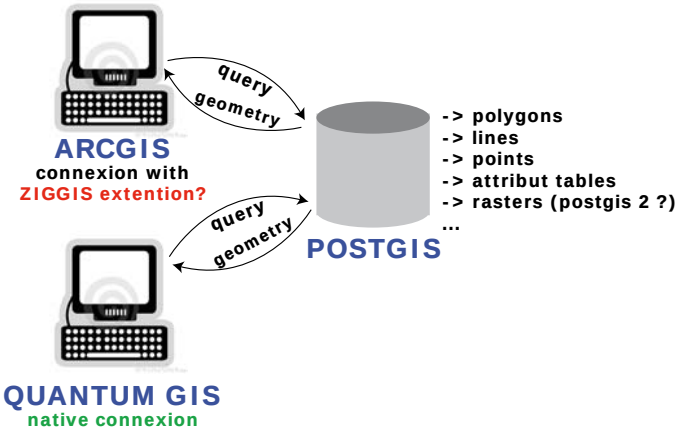


Figure 20 – Proposal for ease the diffusion of mapkits

Deliveries-Timetable

As a general rule, updates of the mapkit will be delivered before each ESPON Seminar.

December 2011

- Provisional mapkit on Seas (ESaTDOR), Inclusion of NUTS 2010 delineations (provisional).

June 2012

- Stabilized mapkit on Seas (ESaTDOR), Inclusion of NUTS 2010 delineations (final)
- Update according to M4D improvement and TPGs needs
- Provisional mapKit on ESPON Area and Candidate Countries + European Neighbour Regions (including Bosnia-Herzegovina, Albania, Serbia, Morocco, Algeria and Tunisia).

December 2012

- Update according to M4D improvement and TPGs needs
- Update of the mapping guide; update of the MapKit "ESPON Area and European Neighbour Regions" (including Russia, Belarus, Lybia and Egypt)

June 2013

- Update and stabilization mapkit for the ESPON Area and European Neighbour Regions (including Armenia, Georgia, Syria, Irak, Lebanon, Israel, Occupied Palestinian Territories, Azerbaijan and Jordan

Before each ESPON Seminar

- Update according to M4D improvement, update according to TPGs needs

3.1.4 WP A4 Updates (LIG)

In order to ensure an efficient data and metadata flow through the next ESPON database, two kinds of updates will be undertaken:

First, once data and metadata sets have been entered into the ESPON database through the Upload interface, each partner assigned to the checking of the corresponding type of geographical object will validate or invalidate them. Then, once the data and metadata sets are valid, they can be stored in the database, which means that the content of the ESPON database has to be updated on the ESPON Web site. The frequency of these content updates has to be decided with ESPON CU as it depends on the inflow of data and metadata supplied by the ongoing ESPON Projects.
The integration workflow to insert new datasets into the database has already been discussed in the previous "Database" section, please see also the "Platform" section for further information about the multiple steps of this complex integration workflow.

Second, periodic updates of the ESPON database application (including new developments) will be delivered to the ESPON CU and, after agreement, deployed on the ESPON Web site.

Deliverables

The timetable of this task has been described in the previous "Interface" section.

3.2 Working Package B - Thematic

3.2.1 B.1 Regions (RIATE)

Analysis of the situation

Indicators available in the NUTS codification remain a key element for European studies on spatial planning and measuring policy impact assessment. Concretely, when analyzing the data flow within the period 2007-2010 of the ESPON Program, it is clear that more than 75 % of the data are delivered in the NUTS delineation. As it is a contractual obligation for ESPON projects to use official and most recent versions of NUTS2 or NUTS3 levels, a high involvement of the ESPON M4D project is needed. Exchanges with TPG working on regional data at European level give some conclusions on the needs:

1. **Need of a standard "core package" of regional data at the beginning the project**: Although the regions (NUTS) are the most common level for the collection of statistical data and the level of action for regional policy, the datasets are often incomplete. These gaps are due to changes of geometry of statistical units and the lack of data. The fact that some regional data are missing for very specialized indicator like "*Regional consumption of bretzel*" is acceptable. But what is definitively not possible is to have holes for core information like raw count of population, area and added value. And further, the subdivision of this count variable by main categories like age or sex for population, land use type for area, economic branchs for added value. This variable are not only the most common one, they are also the basis for the construction of the majority of the future specialized indicators produced by ESPON projects where they provide the denominator of the ratio and, more generally the weighting criteria for statistical procedures. In our fictive example of bretzel's consumption, the lead partner of the project will obviously need to compute the ratio of number of bretzel per inhabitant (and need therefore population) or share of bretzel in economic activity (and need therefore GDP). If the lead partner did not received on time the reference count data, he will collect himself the value and introduce in the database a new variable of population which will not necessarily be the same than the one collected before, because updates of EUROSTAT will be made at a different period of time or because his procedure to estimate missing values will be different. This situation has occurred many times during the period 2008-2011 and we have therefore delivered to ESPON CU in January 2012 a technical report called **"Core Database Strategy"** that propose general solutions to this difficulty. This solution has been discussed during a meeting with ESPON CU in Luxembourg the 16th January and, if it is validated, we propose to engage the TPG to adopt it and present results obtained at the Denmark Meeting in June 2012
2. **Explore new hierarchies of regional division for strategic planning purposes**: As explained in the section 1 of this report ESPON is currently in an in-between situation between an official producer of data like Eurostat and an international expert organization like OECD. This ambiguity can be very creative if ESPON is able to implement a solution that facilitates the delivery of the same initial data collection in several types of regional divisions. ESPON should certainly not propose an "ideal" system of regional division from scratch but rather explore "pragmatic" combinations of official regions, like the NUTS2/NUTS3 that appeared as a very relevant level for ESPON community users, with equal agreement of scientists and policy makers. What is at stake is to enlarge this experience of NUTS2/NUTS3 and to build an original hierarchy of territorial unit. This new hierarchy would not replace the current official definition of Eurostat (NUTS) that are 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 work on territorial agenda and cohesion. More details of this

project are described in annex of this report, including the reactions of some ESPON experts (K. Spiekermann, V. Schmidt-Seiwert, M. Lennert) which consider that the subject is interesting but should not be launched by ESPON in an isolated perspective. Considering these different reactions and the immediate pressure of other tasks, we propose to delay the actions expected for this task until the beginning of 2013. The year 2012 should be better employed to improvement of ESPON Interface and ESPON Platform and launching of Core Database Strategy.

Objectives

Time and resources being limited, we introduce a clear rank of priority in the two main tasks:

1. **Absolute priority to the collection of core data** until end of 2012
2. After 2012 only, exploration of new nomenclatures (FUNC) in three steps : (i) Establish contact with interested stakeholder and prepare a review of literature and experience (ii)Realize small experiments on limited area but at various scales (iii)Wait for confirmation of ESPON CU and ESPON MA before to engage more active action.

- ***Deliveries – Timetable***

December 2011

- Discussion paper on the project of creation of new functional division (Annex to first version of Inception report) and provisional conclusion (second version of inception report).
- Technical report on the project of "Core Database Strategy"[14].

June 2012

- Technical report on preliminary results of the Core Database Strategy.
- Basic indicators collected with principle of Core Database Strategy in NUTS 2006 Division.

December 2012

- Basic indicators delivery collected with principle of Core Database Strategy in NUTS 2010 Division

June 2013

- Technical report on the empirical feasibility of revised functional divisions at different scales.
- Basic indicators delivery collected with principle of Core Database Strategy for cities and other delineations (water basin, grids, …)
- Final report on Core Data Base Strategy

3.2.2 WP B2 Cities (Géographie-cités)

Analysis of the situation

ESPON DB1 has focused on different European urban databases, such as Larger Urban Zones (Urban Audit), Urban Morphological Zones (European Environment Agency), Morphological Urban Areas (IGEAT, Belgium) and more recently Functional Urban Areas (IGEAT, Belgium). Two other databases have been created by Urban Audit (City Core and Sub-Urban Districts), and some other projects are currently on going for defining new delineations.

[14] These two documents has been yet delivered to ESPON CU.

The objective of M4D is to propose an integrated perspective for the management of the different urban delineation inside ESPON Database and to choose a sustainable perspective making possible to anticipate new delineations that could be proposed in the future in order to improve or update the current ones.

The most innovative key-point of our proposal is to introduce the possibility to propose two alternative approaches that are (1) to use the same delimitation of cities through time or (2) to describe the same urban object but with changing delimitation through time (see Figure 25)

Objectives

- **Integration of urban databases in ESPON DB**

There is a growing need for urban indicators that would be associated to a variety of city definitions and harmonised from a spatial and a temporal point of view. However, choosing the most appropriate urban data base according to the aim of a study or to a specific approach seems to be more and more complicated for ESPON DB users. Since a few years, different new urban data bases have been created, using morphological agglomeration delineations, or functional area delineations, or other definition criteria. These databases depend on the degree of completeness in metadata and on the availability of the data. The aim of ESPON Data Base project is not to select the "best" urban database, but to try to enlighten the specificities of each one, by working on the metadata and by comparing their construction methods and sources (LAU, NUTs, grids, etc.) in order to facilitate the choice and the use of one of them.

Three steps are involved, from the conceptual basement (comparison of European urban databases and construction of a metadata model) to the data integration process in the ESPON DB platform.

- Step 1: Metadata ontology (Géographie-cités + Expert Data Model)

Time-Schedule: December 2011
The sub-task consists in constructing a metadata model of the selected databases, starting from a general reflection on the different approaches for harmonizing urban data in Europe (Figure 21) and going into more details through the particular specifications of each database.

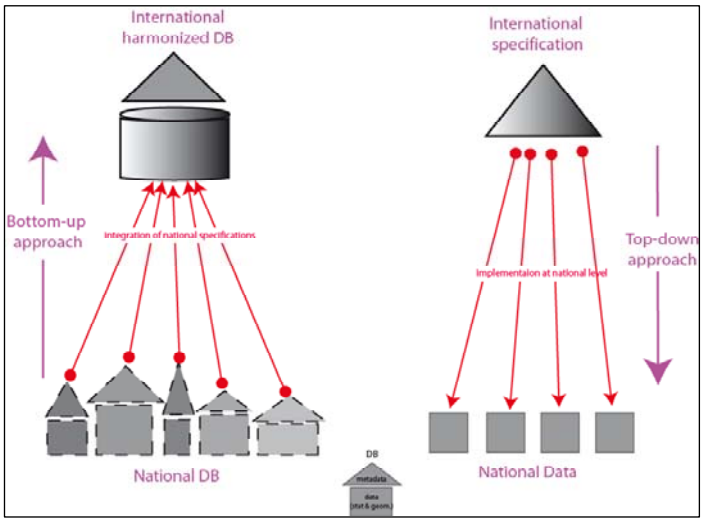


Figure 21: International urban databases and harmonization, two basic approaches

- Step 2 : Database integration (Géographie-cités + Expert Data Model)

The integration of the selected urban databases is based on two different sub-tasks. First, creating interoperability with the other objects of the ESPON Database (LAU, grids). Secondly, creating a zoning correspondence table between the urban databases.

- Step 3: Populating the selected databases (Géographie-cités, UAB, Tigris, Riate and other)

This sub-task consists in an enlargement of urban indicators, towards environmental (grid/raster) and socio-economic data (using, in particular, the "OLAP Cube for Urban analysis" developed by UAB, and indicators collected at LAU2 level by Tigris and aggregated in urban delineations). This task may also concern indicators and data gathered and developed by ESPON projects if they are compatible with our data dictionary.

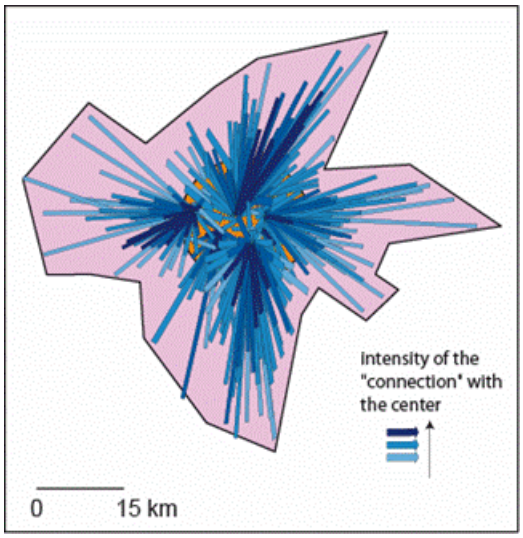
- **Task 2: Expertise on FUA construction methods**

Objectives

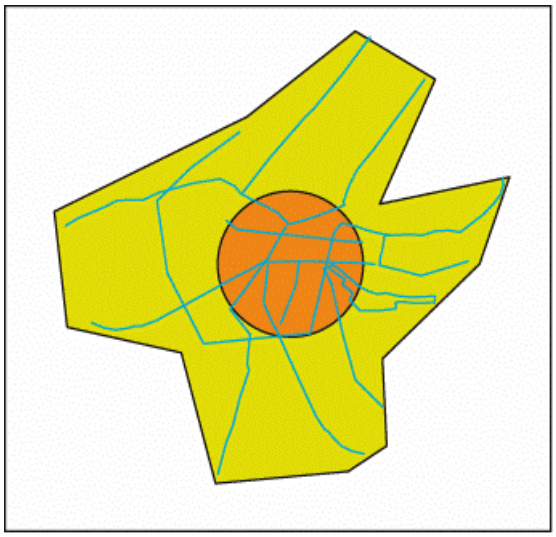
The question of international harmonization is of course fundamental for each European database, but particularly for the construction of functional objects such as FUAs. However, it is a very complex question and the FUAs that have been constructed until now (FUAs from ESPON 1.1.1, FUAs from ESPON 1.4.3., FUAS from IGEAT, LUZ from Urban Audit) don't fully answer it, even if some of them suggest interesting solutions. Other project try to build FUAs from a different approach, based on accessibility isochrones, i.e. from a theoretical model and not empirical data15. The aim of ESPON DB2 is to expertise these two different approaches and find out what processes would be the most relevant for ensuring, or at least improving, international harmonization at the LAU2 level. The question that needs to be answered is: Is it possible and consistent to implement the same methodology in each country, taking into account the variability of data, the heterogeneity of the settlement contexts and the differences in the resolution of LAU2?

In order to answer this question, we propose to build and compare two types of functional areas in different sample zones. The first consists in observed or empirical FUAs. They are based on a model, a certain representation of the city as a system of relationship based on the logics of polarisation, relying here on polarisation by employment and commuter data. The second type of FUAs consists in "potential" FUAs. They are based on a theoretical model coupled with local transportation networks and accessibility measures. These two approaches integrate inputs (commuting flows, transportation networks…) and parameters (levels of attractivity for commuting, transportation time…). The inputs and parameters will take into account the variability of data, settlement contexts or resolutions between countries, and the major challenge is to try to build a unique methodology that integrate a diversity of parameters and that could be applied for all the countries (Figure 22).

[15] Federal Institute for Research on Building, Urban Affairs and Spatial Development (2010), Metropolitan Areas in Europe. Abstract of a new BBSR study. *BBSR-Berichte Kompakt* 7/2010. Bretagnolle A., Giraud T., Mathian H. (2008), Measuring urbanization in United States, from the first trading post to the *Metropolitan Areas* (1790-2000). *Cybergeo*, 427, http://cybergeo.revues.org/index19683.html



FUAs based on connection and a selection of major flows (empirical approach)



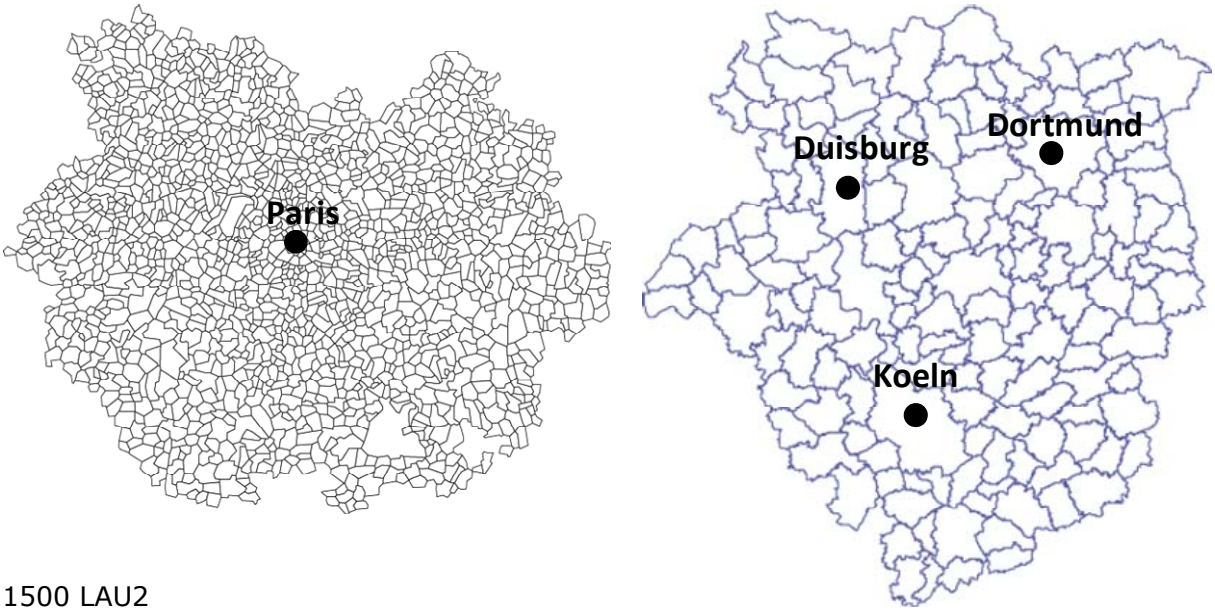
FUAs based on transportation network and a one hour isochron (theoretical approach)

Figure 22: Crossing two different approaches for constructing FUAs

- Step 1: Data collection for the selected sample zones (Géographie-cités + Expert Transport)

Collection of available data (transportation, commuters, LAU2, etc.). The sample of countries and cities will be representative of differences in settlement context (dense or sparse urbanization, conurbations or large monocentric cities, major or low urban sprawl…) and/or characterised by a large number of data. Data related to administrative units are also very important to consider, on regards to the MAUP (Modifiable Areal Unit Problem). For instance (Figure 23), a regional comparison between France (Paris area) and Germany (Rhein-Ruhr cities) shows that whereas the surface of these two regions is quite the same, the French LAU2 (communes) are 10 times smaller than the German LAU2 (Gemeinde). Moreover, we observe a local heterogeneity of administrative units in Rhein-Ruhr (around city cores, NUTS3 instead of LAU2)16. This question of spatial resolution is of high importance, as the structures of flows strongly depend on the spatial resolution of statistical units.
In a first phase, we will choose a few « core » samples from a first expertize: Barcelona, Paris, Rhein-Ruhr (around Koeln-Bonn, Dusseldorf, Dortmund), Roma, and Budapest.
Then, in a second phase, we will choose a larger sample (about 10 cities) to be defined from the conclusions of Expert1 (Transport) and regarding the following criteria:
- Diversity and usability of data (metadata, completeness degree, updates…)
- Complementarity of case studies: monocentric/polycentric pattern, urbanization context (ie densities), sprawl temporality (West./Eastern Europe).

[16] Le Nechet F. (2010), Public Transport and shape of European cities. PhD, University Paris-Est, Laboratoire Villes et Mobilité and Géographie-cités.



1500 LAU2

154 LAU2

Figure 23: Heterogeneity of spatial resolution, France / Germany (Source: F. Le Nechet, 2010)

- Step 2 : Data model, tests and calibration (Géographie-cités + Expert Data Model + Expert Transport)

- Test and calibrations (FUAs based on commuter data): definition of the cores of FUAs, using UMZ and minimal thresholds of employees or population.
- Tests and exploration of the question of international variability in resolution, which affects the measure of commuter flows17.
- Tests and exploration of the minimal thresholds associated to different configurations18 (mono-polarization and multi-polarization). The choice of one threshold rather than another cannot be justified by empirical observation, as there is no break or even slight discontinuity in the statistical distribution of the dependency ratio values that would be in favour of a particular threshold, as illustrated in the example of the western part of the Parisian suburb (Figure 24).
- Construction and validation of a data model integrating network and flows data (FUAs based on accessibility).
- These tests, especially for the eastern urban areas, will also allow the validation of the FUAs database of the previous ESPON database project. (FUA-IGEAT) and its integration if the ESPON DB. Indeed this database has been elaborated using a sole dependency ratio for the whole Europe (10% commuters)."

[17] This work will start from results obtained in Laboratory research Géographie-cités (Sandrine Berroir, Hélène Mathian, Thérèse Saint-Julien, Lena Sanders) on the metropolitan region of Paris. A first illustration shows that when a pole (Marne La Vallée for example) is considered as a set of LAU2 (26 entities), the delineation of the attraction area differs considerably from the one obtained with a unique pole (one sole entity), even with the same dependence ratio. The polarised area is about twice larger in the second case.

[18] Also coming from the study quoted above, the observation of commuter data enlightens a great sensibility to the threshold of the "dependency ratio" parameter (Select Flow(i,j) where Flow(i,j)/ActivePopulation(i) > K %). Low differences in the threshold value bring completely different spatial extensions of the attraction basin. We can also obtain isolated units (not in continuity with the main area of influence), which raises the problem of how to delineate the attracted area.

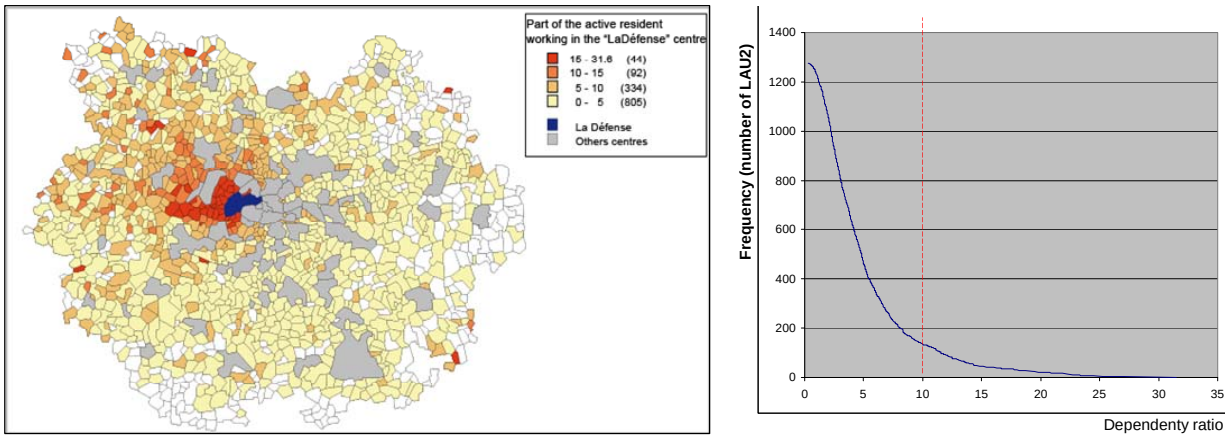


Figure 24: The absence of break in the statistical data around such or such threshold…An illustration with the case of La Defense (Western suburb of Paris, France)

- Step 3: Construction of FUAs for selected cities (Géographie-cités + Expert Data Model + Expert Transport)

Building FUAs from commuter based method for the selected cities, using an automatic process.

Building the one-hour time budget delineations for the selected cities, through computation of multimodal accessibility and estimation of the delineation associated to the one-hour time-budget (FUAs based on accessibility).

Comparison of the two approaches (accessibility/commuter data).

- **Task 3: Time-series issues on urban data**

Objectives

European urban databases are generally based on static delineations and/or static indicators. However, working on an anticipation of the evolution is important, not only for the updating of the indicators through time, but also for the updating of the delineations or of the criteria used to define urban objects (thresholds in minimal populations, minimal density, in commuter levels…).

- Data Model (Géographie-cités + Expert Data Model)

Concerning in the conceptualization of the data model, several approaches could be relevant for modelizing the temporal change in urban data (Figure 25): the top image corresponds to the situation where urban indicators can be followed through without any artificial jump due to a fusion of urban objects, whereas the bottom image displays the case where the spatial reality of urban objects can be followed through at each time step, without overestimating prior extensions. Other approaches can be followed (See ESPON DB1, Luxembourg meeting on time-series issues, May 2010). According to the urban databases under the study, the expertise will select the most appropriate data model for time-series issues.

Concerning the implementation work, an evaluation of the possible ways for identifying the different types of change will be done.

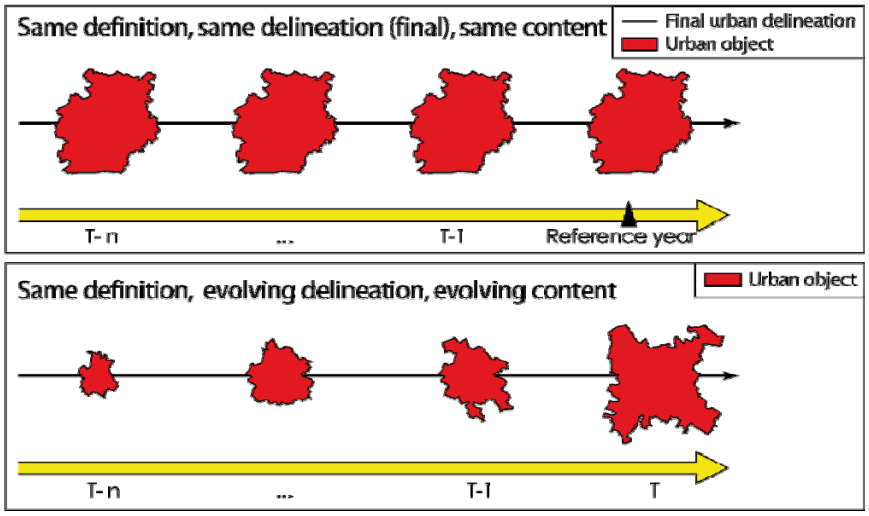


Figure 25: Two complementary urban data modelling schemes for a temporal following urban databases

- ***Deliveries – Timetable***

December 2011

- Urban databases – metadata ontology

June 2012

- Urban databases – database integration (interoperability)
- FUA – Data collection for the selected sample zones

December 2012

- Urban databases – database integration (table of correspondance)
- FUA – Data model, tests and calibration

June 2013

- Populating the selected urban databases
- Construction of FUAs for selected cities
- Time series on urban data – data model

3.2.3 WP B3 Grids (UAB)

Analysis of the situation

The **seamless integration of social and environmental dimensions** by combining several types of datasets, both continuous and discrete, is an essential element in the process of building sound and reliable indicators as the basis for European research projects and policy-making at a European level.

Most of the environmental data, such as the air quality or land cover, do not follow the administrative boundaries, but natural units or regular grid cells.

Under the ESPON M4D project, we propose to continue the work undertaken in ESPON 2013 Database, about defining and applying the most suitable methodology to make possible the integration of these different geographical objects. Thanks to the follow-up of the last European approaches to this specific topic it will be possible, on one hand, to continue with the disaggregation methodologies implemented so far, and, on the other hand, to introduce innovative improvements to define the best solutions.

The **needs of the ESPON community** will be crucial in order to define and provide the most useful indicators in terms of territorial development and planning. The integration of different geographical objects at different scales will be crucial in order to study and understand specific European trends like, for example, urban sprawl.

In the framework of the Core Database Strategy that we have established with agreement of our project officer in December 2011, grid data are playing a major role because they offer a potential solution for the exchange of data between geographical objects of different types (urban, LAU, NUTS of different versions, water basin, …).

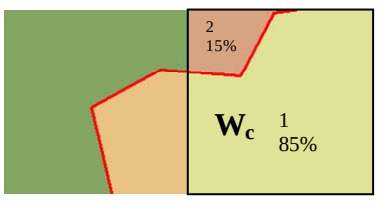
Objectives

Considering the results obtained during the ESPON 2013 Database in reference to the integration of socio-economic and environmental data19, we propose to continue the same line of work by integrating these two different geographical objects by redistributing them by a regular grid, so it was proved the best solution in order to downscale information reported by administrative areas.

The **1 km2 European Reference Grid**20 will continue being used, as it was seen as a very good option due to its European coverage, its compliance with the INSPIRE specifications, and its optimal resolution in order not to lose data precision. In this second phase, we also propose to use other grids at a higher resolution for some specific purposes, such as urban sprawl analysis.

Following the **methodology** defined by the UAB in the ESPON 2013 Database project, the first methodological step to be carried out should be the creation of a new layer by means of the intersection between the 1 km European Reference Grid and the administrative units by which the data is given.

Once the intersection has been computed, the third method "Proportional and Weighted Calculation" will be applied:



Cell value = $W_c \Sigma (V_i * Share_i)$
Where: V_i = Value of unit i
$Share_i$ = Share of unit i within the cell
W_c = weight assigned to cell c
In the example: $W_c * (V_1 * 0.85 + V_2 * 0.15)$

[19] "Disaggregation of socioeconomic data into a regular grid: Results of the methodology testing phase" available at http://www.espon.eu/export/sites/default/Documents/Projects/ScientificPlatform/ESPONDatabase2013/Technical_Reports.pdf

[20] http://www.eea.europa.eu/data-and-maps/data/eea-reference-grids

This method was proven to give the best results, plus some benefit to the downscaling, thus improving the territorial distribution of the socioeconomic indicator. In this second phase, new weighting variables, besides population density, will be tested depending on data availability. Some of these new variables could be, for instance, nighttime light imagery or accessibility matrix.

New and better datasets for population density should be obtained and tested, such as results from the GEOSTAT project[21].

We propose the use of the **OLAP22 technology** and the development of web tools, which will allow the different types of users (policy makers, practitioners, stakeholders and researchers) to use, discover and work with the integrated data (environmental and socio-economic data) in a high user-friendly interface.

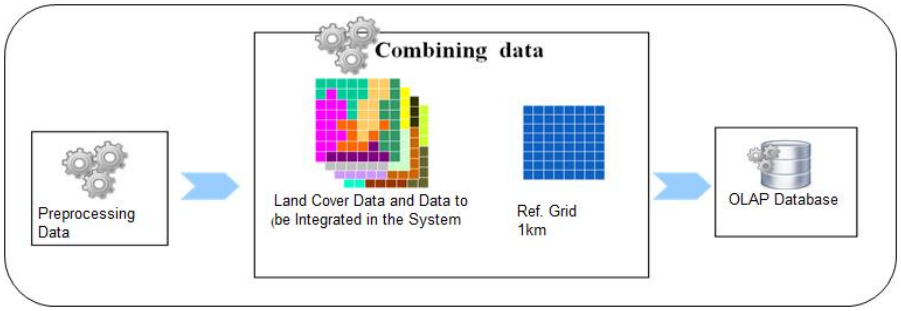


Figure 26: Simplified schema of the OLAP Database built-up

A very first step will consist on a survey carried out amongst all the ESPON projects, in order to explain them well the possibilities of the integration of data into the OLAP Cube, and ask them which data they are using and which data they would like to have integrated within the ESPON OLAP Cube(s) so that they could benefit from combining data with other data sources to produce sound and relevant results for their assessments.

We will follow up the outcomes of the **European Forum for Geostatistics** in order to contrast the proposed method, and, eventually, improve it. Furthermore, we should provide a clear picture of the deviations caused by the methodology with respect to the source data.

Finally, several ESPON OLAP Cubes will be built up in order to integrate the gridded data and make them ready to be queried and analysed together with other types of data, like land cover, natural units, protected areas, and so on. A user-friendly visualisation tool will integrate the ESPON OLAP Cubes and all their potential, allowing the production of maps, tables and charts.

The ESPON OLAP Cube(s) will be updated as soon as new data are being available (e.g. NUTS 2010, new Eurostat statistics...) and new data from the ESPON projects which showed interest in having their data combined are being released.

❖ ***Deliveries – Timetable***

December 2011

- Survey about the ESPON OLAP Cube use carried out
- Delivery of an updated ESPON OLAP Cube (v.4) and updated ESPON OLAP Cube Tutorial: ESPON Seminar in Krakow

Before each ESPON Seminar

- Update foreseen according to M4D improvements

[21] http://www.efgs.info/data
[22] http://en.wikipedia.org/wiki/Online_analytical_processing

3.2.4 WP B4 Local data (TIGRIS)

Analysis of the situation

Different studies and our experience confirm that the main difficulties in the manipulation of the local data and indicators come from the fluidity of their geometry and from the heterogeneity of the information sources. At the request of ESPON CU, the TIGRIS team has proposed a strategic document (Annex 1) that will be improved and completed for June 2012, in particular after a discussion with EUROSTAT services in charge of SIRE and GISCIO database. We summarize just briefly the main conclusion here, and let the reader move to Annex 1 for more details.

It is important to notice that LAU data are a key-component of the Core Database Strategy because they offer precise data that can be further transferred to grid and use for the building of new indicators at urban or regional level.

Objectives

The main activities that the TIGRIS team will implement in the time frame of this project refer to the scientific exploration of new relevant indicators at local scale, to the data and indicator collection and to the networking activities. As the main philosophy of the M4D project focus on the quality of the indicators to be offered, our strategy will follow this line of action and it will insist on the reliability of the local data to be integrated.

1. The first component in the scientific exploration of data at local scale will **put into relation the VIGO (the very important geographical objects) with the LAU2 frame and data**. (see. Annex 1 for more details) .

2. The second component of the scientific exploration of data at local scale will be based on the **construction of an alternative geometry,** relevant for the local scale and able to bypass the issues of the mass effect (differences of surface). (see. Annex 1 for more details)

3. **Integrating data concerning the land use patterns** at local scale will be the continuation of the process started in the previous database project. For the moment, using the LAU2 geometry and the Corine Land Cover data, only 5 countries from the Eastern Europe present indicators. We will enlarge the collection of this data for other ESPON countries, aiming to cover relevant transnational regions with specific land use patterns. (see. Annex 1 for more details)

❖ ***Deliveries – Timetable***

December 2011

- Methodological and conceptual exploration of the objectives proposed in relation with this task (draft version in Annex of this inception report).

June 2012

- Technical report: strategy for creation of data and geometries at local level.
- Data: Area and share of the artificial surfaces => location quotient of the artificial surfaces. This indicator will provide information on the degree of human pressure on the landscapes.
- Data: Area and share of arable surfaces => location quotient of the arable surfaces. This indicator will provide information related to the organization of the agricultural system at local level.
- Data The road density at LAU2 scale. It can offer indirect information related to the degree of local connectivity; it is sensible to the LAU2 surface. The map will

eventually depict areas with strong territorial endowment, from a transportation point of view.

December 2012

- Data: Area and share of permanent cultures =>location quotient of permanent cultures. This indicator describes local territories with a long stability of the agricultural systems.
- Data: Typology of LAU2 based on the proximity to major rail-road transportation corridors.
- Average distance towards the closest 3 universities

June 2013

- Data: Distance to the closest FUA. As the FUA are classified in function of their territorial importance, measuring the distances to different categories of FUA is the base for a more complex typology of the LAU2 units.
- Data: Other indicators and products provided during the work on local data.
- Proposal of alternative functional geometry at local level

3.2.5 WP B5 Neighborhood (RIATE-Géographie-cités-UAB)

3.2.5.1 Regions (RIATE)

Analysis of the situation

Relation between Europe and its neighbourhood become a strong thematic axis of the policy of the European Commission. The policy was first outlined in March 2003 in Commission Communication on Wider Europe23 and followed by a Strategy Paper on European neighbourhood policy in May 2004.

The European Neighbouring Policy was launched in December 2006 and has as a main objective to improve relationships - in term of governance, market economy and sustainable development namely - between Europe and its Neighbouring countries, in the Eastern part of Europe and in the South of Mediterranean as well.

However, improving relations in political term imply to have common information to better understand differences and similarities between the European Union and the Neighbouring countries. At the moment, the only information concerning this study area is available at State level (databases from United Nations and papers published by the European Neighbouring Policy) which may be useful for policy makers but certainly insufficient for building key elements for a sustainable cooperation.

Some impressive attempts have already made, such as figures presented in the green paper on territorial cohesion[24] (figure 27). This interesting demarche taking into account of the Euro-Mediterranean Area should be extended: Land use, through the use of grids, information concerning basic indicators at regional level, better knowledge on the structure and dynamics of cities in the Eastern and Southern of Europe, are a first domain of interest which have to be explored.

In ESPON DB1, a first attempt of integration of data on Western Balkans and Turkey has been realised. For these countries, the project has proposed an innovative hierarchy (the SNUTS for "Similar to NUTS") in order to make comparable the SNUTS and the NUTS units. However, the project has collected few data on this area.

[23] " http://ec.europa.eu/world/enp/pdf/com03_104_en.pdf
[24] http://ec.europa.eu/regional_policy/consultation/terco/index_en.htm

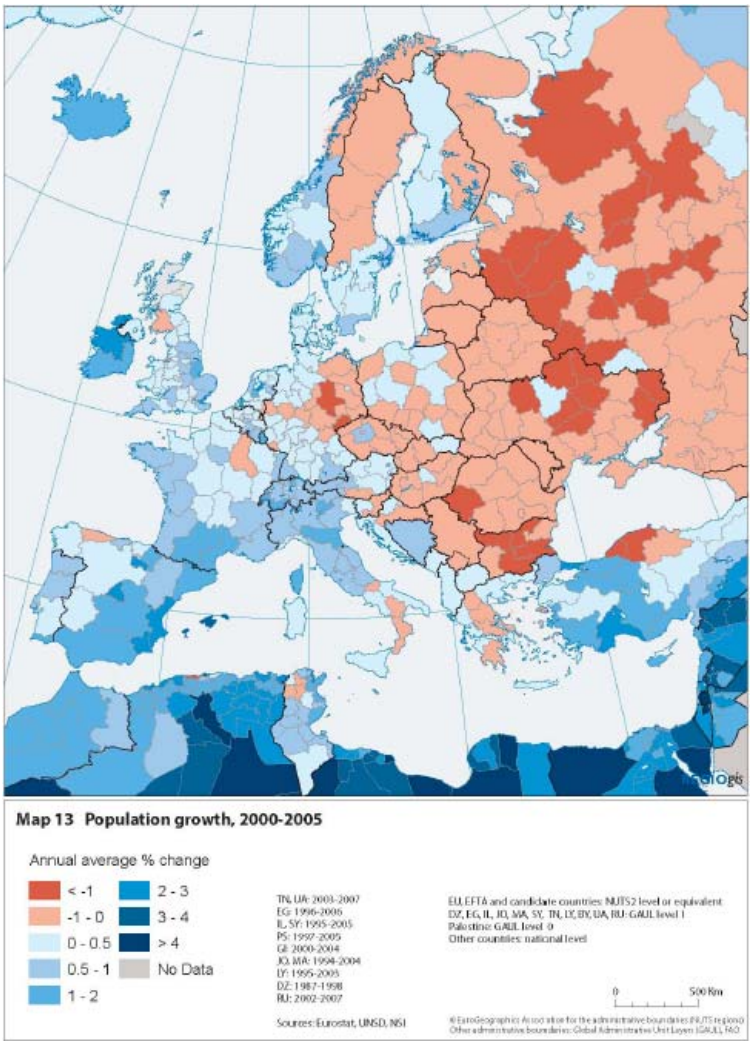


Figure 27: Evolution of population 2000-2005 in European Union and its peripheries (annexes of the Green Paper on Territorial Cohesion), 2009

As regard to priority in data collection, we will focus on the time series of basic count variables selected in the document called Core Database Strategy delivered in December 2012 to ESPON CU. We consider indeed that this focus will help to cover a largest area around EU territories with similar indicators. And also provide basic information for more specialized project working on the neighbourhood or external cross-border areas.

Objectives

The aim of the WP B.5.1 is to provide information of interest for ESPON Project and policy makers at regional level (SNUTS). We consider that it includes the collection of:

- **Seamless geometries** at different degree of map generalization
- **Statistical** information, compatible with the geometries
- **Documentation** allowing to understand and to interpret statistical information.

ESPON 3.4.1 Project has defined a specific area for analysing Europe and its neighbourhood, which is called "Euromed". It includes 74 countries, in other terms 33 countries in the ESPON neighbourhood. Collecting systematically regional data in these

33 countries is a very long term process; this is the reason why we propose firstly to prioritize the data collection (figure 28):

- **Targeted area - priority 1**: It corresponds to countries which share a boundary with the ESPON Area (Morocco, Russia, Belarus, Ukraine and Moldova) or are located around the Southern cost of the Mediterranean area (Algeria, Tunisia, Libya and Egypt). Basically, it is where the relation between the ESPON Area and its neighbourhood are the strongest in mass.
- **Targeted area - priority 2:** We have decided to put here countries which share a boundary with Candidate Countries (Armenia, Georgia, Syria, Irak) or are located in the Eastern Cost of the Mediterranean area (Lebanon, Israel, Occupied Palestinian Territories). We have also added in this category Azerbaijan and Jordan)
- **Targeted area - priority 3:** It includes countries where immediate links with the ESPON Area are relatively less important: countries located in the South of Sahara desert (Western Sahara, Mauritania, Niger, Chad, Sudan) or at great geographical distance of the ESPON Area (Saudi Arabia, Yemen, Oman, United Arab Emirates, Bahrain, Kuwait, Iran, Kazakhstan)
- We exclude of the objective of this work the **ESPON Area and Western Balkans & Turkey** are not covered by this Working package (respectively WP B1 & B.5.4)

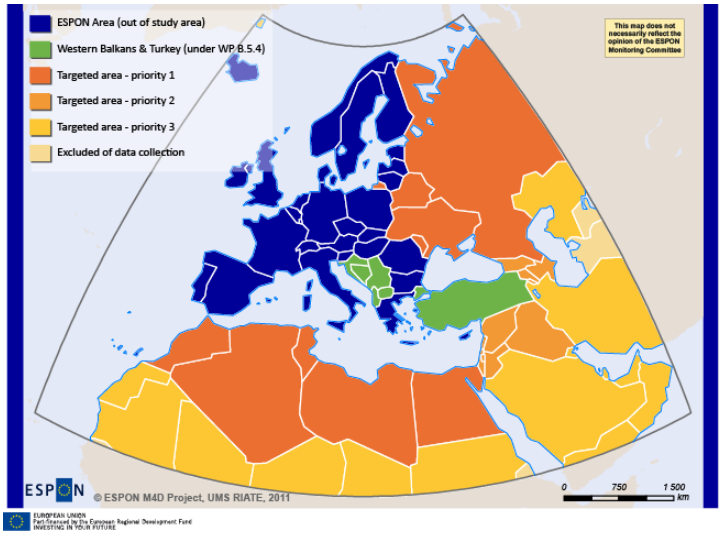


Figure 28 - Data collection in neighbouring countries targeted areas

Taking into account of the priorities of data collection, the building of the Euromed database will be organised in four steps (figure 29)

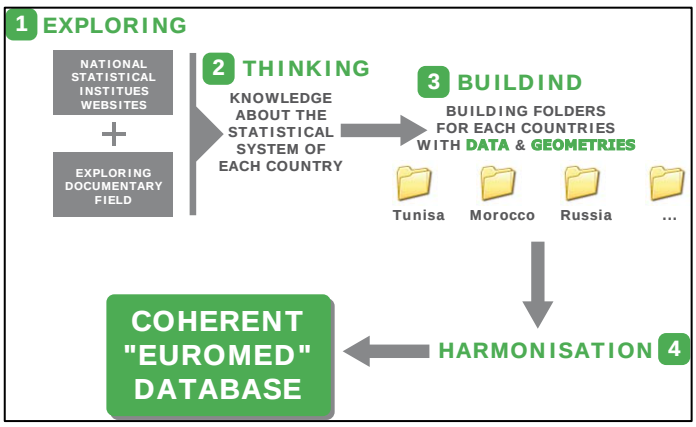
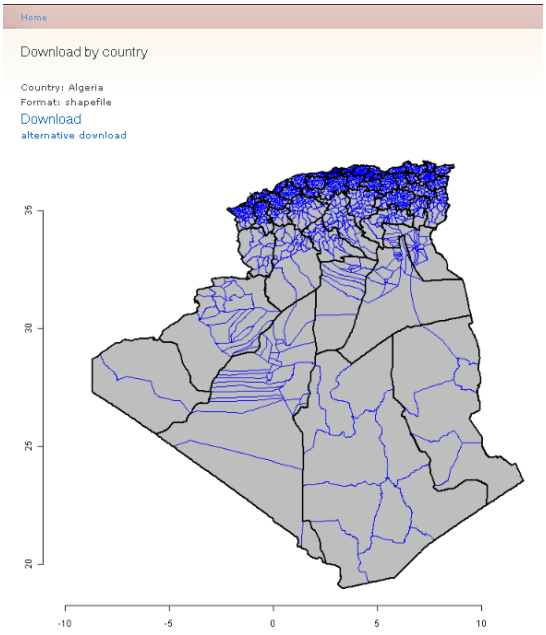


Figure 29 - Data collection strategy in Neighboring countries

➢ Step 1: Exploring

We will set up a two approaches strategy for exploring available material in neighbouring countries:

- **Top down approach**: It implies to use existing harmonised resources. According to our first researches, it concerns mainly **maps and territorial history** of countries of the World. In this perspective, the Second Administrative Level Boundary of the United Nations (SALB)25 provides information concerning territorial changes in some countries of the World. The Global Administrative Unit Layers of the FAO (GAUL)26 maintains historical layers with a unified coding system at country at the first and second administrative levels. The Global Administrative Areas database, leaded by different researchers (GADM)27 is also a very interesting resource since it provides layers for each country of the World at different regional scales. For instance, GADM gives geometries in a very precise way for Algerian Wilayas and municipalities (figure). We can also note that this database is freely available for academic and other non-commercial use.



- **Bottom up approach** : Even if some resources exist for collecting statistical information at regional level for the countries of the World (World Gazetteer, City population), the content of these online databases concerns generally few indicators (area and population only) and are not necessarily harmonized (different definition of cities depending of the country for instance). Consequently, it is also very important to explore available resources in each country of the European neighbouring, by downloading information and documentation from the National Statistical Institutes website and accessing to National Censuses for **improving the quality and the quantity of data collection**.

➢ Step 2: Thinking

Once the step 1 finalized, data needs expertise and validation. Consequently, we propose in the "thinking step" to set up three different elements:

- For each country, to **define a hierarchy "Similar to NUTS"** for making possible comparisons between European regions and Neighbouring regions. We will follow the 3 principles of NUTS definition, defined by Eurostat and summarized in the Technical Report on Western Balkans & Turkey, provided by ESPON DB1 Project[28].
- **Check, completion (if necessary) and improvement of the quality of metadata**. This task will be externalised to the analysis of experts specialised in different geographical area. We expect to take an expert specialised on the

[25] *http://www.unsalb.org/*
[26] *http://www.fao.org/giews/english/shortnews/GAUL1.pdf*
[27] http://www.gadm.org
[28] Principle 1 : The NUTS regulation defines minimum and maximum population thresholds for the size of the NUTS regions; principle 2: NUTS favors administrative regions; principle 3: NUTS favors general geographical units.

Southern Neighbourhood another one on the Eastern Neighbourhood of the European Union.

- For each country of the Neighbourhood, we propose to deliver an **inventory of available information** according to a defined template. This information will be built in two parts, one dedicated to the territorial divisions identified and another one on data availability on National Statistical Institute websites (figure 30a and 30b).

➢ Step 3: Building

The ESPON Neighbourhood database will be organised and delivered in three different folders.

- **Data**: Integration of data following the data and metadata template. In term of data, we fix the priority to count data instead of ratios. High importance will be made on the fulfilment of metadata since it is highly probable that the data collection will be based on several data sources (surveys, census, websites etc.)
- **Geometries**: Delivery of seamless geometries with names and SNUTS codes. There will be two map templates: one generalized for mapping and another less generalised for GIS calculations. We expect to integrate these new geometries in a dedicated MapKit which will allow to display ESPON territorial units and European Neighbour Regions commonly.
- **Documentation**: All information collected is saved, linked to the metadata. We expect to work on the best way to diffuse such material.

➢ Step 4: Harmonisation

While the database will be built, we want a last step of data validation/harmonisation. It implies both to pay attention to data and metadata.

- **Metadata**: seamless definition of indicators between countries (total population/resident population).
- **Geometries**: Proposing solution for territorial incoherencies (Western Sahara, Occupied Palestinian Territories/Israel) and ensuring a total compatibility with European mapkits.
- **Data**: In problematic cases, value smoothing with external data providers at country level (UNEP data for instance).

❖ ***Deliveries – Timetable[29]***

December 2011

- Three country reports for Maghreb area (Tunisia, Morocco, Algeria)

June 2012

- Country reports for Russia, Ukraine, Belarus, Libya and Egypt
- Geometries, documentation and data for Tunisia, Morocco and Algeria (included in a dedicated MapKit "ESPON Area and European Neighbour Regions")

[29] This timetable will be probably adjusted as regard to the data collection strategy which will be defined new ESPON Project on European Neighbour Regions.

December 2012

- Geometries, documentation and data for Russia, Belarus, Libya and Egypt (update of the MapKit "ESPON Area and European Neighbour Regions")
- Country reports of Armenia, Georgia, Syria, Irak, Lebanon, Israel, Occupied Palestinian Territories, Azerbaijan and Jordan.

June 2013

- Geometries, documentation and data for Armenia, Georgia, Syria, Irak, Lebanon, Israel, Occupied Palestinian Territories, Azerbaijan and Jordan (update of the MapKit "ESPON Area and European Neighbour Regions")

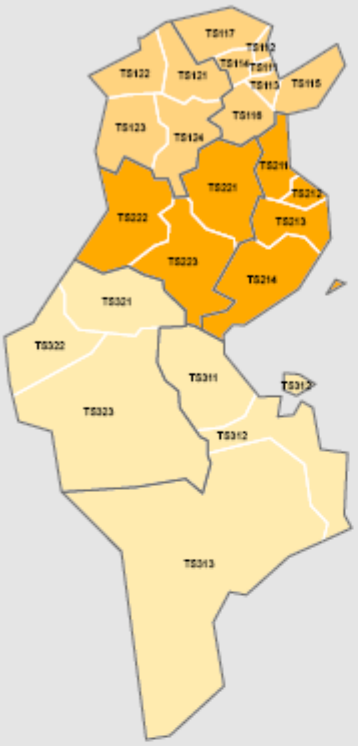
December 2013

- Draft Technical Report on the ESPON Neighbourhood database

June 2014

- Final Technical Report on the ESPON Neighbourhood database
- Data harmonization
- Integration in the web interface

TUNISIA - TERRITORIAL DIVISIONS



- **Territorial Units (T.U.) listing**

SNUTS3	NOM	pop 2004
TS111	Tunis	984
TS112	Ariana	422
TS113	Ben Arous (Tunis Sud)	506
TS114	Manubah	336
TS115	Nabeul	694
TS116	Zaghouan	161
TS117	Bizerte	524
TS121	Beja	305
TS122	Jendouba	417
TS123	Le Kef	259
TS124	Siliana	234
TS211	Sousse	544
TS212	Monastir	456
TS213	Mahdia	378
TS214	Sfax	855
TS221	Kairouan	546
TS222	Kasserine	412
TS223	Sidi Bou Zid	396
TS311	Gabès	343
TS312	Médenine	433
TS313	Tataouine	144
TS321	Gafsa	324
TS322	Tozeur	98
TS323	Kebili	143

SNUTS2	NAME	pop 2004
TS11	Nord-est	3627
TS12	Nord-ouest	1214
TS21	Centre-est	2233
TS22	Centre-ouest	1354
TS31	Sud-est	919
TS32	Sud-ouest	564

SNUTS1	NAME	pop 2004
TS1	Nord	4841
TS2	Centre	3587
TS3	Sud	1483

- **Territorial history**

Before 1881: 70 caïdats (tribal division)
1881-1956: French protectorate - 36 caïdats
1956: 14 governorates
1970's: The governorates of Sidi Bou Zid (1973), Siliana (1974), Sousse, Monastir, Mahdia (1974, split of the governorate of Sousse) and Zaghouan (1976) are created
1981: The governorates of Kebili, Tozeur and Tataouine are created (split of the governorates of Gabès, Gafsa and Medenine)
1984: The governorates of Tunis, Ben Arous and Ariana are created (split of the governorate of Tunis)
2000: Creation of the governorate of Manubah (split of the governorate of Ariana)

- **Other territorial divisions known**

Delegations: 263 territorial units (geometries available, codified in the SNUTS delineation)
Municipalities: Approx. 270 territorial units
Secteurs: More than 2000 territorial units

- **Specific notes**

At SNUTS4 level, the delegation of Douz has recently been split in two governorates: Douz Nord and Douz Sud. Data exists, but no official geographical delineations of these territories has been published at the moment. Consequently, these territorial units have been split in an aleatory way.

- **Source (map)**

GADM database (http://www.gadm.org/)

	NAME	UNITS	AREA	POP	STATUS
SNUTS1	Macro regions	3	49 784	3 304	Creation
SNUTS2	Planning regions	6	23 100	1 662	Official
SNUTS3	Governorate	24	6 222	413	Official
SNUTS4	Delegation	264	568	38	Official
SNUTS5	✖	✖	✖	✖	✖

Date of the report: 22/12/2011

Figure 30a - Country report of Tunisia (delivered in december 2011) – Territorial division part

TUNISIA- DATA

• Situation of the data collection

TARGETED INDICATORS MAIN COMPONENTS	TARGETED INDICATORS		YEARS	LEVEL	NOTES
POPULATION	TOTAL	✓	1984, 1994, 2004-2010	SNUTS 0-1-2-3-4	2004 only for SNUTS4
	By sex	✓	2004	SNUTS 0-1-2-3-4	
	By sex & age (pyramid)	✓	2004, 2008	SNUTS 0-1-2-3-4	* SNUTS 0-1-2-3: Five years age-class Maximum age-class: 80 years and more * SNUTS4: 5 years age-class under 20 years, 10 years age-class after (maximum: 60 years and more)
	By matrimonial status	✓	2004	SNUTS 0-1-2-3	Single, married, widowed, divorced
	By urban/rural areas	✓	2004	SNUTS 0-1-2-3-4	Population living in Tunisian communes
DEATHS	TOTAL	✓	2000-2009	SNUTS 0-1-2-3	Data derived from surveys
	By sex	✓	2000-2009	SNUTS 0-1-2-3	Data derived from surveys
BIRTHS	TOTAL	✓	2000-2009	SNUTS0-1-2-3	
	By age of the mother	▬	2004-2009	SNUTS0-1-2-3	Available but not collected (uncertain values)
MIGRATION	TOTAL	✓	1987-1994 1999-2004	SNUTS 0-1-2-3-4	Emmigration and immigration/ internal migratory balance
	By sex	✓	1999-2004	SNUTS 0-1-2-3-4	Internal migratory balance
	By age	✓	1999-2004	SNUTS 0-1-2-3-4	Internal migratory balance
	By education level	✓	1999-2004	SNUTS 0-1-2-3-4	Internal migratory balance
	By cause	✓	1999-2004	SNUTS 0-1-2-3-4	Internal migratory balance
ACTIVE POPULATION	TOTAL	✓	2004	SNUTS 0-1-2-3-4	
	By sex	✓	2004	SNUTS 0-1-2-3-4	
	By age	✓	2004	SNUTS 0-1-2-3-4	18-59 years
UNEMPLOYED POPULATION	TOTAL	✓	2004	SNUTS 0-1-2-3-4	
	By sex	✓	2004	SNUTS 0-1-2-3-4	
	By age	✓	2004	SNUTS 0-1-2-3-4	5 years age-class
	By education level	✓	2004	SNUTS 0-1-2-3-4	Available by sex also
GROSS DOMESTICAL PRODUCT	TOTAL	✓	2000-2010	SNUTS 0	In current and in constant prices (basis 2005)
	By economic branch	✓	2000-2010	SNUTS 0	In current and in constant prices (basis 2005)
INCOME	TOTAL	✗			
EMPLOYED POPULATION	TOTAL	✓	2004	SNUTS 0-1-2-3-4	
	By economic branch	✓	2004	SNUTS 0-1-2-3-4	Agriculture-fisheries-forestry, mining&energy, industry&manufacturing, building, shopping, transport & communication, services, administration&education, non delcared Available by sex
	By sex	✓	2004	SNUTS 0-1-2-3-4	
SCOLARISATION (YOUNGS)	TOTAL	✗			
EDUCATION (ADULTS)	TOTAL	✓	2004	SNUTS 0-1-2-3-4	Population aged above 10 years
	By level	✓	2004	SNUTS 0-1-2-3-4	No instruction, primary, primary (elementary), secondary, higher level
	By sex	✓	2004	SNUTS 0-1-2-3-4	Data available by level of education also
LAND AREA	TOTAL	✓	2004	SNUTS 0-1-2-3-4	GIS calculation at SNUTS4 level

• Main source of information and references

Institut National de la Statistique (http://www.ins.nat.tn/)
Almost data are derived from the census of 2004 (RGPH). Since 2004, the national statistical office of Tunisia publishes every year an annual report where most of the data provided in the RGPH are estimated for the governorates.

• Years of national census

1946, 1956, 1966, 1975, 1984, 1994, 2004, 2014

Figure 30b - Country report of Tunisia (delivered in december 2011) – Data availability part

3.2.5.2 Cities (WP B.5.2, Géographie-cités)

Analysis of the situation

In order to enlarge in the future the urban databases to the neighbourhood, we propose to expertise some of the existent urban databases at the world scale. The objective is not to have a bottom up approach but to experiment a top down one, validated on the European countries.

For instance, the World Urbanization Prospect 2009 of the United Nations provides the total population 1950-2025 of all the urban agglomerations populated above 750 000 inhabitants[30] in the World but using national definitions of cities, without any international harmonization (some cities are defined as urban agglomerations, such as in France, whereas other are defined as functional urban areas, such as in United States). The GRUMP project[31] (CIESIN) uses satellite images taken at night, which provides interesting perspectives for the developing countries, characterised by dense vegetation cover, but their contribution is more debatable for Europe. The Geopolis database[32] gives populations for morphological agglomerations (continuous built-up area) larger than 10 000 inhabitants, but the tables giving populations for the most recent date (2000) are accessible only for the largest ones (Figure 31).

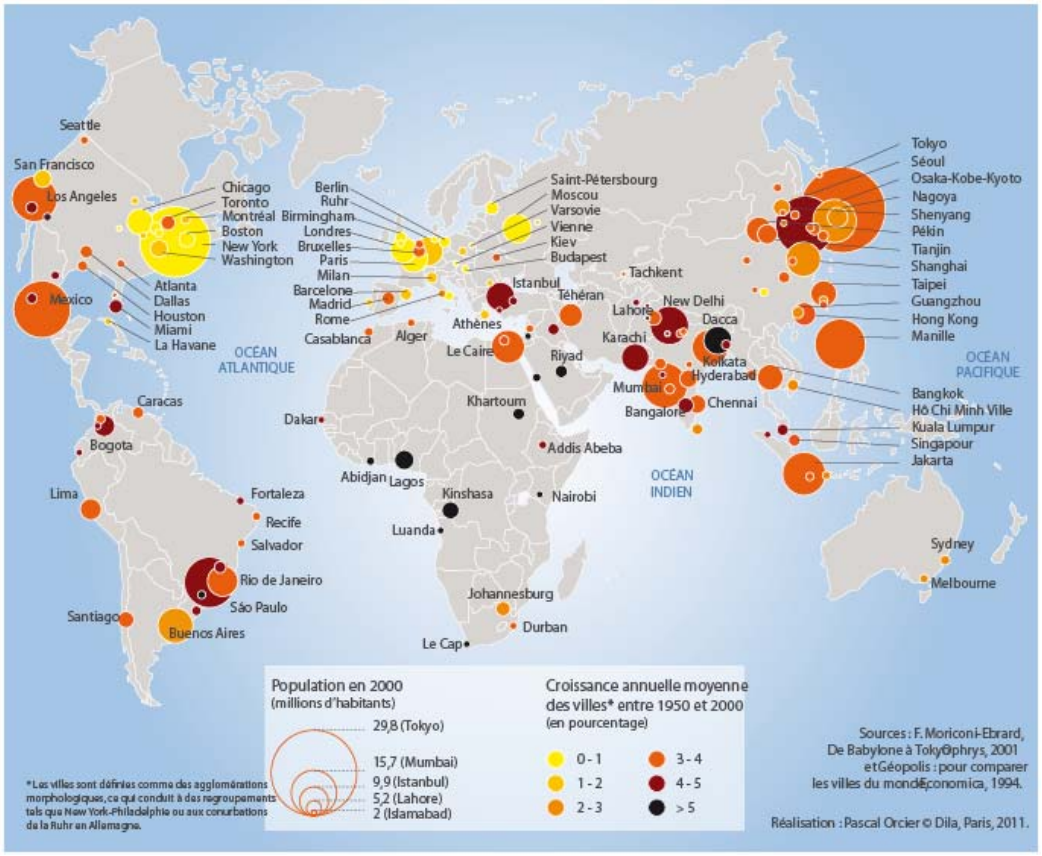


Figure 31 - European and Neighbouring countries cities at the World level (source: Bretagnolle et al. 2011, Métropoles et mondialisation, La Documentation Photographique n°8082, La Documentation Française)

[30] http://esa.un.org/unpd/wup/index.htm

[31] CIESIN (Center for International Earth Science Information Network) (2004), *Global Rural-Urban Mapping Project (GRUMP)* Alpha Version: Urban Extents http://sedac.ciesin.columbia.edu/gpw last accessed 17 November 2010.

[32] Moriconi-Ebrard F. (1994), *Géopolis, Pour comparer les villes du monde*, Paris, Anthropos, Economica, Collection Villes ; Moriconi-Ebrard F. (2001), *De Babylon à Tokyo*, Paris, Ophrys.

Objective

Taking into account the literature and database available on the topic of cities in the World, we aim to deliver a guidance paper for project interested in collecting and displaying data in the European Neighbourhood area. It could include namely:

- A review of existing information on that topic.
- A study of city and urban/rural definitions in the neighbourhood countries where information is accessible (both in National Statistical Institutes and harmonized databases).
- As far as possible, some advices on how proceed to collect data in these areas.
- As far as possible, an harmonized GIS layer of cities (points with X/Y attributes) in these areas.

The delivery of this Working Package will take the form of a Technical Report.

❖ ***Deliveries – timetable***

December 2012

- Draft version of the Technical Report on "Data collection on European Neighboring cities".
- Draft version of the GIS layer (point locations) on European Neighboring cities

June 2013

- Final version of the Technical Report on "Data collection on European Neighboring cities".
- Final version of the GIS layer (point locations) on European Neighboring cities

3.2.5.3 Grids (WP B.5.3, UAB)

Analysis of the situation

Corine Land Cover covers only EEA Member States and it has not been extended to the neighbourhood. Nevertheless, a project called GlobCover exists. It is an ESA initiative which began in 2005 in partnership with JRC, EEA, FAO, UNEP, GOFC-GOLD and IGBP.

Globcover is based on input observations from the 300m MERIS sensor on board the ENVISAT satellite mission, and collects landcover worldwide, producing raster data with an approximate resolution of 300x300m. The last version released covers the period January-December 2009.

On the other hand, the European Reference Grid has not been extended in the neighbourhood, in order to facilitate the combination and integration of different data types, like it has been done so far within the ESPON extent by the ESPON Database project.

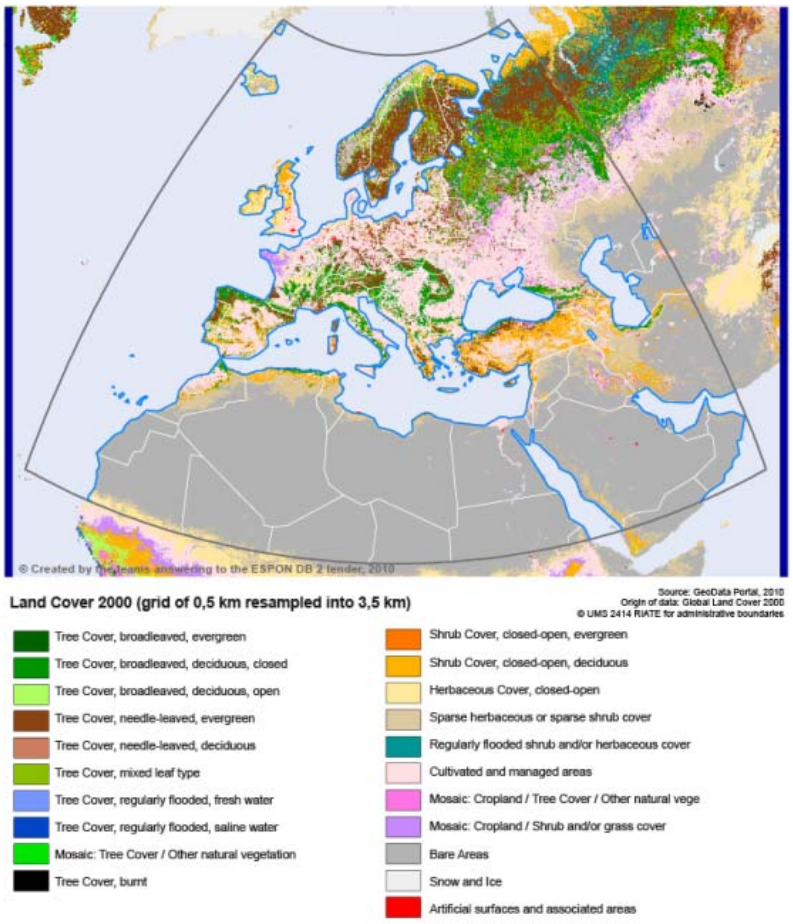


Figure 32 - Land Cover in the Euromed area

Objectives

We aim to carry out the distribution of GlobCover data in SNUTS regions defined by the WP B.5.1.

Before making possible such process, it would be needed to get the precise geometries coming from the WP B.5.1.

On the other hand, it will be tested how to enlarge the European Reference Grid extent onto the neighbourhood areas, and Globcover data could also be distributed by grid (figure 32), opening the door to future data integration and combination in the framework of neighbourhood areas analysis.

- ***Deliveries – timetable***

December 2012
First Draft version of the GlobCover 2009 by SNUTS units and draft proposal for Grid enlargement onto the neighborhood.

December 2013
Second Draft version of the GlobCover 2009 by SNUTS units and by enlarged grid cells and Delivery of first version of enlarged Grid.

June 2014
Final version of the GlobCover 2009 by SNUTS units and by grid.

3.2.5.4 Western Balkans & Turkey (WP B.5.4, RIATE + 1 expert)

Analysis of the situation

TEMPORAL_START TEMPORAL_END		
trb14	NUTS3	Tunceli
trb21	NUTS3	Van
trb22	NUTS3	Mus
trb23	NUTS3	Bitlis
trb24	NUTS3	Hakkari
trc11	NUTS3	Gaziantep
trc12	NUTS3	Adiyaman
trc13	NUTS3	Kilis
trc21	NUTS3	Sanliurfa
trc22	NUTS3	Diyarbakir
trc31	NUTS3	Mardin
trc32	NUTS3	Batman
trc33	NUTS3	Sirnak
trc34	NUTS3	Siirt
ba1	similar to NUTS 1	Bosne i Hercegovine
al001	similar to NUTS 3	Tiranë
al002	similar to NUTS 3	Fier
al003	similar to NUTS 3	Elbasan
al004	similar to NUTS 3	Shkodër
al005	similar to NUTS 3	Durrës
al006	similar to NUTS 3	Vlorë
al007	similar to NUTS 3	Korçë
al008	similar to NUTS 3	Berat
al009	similar to NUTS 3	Dibër
al010	similar to NUTS 3	Gjirokastër
al011	similar to NUTS 3	Kukës
al012	similar to NUTS 3	Lezhë
ba001	similar to NUTS 3	Unsko-sanski kanton
ba002	similar to NUTS 3	Kanton Posavski
ba003	similar to NUTS 3	Tuzlanski kanton
ba004	similar to NUTS 3	Zeničko-dobojski kanton
ba005	similar to NUTS 3	Bosanskopodrinjski kanton
ba006	similar to NUTS 3	Srednjobosanski kanton
ba007	similar to NUTS 3	Hercegovačko-neretvanski kanton
ba008	similar to NUTS 3	Zapadno-hercegovački kanton
ba009	similar to NUTS 3	Kanton Sarajevo

In the current situation, data on Western Balkans & Turkey are stocked in a zip file in the ESPON Database. It is indeed impossible to query data from the regional interface data coming from neighbouring countries.

On top of that, indicators provided by the ESPON DB 1 Project contains a lot of missing values and relatively few indicators, which make this database difficult to use.

Of course, Western Balkans & Turkey data contains both Candidate Countries (Croatia, Turkey, FYROM, Montenegro) which are already defined in the NUTS nomenclature[33] and non-candidate countries (Albania, Serbia and Bosnia-Herzegovina). For these countries the ESPON Database 1 project has defined a "Similar to NUTS" nomenclature (SNUTS), which allow to compare these territorial units with the European one[34].

Figure 33 - NUTS and SNUTS nomenclatures for Western Balkans & Turkey defined by the ESPON Database 1 Project

Objectives

Firstly, we propose to make possible to query SNUTS units from the regional web interface add this functionality and to include non-Candidate countries of the Western Balkans in the MapKit dedicated to ESPON Area and European Neighbour Regions.
Secondly, we propose to update the data already collected by the ESPON DB1 Project, by following the same data collection strategy shown in WP B1 (regional data) and using results produced by other TPGs for these countries. In that order, we aim to propose more indicators for these countries and, for indicators which are currently available, try to estimate missing values.

We only notice that we will not consider Candidate Countries in this Working Package, considering that these countries are taken into account in the WP B1 (regional data).

- ***Deliveries – timetable***

June 2012

- It is possible to query SNUTS data from the regional web interface
- Deliveries of geometries for Albania, Serbia and Bosnia-Herzegovina (included in a dedicated MapKit "ESPON Area and European Neighbour Regions")

December 2012

- Update of the content of the Western Balkans & Turkey database

[33] Regions in the European Union: Nomenclature of territorial units for statistics NUTS 2010/EU-27, Eurostat
34 Analysis of the availability and the quality of data on Western Balkans and Turkey, ESPON Database Project, March 2011, 50p.

3.2.6 WP B6 Surveys (RIATE)

Analysis of the situation

The first ESPON Database Project has yet delivered a technical report on the use of large official European surveys (Labour Force Survey, Farm survey, …) and it does not appear interesting to duplicate this work in ESPON M4D, especially considering the limited amount of funds allocated to this task (less than 10 k€). It appears more interesting for ESPON program to explore the interest of new types of survey, in particular related to the external image of EU territories and their related capacity to improve their "brand" and more generally to attract qualified migrants or creative working class.

Objective

We suggest therefore developing the partnership between ESPON and DG Research that has been established under the project FP7 EuroBroadMap, which was directly based on previous results obtained by ESPON 2006 project 3.4.1. Europe in the World. A very precious database has been established in 2009 at world scale concerning, among other, the cities and states where 9300 students from 18 countries would like to live (or not like to live) in a near future (figure 34a). But this level of countries is probably too general in terms of scale for ESPON program and we propose to focus also the analysis of the attractiveness of cities (figure 34b) that provides more details on internal differences of national territories. Bot necessary to precise that the combination of both geometries can reveal particularly interesting information (cities more or less attractive than the country where they are located …).
What we propose is to realize a focus on ESPON territory based on this data (that has not been done by the FP project) but also, if funds are sufficient, to update the survey on a limited sample of students in order to analyze what are the new trends after the big economic depression. Not necessary to precise that such an analysis could feed the next synthesis report of ESPON with very original illustrations and discoveries.

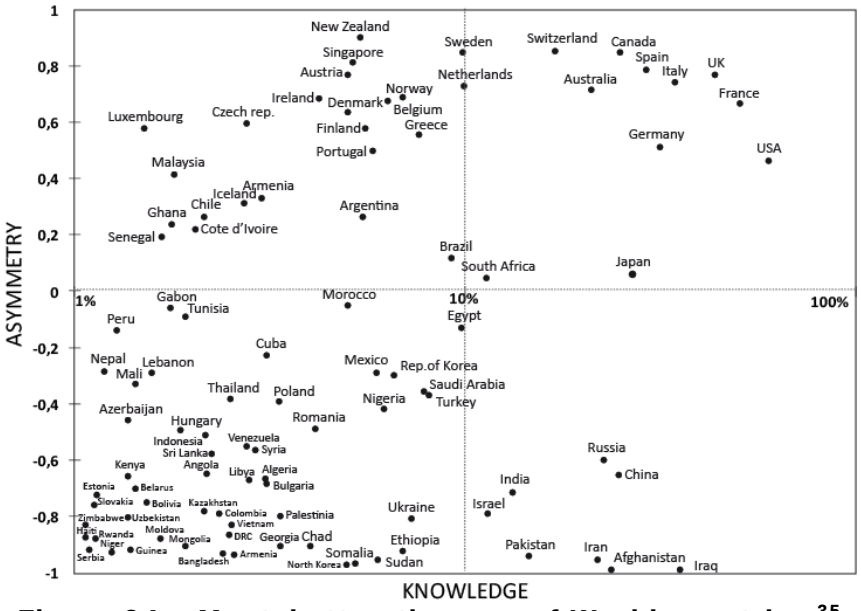


Figure 34 – Mental attractiveness of World countries [35]

This graphics is based on the comments on the maps on which the students noted where they would or would not like to live in the near future. It allows to distinguish:

[35] Derived from the results of the EuroBroadmap survey.

- ***Attractive areas*** *(asymmetry index significantly above 0), which are the places in the world where students declare that they would like to live in the near future, even if it is not possible at the present time. This area may be related to the countries they have visited, countries where migration linkages exist, etc.(Canada, Spain, Switzerland)*
- ***Repulsive areas*** *(asymmetry index significantly under 0), which are the places in the world where students would not like to live, at least at the present time. This does not mean that this repulsion is related to political hostility; in many cases, it simply depends on the country's lack of economic opportunities, wars and negative images disseminated by the media (Iraq, Iran, Afghanistan)*
- ***Ambiguous areas*** *(asymmetry index around 0) are interesting in the sense that they are parts of the world that are attractive and repulsive to a more or less equal proportion of students. This can reveal a structural conflict in visions (this is generally the case for the USA, which is considered as both attractive and repulsive by students in many countries), but also a dynamic process of changing appreciation, which is unfortunately impossible to demonstrate on the basis of our survey, as we have only one historical point of measurement;*
- ***Ignored areas*** *(characterized by a low knowledge index) must also be identified because they are parts of the world of which students are not consciously aware, which is, in a sense, worse than ignorance.*

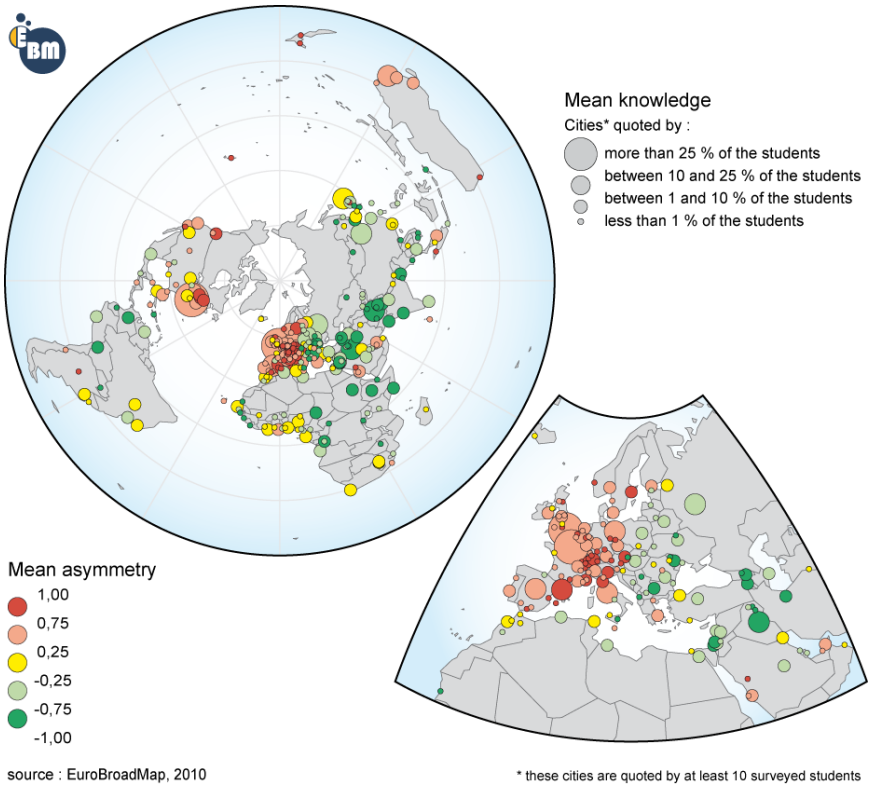


Figure 34a – Mental attractiveness of World cities

❖ ***Deliveries – timetable***

December 2013

- Technical report presenting a focus on the attractiveness of ESPON cities and states derived from results of EBM survey 2008-2009.

June 2014

- Second Technical report presenting the evolutions of attractiveness between 2009-2012 on a sample of students located inside and outside EU.

3.3 Working Package C Networking

We see networking as an important component of Phase II of the Database project and which has several objectives: efficient flow of data and information in the ESPON 2013 network; cooperation with the ESPON web service provider; updating of the Map Kit tool; supporting the ESPON Coordination Unit; and external networking. And as part of the ESPON triad under the Scientific Platform, there must be close cooperation with the ongoing project on "Territorial Indicators and Indices". Four levels of networking have been identified: networking with ESPON priority 1 Projects, with priority 2 Projects, with Priority 3 Projects and External Organizations.

3.3.1 WP C1 Platform (RIATE / LIG)

Objectives

Whilst a central concern of the M4D Project is to create a comprehensive, accurate and reliable Database, the users require support in order to feed and to use this Database. They need advices, support but also information about tools they may or have to use.

Into the existing application, we will develop new functionalities that were previously called **Platform but from now is called Portal**. The main aim is to inform other ESPON Projects and users on the DATABASE project and on the important information they may have to know concerning data AND to manage the data flow and its very high complexity within the ESPON community.

We had then intensive work designing a user-friendly and useful tool through new functionalities and a new organization of the existing application.

New functionalities are: a **Database homepage**, with News and a bi-yearly newsletter, a page dedicated to the **query interface**, a page to **follow the process** from the deposit of datasets by projects to the integration of their datasets into the Database (upload page, check in progress), a page where users will find **resources** (mapkit, reports), and a **help page** (FAQ, how to, guidelines, specifications).

This revised **interface will be user-friendly and useful for the whole ESPON community** but will be mainly developed to support the Priority 1, 2, and 3 Projects.

Detailed specifications will be delivered in March 2012. Implementation will be done for June 2012.

Figure 35 - Draft version of the ESPON Database Portal homepage

- ***Deliveries – timetable***

March 2012

- Detailed specifications of the portal

June 2012

- Implementation of main functionalities of the portal

3.3.2 WP C2 Follow-up of Priority1 projects

Analysis

Establishing an operational information and data flow with the ESPON projects is a real challenge for ESPON M4D project. In fact, ESPON M4D has to provide guidelines in terms of data collection, metadata editing and mapping; has to check that these ones are followed; has to collect the data and metadata created by each ESPON Project and store it in an efficient way in the ESPON Database. However, the needs are not the same for ESPON Priority 1 projects, using and feeding the ESPON database for the ESPON area (this section) or for ESPON Priority 2 projects (or for case studies of ESPON Priority 1 projects), using and feeding the ESPON database with zoom on specific territories (section 3.3.3).

Objectives

There are two critical phases during the lifetime of an ESPON Project, the beginning (a.), where the project has to find the material and guidance for beginning its investigations (mapkit, basic data); and the ending (b.), where the project delivers its data and metadata in the specified ESPON format.
The aim of the M4D Project is also to provide help to the various ESPON Projects:

a. **The guidance phase** is coordinated by the ESPON M4D contact team (cf part 3.4.3) and consists by the following elements:
   - Ensuring that each Priority 1 Project have **access to the entire ESPON Database** (public and private part). It implies to give a login and password pair to each project.
   - Delivery of a "**Starting Package**" to each Priority 1 Project, which includes guidelines concerning data, metadata[36] and map[37] creation in ESPON. This starting package could be provided with the guidance paper delivered by ESPON CU when the project starts.
   - Each Priority 1 project currently launched by ESPON has a **contact team**, which is involved in ESPON M4D Project (cf part 3.4.2). The relation between contact team and the priority 1 Project has been built in order to create the best linkage possible with teams involved in ESPON M4D Project (knowledge, fields of development of M4D Project etc). The contact team is the reference entity which has to be considered by each Priority 1 Project: in terms of data and metadata management, data collection, data delivery. In order to create a synergy between ESPON M4D Project and Priority 1 Projects, a face-to-face meeting is needed. It can take place during a TPG meeting, or at least during an ESPON Seminar.
   - During the first period of the project, the contact team have to answer (or at least transfer to the right person) to specific questions of priority 1 project.

When the **portal** is operational (June 2012), this process of guidance and information will be improved through the new functionalities (news, documentations, FAQ). A newsletter will inform the ESPON Community on the last developments of the ESPON Database.

b. **The data integration process** aims to apply a very steady quality control of datasets delivered by ESPON projects. This process concerns only the so called **"10 best indicators"** and is divided in 5 steps which will be summarized in the Tracking tool which will be available on the platform both for the concerned ESPON Project, the ESPON Coordination Unit and the M4D Team (figure 36a and 36b).

[36] Metadata specifications (http://database.espon.eu/metaspecifs)
[37] Mapping guide (http://database.espon.eu/reports)

Figure 36a - Data integration process

1. **Data and metadata upload**: When a project is ready for delivering its data and metadata, it activates a dedicated module in the metadata editor. It means that the data integration process is started and the tracking tool (in the platform) is activated for the project. A notification is sent both to the ESPON Coordination Unit and to the M4D team.

2. **Syntaxic checking:** aims to check the compliance of data and metadata delivered with the ESPON Database profile. In concrete terms, it checks if all the mandatory fields of the ESPON data and metadata are correctly filled. This control is done automatically when the project uploads it datasets in ESPON CU website by the upload interface developed by ESPON Database project. This is the only compulsory step of the data integration process. After the check, a notification is sent to ESPON CU and to the concerned project in order to summarize the result of the check.

3. **Semantic checking**: The dataset is transferred to the M4D contact team in order to analyze the content of the data and metadata (and namely the free-text fields). The aim of this step is to analyze if all the indicators of the dataset are correctly described and understandable by a large public. As a consequence, a short report is sent to the ESPON Coordination Unit and to the ESPON Project concerned.

4. **The quality control** At this step, an outlier detection tool will be run on the datasets. The results of the check will be reported to the ESPON CU and the project provider of data.

5. **Integration into the database**: those three checks allow giving a strong expertise on the quality of the datasets delivered by the project. For integrating the dataset into the database, the ESPON M4D Project needs the agreement both of the ESPON project concerned and the ESPON Coordination Unit. After that, the data will be available for download through the web interface.

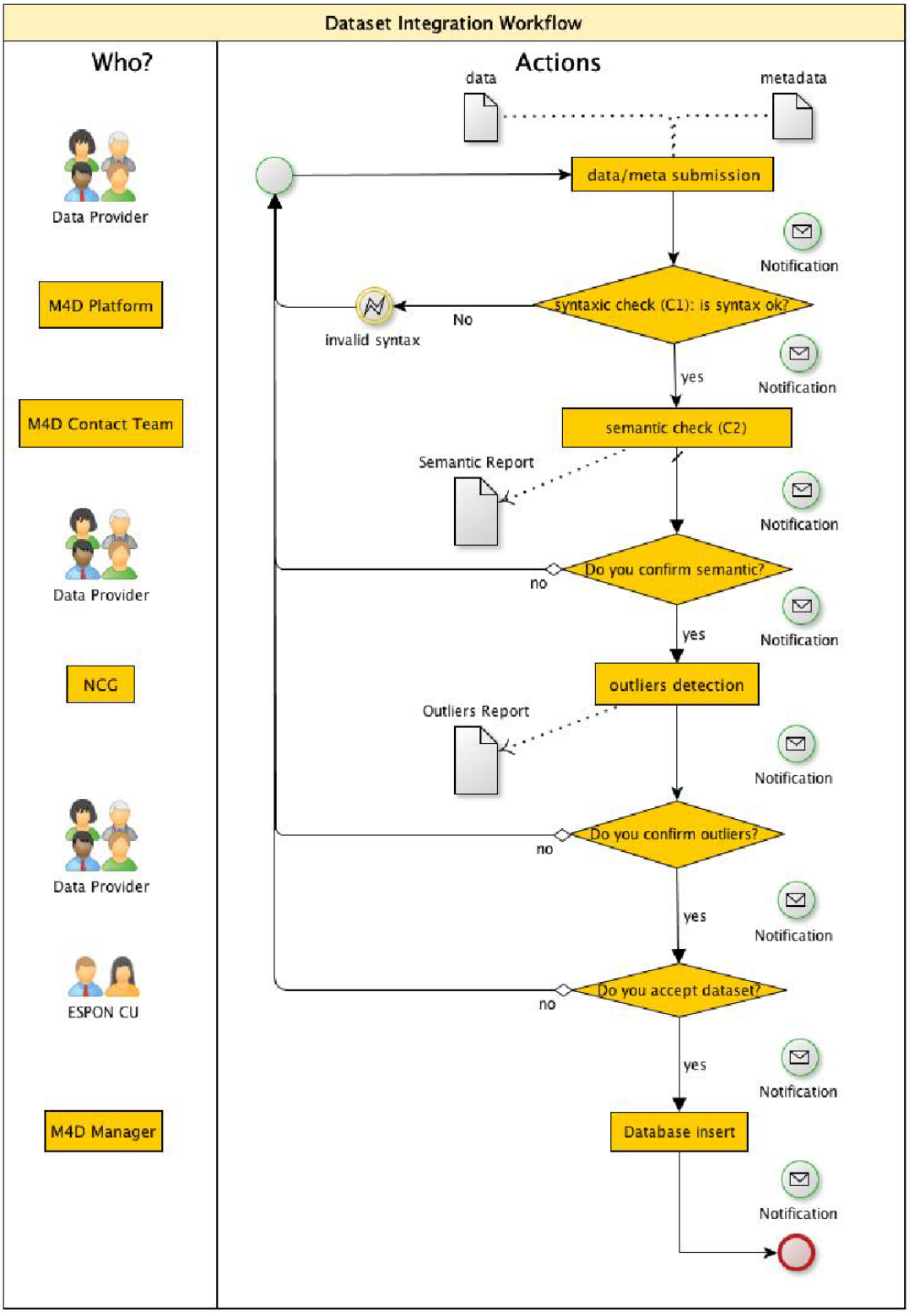


Figure 36b - Diagram on the data integration process

3.3.3 WP C3 Follow-up of Priority 2 projects (TIGRIS)

Analysis

The **networking activities with the Priority 2 projects** will be in the center of our interest. It is crucial for the ESPON database to integrate all the data and indicators provided in the framework of these projects, even if the information does not cover all the ESPON space.
However, the characteristics of data provided at local scale make impossible a homogeneous integration of such information in the query interface described above, and this for different reasons:

- **Too precise information:** One of the aims of the web interface consists by providing datasets for the all ESPON Space. Making available indicators for only a couple of NUTS2 at local space will produce noise in the database.
- **Heterogeneous nomenclatures:** Some datasets can be produced in heterogeneous geographical delineation, out of the NUTS or the LAU nomenclatures (*bassin de vie* in France, *Super Output Area* in UK). It will be very difficult to store systematically all the nomenclatures provided.
- **Too specific indicators:** When analyzing territorial dynamics at local scale, some indicators of high interest may be collected for these case studies, but are totally useless at the ESPON scale (for instance, number of commuters going from Germany to Luxemburg in the Grande Region).
- **Difficulty to identify easily what is available**: When multiplying the case studies, sometimes at a very local scale, making possible to have an overview of what is available is a challenge. The query interface is clearly not adapted to this kind of request.

The data storage of data coming from ESPON Priority 2 projects raises a lot of conceptual and practical problems, which could be solved by proposing an alternative solution only.

Objective

Our proposal consists by building a specific interface for querying data coming from ESPON Priority 2 projects. The data will be stored "as such" (in a zip format) and will downloaded following more or less the same principles than the one proposed by the Google Map technology (figure 37).

This proposal (which is a draft version and will be improved until the First Interim Report) presents some clear advantages for the users:

1. Google technology is known, used and shared by a large public.
2. It is possible to have a clear overview of the location of ESPON priority 2 projects study areas.
3. Data integration is not limited to Europe and it is easily possible to integrate data coming from case studies outside Europe (USA, China…)
4. It is a solution to display in an homogeneous way the heterogeneity this kind of ESPON Projects.

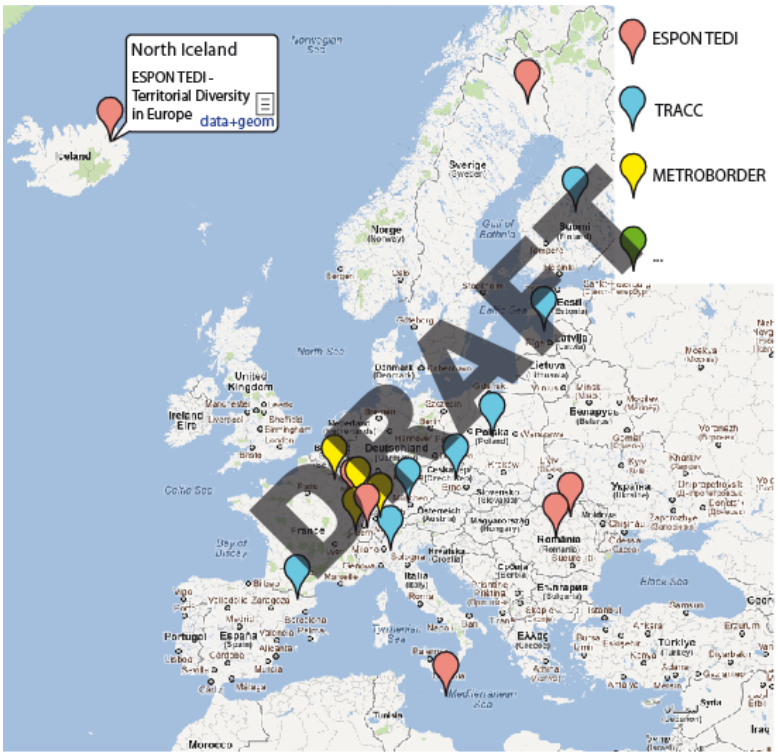


Figure 37 – First proposal for querying ESPON Priority 2 data

As regard to ESPON Priority 1 projects, this strategy does not imply to follow the data check process described above. Nevertheless, in order to integrate in an harmonized way this information, we will ask to the ESPON Priority 2 projects the following elements:

- Name of the project
- A short abstract describing the dataset content
- A shapefile linked to the dataset allowing to map the data (compliance with codes contained in the dataset)
- Listing of the case studies contained in the dataset and short description of them (name, country of belonging, X/Y coordinates, type of nomenclature)
- A contact person (name, organization of belonging and email address)
- The name and the temporal extent of the indicators contained
- A listing of the data source with their URL.

A template of Case studies metadata will be created until June 2012 and diffused to ESPON Priority 1 (for case studies) and 2 Projects for organizing the data flow with these projects.

❖ ***Deliveries – timetable***

June 2012
Specifications for the data delivery of ESPON Priority 2 projects.

June 2014
Implementation of the web-interface dedicated to ESPON Priority 2 projects.

3.3.4 WP C4 Follow-up of Priority 3 projects

The aim of this activity is to ensure the networking between the ESPON M4D Project and other ESPON Priority 3 Projects: the ongoing INTERCO Project on Territorial Indicators and Indices, and the forthcoming Project on Territorial Monitoring. UNEP/GRID (University of Geneva), which is Lead Partner on the INTERCO Project, will play the role of expert for partners involved in this sub-activity.

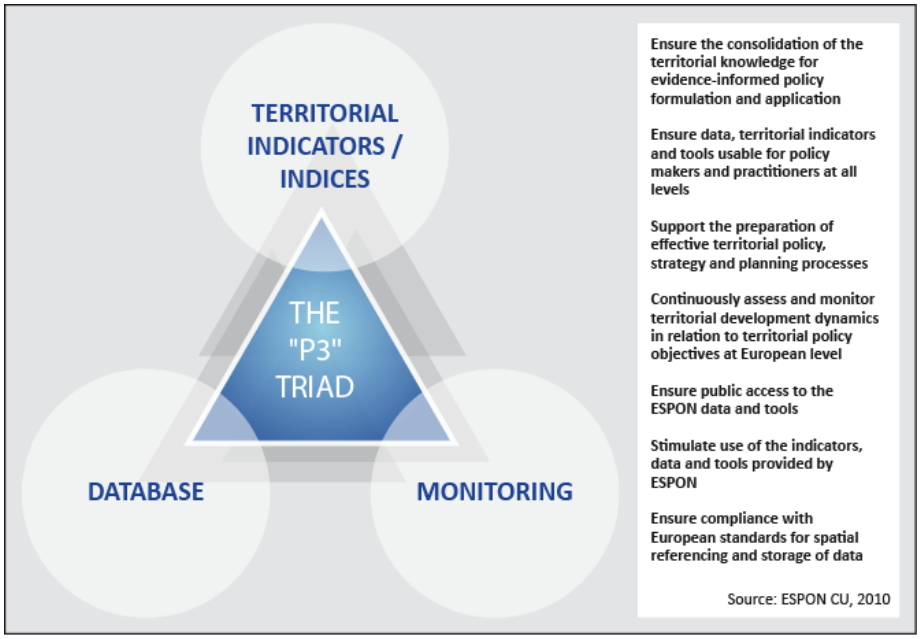


Figure 38 – The "emblematic ESPON P3 Triad"

These crucial point has been discussed many time in previous sections. So, we just summarize here what have be done in the past and what we planned to do in the future in order to insure better integration of Priority 3 projects :

- **INTERCO :** this project is now achieved but we have make an intensive use of their draft final report in order to improve our different classifications of indicators involved in the database.
- **MONITORING :** This project is unfortunately delayed for the second time, but we have engaged an important revision of the codification of ESPON database in order to make more easy the future development of this project. We have anticipate the fact that policymakers would like to obtain regularly synthetic notes on their territories (cities, regions, states) with no more than 15-25 top criteria, clearly related to political objectives. We also anticipate the need to update this criteria through time and to make possible the benchmarking of different places by diagram, maps, tables, … We strongly underline that it is not the mission of ESPON M4D project to do this job but we do our best to anticipate the needs of the future Monitoring Project in charge of this crucial task.
- **WEBMAPPING TOOL:** We elaborate a revised version of metadata where the statistical type of indicator will be precised, in order to avoid any semiotic fault in the realization of maps. We offer also a revised version of thematic classification that could be usefull. Last but not least, we work on the elaboration of long term time series for core indicators and the introduction of flags for outlier value. All of this should contribute to improve the quality of the application.
- **ESPON ATLAS :** In the criteria of selection of the most important indicators delivered by ESPON project, we have considered to introduce a specific flag for

statistical data that has been used for the production of maps in final reports of TPG's. Maps are indeed circulating a lot and contribute strongly to ESPON diffusion? But users of maps often want to get data behind the map in order to realize their own analysis. We try therefore to make possible an easy access to data related to maps published by ESPON.

- **HYPERATLAS V3 ?** : We are ready to introduce connections between the ESPON Database and a new version of HyperAtlas that could focus on the monitoring criteria. We have indeed observed that 80% of the existing indicators of ESPON Database (out of typologies and count variable) are single ratio (A/B where A and B are two count variables) which means that 80% of indicators could be theoretically put in form of files for HyperAtlas, making possible better visualization and creation of new indicators like deviations to European, territorial and Spatial contexts …

3.3.5 WP C5 External Networking

It is of high interest to establish more direct contacts with statistical institutes at European level. Contacts between ESPON M4D project and official institutes will be more flexible and give more degrees of initiative to ESPON M4D project. For instance, ESPON Database phase1 has organized a workshop involving Eurostat, EEA… on time-series issues. This kind of activity would be followed in this phase

As we have the ambition to help the ESPON programme to become an integrated actor of the European statistical system, we will work hard on the external networking. ESPON has developed results of high interest and they have to be spread at large. In the same time we believe we can gain in sharing experiences and methodology with other statistical institutes.

1. Sharing methodological approach such as outlier detection, estimation value process, data collection, interface, etc. Engineers and researchers involved in M4D project would like to exchange technical subjects and methods with their counterparts from other European institutes.

2. Sharing data and metadata. As most of data is free of use and easily disseminated, it is the ESPON interest to diffuse it outside ESPON world. In addition, metadata editor is added-value in European statistic.

3. A specific action could be launched on the ESPON project of elaboration of a hierarchy of functional divisions, in partnership with OECD and EUROSTAT and interested ESPON projects. It is indeed clear that such a proposal can probably not be achieved in the time-span of ESPON 2013 program, but could be a milestone for the future period 2014-2020.

N.B. This crucial question of external networking is further detailed in the conclusion of this report where we propose a general SWOT analysis of internal and external relations to be developed by M4D project.

3.4 Working Package D Coordination

The first semester of the WPD activities was mainly focused on how to launch the project in a way that allowed us to be sure that every goal will be reached. In fact, the start-up was concentrating a lot of efforts from Lead partner to organize and launch activities. Of course, efforts will be on-going throughout the lifetime of the M4D project and we will also take a special attention to dissemination and diffusion of results.

3.4.1 WP D1 Reporting & deliveries

Please find below a summary of the continuous deliveries of the ESPON M4D Project and their respective timetable mentioned in the part 3 of this inception report.

As regard to the advancement of the M4D activities, additional technical reports or short reorientations of the work done in the different Working Packages may not be excluded. In case of modification of this timetable, the ESPON Coordination Unit will be informed and changes which may be occurring will be justified.

The December 2011 deliveries have not been mentioned in the listing.

March 2012

- WPA1: Detailed specifications of the new functionalities of the web interface
- WPC1: Detailed specifications of the platform

June 2012

- WPA1: Design and develop JAVA API for metadata
- WPA1: Update the implementation of the metadata editor (and data check).
- WPA1: Design and extend the database model for new geographical objects
- WPA1: Design and create some software for integrating and checking more efficiently data from all ESPON Projects (ESPON DB Admin)
- WPA1: Delivery of a version of the web application including new functionalities
- WPA1-B2: Based on the expected delivery in December 2011 by the Géographie-Cités partner, the schema of the ESPON Database will be able to support some nomenclatures of Urban Geographic Objects, as well as a glossary of city names used by the Metadata Editor.
- WPA1: Update of the query and download application
- WPA1: Construction of an extended classification covering the three types of criteria defined: statistical, thematic and political challenge.
- WPA1: Draft version of the ESPON OLAP Visualisation tool
- WPA2: Data quality methods for time series data
- WPA2: Technical report: Outlier detection technique for spatially normalized values (completion)
- WPA2: Technical report: Detection techniques for spatially un-normalized data, interval, ordinal and nominal (completion)
- WPA2: Web function: Operationalisation of data quality check on platform and reporting procedures
- WPA3: Stabilized mapkit on Seas (ESaTDOR), final version of the mapkit with NUTS 2010 delineations, Provisional mapkit on ESPON Area and Candidate Countries + European Neighbour Regions (including Bosnia Herzegovina, Albania, Serbia, Morocco, Algeria and Tunisia)
- WPB1: Technical report on preliminary results of the Core Database Strategy.
- WPB1: Basic indicators collected with principle of Core Database Strategy in NUTS 2006 Division.

- WPB2: Data collection for the selected sample zones
- WPB3: Update of the ESPON Olap Cube
- WPB4: Technical report: strategy for creation of data and geometries at local level
- WPB4: Data, area and share of artificial and arable surface (location quotient); road density at LAU2 scale
- WPB5: Country reports for Russia, Ukraine, Belarus, Libya and Egypt
- WPB5: Geometries, documentation and data for Tunisia, Morocco, Algeria, Albania, Serbia and Bosnia-Herzegovina
- WPC1: Implementation of the M4D Platform
- WPC1: M4D Newsletter 1
- WPC3: Specifications for the data delivery of ESPON Priority 2 projects
- Cross WP activities: First Interim Report

December 2012

- WPA1: A technically study based on the SWOT Analysis shown in table 1 and conclusion for querying the database from external third party application via the Web Services technology
- WPA1: improvement of the Metadata editor and the upload interface to take into account LAU datasets and Grid data
- WPA1: Update of the application taken into account remarks and feedbacks from users
- WPA1: Update of the thematic classification with other Priority 3 projects
- WPA1: Update of the ESPON OLAP Visualisation tool
- WPA3: Mapkit on ESPON Area and Candidate Countries + European Neighbour Regions (integration of Russia, Belarus, Libya and Egypt).
- WPB1: Basic indicators delivery collected with principle of Core Database Strategy in NUTS 2010 Division
- WPB2: Database integration (table of correspondence)
- WPB3: Update of the ESPON Olap Cube
- WPB4: Data, area and share of permanent cultures (location quotient); typology of LAU2 based on the proximity to major rail-road transportation corridors, average distance towards the closest 3 universities.
- WPB5: Geometries, documentation and data for Russia, Belarus, Libya and Egypt
- WPB5: Update of the content of the Western Balkans & Turkey database
- WPB5: Country reports of Armenia, Georgia, Syria, Iraq, Lebanon, Israel, Occupied Palestinian Territories, Azerbaijan and Jordan
- WPB5: Technical report, draft version of the data collection on European Neighboring cities
- WPB5: Draft version of the GIS layer (point locations) on European Neighboring cities
- WPB5: First draft version of the GlobCover 2009 by SNUTS units and draft proposal for Grid enlargement into the neighborhood
- WPC1: M4D Newsletter 2

June 2013

- WPA1: Delivery of the final query and download application
- WPA1: Maintenance and minor evolutions
- WPA1: Implementation of one of the solutions studied in the technical study delivered in December 2012
- WPA1: Final version of the ESPON OLAP Visualisation tool and integration of new indicators
- WPA1: Update of the thematic classification with other Priority 3 projects
- WPA1: Final version of the ESPON OLAP Visualisation tool
- WPA3: Update and stabilization of the mapkit on ESPON Area and Candidate Countries + European Neighbour Regions (integration of Armenia, Georgia, Syria, Irak, Lebanon, Israel, Occupied Palestinian Territories, Azerbaidjan and Jordan).

- WPB1: Technical report on the empirical feasibility of revised functional divisions at different scales.
- WPB1: Basic indicators delivery collected with principle of Core Database Strategy for cities and other delineations (water basin, grids, …)
- WPB1: Final report on Core Database Strategy
- WPB2: Populating the selected urban databases
- WPB2: Construction of FUAs for selected cities
- WPB2: Time series on urban data – data model
- WPB3: Update of the ESPON Olap Cube
- WPB4: Data: distance to the closest FUA; and other indicators provided during the lifetime of the project
- WPB4: Proposal of alternative functional geometry at local level
- WPB5: Geometries, documentation and data for Armenia, Georgia, Syria, Iraq, Lebanon, Israel, Occupied Palestinian Territories, Azerbaijan and Jordan
- WPB5: Technical Report - Final version of the data collection on European Neighboring cities
- WPB5: Final version of the GIS layer (point locations) on European Neighboring cities
- WPC1: M4D Newsletter 3
- Cross WP activities: Second Interim Report

December 2013

- WPB5: Draft technical report on the ESPON Neighborhood database
- WPB5: Second draft version of the GlobCover 2009 by SNUTS units and by enlarged grids and first version of enlarged Grid
- WPB6: Technical report presenting a focus of the attractiveness of ESPON cities and states derived from results of EBM surveys 2008-2009

June 2014

- WPB5: Final technical report on the ESPON Neighborhood database
- WPB5: Data harmonization and final integration in the web interface
- WPB5: Final version of the GlobCover 2009 by SNUTS units and by grid
- WPB6: Technical report on the evolutions of attractiveness between 2009-2012 and a sample of students located inside and outside EU
- WPC3: Implementation of the web-interface dedicated to ESPON Priority 2 projects
- Cross WP activities: Draft Final Report

December 2014

- Cross WP activities: Final Report

N.B. We take the opportunity of this very long list of deliverables to underline the number of tasks that is expected from M4D project as compare to other projects. *All things being equal with the amount of funds received*[38], it is expected from this project much more deliverables than usual (technical papers, data, software, …).

[38] The funds received by M4D (€1 300 000) are more or less equivalent to the funds of one and half priority 1 project or three priority 2 projects.

3.4.2 WP D2 Meetings

This sub-activity manages the organization of regular meetings inside the project (1) and outside the project (2).
Inside the Project (1): In order to establish a necessary synergy inside the ESPON M4D Project, regular meetings are planned. We will then plan at least one meeting per year involving all the partners of the project, but also regular meetings on a monthly basis, with the participants of the operational Working Packages A and B in order to organize results and key-findings in a more operational way.
Outside the project (2): Every six months, partners will be involved in ESPON seminars (presentations, parallel sessions) and will show the main advancements of ESPON 2013 DataBase taking also into account the feedbacks from the attendees.

3.4.3 WP D3 Management

The management task is a multilevel sub-activity since it will ensure to fill the specific needs of the project. The Lead partner bestows a Scientific coordinator and two co-managers. They will ensure the good management of the project in a multi-level approach. To facilitate the management, one e-mail address for all the queries, questions and deliveries has been chosen: manager@espondb.eu

Either to plan activities during the lifetime of the project, or to ensure the good delivery of material, the co-managers are in charge of all the aspect of the management of the project. In addition, the financial manager is dedicated to the project and for being also the co-manager of the project, she will work in close contact with the consortium following the execution and the achievement of the project. Scientific activities will be linked to financial matters at all time of the project lifetime.

Scientific coordinator	
Co-Manager #1	**Co-Manager #2**
General duties + Scientific management	General duties + Financial management

In addition, a team from the M4D project is dedicated to, and then, in charge of a certain number of ESPON Projects (P1, P2, P3). This team is the "representative" of M4D for the following of any ESPON Project and has to answer to questions linked to the database (last updates, availability of indicators, metadata, deliveries…). A detailed list of contact persons is given below.

P1 project	Team in charge of it
ATTREG	GEO-CITES
EU LUPA	UAB
TERCO	NCG
TRACC	GEO-CITES
SGPTDE	GEO-CITES
GEOSPECS	UAB
KIT	NCG
TIGER	RIATE
ESaTDOR	UAB
SeGI	NCG
ARTS	TIGRIS

P2 project	Team in charge of it
all	TIGRIS

P3 project	Team in charge of it
all	LIG

3.4.4 WP D4 Support to ESPON CU

This sub-activity aims to organize in an efficient way the support to ESPON Coordination Unit for questions related to data collection, data harmonization, data expertise or mapping for ESPON reports, publications, press releases or any presentations.

N.B. This support is limited by the workforce allocated to the task in the general budget of ESPON M4D.

CONCLUSION: ESPON M4D IS NOT THE ENTIRE ESPON SCIENTIFIC PLATFORM (P3)…

In conclusion of this inception report, and in order to summarize the different proposals that have been made, we have tried to propose a SWOT analysis that covers not only the questions of external networking (expected deliverable) but also more generally the internal situation of ESPON database and Priority 3 projects in the general strategy of the ESPON Program.

In our opinion, there has been some misunderstanding since the beginning of ESPON 2013 on the respective role of the ESPON Database on one hand, and the whole set of priority 3 projects defining the EPSON Scientific Platform (Priority 3), on the other hand. This confusion is a source of prejudice for both ESPON M4D project and for the ESPON program as a whole, because it introduces pressure on the single ESPON M4D project to realize tasks that are not in the contract, without considering the whole picture of the ESPON Scientific Platform (Priority 3).

Criteria	Discussion
STRENGTH	• ESPON has produced many original indicators at regional level that cannot be found in the equivalent databases such as OECD or Eurostat • ESPON proposes data for various types of geographical objects, not only regions but also cities, grid, … • ESPON proposes information at different scales, from local to global, making possible the realization of a 5-level approach • ESPON has introduced strong metadata model that explains clearly the origin of data and the methods of estimation • ESPON is developing a tool for outlier checking that is very promising. • ESPON is elaborating time series and methods of update despite the changing territorial division • ESPON is a Program of reference concerning the evaluation of the effects of MAUP on statistical and cartographical results ⇨ **As a whole, ESPON proposes a very original and innovative material in the European statistical landscape**.
WEAKNESS	• The interface of the ESPON database is not as elegant as the one of other competitors. • The ESPON database is filled with indicators of various interest • The ESPON database is filled with duplicate of the same indicator which is a factor of confusion • Many data actually present in the ESPON database are not original and are just extracted from Eurostat or EEA. • Some data provided to ESPON are not precisely documented in terms of metadata (typically, the tables proposed by DG Regio) • The ESPON database is not currently and directly connected to cartographic or graphic application, like the OECD explorer. • The profile of users is not simple (scientific, policy makers, …) and it is difficult to find efficient solution for all of them • The copyright problems limit the diffusion of some

	interesting data (typically LAU2 data and their geometries) ⇨ **As a whole, the ESPON database suffers from an internal complexity inherited from the obligation to store too much variables delivered by project.**
OPPORTUNITY	• Inspire directive should help ESPON to remove some copyright problems • The political debate on the future of the Structural Funds has produced growing interest for ESPON data • The connections between maps and data will be reinforced by new projects on WEBMAPPING. • HyperAtlas is an original complement to ESPON database as it makes possible the creation of new variant of indicators. • The INTERCO and the MONITORING projects should increase the interest of policymakers for ESPON database. • The use of standard keywords (GMES) should facilitate the external diffusion of ESPON metadata and the attractiveness of the database. • The future publication of an ESPON Atlas (but also the regular ESPON Synthesis Reports or ESPON updates) is crucial factor of attraction, as long as data related to this publication are easily accessible. • Scientific reputation of ESPON database is high in the field of Computer Science, Cartography and Statistics and could be reinforced in the future. ⇨ **ESPON has a strong potential for becoming an international reference in the field of databases at European but also at World level.**
THREAT	• Contradiction between the external target of excellence and the internal target of cohesion. ESPON database should be more selective in the storage of data (internal excellence) but it would create problems if all the ESPON projects are not allowed to contribute (internal cohesion). • Contradiction between the objectives of scientific excellence and policy relevance. Concerning the selection of variables as well as the design of the interface, it appears difficult to fill at the same time the two objectives. A dual interface is not necessary a solution easy to implement. • Indetermination of the choice between two models for ESPON database: (1) the EUROSTAT model of *normative statistical power* (priority to official data, territorial units and typologies) and (2) the OECD model of *soft statistical power* (priority to innovation in terms of territorial division – non official mixture of NUTS – building of original typologies – urban-rural, innovative cartographic and graphic tools offered to general audience. • Difficulty to harmonize and connect the production and realization of the different Priority 3 projects of the ESPON Scientific Platform. Out of the problem of delay in the starting of the MONITORING project, too much pressure is put on the ESPON DATABASE project (phase I and II) to realize alone some tasks that are in practical terms from the responsibility of other P3 projects (webmapping, Interco, atlas, HyperAtlas). • Underestimation of the cost of platform integration and shortage of funds allocated to Priority 3 projects as compared to Priority 1 and Priority 2. It is for example very surprising that, as a whole, the projects HyperAtlas (75 k€),

	ESPON ATLAS (150 k€) and Webmapping tool (150 k€) represent no more than the majority of new priority 2 projects (360 k€) and less than half of all Priority 1 projects. The same is true for the flagship project INTERCO (400 k€) and MONITORING (600 k€) that represent together no more than some Priority 1 projects like SEGI. The important allocation of funds received by ESPON database (two times 1.300 k€) appears not so much when considering the amount of work expected and the duration of the period of activity. More, this project is unable to integrate alone the Scientific Platform without strong partners inside Priority 3. ⇨ **Until now, the ESPON Program has underestimated the importance of the ESPON Scientific Platform as a whole for the external promotion of ESPON program results. The major threat is to accumulate results (thanks to regular funding of new Priority 1 and 2 projects) without corporate strategy of storage and dissemination of data and maps.**

Table 2 - SWOT analysis on the external valorization of the ESPON Database and ESPON Scientific Platform (Priority 3)

As a whole, the conclusion of this SWOT analysis lead to the conclusion that ESPON Program should probably put more attention to the ESPON Platform which is recognized as one of its major output by a recent report of the European Parliament on the future of Cohesion Policy after 2013:

"*A final aspect of the territorial dimension concerns the generation of territorial data,analysis and learning. A key source is the ESPON 2013 Programme, managed by the European Observation Network for Territorial Development and Cohesion (ESPON). It aims to support policy development in relation to territorial cohesion by providing comparable Information, evidence, analyses and scenarios on territorial dynamics across the EU". (p.64) "The statistical and knowledge base on territorial development issues has been considerably advanced by the work of ESPON, providing valuable contributions to key policy documents, including the EU Territorial Agenda and the Third, Fourth and Fifth Cohesion Reports. The main criticism is that the quality of the study reports is variable and that their conclusions and policy implications are not always formulated in a user-friendly manner". (p.66)*

***Source** : **Mendez C., Bachtler J., Wishlade F., 2011**, "Comparative study on the visions and options for cohesion policy after 2013", European Policies Research Centre, Study realized for European Parliament's Committee on Regional Development. Administrative Coordination by Dr Esther Kramer Policy Department B: Structural and Cohesion Policies*

Looking at this document, they are no doubts on the fact that a success or failure of the Priority 3 projects would have a major consequence for the future of the ESPON Program after 2013…

The ESPON M4D project has proposed clear solutions to improve the external image of ESPON Program, through an agenda of quick and deep improvement of the database for both policy makers and scientists. But we cannot do the job alone and we need more cooperation with the other projects of Priority 3 in charge of development of cartographic tools, political monitoring, atlas … And we strongly suggest improving the support for

these Priority 3 projects that has received incredibly low amounts as compared to projects of less strategic interest for ESPON!

ANNEX : STRATEGY FOR THE CREATION OF NEW INDICATORS AND GEOMETRIES AT LOCAL LEVEL

N.B. This annex is the draft version of a technical report to be completed and delivered by TIGRIS team in June 2012. It is presented in advance on the request of ESPON CU who requested for more details on the Inception Report

Analysis of the situation

Different studies and our experience confirm that the main difficulties in the manipulation of the local data and indicators come from the fluidity of their geometry and from the heterogeneity of the information sources. One change in the geometry starts a chain-reaction in the list of indicators, starting with the basic one - the surface. In the case where we need to weight variables with some spatial relations (a neighbourhood matrix, for example), a new LAU2 in the list will ask a new set of spatial relations. When this geometry changes occur, new geographical information tables are needed, new stocking facilities in the computer, new formulas or calculus implementation; that's why we called it a "chain reaction".

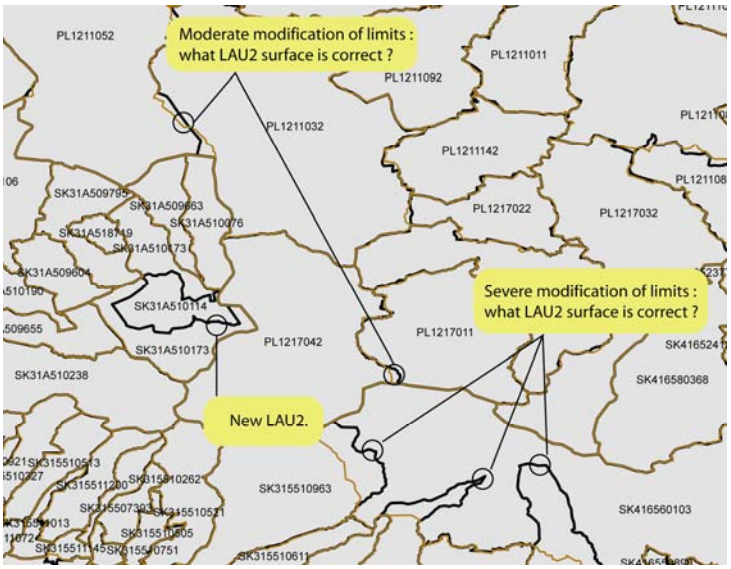


Figure A: Different geometries at LAU2 level. Green limits reflect the situation in 2004 and black limits the administrative frame in 2008. The illustration is extracted from the frontier zone between Poland and Slovakia.

When the administrative geometry is modified, interventions in the indicator's structure are needed. However, these changes of geometry don't occur with the same pattern and frequency in all the ESPON space, the differences between the countries and the administrative systems being considerable. One good example for this issue is Bulgaria, where the modifications of the LAU2 frame interfered with the changes of the LAU1 geometry and the NUTS2/3 polygons. Consequently, the coherence of the indicators was affected. Linking data to these fluid polygons is an issue just as complicated as the administrative modification's management. Previous explorations of official data providers indicate a high degree of heterogeneity in data accessibility. Some National Statistical Institutes are more advanced than others in the quantity and quality of data offered. Nevertheless, there are many limitations in the data collection (availability of data at a certain geographical scale, format of data, indicators generally provided for 2001 and limited update) and no synthetic harmonization concerning the sense of the indicator, the time reference and the geographical level. For the moment, the available geometries at local level refer to the administrative frames of 2001, 2004, 2006 and 2008. Naturally, any link between data collected from the NSI and the basemap will have to comply with these chronological reference points. In this case, the components of our approach (geometries, indicators at local level and data availability) start interfering and the eventual result can be compared with a cluster of indicators projected in a 3D graph with three axis: time (2001-2011), space (all countries in the ESPON space and all the LAU2) and thematic (land use,

demography or other indicators). This cluster will present "regions" of data and indicators' concentration and empty spaces in the graph, where no link was possible between the three components.

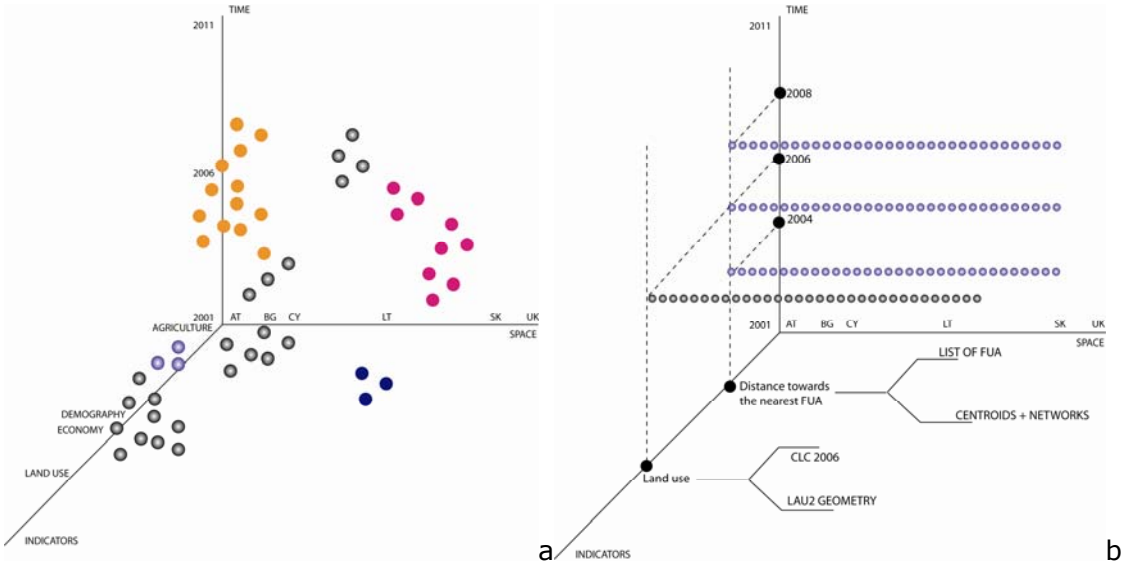


Figure B: Synthetic illustration of three components in the local data collection: time, space, indicators. Their interference will determine the quality and quantity of a local database. Amorphous data integration (a) vs. a more structural strategy (b).

Taking into account the fact that we have only 4 or 5 basic geometries and a large amount of combinations between indicators, countries and local scales of geographical representation (LAU1 or LAU2), we can imagine that the shape of the point's cloud will be amorphous, just like the previous figure illustrate. A different approach, based on a more structural strategy of the data collection, leads us to three new possibilities concerning the local database:

- the integration of a coherent indicator, independent to the national definitions (for example, distance towards the nearest FUA)

- a complete spatial cover of the ESPON space, using the LAU2 as spatial reference

- the chronological problems are eliminated (we no longer need to mix data from two different years)

Of course, there is always the possibility to see the local database as a mix between amorphous data collection and structurally built indicators. The distinction we mentioned is just the reflexion of methodological choices that we will make during the work.

Having in mind this particularity of the local data manipulation, we will canalize our approach on three main objectives.

Objectives

The main activities that the TIGRIS team will implement in the time frame of this project refer to the scientific exploration of new relevant indicators at local scale, to the data and indicator collection and to the networking activities. As the main philosophy of the M4D project focus on the quality of the indicators to be offered, our strategy will follow this line of action and it will insist on the reliability of the local data to be integrated.

1. The first component in the scientific exploration of data at local scale will **put into relation the VIGO (the very important geographical objects) with the LAU2**

frame and data. By VIGO we understand spatial units, such as the European network of FUA, some major transportation corridors and equipments. The output of this first exploration will provide indicators related to the potential of polarization by the VIGO, at local scale, together with some insights concerning their local proximity. When the results will be compared at different scales (regional, national or in a cross-border context), the image that we have about local territories will be better filtered. One difficulty that is easy to imagine stays in the implementation of the calculus models for areas with a high fragmentation of the administrative frame (the case of France, but not only). Regionalized and dedicated models will be provided in order to outcome this difficulty.

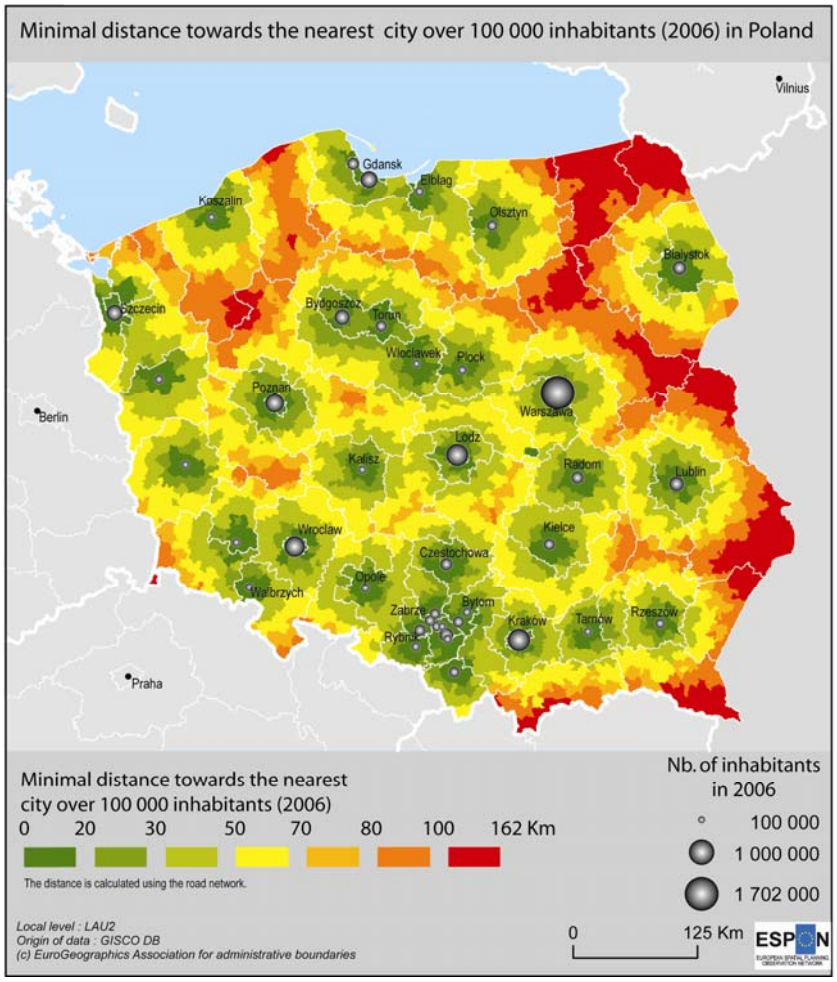


Figure C: The closest VIGO (cities with more than 100000 inhab. In 2006) in Poland - distances in the road network (draft map)

From the perspective of the database construction at local level, the ESPON space is basically divided in points and polygons (geometries). These points and polygons are linked with "lines" which are commonly known as networks. As a matter of fact, these networks have no reason to exist but to put into relation the basic spatial structures already mentioned. To these geometries we will attach data - core variables and indicators derived from these variables (location quotient derived from the built-up area by LAU2). In the space of the nowadays Europe some points behave like very important geographical objects (VIGO). The word "important" is the key-term that shapes our approach because all the geographical objects could look important. Consequently, its definition/s should clarify at least two issues:

1) Why this subject (VIGO) should be taken into consideration by the policy makers?

2) How do we create a hierarchy between the geographical objects that populate the ESPON space?

The answer to the first question is crucial for the development of a classification concerning the hierarchy of the VIGO (the 2nd topic in this argumentation). If one defines the territorial cohesion as a principle of equitable access to the providers of public services having social and economic impact, despite the territorial belonging of the consumers, she/he might find that these services are very important geographical objects, both for policy design and territory itself. The maps can clarify this statement.

In the national context, the relation between VIGO (universities) and the LAU2 geometry allows us to better understand **where** the lack of territorial cohesion manifests, in this case the red regions. This is not necessarily a map of accessibility because it does not intend to put into relation a resource (universities) with the population aged between 18 and 32 years (a plausible *bassin de clientele* for this service).

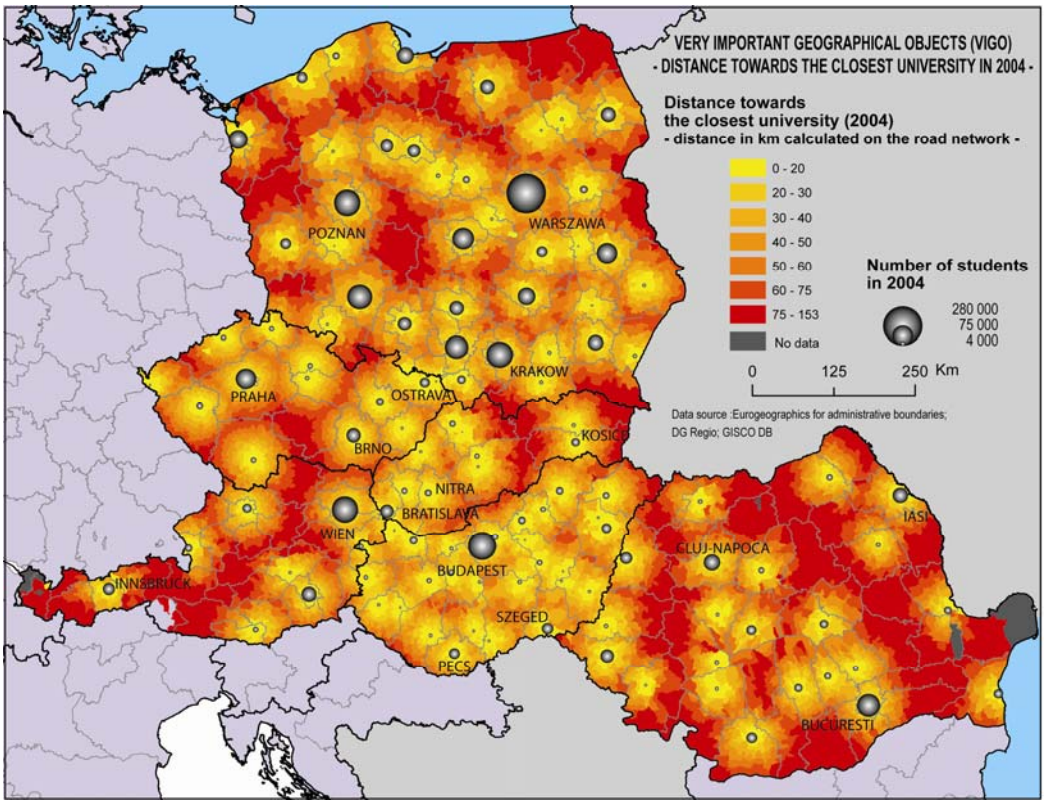


Figure D: Universities as VIGO in selected countries from the Eastern Europe (data for the situation of 2004 - draft map)

Nevertheless, it is quite clear that persons belonging to a territory covered by the red class of distances are penalized by a higher cost of transportation or a higher time of commuting to the universities. At a superior territorial level, we might observe that these red regions have a lower level of tertiary education attendance and difficulties in coping with some of the competitiveness objectives. As the population distribution in space largely overlays the proximity of the metropolitan areas endowed with universities, one might think that the territorial cohesion issue is not an issue at all, at least from the perspective of the distance to the nearest universities.

If we change the scale of analysis, the problems induced by the relation between VIGO and LAU2 frame show a transnational pattern of manifestation. Poland will look as a balanced case if compared with Romania but not if compared with Hungary. While Slovakia is longitudinally divided, the Czech Republic concentrates its problems in the regions between Ceske Budejovice and Brno. Again, red regions behave like areas where the lack of territorial cohesion can be flagged as an issue, at least from the perspective of the relation between VIGO (universities) and LAU2. Different national patterns interfere

on this map and different types of local spaces are present here. Consequently, different approaches are legitimated in order to better understand the issues of the territorial cohesion, as long as the quality and quantity of data allows it. In the same logic, other VIGO can be mobilized in order to explore the organization of local territories - airports, FUA, transportation nodes etc.

The second question (how to create a hierarchy of VIGO) complicates the situation. An eventual answer will have to balance the feasibility of the approach (health infrastructure is essential but almost impossible to obtain) and relevance (some VIGO acting like the touristic points of interest are easy to find but not relevant, being VIGO only for a minority). Both theory and empiric studies agree that there is a strong correlation between the spatial distributions of different services that act like VIGO and the spatial distribution of cities.

Fig. Potential interaction with cities over 7000 inhabitants - a case study on NUTS2 North-East in Romania (draft map)

Different kinds of cities concentrate hierarchical services. It is the distribution of cities in the territory that can be used as a frame of analysis for the relation between LAU2 and territorial services providers. The quantity and quality of services provided by cities is different from country to country. There is some regularity to observe: superior services concentrate in metropolitan areas, the urban entities with smaller population will present a limited offer. Choosing a good method of discretization within this hierarchy allows us to create new and relevant indicators regarding the relations between the LAU2 and VIGO. It is not always a good option to map these indicators in a classic way (choropleth). Using combinations of different indicators on the map, one can better seize the connection between densities, urbanization and distances on the road network, all combined at regional or national scale. In the same time, using the VIGO as a reference frame in the construction of indicators will help better understand how the territories are organized. The distribution of the local economic performance (estimated GDP at LAU2 scale in 2006) is not depending only on the demographic mass (population, employment or number of firms). Analysis might show that the spatial distribution of this indicator can be explained also by other geographical aspects. In the case of Poland, the local economic performance might be affected by several factors - the average distance towards the nearest university and the road density at LAU2 level. As the location of universities is generally a matter of metropolization and the road density is a component of the territorial endowment, we can interrogate how strong the relation between these three indicators is. A classical regression will show that this relation is rather weak ($R2 = 0.27$) and that the distribution of the estimated local GDP in 2006 is only partially explained by the territorial endowment and distances to VIGO (universities or selected metropolitan areas). Changing the scale of analysis can offer more information. If we implement our regression in a moving window of the first 100 neighbors for every spatial unit on the map (LAU2), we can measure the performance of the relationship at local scale. The cartographic output shows that the distribution of the economic performance is a matter of sub-regional territorial structures, with three cases of figure:

a) a weak relation will indicate that the local level of administration is indifferent to the territorial endowment and to the presence of metropolitan areas nearby. The economic performance (or the lack of economic performance) follows other explanation schemas.

b) a moderate or fair relation will show that for some LAU2 the proximity to an university (a superior service) and the road density (an indicator of territorial endowment) can be interpreted as a local stake. Basically, for these regions the objective is to become a "positive residual" - more economic performance that the road density and the distance to a VIGO (university, in this case) let us predict.

c) a strong local coefficient of statistic determination shows that there are territories where the local economic performance is strongly determined by the combination between distances towards universities and the territorial endowment, at least from a statistical point of view. In these sensitive territories, being far or being closed, having a superior density or a reduced one, explain why the LAU2 lack or benefit of economic performance.

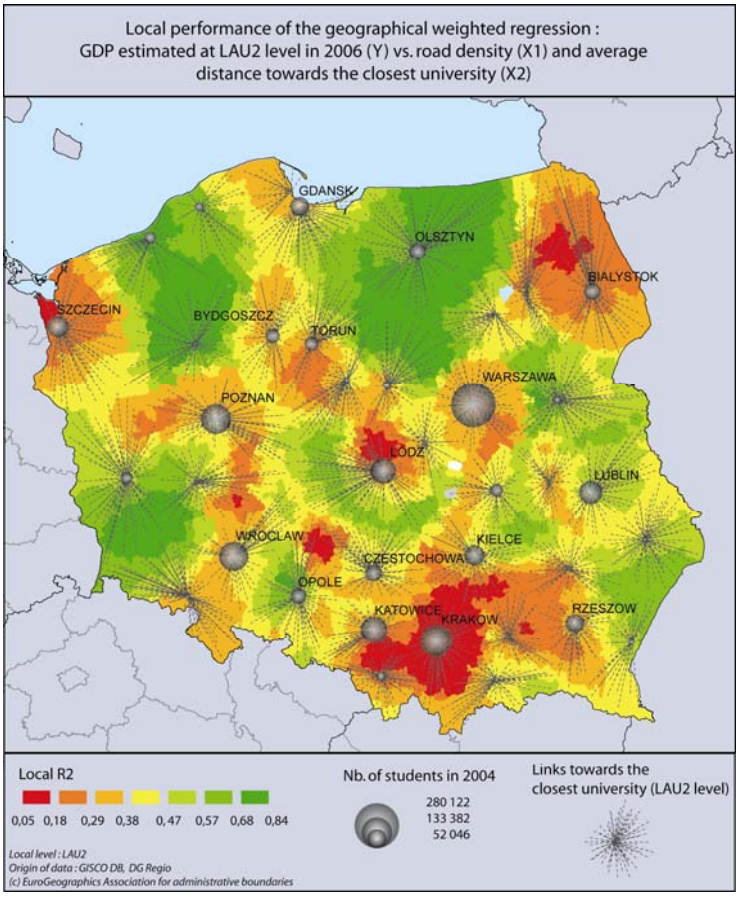


Figure E: Local performance of a geographical weighted regression - economic performance, distance towards VIGO (universities) and territorial endowment (road density)

At this moment, the exploration of the relation between VIGO and the LAU2 frame invites us to propose a hypothetical classification of the possible outputs of our work. Every case in this table represents a set of indicators that can be extracted after the implementation of some specific methods of data collection and analysis.

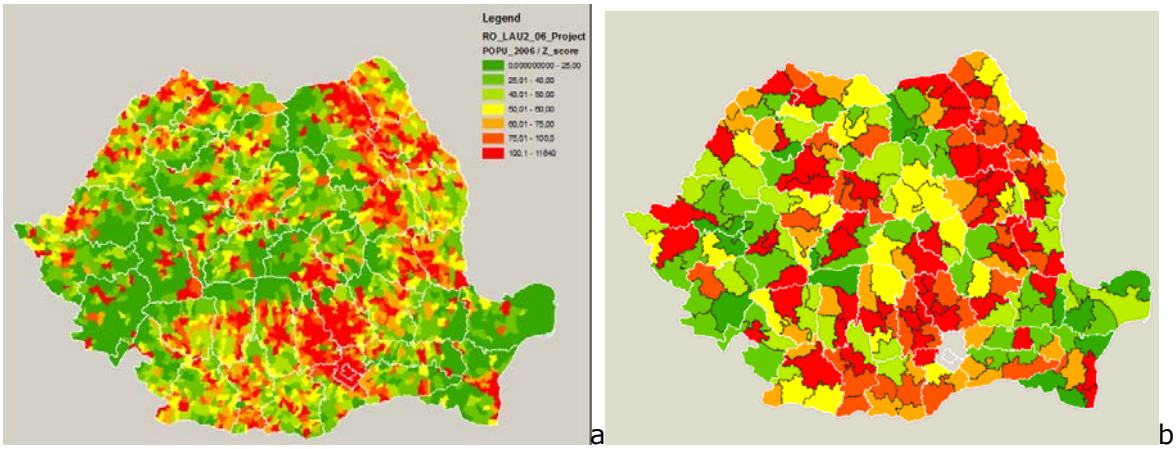
OUTPUT/VIGO	Independent spatial point structures	VIGO based on LAU2 frame	Output of ESPON typologies and projects
Descriptive map as output	- distance towards the nearest university or airport; - distances towards the nearest major road nodes	- distances towards the nearest cities over 15 000 inhabitants (for example). - distances towards the nearest LAU2 with administrative functions (NUTS2/3)	- distance towards nearest FUA or MEGA - time travel towards the nearest FUA or MEGA - average time travel or distance between the LAU2 and the n closest FUA or MEGA
Typologies based on the relation between VIGO and LAU2 frame	- typologies of LAU2 based on distances towards the closest VIGO	- typologies of LAU2 based on distances towards the closest VIGO with administrative functions	- typologies based on cluster analysis (average time travel towards the nearest FUA, in different territorial context)
Explanatory variables	- construction of variables to be used in different statistical models (OLS, GWR, theoretical areas of polarization)	- construction of variables to be used in different statistical models (OLS, GWR, theoretical areas of polarization)	- construction of variables to be used in different statistical models (OLS, GWR, theoretical areas of polarization)

Table A: possible output of an analysis based on the relation between VIGO and LAU2 frame

As the definition of VIGO must take into account some scientific and methodological aspects, combined with an effort of data and sources exploration (availability of data, new geometries such as the universities, the airports, the transportation nodes or updated typologies of FUA), it would be premature to offer one or more definitions of VIGO. The core of an eventual definition will focus on the meanings of the term *providers of public services with social and economic impact* and the possibility to delineate the VIGO as its direct application.

2. The second component of the scientific exploration of data at local scale will be based on the **construction of an alternative geometry,** relevant for the local scale and able to bypass the issues of the mass effect (differences of surface). The LAU2 scale might not be always relevant in the study of some territorial aspects, especially when the analyzed phenomena are independent to the national administrative boundaries. The NUTS3 scale is also problematic (MAUP), but an intermediate level of analysis that combines different geometries has many chances to better capture the spatial dimension of the indicator that one will want to map. Basically, it is a regionalization problem at local scale and, as in any regionalization problem, choosing the good set of criteria will prove crucial. This alternative level of geographical geometry works like a visualisation tool rather than an operational one and there is no intention to replace the NUTS/LAU level with it. Its utility stays in the possibility to compare different patterns of spatial distributions in two different frames, in order to extract different and (maybe) better scientific conclusions.

With more than 120000 LAU2 in the ESPON space, the methodological aspects in the construction of this alternative geometry are crucial. Our strategy aims to choose a proper mix of NUTS3, LAU1 and groups of LAU2 that can be relevant as an alternative geometry. The NUTS3 are related to a problem of size (surface). At this moment, the NUTS3 geometry of Belgium, Netherland or Germany looks like the best choice we have. Not all the countries present a LAU1 frame, complicating our decision steps. However, even in the countries endowed with LAU1 geometry, we will encounter problems induced by the equivalence between the NUT3 and the LAU1, in terms of size or surface. Grouping LAU2 in an alternative geometry might look like the most complicated part of our approach. As a matter of fact, grouping is the simplest part of the work. Choosing the good set of grouping criteria involves a choice between two methods: minimization of internal dissimilarities and identification of centres of theoretical polarization; it is an illustration of an old, but still actual, geographical debate between the sense and the role of homogeneous vs. polarized regions. In the next cartographic example, we grouped the LAU2 units taking into account the distance towards the closest city or LAU2 with polarization potential, letting the "expert opinion" being the criteria which leads to this spatial configuration. As an alternative method, working with a location-allocation model that minimizes the distances towards the nearest facility (central place) or maximizes the number of theoretically attracted individuals is a possibility that deserves a test.



a b

Figure F Density of population at LAU2 and FUNC_local levels

Reducing the "noise" - LAU2 aggregation in superior spatial units (LAU1 like) based on the proximity between LAU2 and local poles of theoretical attraction. Density of population in 2006 - LAU2 scale (a); Density of population in 2006 - "LAU1 like units" based on theoretical polarization (b) Draft maps.

The method we proposed relies on two elements: the definition of distances and the definition of attraction/polarization, the last one depending on the distances. Euclidean distances are easier to estimate, the distances in the road network demands for a larger quantity of time; beside the files must be properly projected. The polarization is not only a function of distance; it also involves demographic or economic mass to be included into the equation. A classical model (Reilly's model of market area estimation) is not recommended in spaces where the regional distribution of the rank-size rule is obviously macro-cephalic because the results will be in contradiction with the major goals of the regionalization - an alternative geometry with almost the same surface. As the polarization centres seem to be homogeneously spread in the geographic space, this method has some advantages - easy to calculate (if the centres were properly chosen), equal size of the output - LAU1/NUTS3 like alternative geometry, easy to theorize. However, blindly applying this method is a risk. Equal surface size does not necessarily mean equal population and the demographic or economic differences between the resulted geometries can become an issue of major importance, in the data aggregation.

The second method is less depending on the polarization centres; it functions like a regionalization based on the concept of spatial dissimilarity (absolute difference between two contiguous/neighbour spatial units). Homogeneous LAU1/NUTS3 like regions should minimize the internal differences, respecting one criterion (one indicator) or a set of indicators. Easy to implement with only one constraint (dissimilarity minimization), it leads to difficult solutions when we add one more constraint - the contiguity of the spatial units should be respected during the regionalization. Basically, we need homogeneous regions that respect a topological rule (no "archipelago" region allowed). In this case, the definition of the dissimilarity via a matrix might be less interesting than the use of a dissimilarity tree, where double interrogations related to proximity and differences are possible. As the tree has a starting point (spatial unit A, in our case), changing this starting point will provide different solutions. Testing these solutions with a territorial auto-correlation test (the closest to 1, the better the solution) will help choosing the best option.

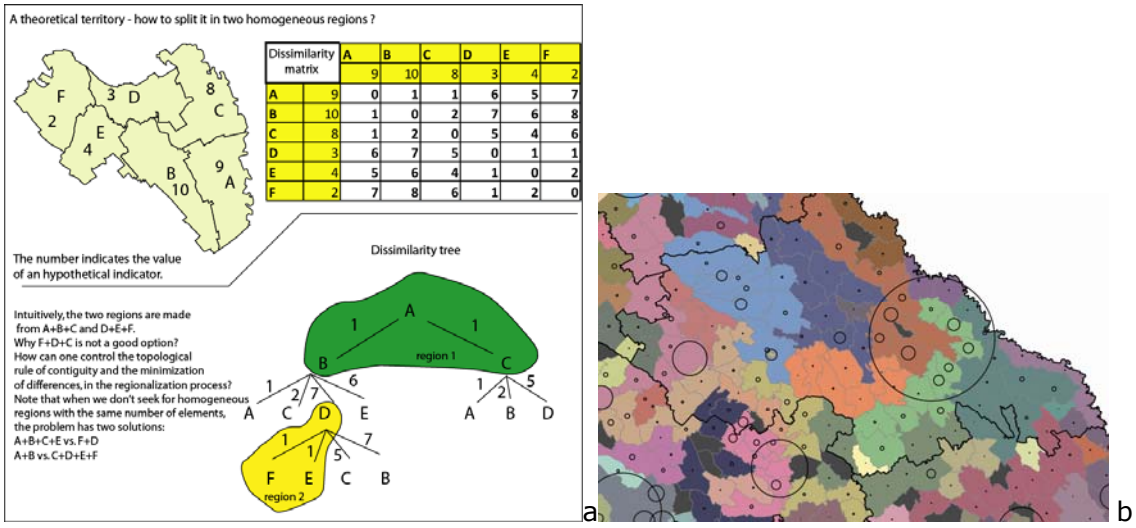


Figure G – A simple regionalization problem

a. Illustration of a simple regionalization problem: how to cluster spatial units in two regions, minimizing differences and conserving contiguity? b. The size of the proportional symbol indicates the magnitude of the indicator included in the analysis (Gross turnover in 2006 - Romania). The image is the output of a regionalization algorithm based on a dissimilarity tree.

Excepting the contiguity rule, new constraints can be introduced in order to better seize the territorial structures at local level. For example, one new rule could forbid the LAU2 that belong to a certain NUTS3/NUTS2 to make a new region with LAU2 included in a different NUTS2/3 (territorial belonging). Major spatial discontinuities (rivers, frontiers, different historical regions) can supplementary be included. As the strategic goal is to provide an alternative geometry, the tactical solution on the "field of operations" (different countries and different spatial patterns of LAU2 manifestation in the analysis) will involve the application of one of the two methods. Theoretically, the case were we will be forced to find a new solution (homogeneously polarized regions) is not to be excluded, as a distinct possibility. The algorithm we described is still in preparation, working with a limited number of LAU2 (about 1000 spatial units, one indicator of dissimilarity). The examples we proposed illustrates that the construction of alternative geometry is not so exotic, after all. The output will not replace official frames of NUTS and LAU, it will only provide a tool able to verify where the maps and the indicators are distorted by a geometry effect.

3. **Integrating data concerning the land use patterns** at local scale will be the continuation of the process started in the previous database project. For the moment, using the LAU2 geometry and the Corine Land Cover data, only 5 countries from the Eastern Europe present indicators. We will enlarge the collection of this data for other ESPON countries, aiming to cover relevant transnational regions with specific land use patterns. The indicators can then be used in order to perform typologies at local scale, offering also basis for an alternative analysis of the specific trends in the European rural spaces.

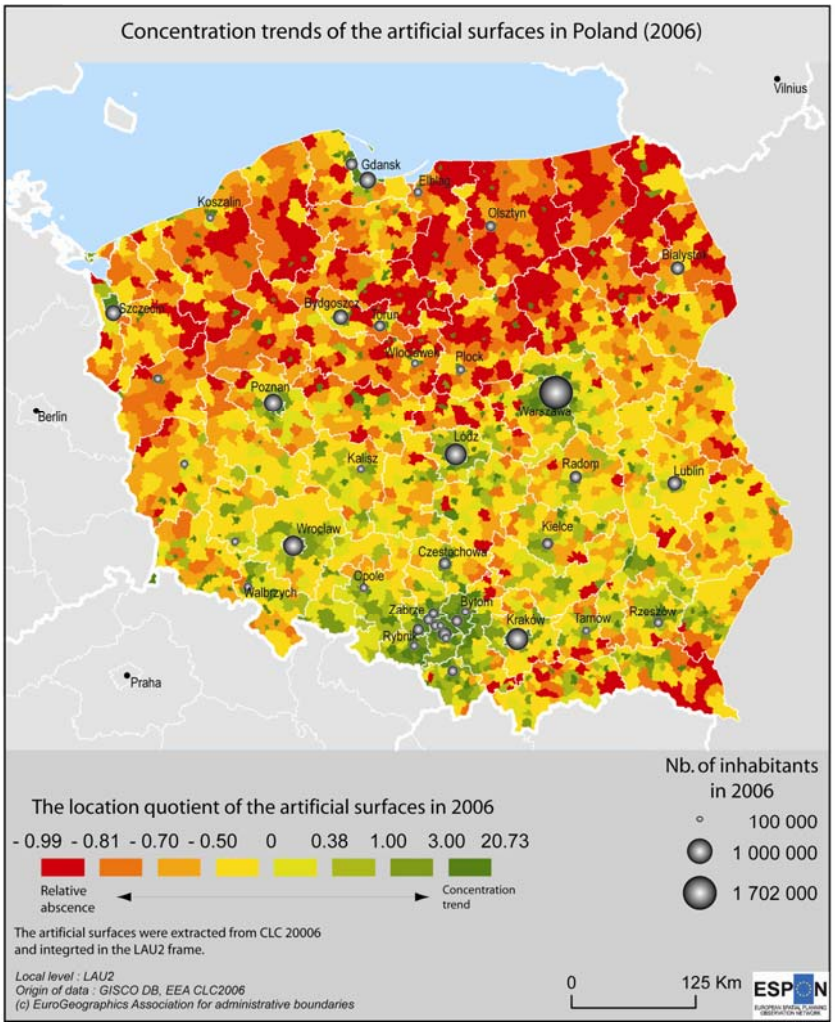


Figure I: Concentration trends of the artificial surfaces in Poland (2006)

Concentration trends of the artificial surfaces - Location Quotient applied after data integration from CLC 2006 in the LAU2 frame in Poland

Major steps in the construction of indicators related to the land cover at LAU2 scale:

- Data and geometry check; proper projection of the GIS files.
- Intersection of CLC 2006 vectors with the LAU2 geometry.
- Calculation of land cover surfaces (artificial surfaces in sq.m., for example).
- Calculation of basic relative indicators (share in % of the artificial surfaces at LAU2 scale).
- Construction of new indicators based on surfaces - location quotient of artificial surfaces.
- Mapping the indicator - location quotient of artificial surfaces.
- Error check and aberrant values delineation.

This approach is not overlapping work already in progress (ESPON EU-LUPA); the idea of integrating land use indicators is more likely to work as a paradigmatic goal in many projects, due to the stakes that different teams attach to this concept. As a matter of fact, the scientific convergence on this subject is the output of one of the numerous intersections between scientific and policy interest that prioritize integration of data describing the land use patterns. In this logic, as an overlapping of work is excluded, intersecting different approaches on the subject presents a double interest, for the teams involved in this topic. The general aim of the ESPON EU-LUPA project is to understand a complex mechanism such as the interaction between spatial and temporal land-use dynamics and the socio-economic changes in Europe. Working at different NUTS levels of analysis and with selected case studies, this project will identify the main challenges derived from the land cover patterns and it will provide relevant policy options and recommendations. The major opportunity for networking, in this case, is to integrate indicators and data derived by EU-LUPA at local scale or to propose alternative methodologies and geographical scales of investigation. The matter of scale of analysis is one major methodological difference in the approach that we propose. In the ESPON M4D project, we intent to provide indicators of land-use at LAU2 scale only, even if they can eventually be aggregated at superior LAU1 or NUTS3 scale. In many aspects, we use the same data sources as the ESPON LUPA project (CLC) for different geometries. One other aspect that makes our work different is the fact that we lack the dynamic vision of land use that guides LUPA's intentions. The recommendation is to contact the EU-LUPA project in order to explore common visions and methodologies regarding the subject of land use patterns. This contact will be executed as soon as we will advance with the data integration for some new European countries (for moment, only data on Poland, Czech Republic, Slovakia, Hungary, Romania, Bulgaria, Spain and Portugal are available). Covering a greater territory will certainly smooth the methodology we used and will be a topic of interest for EU-LUPA.

❖ *Deliveries – Timetable*

Despite the advanced collection of data at local level (SIRE, Eurostat, ESPON DB) and despite the existence of several geometries that can eventually fit these data, the systematic analysis of how this data can be used and modeled is still in progress. One major issue related to the exploitation of data at LAU2 level is the constant check of the results (proper match between geometry and data, topological rules to respect, projection issues, a good choice of the LAU2 "markers" of results), a time consuming operation that can be estimated only in a limited number of cases.

When dealing with the land use integration as data describing local geometries, we cannot properly estimate the time that takes to collect data for one country, for a simple reason: we cannot extrapolate from 9 countries (already integrated) to another

one. If we want to integrate Corine Land Cover 2006 data for France, we have to deal with 36 000 spatial units and different patterns of land use (very complex in some rural areas, more simple in other regions). The basic operation for this kind of data collection is an intersection between two vector files, an operation that technically is not a problem on reasonable spatial samples. France might need to be split in regions and the new files will provide the basis for the application of the methodology. Let's imagine that we split France into four regions. There are chances that the algorithm we use works for the first three and not for the last. Consequently, we will be forced to split again the last region in order to make it work. This readjustment is also time-consuming because merging all the information after the operation was done requires new steps in the data management.

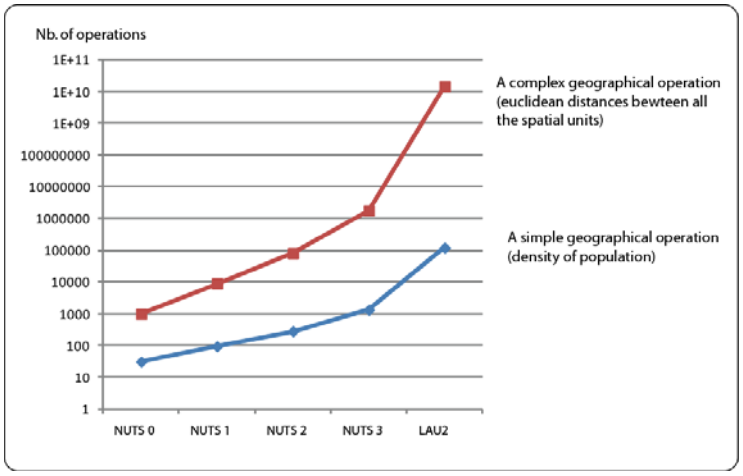


Figure J : Comparing simple and complex geographical operations at different scales

That's the main reason for which we have preferred to structure the deliveries in function of the main events in the project's lifetime - FIR, SIR, Final Report and the intermediate periods. At the moment we wrote our contribution, we considered that this method could be also an opportunity to prioritize the data inflow in the database. The list of deliverables could also be a topic of discussion with the Lead Partner and the TPG, taking into account the strong networking dimension of our project.

The set of deliveries will be based on a core of indicators that we consider relevant for a better description of the territorial structures, at local level. These indicators are organized on thematic topics of interest:

Land use data integration in the LAU2 frame

1. Area and share of the artificial surfaces => location quotient of the artificial surfaces. This indicator will provide information on the degree of human pressure on the landscapes.

2. Area and share of arable surfaces => location quotient of the arable surfaces. This indicator will provide information related to the organization of the agricultural system at local level.

3. Area and share of permanent cultures =>location quotient of permanent cultures. This indicator describes local territories with a long stability of the agricultural systems.

As the location quotients put into relation the local share with the regional one, different types of analysis can be implemented, in order to seize the concentration trends at NUTS2, national or supra-national scale.

Integration of data related to the networks

1. The road density at LAU2 scale. It can offer indirect information related to the degree of local connectivity; it is sensible to the LAU2 surface. The map will eventually depict areas with strong territorial endowment, from a transportation point of view.

2. Typology of LAU2 based on the proximity to major road transportation corridors. The indicator will provide information about the capacity of the local territories to link with these corridors, in a national and supra-national context. Different

approaches of the proximity are possible (distance or contiguity). The indicator is sensible to the size of the LAU2 and to the degree of administrative fragmentation.

3. Typology of LAU2 based on the proximity to major rail-road transportation corridors. A combination between these two typologies will help to identify the local territories where the remoteness and the lack of connectivity is a major stake.

Integration of data related to the VIGO

1. Average distance towards the closest 3 universities. The universities are superior services that can be found in privileged FUA or MEGA. Their distribution in space can be related to the repartition of potential users (individual and firms). The indicator and the maps will depict this relation, taking into account the presence of "interposed opportunities" - more than 1 university.

2. Distance to the closest FUA. As the FUA are classified in function of their territorial importance, measuring the distances to different categories of FUA is the base for a more complex typology of the LAU2 units.

The list of core indicators is a method to prioritize the work on the local data; it does not mean that we will reduce our work only to this set of indicators.