ESPON 2013 DATABASE

FIRST INTERIM REPORT

2009 February 27

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

This first interim 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 Liechteinstein. 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 ongoing ESPON projects.

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

List of contributors to the first interim report

UMS RIATE (FR)
Claude Grasland*
Ben Rebah Maher
Ronan Ysebaert
Christine Zanin
Nicolas Lambert
Bernard Corminboeuf
Chloe Didelon

LIG (FR)
Jérôme Gensel*
Bogdan Moisuc
Christine Plumejeaud
Marlène Villanova-Oliver

UAB (ES)
Andreas Littkopf
Juan Arevalo
Roger Milego

IGEAT (BE)
Moritz Lennert
Didier Peeters

UMR Géographie-cités (FR)
Anne Bretagnolle
Hélène Mathian
Joël Boulier
Timothée Giraud
Marianne Guerois

TIGRIS (RO)
Octavian Groza
Alexandru Rusu

Université du Luxembourg (LU)
Geoffrey Caruso

National University of Ireland (IE)**
Martin Charlton
Paul Harris

National Technical University of Athens (GR)**
Minas Angelidis

Umeå University (SE)**
Einar Holm
Magnus Strömgren

UNEP/GRID (CH)**
Hy Dao
Andrea De Bono

* Scientific coordinators of the project

** Experts

TABLE OF CONTENT

Organisation of the first interim report

At first, and after consultation with the ESPON Coordination Unit (CU), the aim was to produce a short report (max. 60) where only major information is reported and where details that are not of prime interest are rejected to different annexes. But we deceided to overcome this limit for 2 reasons: (1) inclusion of illustrations making the document more attractive. (2) in depth discussion of important cross-challenge topics like metadata and map-kit tool.

The aim of the first interim Report (Part 1) is an introduction where we precise the legal expectations to be fulfilled by the project and to addresse the specific request made by the ESPON CU after the delivery of the first Interim Report (1.1). It also describes what are the most important evolutions of the project that have been decided since the inception report in order to reach the objectives and answer to ESPON CU requests (1.2).

The review of challenges (Part 2) is the core part of the report that provides synthetic information on the work done so far. Each challenge is organised in the same way (objectives, results, difficulties, workplan) and can be read independently. Connexions between challenges are clearly identified and help the reader to navigate between each of them[1]. A first group of challenges is related to the production of specific datasets or specific expertise on different types of geographical objects: collection of basic data at regional level (2.1), harmonisation of time series (2.2), enlargement of regional data toward global (2.3) or local (2.4) levels, combination of social and environmental data (2.5), and collection of urban data (2.6). A second group of challenges is more closely srelated to data flows, both external (2.7) and internal (2.8), with the target of production of an integrated data model that can be implemented as a computer application (2.9). The involvement of the expert team is related to the specific description of new challenges that are related to spatial analysis tools for quality control (2.10), collection of data on neighbouring countries (2.11) and exploration of individual data and surveys (2.12).

The transversal questions (Part 3) are related to specific deliveries of the project like the ESPON Mapkit tool (3.1) or to questions of common interest that involves all partner teams, like the elaboration of a common strategy for metadata (3.2).

The conclusion (Part 4) defines firstly the agenda of the project for the next period of 12 months until second interim report in February 2010. Special attention is paid to the ESPON seminars of Prague (June 2009) and Sweden (December 2009) that are crucial milestones for the publication or the dissemination of new results. It proposes

[1] Due to contractual obligation, the report has to be delivered in paper format, but an HTML file would be more convenient for an easier "navigation" between challenges.

some synthetic tables of objectives and deliverables and addresses finally some specific questions to the ESPON CU.

The Annexes (Part 5) provides more details on specific topics.

1 Aim of the first interim report

1.1 Expected content (legal obligations)

The content of the first interim report is firstly delineated by the legal obligations defined in the Subsidy Contract (SC) and the Response on Inception Report (RI) sent by ESPON CU the 24 October 2008. This points are quoted below as **SC1** to **SC5** and **RI1** to **RI7**

February 2009 (1st Interim Report)

[SC1] Presentation of the results of the test to be undertaken within the ESPON community in order to assess the database compliance with the objectives initially defined and its user friendliness towards researchers, policy makers and practitioners working at different geographical levels. (cf. point V, 3).

[SC2] Delivery of a consolidated version of the ESPON 2013 Database (internal and public versions) and of a compatible ESPON map kit tool, taking also in consideration the results of the test and evaluation stage (cf. point V, 3).

[SC3] Presentation of a timetable for regular updating and ESPON 2013 Database, including statistical validation of data sets delivered by other ESPON projects, updating of data and indicators, delivery of data for ESPON publications and possible update or adjustments of the ESPON map-kit tool.

[SC4] Short reporting of the networking activities, both planned and realised, at internal (with ESPON 2013 projects) and external level (with European and international organisations with relevant data for ESPON).

[SC5] Work plan until 2nd Interim Report.

Points to be improved during the project implementation and to be addressed in the First Interim Report

[RI1] Presentation of an overall work plan including a more detailed overview on the activities and the expert teams involved, as well as the respective timetable.

[RI2] On challenge 1 (page 12-14). The Lead Partner is requested to precise the list of indicators considered as "basic indicators". In addition, the Lead Partner is asked to present the current situation of the ESPON 2006 database and define immediate needs for updating (cf. annex III to the contract, point k)

[RI3] On challenge 3 (page 16). The Lead Partner is considering improving the WUTS System provided by ESPON 2006 project 3.4.1 – Europe in the world. It is important to mention that it is envisaged in the near future to open a call for an ESPON project dealing with the world scale. Therefore, the Lead Partner of the ESPON database is requested to take this information into consideration and to cooperate with this project in order to avoid an overlap of work.

[RI4] With regard to challenge 5 (page 18), the Lead Partner is asked to better explain it. The objectives are not given; the cooperation envisaged between ESPON and EEA is not clear, in particular the practical meaning of the following sentence needs to be clarified: "Therefore, the problem is not to duplicate the work realised by EEA but to introduce a flow of data exchange between ESPON and EEA and to build common data infrastructure in order to ensure full compatibility of database on each side".

[RI5] Challenge 6 (page 19-20). The construction of complex geographical objects of higher level is aimed. This challenge is explained using cities. No other examples are mentioned. Considering the time frame and the complexity of the object "cities", it is suggest that this challenge will be focussed only on cities.

[RI6] Challenge 7 (page 21), it would be important to have a more concrete idea on the networking activities to be developed with the different organisations mentioned. In addition, the repartition of tasks between UMR RIATE and UL should be made clearer.

[RI7] Challenge 9 (page 34). It should better describe. It has no name, no objective, no timetable.

[RI8] Components of the application (page 31)

i. The description of the import pool seems too ambitious. Please check that all the verifications mentioned for importing data will really be undertaken.

ii. On page 33 it reads: "In order to overcome these issues, a simplified database will be set up in the more advanced stages of the project". What do you mean with "simplified version" and with "advanced stages of the project"? Please be aware that a public version of the ESPON database should already be delivered by November 2008.

iii. In addition and according to the project specification, the Lead Partner should ensure "usability" to the ESPON 2013 Database. In particular "the application should be user-friendly and make the users understand which data is available". In particular for "non-experts" on data issues.

iv. In relation to the hosting of the application and management of the server resources, the Lead Partner is requested to consider the following: The ESPON Programme will host the application developed in all stages of the project and access to the ESPON 2013 database will only be given through the ESPON website (public database) and the ESPON intranet (internal database). In relation to this issue, the Lead Partner is requested to comply with point f) of the Annex III to the contract, which says: "the project will provide, as soon as possible, a more detailed technical description of the requirements for hosting the database. Furthermore, the project will describe, in the inception report, a procedure with a time table to keep the database on the ESPON server up to date".

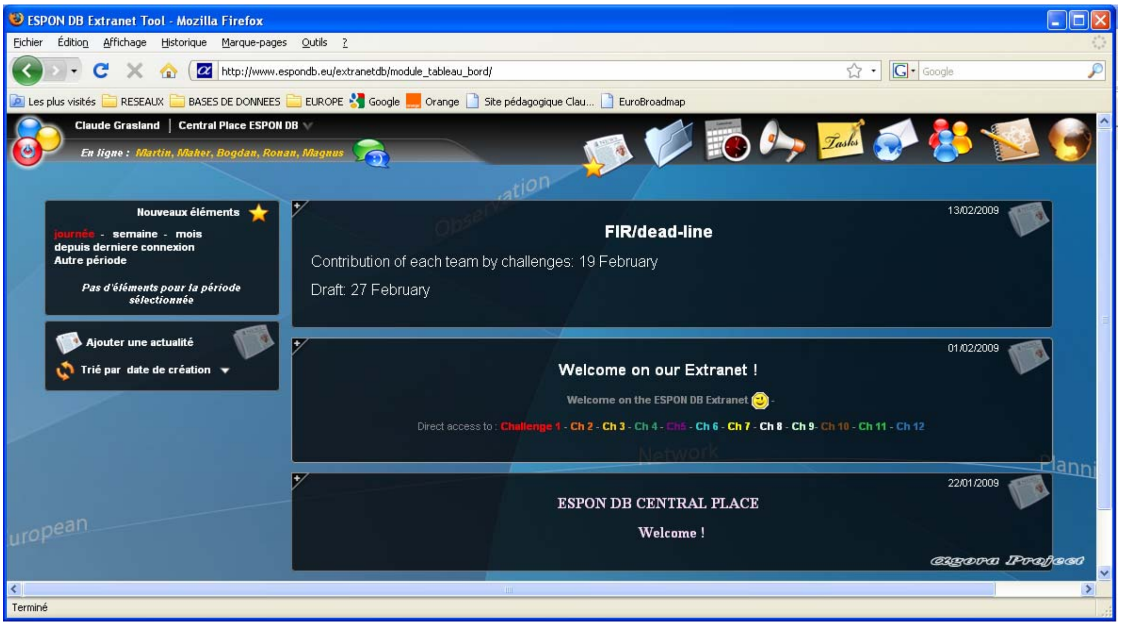
1.2 Clarifications of ESPON DB's objectives

An internal meeting has been organised in Paris the 2-3 Feb. 2008 with all the project partners and the expert teams, in order to summarize the results of the work done so far, to prepare efficiently the First Interim Report (FIR) and to organize the work for the next 12 months until the Second Interim Report (SIR). The ESPON seminar of Bordeaux in December 2008 has been a first opportunity for the project partners of ESPON DB to meet each other and to exchange with the other ESPON projects under Priority 1 and Priority 2. In this section, we summarize the main conclusions of the internal meeting and the way they have contributed to clarify the orientations of the project and to provide answers to the questions to be addressed in the FIR (see. 1.1).

1.2.1 An internal organisation by challenge

The presentation of the results of ESPON DB project by challenge (Bordeaux Seminar, Paris meeting) has proven to be very efficient. It gives a clear idea of results of the test phase in order to assess the database compliance with the objectives initially defined and its user friendliness towards researchers, policy makers and practitioners working at different geographical levels **[SC1]**. As each project partner is responsible for at less one challenge, its contribution is more visible and the internal and external networking of the ESPON DB project is more visible and efficient **[SC4]**. Moreover, it is easier to define the workplan and the objectives of the project for the next period **[SC5]** because each project partner has to identify the contributions and deliverables that are under its direct responsibility. It is also easier to provide answers to request of clarifications addressed by ESPON CU to specific challenges **[RI2, RI3, RI4, RI5, RI6, RI7].**

One possible danger of this organisation by challenge could be a lack of integration of results at project level. But it is not the case because the internal seminars but also the Extranet (opened in Feb. 2009, see Figure 1) give to partners the opportunity to exchange their discoveries and to identify connexions and areas of common work between challenges (as shown in Figure 2).


ESPON DB Extranet Tool - Mozilla Firefox
Fichier Édition Affichage Historique Marque-pages Outils ?
http://www.espondb.eu/extranetdb/module_tableau_bord/
Google
Les plus visités RESEAUX BASES DE DONNEES EUROPE Google Orange Site pédagogique Clau... EuroBroadmap
Claude Grasland | Central Place ESPON DB
En ligne : Martin, Maher, Bogdan, Ronan, Magnus
Tasks
Nouveaux éléments
journée - semaine - mois
depuis derniere connexion
Autre période
Ajouter une actualité
Trié par date de création
Observation
Network
Planni
uropean
Agora Project
Terminé

Pas d'éléments pour la période sélectionnée

13/02/2009

FIR/dead-line

Contribution of each team by challenges: 19 February

Draft: 27 February

01/02/2009

Welcome on our Extranet !

Welcome on the ESPON DB Extranet 🙂 -

Direct access to : Challenge 1 - Ch 2 - Ch 3 - Ch 4 - Ch5 - Ch 6 - Ch 7 - Ch 8 - Ch 9- Ch 10 - Ch 11 - Ch 12

22/01/2009

ESPON DB CENTRAL PLACE

Welcome !

Figure 1 - The Extranet of the ESPON DB project (Feb. 2009)

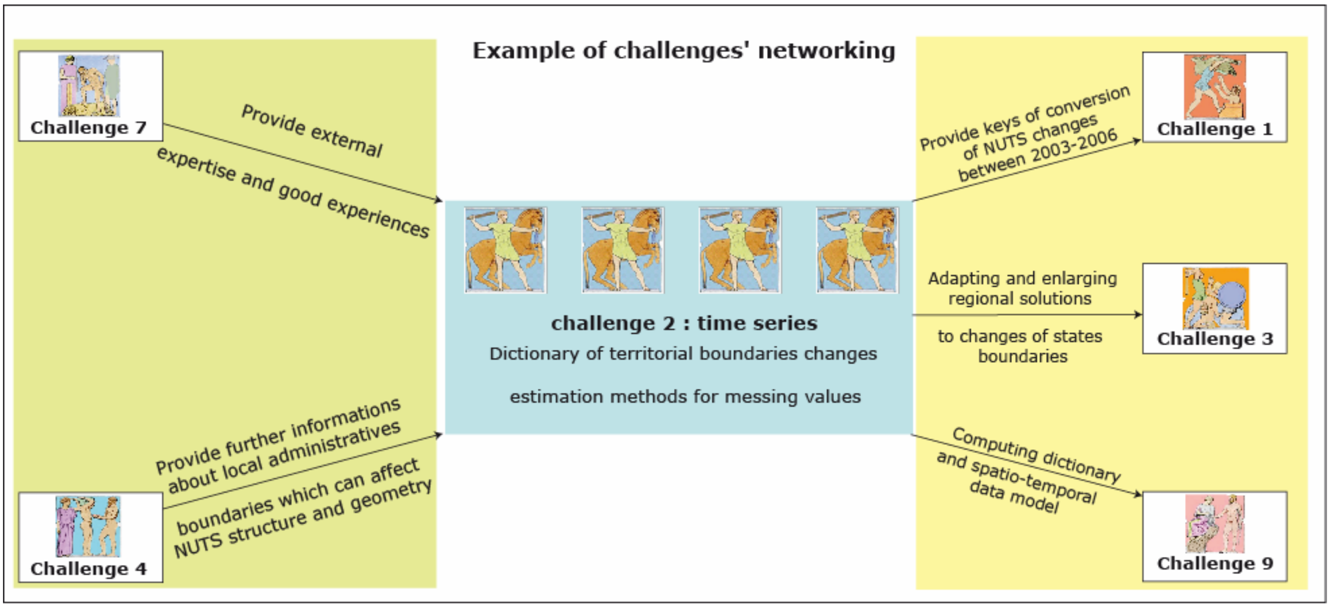

Example of challenges' networking
Challenge 7
Provide external expertise and good experiences
Challenge 4
Provide further informations about local administratives boundaries which can affect NUTS structure and geometry
challenge 2 : time series
Dictionary of territorial boundaries changes
estimation methods for messing values
Provide keys of conversion of NUTS changes between 2003-2006
Challenge 1
Adapting and enlarging regional solutions to changes of states boundaries
Challenge 3
Computing dictionary and spatio-temporal data model
Challenge 9

Figure 2 - Example of challenges' networking (Feb. 2009)

1.2.2 Two types of deliverables : Indicators and Technical Report

Since the meeting in Paris, some clarification has been made about what can be delivered by the ESPON DB project to the ESPON community and to external world. Until the Paris meeting, it was admitted that deliverables were mainly "databases" with different components (statistical information, geometries, computer application for data management).

More precisely, it was admitted that one indicator of performance of the project ESPON DB should be the elaboration of "indicators", but this word was relatively unclear as it can cover different meanings. For some researchers, "indicators" can be understood as an opposition between "raw count data" (e.g. population, GDP, area, …) and "relative measure of intensity" (e.g. population density, GDP per capita, …) that can be used for the measure of territorial units of different sizes. But we can object to this point of view that size criteria like population and GDP can be sometimes precious criteria for the evaluation of regional trends. Another point of view could be to consider "indicators" as new data elaborated by an organization, that were not previously available or that have undergone some transformation resulting in a clear added value. It is clearly the semantic point of view of OECD that publishes datasets of "regional statistics and indicators". These data are generally derived from national or international agencies, but their added value is related to the harmonization done by OECD, in particular through the definition of harmonized regional levels. If we adopt this point of view, **an ESPON indicator** could be defined as "**an integrated set of <u>statistical data</u> and <u>geometries</u> harmonized by ESPON, documented by <u>metadata</u>, with a clear added value as compared to initial informations**".

But it was also clear that the deliverables of the project ESPON DB can not be limited to "data" and are also related to the "Know how" of how to integrate data (Figure 3). That is the reason why an important decision of the Paris meeting was to launch a collection of **ESPON DB Technical Reports** that describe **how to solve specific problems of data integration** that can not be fully explained in the very brief description that are usually given in metadata files. In the elaboration of a timetable for regular updating of the ESPON database **[SC3]** and in the definition of the Workplan **[WP4],** we have clearly introduced the delivery of Technical Reports as important milestones (see conclusion 4.2).

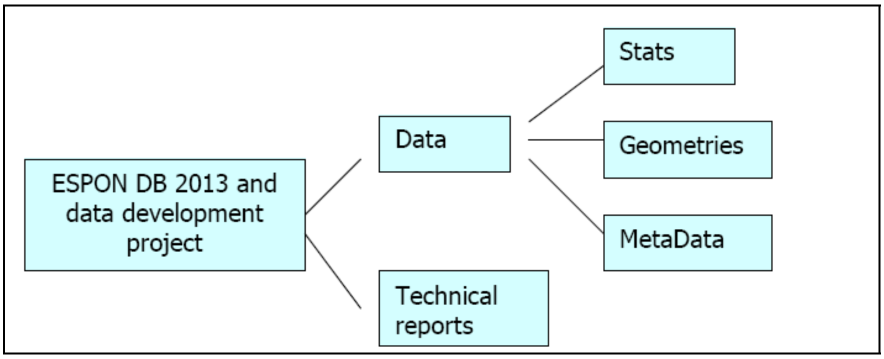


Figure 3 - The two types of deliverables of ESPON DB project

1.2.3 Dataflows and metadata

In the inception report as in the presentation of the ESPON DB project made at the ESPON seminar in Bordeaux, the CU pointed some ambiguities in the definition of the so-called "Internal" and "External" database **[SC2, RI8]**. More generally, the question of metadata was considered as crucial, both for input in the ESPON database (from other ESPON projects, other organisation) and for output (toward other ESPON projects, other organisations) and it appeared urgent to provide strong guidelines on this issue **[SC4, RI6]**.

The distinction between "Internal" and "External" database was clarified by ESPON CU that explained during the Paris meeting that the distinction between the two databases is firstly related to **copyright issue**. The external data are the one that are not protected by copyright and can be therefore disseminated out of the ESPON community. At the same time, it appeared also that the content of the "External" database can be considered as an **ESPON publication**, subject to quality control and a form of official stamp as it engages the collective responsibility and the reputation of the ESPON program. The **metadata** that are related to external publications of ESPON data should be therefore extremely precise and fully INSPIRE compliant, in order to make possible their dissemination. On the basis of this discussion, it was decided that external database should be based, in the initial period, on the publication of **fixed tables** and not on an interactive computer application where users can download data without any pre-definite form. The interactive consultation of data stored in the ESPON Database will define the "Internal database" where the access is limited to ESPON members.

Based on the need of the final users (internal and external databases) we have redesigned the organisation of dataflow (see Figure 4) and launch a working group on metadata that has provided efficient guidelines for integration of new data in the ESPON database, either from external organisation or from other ESPON projects. In order to test the efficiency of this rules for metadata and data checks, we have decided that each responsible of challenges 1 to 6 will introduce himself a set of basic data in order to provide models of each type (regional, world, local, cities, grid) for other ESPON projects.

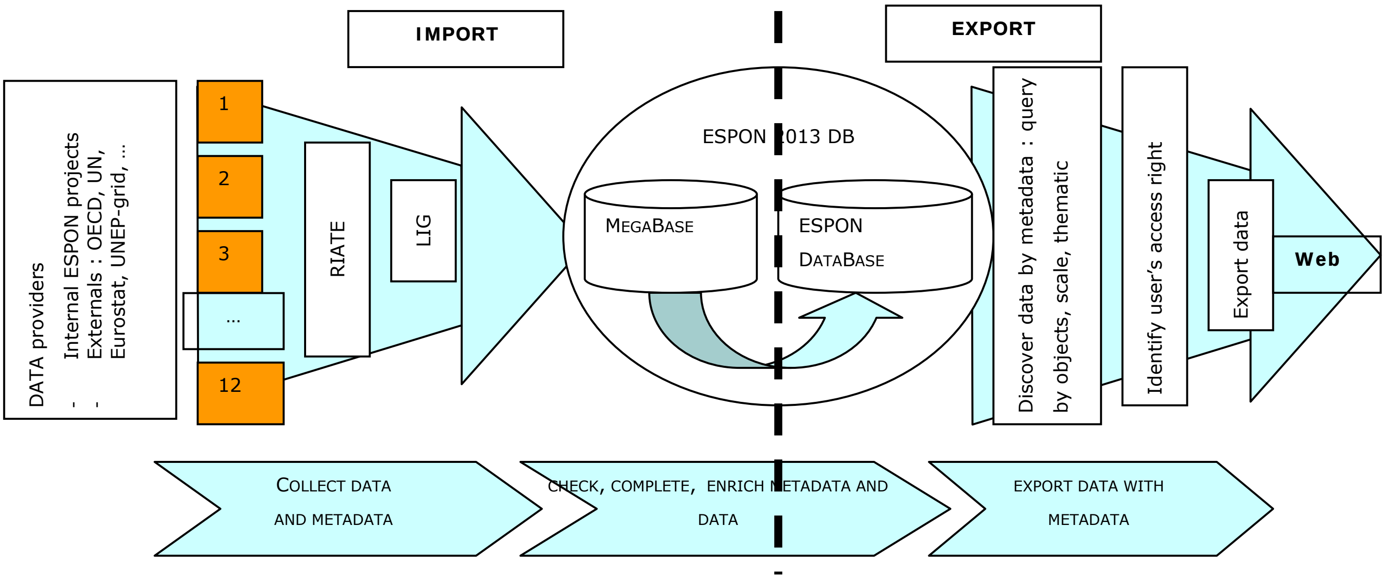

IMPORT
EXPORT
DATA providers
- Internal ESPON projects
- Externals : OECD, UN, Eurostat, UNEP-grid, …
1
2
3
…
12
RIATE
LIG
ESPON 2013 DB
MEGABASE
ESPON DATABASE
Discover data by metadata : query by objects, scale, thematic
Identify user's access right
Export data
Web
COLLECT DATA AND METADATA
CHECK, COMPLETE, ENRICH METADATA AND DATA
EXPORT DATA WITH METADATA

Figure 4 - Overview of data flows

2 Review of the challenges

2.1 Challenge 1: Collection of basic regional data

Coordinator: RIATE

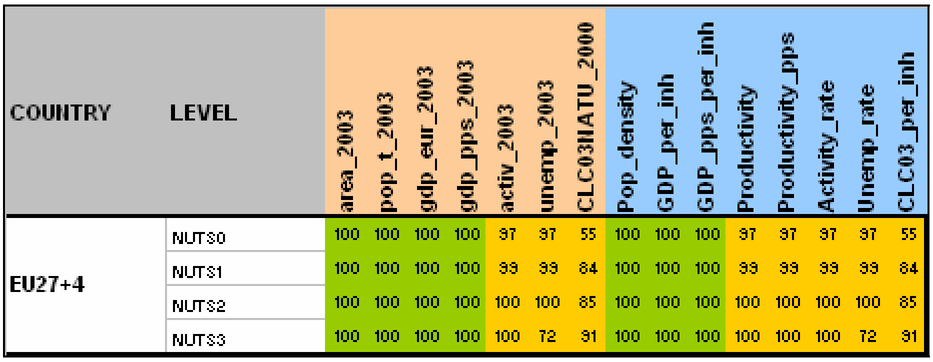
Delivery of basic datasets derived from EUROSTAT and EEA at NUTS2 and NUTS3 levels according to NUTS2003 and NUTS2006 divisions.

2.1.1 Objectives

The production of harmonized datasets covering all the ESPON space (31 countries) at NUTS 2 or NUTS 3 level has been recognized as the first challenge to be solved with an absolute priority as it is a condition of continuity with previous work realized in ESPON 2006 program. It is obvious that the new ESPON 2013 project needs immediately basic information at this level like area, population, GDP, employment, which will be used as reference for more sophisticated analysis where these projects will produce more precise information in their specific fields. Moreover, the map kit tool that will be sent to these projects (see. Section 4) should not be limited to purely geometric information and should involved this basic data sets as starting point and model for more elaborated data collections. Finally, we should be able in a short delay to connect the new information elaborated by ESPON 2013 Program with former datasets elaborated by ESPON 2006 Program in order to produce time series of indicator, with the objective to support projects on the monitoring of European territory.

2.1.2 Work done

The data collection has begun in the NUTS 2003 version, where the data availability was the most important thanks to last downloads from Eurostat centralized at UMS RIATE and the previous ESPON database. Some basic indicators have been collected: GDP, population, area, unemployment, active population and land use in 2003. The collection of this information has made it possible to compute them in order to develop some basic ratios: GDP per inhabitant, population density, unemployment rate etc. The variety of the sources existing concerning NUTS 2003 version allows having a good quality of completeness of data (fig. 5).



COUNTRY	LEVEL	area_2003	pop_t_2003	gdp_eur_2003	gdp_pps_2003	activ_2003	unemp_2003	CLC03NATU_2000	Pop_density	GDP_per_inh	GDP_pps_per_inh	Productivity	Productivity_pps	Activity_rate	Unemp_rate	CLC03_per_inh
EU27+4	NUTS0	100	100	100	100	97	97	55	100	100	100	97	97	97	97	55
	NUTS1	100	100	100	100	99	99	84	100	100	100	99	99	99	99	84
	NUTS2	100	100	100	100	100	100	85	100	100	100	100	100	100	100	85
	NUTS3	100	100	100	100	100	72	91	100	100	100	100	100	100	72	91

Figure 5 - Degree of completeness of the indicators collected in NUTS 2003 version

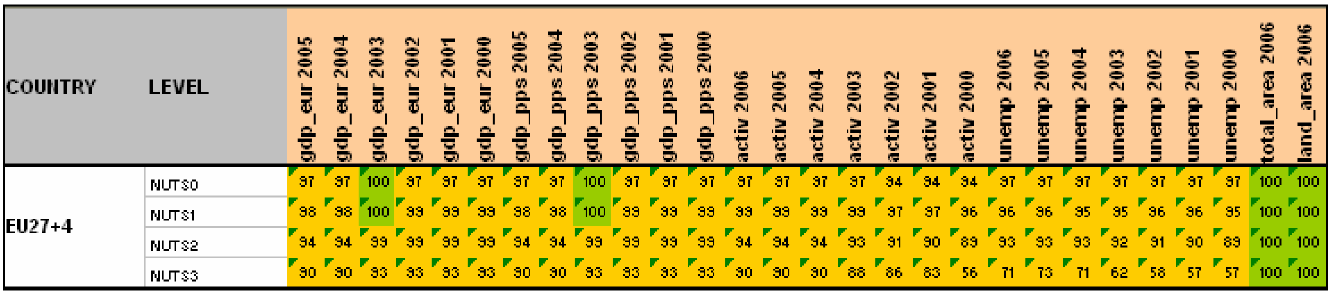
The next step of the work has been to extend the data collection at NUTS 2006 version. Three main ways have been investigated:

A) Download on Eurostat of the same basic indicators (GDP, Unemployment, area) and its evolution on a time-period of 5 years (2000-2005 or 2006).

B) Try to have a complete dataset from NUTS3 to NUTS0 for total population 2000-2006. It implies to overcome the problem of missing values and making some data estimations.

C) Check and integration of data from ESPON Territorial Observation No.1 with computing the results obtained at different NUTS level.

A) The idea of the download of the basic indicators was to follow and extend the previous integration in NUTS3 division. Follow, because the same stock indicators were uploaded and extended considering that it was tried to make possible the calculation of evolution. No estimations have been implemented here (except for land use); i.e. the table down (Figure 6) is a sum up of the availability of the data on Eurostat website in February 2009. The fact is that it is very difficult to have complete dataset for these indicators for the moment.



COUNTRY	LEVEL	gdp_eur 2005	gdp_eur 2004	gdp_eur 2003	gdp_eur 2002	gdp_eur 2001	gdp_eur 2000	gdp_pps 2005	gdp_pps 2004	gdp_pps 2003	gdp_pps 2002	gdp_pps 2001	gdp_pps 2000	activ 2006	activ 2005	activ 2004	activ 2003	activ 2002	activ 2001	activ 2000	unemp 2006	unemp 2005	unemp 2004	unemp 2003	unemp 2002	unemp 2001	unemp 2000	total_area 2006	land_area 2006
EU27+4	NUTS0	97	97	100	97	97	97	97	97	100	97	97	97	97	97	97	97	94	94	94	97	97	97	97	97	97	97	100	100
	NUTS1	98	98	100	99	99	99	98	98	100	99	99	99	99	99	99	99	97	97	96	96	96	95	95	96	96	95	100	100
	NUTS2	94	94	99	99	99	99	94	94	99	99	99	99	94	94	94	93	91	90	89	93	93	93	92	91	90	89	100	100
	NUTS3	90	90	93	93	93	93	90	90	93	93	93	93	90	90	90	88	86	83	56	71	73	71	62	58	57	57	100	100

Figure 6 - Degree of completeness of the indicators collected in NUTS 2006 version.

B) The Eurostat data on population development (2000-2006) were lacking in some cases (DK, UK, PL…), namely at NUTS2 and NUTS3 level. On top of that, some values appeared probably false (discontinuities in time series, cf annex 1). The work of the ESPON 2013 Database project has been first to estimate missing values. Secondly, to identify some discontinuities of values in the evolution of population for each NUTS in order to point out some strange behaviour. In deed, the ESPON 2013 Database project proposes full dataset at NUTS3 (figure 7), NUTS23, NUTS2, NUTS1 and NUTS0 for total population from 2000 to 2006 and has marked strange values with flags in the dataset.

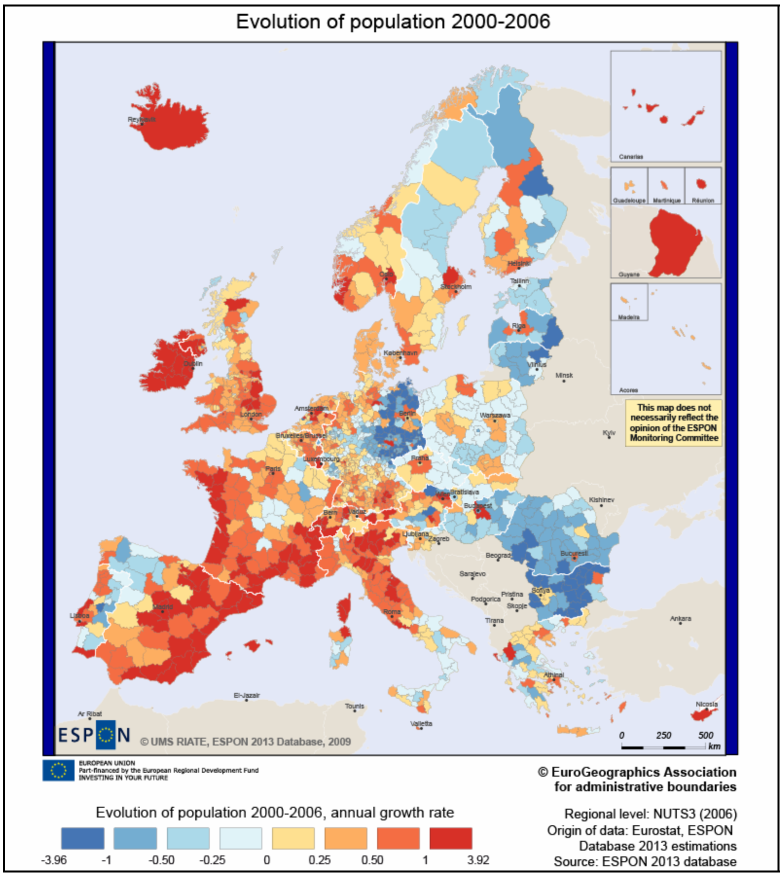


Figure 7 - Evolution of population (2000-2006), NUTS3

C) The integration of data from other ESPON projects is a fundamental point for ESPON 2013 Database. That has been done with data coming from ESPON Territorial Observation (see figure 8). The first step has consisted to check carefully data then some mistakes have appeared (cf annexes 1). After exchanging views with the data provider, the problems encountered has been corrected. After this, the aim has been to re-estimate the indicators created at NUTS23 level in the other official level of NUTS: (NUTS2, NUTS1 and NUTS 0).

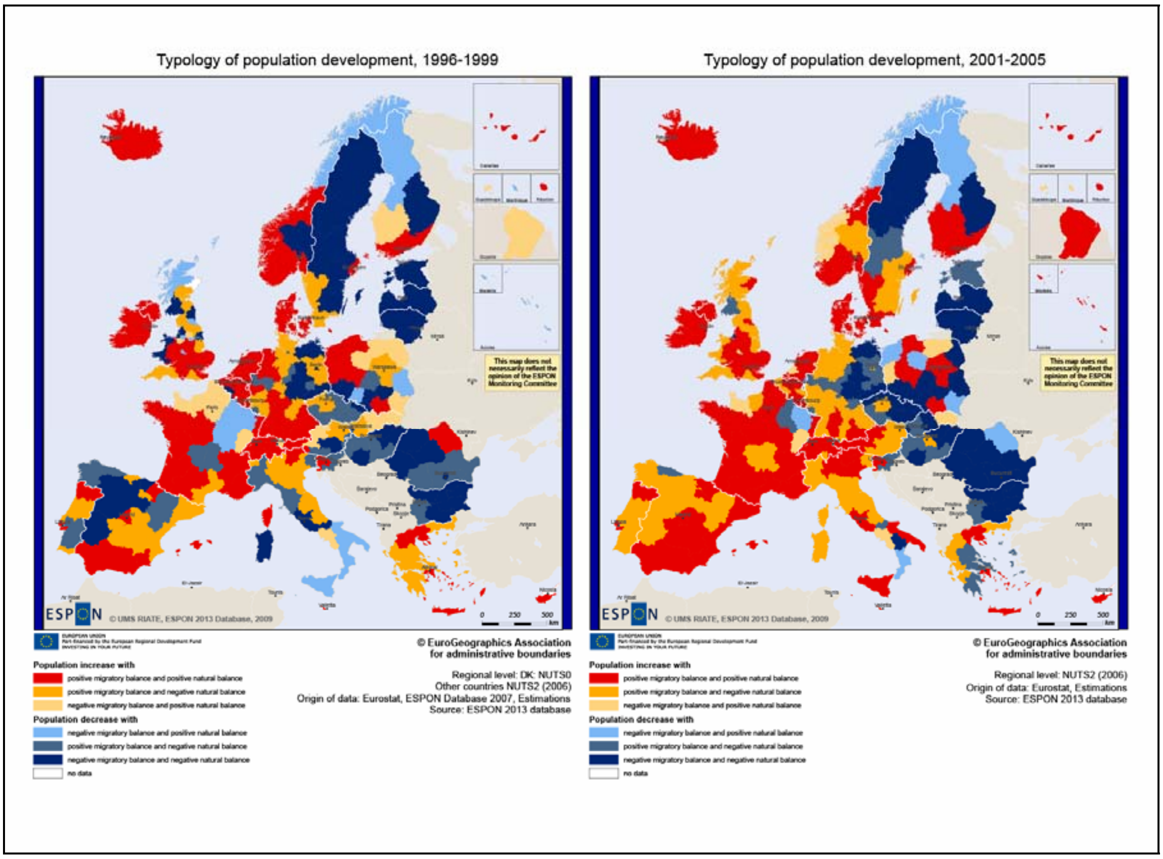
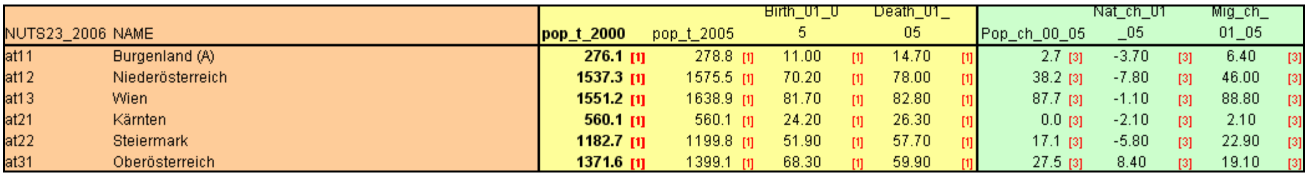

Typology of population development, 1996-1999
Typology of population development, 2001-2005
ESPON
© UMS RIATE, ESPON 2013 Database, 2009
This map does not necessarily reflect the opinion of the ESPON Monitoring Committee
0 250 500 km
© EuroGeographics Association for administrative boundaries
Regional level: DK: NUTS0
Other countries NUTS2 (2006)
Origin of data: Eurostat, ESPON Database 2007, Estimations
Source: ESPON 2013 database
Regional level: NUTS2 (2006)
Origin of data: Eurostat, Estimations
Source: ESPON 2013 database
Population increase with
positive migratory balance and positive natural balance
positive migratory balance and negative natural balance
negative migratory balance and positive natural balance
Population decrease with
negative migratory balance and positive natural balance
positive migratory balance and negative natural balance
negative migratory balance and negative natural balance
no data

Figure 8 -Typology of population development at NUTS2 level

This information has been integrated in the internal database. The metadata is described at the level of the value in order to see immediately which values are official (Eurostat) and which values have been estimated (ESPON projects). The tables that have been checked will be presented in the external database as a form of synthetic tables available at different geographical scales (Figure 9).



NUTS23_2006	NAME	pop_t_2000		pop_t_2005		Birth_01_0 5		Death_01_ 05		Pop_ch_00_05		Nat_ch_01 _05		Mig_ch_ 01_05	
at11	Burgenland (A)	**276.1**	**[1]**	278.8	[1]	11.00	[1]	14.70	[1]	2.7	[3]	-3.70	[3]	6.40	[3]
at12	Niederösterreich	**1537.3**	**[1]**	1575.5	[1]	70.20	[1]	78.00	[1]	38.2	[3]	-7.80	[3]	46.00	[3]
at13	Wien	**1551.2**	**[1]**	1638.9	[1]	81.70	[1]	82.80	[1]	87.7	[3]	-1.10	[3]	88.80	[3]
at21	Kärnten	**560.1**	**[1]**	560.1	[1]	24.20	[1]	26.30	[1]	0.0	[3]	-2.10	[3]	2.10	[3]
at22	Steiermark	**1182.7**	**[1]**	1199.8	[1]	51.90	[1]	57.70	[1]	17.1	[3]	-5.80	[3]	22.90	[3]
at31	Oberösterreich	**1371.6**	**[1]**	1399.1	[1]	68.30	[1]	59.90	[1]	27.5	[3]	8.40	[3]	19.10	[3]

Figure 9 - Example of diffusion table

2.1.3 Identified difficulties

Even if this challenge has tried to overcome the difficulties raised by missing values in NUTS 2006 division for the most common indicator (total population), some questions or problems are still not solved concerning this point:

It will be difficult to guaranty the estimations of missing values of the other basic, indicators, because it implies both a long treatment chain and to ensure the compatibility between the different tables (for example, if we estimate the age-pyramid of each region of ESPON space, it is important to take care of the equality of values between the different tables).

An estimation method has been chosen for total population, based on spatial and temporal extrapolation from a thematic point of view and on linear trends from statistical point of view. It is not the single method which can be used.

What strategy adopting for official values which introduce mistakes in the dataset? The annex 1 proposes some possible solutions but the answer is still open.

Then, considering the intra-ESPON data exchanges, some dangerous practices have been noticed. In order to avoid this, it is fundamental to define a protocol of data downloading and indicator building.

2.1.4 Work plan

In order to follow the results and problems raised by the work done, four main fields will be tested and improved for the Second Interim Report (February 2010).

[June 2009]

Delivery of complete indicators at NUTS2 level (GDP, Population, Unemployment, …)

Continue to check and integrate dataset from other ESPON projects or expertises (ESPON Territorial Observation No.2?)

[Dec 2009]

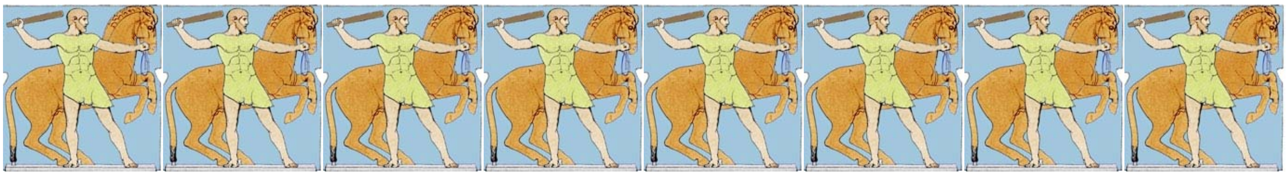
Integration of accessibility indicators and at least, road time distance matrix (if received)

Try to enlarge the integration of two basic data and area - to other geographical objects and scales: World, cities, grids (exchanges with challenges 3, 5 and 6).

[Feb 2010]

Try to define a methodology to detect spatial and statistical outlier in these basic datasets to point out extraordinary values (exchanges with challenge 10)

2.2 Challenge 2: Harmonization of time series



Coordinator: IGEAT

Harmonization of time series for basic socio-economic indicator at regional level for the period 1995-2006.

2.2.1 Objectives

Based on the result of challenge 1, we propose to elaborate a methodology for the harmonization of time series covering ESPON territory at regional level for the period 1995-2006 on the basis of simple indicators of regional policy (population, GDP, unemployment, age structure). The problem is not to cover immediately a great number of indicators but to define a methodology that could be implemented in the ESPON 2013 DB and reproduced by different ESPON projects.

2.2.2 Work done inventory and benchmarking (expertise) of sources and experiences

The first step of the work consisted in enumerating and collecting the different sources that could be relevant (interest) to harmonizing temporal NUTS versions. We have also examined some attempts to create temporal GIS of administrative boundaries' changes. We have focused on how these projects had approached the problem of creating-variant GIS of changing boundaries and how they storage changes.

The harmonization of NUTS geometries is based on a meticulous combination of several sources. The most important are:

The Official Journal of the European Union is the legal source. It constitutes the juridical framework of regulation of NUTS since 2003[2] (see annexe 2).

EUROSTAT provides the most important databases of NUTS versions[3]. It describes the changes occuring between each version.

National statistical institutes[4] can provide historical databases of national administrative boundaries. This source is very important to understand local changes affecting the geometry or structure of NUTS. It is also very useful in the case of the accessing of new countries (E15, E25, and E27) because EUROSTAT databases do not

[2] http://eurlex.europa.eu/JOIndex.do?year=2003&serie=L&textfield2=154&Submit=Search&_submit=Search&ihmlang=en

[3] http://ec.europa.eu/eurostat/ramon/nuts/splash_regions.html

[4] http://www.kyxar.fr/~jalac/europe.html

provide long term information about the historical administrative boundaries of these new members.

Other projects (scientific and operational)[5]: Many countries have attempted to construct temporal databases of their changing administrative boundaries. These experiences can provide databases (in the case of European countries) and methodology (Gregory I.N., 2002). The diversity of proceedings is explained by the specificity of each case.

Based on these different sources, the ESPON Historical GIS NUTS aims to be an innovative operational tool for providing temporal harmonized data series.

2.2.3 Identified difficulties

The Time Series issue can be divided in to three main types of problems which call for different approaches. Fundamentally in each problematic case there is a lack of data for a territorial unit, either because the territorial unit used has changed in the course of time or because data are simply missing for that territorial unit. We summarize below in this first part the three main sources of problems and the usual way to solve them.

2.2.3.1 Changes in NUTS

The "Nomenclature of territorial units for statistics" (NUTS) established by Eurostat for over 30 years is the official territorial subdivision system used in Europe "in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union".

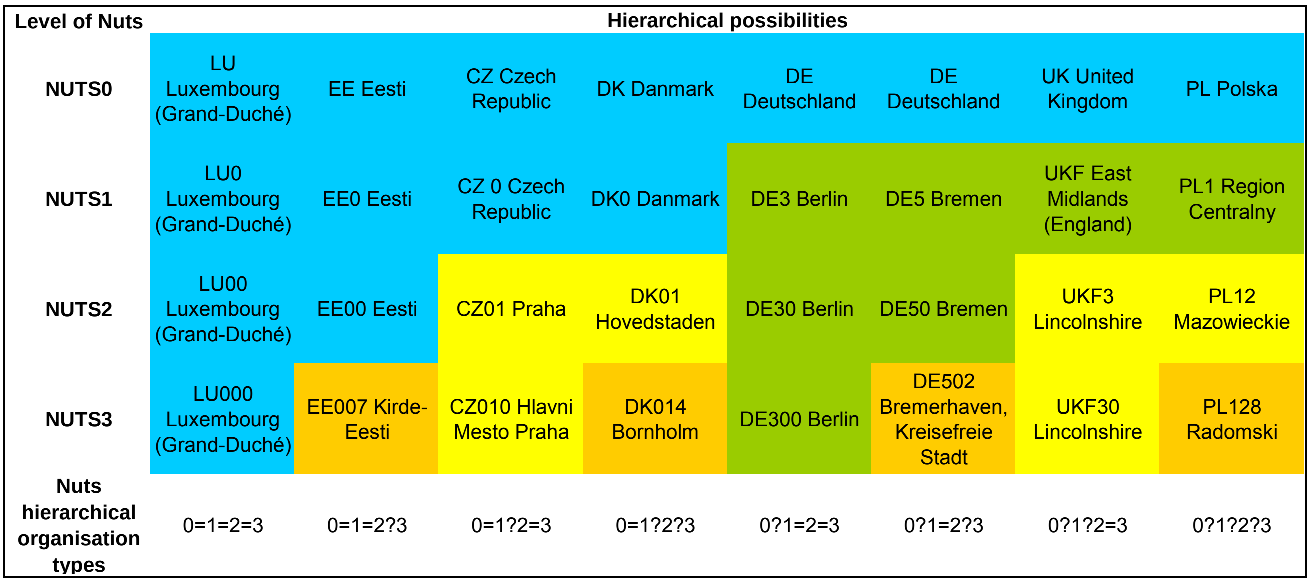
The difficulty to harmonize the geometry of nuts in time can be linked to the specificity of NUTS themselves. It can be explained by:

The degree (level) of hierarchical organization of NUTS is very different (figure 10)

“(2) The NUTS classification is hierarchical. It subdivides each Member State into NUTS level 1 territorial units, each of which is subdivided into NUTS level 2 territorial units, these in turn each being subdivided into NUTS level 3 territorial units” (3). “However, a particular territorial unit may be classified at several NUTS levels” (Regulation EC n° 1059/2003/Official Journal of the European Union L 154/1 of 21/06/2003).

[5] http://www.hgis.org.uk/resources.htm#top

http://www.who.int/whosis/database/gis/salb/salb_coding.aspx#DOCUMENTS%20OF%20INTEREST



Level of Nuts	Hierarchical possibilities							
NUTS0	LU Luxembourg (Grand-Duché)	EE Eesti	CZ Czech Republic	DK Danmark	DE Deutschland	DE Deutschland	UK United Kingdom	PL Polska
NUTS1	LU0 Luxembourg (Grand-Duché)	EE0 Eesti	CZ 0 Czech Republic	DK0 Danmark	DE3 Berlin	DE5 Bremen	UKF East Midlands (England)	PL1 Region Centralny
NUTS2	LU00 Luxembourg (Grand-Duché)	EE00 Eesti	CZ01 Praha	DK01 Hovedstaden	DE30 Berlin	DE50 Bremen	UKF3 Lincolnshire	PL12 Mazowieckie
NUTS3	LU000 Luxembourg (Grand-Duché)	EE007 Kirde-Eesti	CZ010 Hlavni Mesto Praha	DK014 Bornholm	DE300 Berlin	DE502 Bremerhaven, Kreisefreie Stadt	UKF30 Lincolnshire	PL128 Radomski
Nuts hierarchical organisation types	0=1=2=3	0=1=2?3	0=1?2=3	0=1?2?3	0?1=2=3	0?1=2?3	0?1?2=3	0?1?2?3

Figure 10 - Hierachical possibilities of NUTS

The NUTS divisions do not necessarily correspond to administrative divisions within the country, which can affect the degree of evolution of NUTS in time and produces very heterogeneous situations. This hypothesis depends on the national political system.

Semantic expertise: how NUTS can change in time?

To formalize temporal versions of NUTS we must identify the different possibilities of NUTS' changes.

As defined by the regulation of No 1059/2003 of 26/05/2003, NUTS is composed by: name, code, geometry and hierarchy, which can change in time. To simplify we propose five elementary kinds of change:

- ➔ Change of name
- ➔ Change of the spelling of the name
- ➔ Change of code
- ➔ Change of geometry
- ➔ Change of hierarchical level

These different elementary changes determine the existence of NUTS, which can be related to 3 main types of events:

- ➔ The creation of new units
- ➔ The breaking of units
- ➔ The disparition of units

However, the evolution of NUTS is more complex. At first, several changes can happen in the same time. Then, changes can affect many spatial units (see Annexe 2). The proposed formalization should be capable of drawing the genealogy of the NUTS which is a fundamental element for the harmonization of the time series.

2.2.3.2 Missing value

Another common source of difficulty is the absence of data for some years or some portion of the territory. Note that missing values are not an issue specific to time series but a universal problem in statistical series, for which statistical approaches exist like those detailed in the "Data Navigator II Report" of the Espon 3.2 project[6]. These statistical methods can be useful in the case of simple gaps in the data series but not for whole sections of the series unavailable, in which case other data should be used as a workaround.

Nature	Usual solution to consider	Example
Missing values	Interpolation or even extrapolation	Population 2003 derived from population 2002 and 2004
	Using proxy indicator (and make a rule of three)	Using employment distribution in economical sectors instead of added value distribution (rule of three)

2.2.3.3 Indicator definition modification

Probably the most dangerous situation is a modification of the definition of an indicator itself. This for instance happened with the GDP indicator at the European level in 1995, but also occurs recurrently with the unemployment indicators produced by the different countries. The mission of a statistical institute like Eurostat involves a normalization process in order to avoid disparities in the data provided by the different countries. But whenever data are found directly in national or regional statistical institutes the researchers must be aware of this risk. As a data collector Espon DB must then either adapt these indicators whenever it is possible or at least warn the user against the possible inconsistencies that might result from an inattentive use and provide as much as possible a methodology to avoid them. This implies to specify the exact definition of the data provided whenever it is relevant.

[6] available at http://www.espon.eu/mmp/online/website/content/projects/260/716/index_EN.html

Nature	Usual solution to consider	Example
Indicator modification	Using homogenized definitions through time	The GDP data provided by Eurostat are homogenized.
	Using another indicator	Using the International Labour Organization unemployment definition instead of the official national statistics

The inconsistencies in times series due to changes of NUTS and statistics are linked. They will be simultaneously approached.

2.2.4 Work plan

The aim of this challenge is to provide a corpus of methodological solutions to build harmonized temporal statistical series. Considering the difficulty and the complexity of historical database mining, our objectives would be organized in to short and long term. A first attempt will be made to define the NUTS dictionary boundaries changes and to integrate basic indicators (population, GDP, unemployment, age structure) between 2006 and 1995. A second step aims to enlarge the scope of changes dictionary to cover large time evolution of nuts and world databases.

The progress of this challenge will be organized according these following steps:

February-June 2009

Diagnostic of time series' availability in the ESPON area. The review of the different sources can provide information about the times databases which can easily build. Many classifications may be relevant: NUTS level, thematic, country, time periods…. This information can be transcribed in a summary table which will be very useful for the projects and which will serve as a guide.

June- September 2009

Elaboration of dictionary NUTS' changes. Based on the review of different sources, the dictionary of changes is a methodological book which consists in:

Typology of changes

Key's conversion of NUTS' version (genealogy of units)

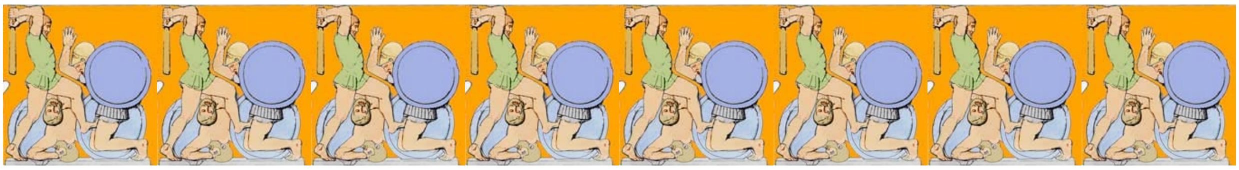
Spatio temporal data models

September 2009-Febrayry 2010

Computing data models and automating some proceedings. The integration of time in layer-based GIS is a real problem for GIS and databases research. Many data models have been proposed to incorporate temporal information into spatial databases but there is no generally accepted model, which can satisfy all temporal GIS requirements. This is due to the diversity of geographic objects' characteristics.

The progress of this challenge should be planned on the networking with other relevant challenges of the project like challenge 1, 3, 4,7 and 9 (Figure 1).

2.3 Challenge 3: World / Regional data



Coordinator: RIATE & UNEP

Harmonization of data at World/Neighbourhood and European/regional levels.

2.3.1 Objectives

Based on the results of ESPON 2006 Program, we propose to examine in a systematic way how to combine datasets at world/neighbourhood levels (where basic territorial units are the states) and datasets at European/Regional levels (where basic territorial units are NUTS2 or NUTS3 units). The interest of such connection is to enlarge the scales of analysis from spatial point of view (situation of ESPON territory in the world, situation of eastern and southern neighbouring countries) but also from historical point of view as time series at state level are generally more easy to obtain on long period (1960-Present) than regional time series (1995-Present).

2.3.2 Work done

The expert team UNEP has established contact with the lead partner RIATE in order to exchange experience on world database and to compare more specifically the Europe in the World database (EIW) realised by ESPON 2006 project 3.4.1 and the Global Environment Outlook database (GEO) realised by UNEP-GRID Genève and available on the internet[7]. After the joint presentation of both databases at the project meeting of 2-3 February 2009, it has been decided to launch specific actions in order to insure compatibility between the new ESPON DB and the GEO database, taking into account the experience gained in ESPON 2006 with the project EIW.

It is important to notice that the GEO database does not cover only socio-economic data and is not limited to state as basic territorial units. Many other ressources are available concerning for example environmental issues and different types of geographical object are covered like grid data, cities, water basin, etc. The challenge 3 will focus in a first step on the elaboration of a territorial database of data at state level, but it will also provide material for challenge 5 (grid data), challenge 6 (cities), etc.

2.3.3 Identified difficulties

[7] http://geodata.grid.unep.ch

Even if we limit our initial ambition to the collection of basic data at state level (population, GDP, land use, CO2 emissions), many difficulties has to be overcome.

Formalisation of the partnership ESPON-UNEP

The data available on the web portal GEO can be normally downloaded for free but many facilities are only available after registration. Moreover, the exchange of data and experiences should be bilateral between ESPON and UNEP which is at that time the most integrated gateway towards UN statistical system. Therefore, we strongly suggest that ESPON sign an agreement in order to become a GEO Collaborating Centre, like the EEA and many other institutions[8].

Data sources

Data collection follows as far as possible the main guidelines:

- ➔ global coverage,
- ➔ time series (1960-2010),
- ➔ primary source of information,
- ➔ public domain (as possible),
- ➔ most recently updates,
- ➔ metadata compiled with the ISO 19115 standard or according with the system that will be used for the ESPON 2013 main database

We propose to assemble our collection starting and testing methodologies on four main groups of variables: population, Gross Domestic Product (GDP), carbon dioxide emissions and land use, that will include in a second stage all the subcategories needed by the ESPON database.

Population:

Authoritative sources are the United Nations/Population Division with the World Population Prospects (WPP) 2008 that will be published in spring 2009 for total population and sub-series, and The World Urbanization Prospects WUPP 2007 (update in spring 2010) for the urban/ rural population.

GDP (two sources need to be evaluated):

World Development Indicators (WDI) from World Bank

National Accounts Main Aggregates Database from UN Statistical Database

Emissions, at least three candidate sources:

UNFCCC data reported by countries (Annex I parties)

[8] The list of collaborating centres of UNEP GEO is available at http://geodata.grid.unep.ch/extras/cc.php

CDIAC data calculated from energy statistics from UN yearbook

IEA / OECD calculated data

Land use:

Main data source will be FAO with its statistical and geospatial databases: FAOStat, SOFO, FRA, …

Elaboration of a common dictionary of states and territorial units

The basic condition for data exchange between UNEP-GEO and ESPON DB is the elaboration of a common dictionary of basic territorial units (states or territories) and the way they can be aggregated toward world regions of different levels. At the moment, the 168 states (or territorial units) of the EIW database are not fully compatible with the 237 states (or territorial units) of the UNEP-GEO database. Some differences can be easily solved by aggregation (ex. France is divided in 5 different units by UNEP-GEO) but other differences are more complex and, in some cases, related to political constraints that are not necessary the same for United Nations (e.g. Taïwan is not available) or European Union (e.g. Western Sahara, Kosovo, …).

Elaboration of common dictionaries of aggregation in world regions

The WUTS system elaborated by ESPON project EIW propose a hierarchical division of the world at 4 levels. UNEP proposes also a hierarchy at 3 levels. And many other levels of aggregation can be proposed by other organisations or can be requested by future ESPON 2006 projects. It is therefore necessary to implement various possibilities of aggregation of states and territorial units, according to the user's need and request (figure 11).

Benchmarking of the definitions of indicators and compatibility problems

Even in the case of very basic data like population. For example, "population 2005" can be defined according to legal status or to effective location. It can also be defined at the beginning of the year (1st January 2005) or in the middle of the year (1st July 2005). It can be based on census data or estimated (with possible revisions of the estimation), etc. The situation is of course increasingly difficult when it comes to more sophisticated indicators like unemployment (different possible definitions), GDP or GNP (different methods of conversion from $ to €, different methods of p.p.a. estimation, etc) or CO2 (different agencies producing different estimations).

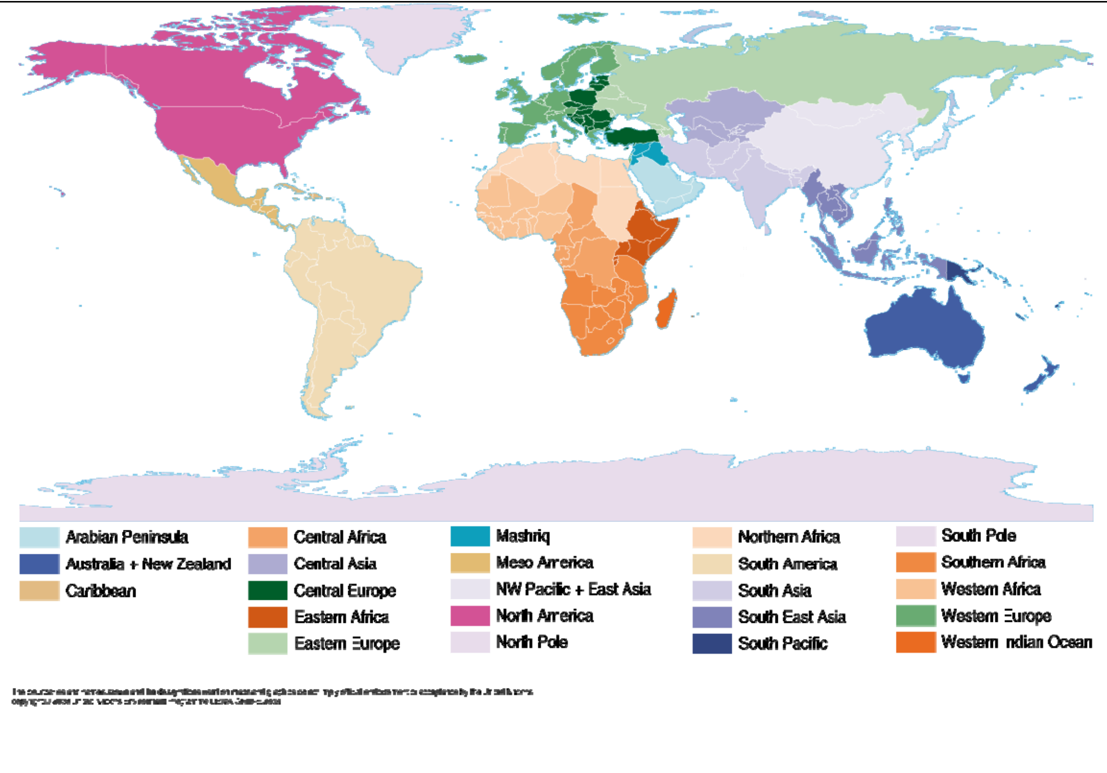


Figure 11 -The GEO sub-regional (2nd level) breakdown

Specific problem of articulation between World/state and ESPON/region databases

One specific but crucial problem is the articulation between world database where states are generally the lower territorial unit and regional databases where states are the upper territorial unit. In order to insure compatibility between the two types of database, we have to examine if the national level is equivalent in the two databases. For example, the mean population of Italy during the period 2001-2005 according to Eurostat regional database is equal to 57.705 millions of inhabitants. But according to UNEP-GEO world database, this population is equal to 58.260 millions of inhabitants (+1.0%). The results are reversed for Belgium where the population is equal to 10.379 millions of inhabitants according to Eurostat but 10.315 millions of inhabitants according to UNEP-GEO (-0.6%). Differences are not always so important (see annex 3) but this problem of articulation of levels is crucial for the scale integration of ESPON DB.

Another possibility for increasing the level of compatibility is to operate at grid level: the disaggregation of demographic data into regular grids at various spatial resolutions (1km, 5km, 10km) that are generally finer that the original census/statistical data. Several methods and products are available for the representation of global demographic data:

CIESIN datasets from Columbia University including GPW (not modeled), and GRUMP (settlement zones),

LandScan from ORNL ("Ambient population")

UNEP data based on the "accessibility index", independent from land cover.

These products are not compatible, essentially in terms of modeling methods and resolution, with the JRC EU Population dataset, that is mainly based on the CORINE

Land Cover. In order to reduce this incompatibility a challenge can be the adaptation of the downscaling methods elaborated by UAB (challenge 5) at global scale.

Elaboration of mapkit tool for World and neighbourhood mapping

The former ESPON project EIW had elaborated different map templates (World, Neighbourhood) that can provide a basis of reflection. But they have to be adapted and upgraded according to new levels of aggregation or new requests of ESPON for benchmarking with other world regions (Cf future projects of priority 4).

Networking with FP7 Eurobroadmap

According to the agreement signed between ESPON and DG-Research, the ESPON DP Project and the FP7-EuroBroadMap project will exchange data at state level. Structural data and geometries will be elaborated by ESPON and sent to FP7-EuroBroadMap. FP7 EuroBroadMap will elaborate distances, flows and network matrixes that will be sent to ESPON. It is of course crucial that both databases follows the same rules of codification and cartography, with metadata fully harmonised. That is the reason why the definition of the dictionary of units is an absolute priority and should be delivered very soon.

Networking with other data providers at world scale

UNEP GEO is per se a node in the statistical system of UN. The expert team will therefore act as the interface between the ESPON DB project and other UN or non-UN organisations producing data, metadata and studies at world scale. ESPON should not duplicate existing works but develop partnerships with existing organisations.

2.3.4 Work plan

The workplan for the year 2009 will focus on the **production of a world database at state level** (app. 200 units) covering basic structural indicators (Population, GDP, CO2,…) for the target period 1960-Present, with eventually projection Present-2050 in the case of demography.

February to June 2009

- ➔ Partnership agreement ESPON-UNEP GEO
- ➔ TECHNICAL REPORT "ESPON World database (I): Dictionary of units and regions"
- ➔ ESPON World Database version 1.0 (Data + Geometry)
- ➔ Networking with FP7-EuroBroadmap

July to December 2009

- ➔ Partnership agreement ESPON-UNEP GEO

- ➔ TECHNICAL REPORT "ESPON World database (II): Integration of national and regional levels"
- ➔ ESPON World Database version 2.0 (Data + Geometry)
- ➔ Support to ESPON project Priority 1 / Globalisation
- ➔ Networking with FP7-EuroBroadmap

January-February 2010

- ➔ Preparation of SIR
- ➔ Integration of results with other challenges, in particular C.1 (basic data), C.2 (time series), C.5 (Grid) and C.6 (Cities).

2.4 Challenge 4: Regional / Local data

Coordinator: TIGRIS

Harmonization of data at European/regional and National/local levels.

2.4.1 Objectives

The ESPON 2006 program has revealed that many questions related to territorial cohesion can not be fully explored at NUTS2 or NUTS3 levels and need further investigations at more local levels LAU1 and LAU2 (former NUTS4 and NUTS5). Case studies providing zoom on specific territories at local level (rural areas, cross border areas, intra-urban differentiation, …) will be more and more requested in the ESPON 2013 program for project of priority 2 and, in certain cases, for project of priority 1. It is therefore of utmost importance to be able to collect such type of data in ESPON 2013 Database and to develop a long term strategy.

2.4.2 Work done

According to the objectives proposed by this challenge, the Tigris team has developed a strategy to explore the range of problems raised by the construction of a database at the smallest level of administrative spatial reference. The strategy is based on the simultaneous approach and problem solving, this being the only viable option in the context of a profound interconnection of the difficulties of data spotting and collecting.

After the identification of the Internet-available national data sources, we have worked for a while on the (mainly systematic) exploration of the LAU 1/2 level information. This stage was necessary for the elaboration of a draft-database with indicators (still in progress) that would allow comparisons regarding the spatial level of data availability (LAU1 vs. LAU2), their chronological harmonization (2001 vs. 2002) and the semantic content of the indicators (e.g. age group of 5-10 years vs. age group < 14 years).

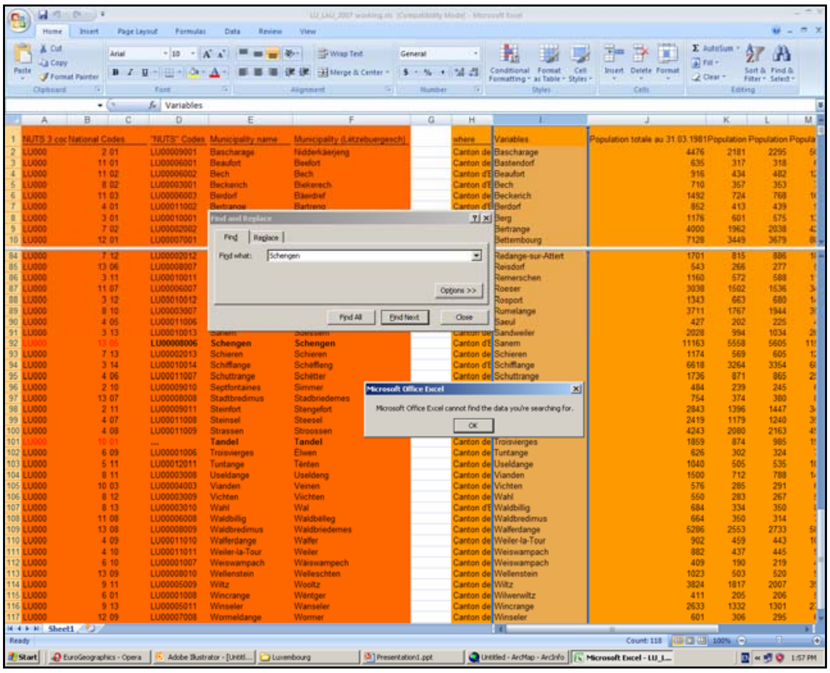
At the same time, a part of the team has dealt with the inventory and testing of a methodology for data collecting only for one state – Romania, in order to identify the occurring problems related to information database management (exceptions introduced by the coming into being of new LAU1/2s, by the changes in the official administrative toponyms, particular situations occurred after the administrative reforms and so on).

In the absence of a base map, the files with the complete nomenclature of the LAU1/2 units obtained from the Eurostat page via Eurogeographics cannot yet be mobilized for the proper start of database construction (figure 12). Nevertheless, they proved to be extremely useful for experimentally testing the collection and the effective populating of the database (in the case of Italy – information extracted from Rec. Ital. 2002). The use of the SABE97 base map and of the SIRE database has been momentarily suspended due to the numerous errors and their inadequacies in relation to the final objective of challenge 4. The Eurogeographics product EBM that was received the 20th of february will be the reference for future work including eventual reconstitution of historical units.

To maintain a certain coherence of the information download sequence, at this moment we preserve the sites on the TIGRIS server, an action quite time-consuming due to the low transfer rates, but useful taking into account the fact that it allows us to obtain a range of chronologically–comparable indicators. In the measure in which this download sequence will be functional, it will help us in the process of elaborating the indicators draft database.

2.4.3 Identified difficulties

One could imagine that building a database and filling its contents represents a quantifiable approach. Reality is different; the quantification of the data collecting process becomes possible only after three simultaneous barriers are outrun: the spatial harmonization, the chronological harmonization and solving the linguistic barriers. so far, the linguistic barrier proves to be the highest drawback, considerably increasing the time needed for information collecting. See proposal to associate ECP to this task in the conclusion of the report (4.3). The lack of a base map and of an attached reference file represents a second impediment, inhibiting the advancement towards the elaboration of a unique identification code for the spatial units. The access to eurogeograhics product will now allow the creation of the respective code and the construction of minimal indicators for administrative hierarchic organization. The decision on the use of a base map might lead in a later stage to the sketching of the transformations occurred in the geometry of the LAU 1/2 units in the ESPON space, by comparing it to the SABE97 base map.



Variables

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	NUTS 3 coc	National Codes		"NUTS" Codes	Municipality name	Municipality (Lëtzebuergesch)		where	Variables	Population totale au 31.03.1981	Population	Population	Popula
2	LU000	2	01	LU00009001	Bascharage	Nidderkäerjeng		Canton de	Bascharage	4476	2181	2295	
3	LU000	11	01	LU00006001	Beaufort	Beefort		Canton de	Bastendorf	635	317	318	
4	LU000	11	02	LU00006002	Bech	Bech		Canton d'E	Beaufort	916	434	482	
5	LU000	8	02	LU00003001	Beckerich	Biekerech		Canton d'E	Bech	710	357	353	
6	LU000	11	03	LU00006003	Berdorf	Bäerdref		Canton de	Beckerich	1492	724	768	
7	LU000	4	01	LU00011002	Bertrange	Bartreng		Canton d'E	Berdorf	852	413	439	
8	LU000	3	01	LU00010001					Berg	1176	601	575	
9	LU000	7	02	LU00002002					Bertrange	4000	1962	2038	
10	LU000	12	01	LU00007001					Bettembourg	7128	3449	3679	
84	LU000	7	12	LU00002012					Redange-sur-Attert	1701	815	886	
85	LU000	13	06	LU00008007					Reisdorf	543	266	277	
86	LU000	3	11	LU00010011					Remerschen	1160	572	588	
87	LU000	11	07	LU00006007					Roeser	3038	1502	1536	
88	LU000	3	12	LU00010012					Rosport	1343	663	680	
89	LU000	8	10	LU00003007					Rumelange	3711	1767	1944	
90	LU000	4	05	LU00011006					Saeul	427	202	225	
91	LU000	3	13	LU00010013	Sanem	Suessem		Canton de	Sandweiler	2028	994	1034	
92	LU000	13	05	**LU00008006**	**Schengen**	**Schengen**		Canton d'E	Sanem	11163	5558	5605	
93	LU000	7	13	LU00002013	Schieren	Schieren		Canton de	Schieren	1174	569	605	
94	LU000	3	14	LU00010014	Schifflange	Schëffleng		Canton d'E	Schifflange	6618	3264	3354	
95	LU000	4	06	LU00011007	Schuttrange	Schëtter		Canton de	Schuttrange	1736	871	865	
96	LU000	2	10	LU00009010	Septfontaines	Simmer				484	239	245	
97	LU000	13	07	LU00008008	Stadtbredimus	Stadbriedemes				754	374	380	
98	LU000	2	11	LU00009011	Steinfort	Stengefort				2843	1396	1447	
99	LU000	4	07	LU00011008	Steinsel	Steesel				2419	1179	1240	
100	LU000	4	08	LU00011009	Strassen	Stroossen				4243	2080	2163	
101	LU000	10	01	....	**Tandel**	**Tandel**		Canton de	Troisvierges	1859	874	985	
102	LU000	6	09	LU00001006	Troisvierges	Ëlwen		Canton de	Tuntange	626	302	324	
103	LU000	5	11	LU00012011	Tuntange	Tënten		Canton de	Useldange	1040	505	535	
104	LU000	8	11	LU00003008	Useldange	Useldeng		Canton de	Vianden	1500	712	788	
105	LU000	10	03	LU00004003	Vianden	Veianen		Canton de	Vichten	576	285	291	
106	LU000	8	12	LU00003009	Vichten	Viichten		Canton de	Wahl	550	283	267	
107	LU000	8	13	LU00003010	Wahl	Wal		Canton d'E	Waldbillig	684	334	350	
108	LU000	11	08	LU00006008	Waldbillig	Waldbëlleg		Canton de	Waldbredimus	664	350	314	
109	LU000	13	08	LU00008009	Waldbredimus	Waldbriedemes		Canton de	Walferdange	5286	2553	2733	
110	LU000	4	09	LU00011010	Walferdange	Walfer		Canton de	Weiler-la-Tour	902	459	443	
111	LU000	4	10	LU00011011	Weiler-la-Tour	Weiler		Canton de	Weiswampach	882	437	445	
112	LU000	6	10	LU00001007	Weiswampach	Wäiswampech		Canton de	Weiswampach	409	190	219	
113	LU000	13	09	LU00008010	Wellenstein	Welleschten		Canton de	Wellenstein	1023	503	520	
114	LU000	9	11	LU00005009	Wiltz	Wooltz		Canton de	Wiltz	3824	1817	2007	
115	LU000	6	01	LU00001008	Wincrange	Wentger		Canton de	Wilwerwiltz	411	205	206	
116	LU000	9	13	LU00005011	Winseler	Wanseler		Canton de	Wincrange	2633	1332	1301	
117	LU000	12	09	LU00007008	Wormeldange	Wormer		Canton de	Winseler	601	306	295	

Find what: Schengen

Microsoft Office Excel

Microsoft Office Excel cannot find the data you're searching for.

Figure 12 -An example of incongruence between the working files obtained from Eurostat via Eurogeographics and the national statistics - Luxembourg

2.4.4. Workplan

The future efforts of the Tigris team will be focused on **six major objectives**, declinable on chronological sequences as it follows:

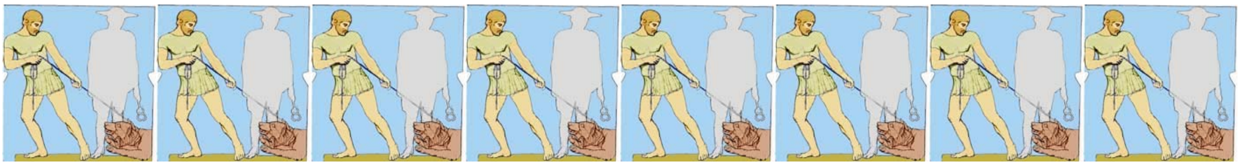
Until June 2009 we will provide a finalized sample database with indicators for at least two neighboring countries (e.g. Romania and Bulgaria). We will also try to complete the database with available indicators at LAU1/2 level for the ESPON space. The first objective largely depends on the proper linkage between the geometry of the base map and the list of LAU codes; otherwise we will be forced to furnish some corrections for the administrative frame and for the attribute table, which is a time consuming problem.

Between June and September 2009 we wish to finalize the **indicator database for most of the countries** and **to derive a short history of the modifications** in the LAU1/2 units' geometry or in the official denominations. Choosing the countries for the first objective is a function of a double constraint: the chronological and spatial harmonization of the indicators and the research priorities of other ESPON contracts. Consequently, we will try to focus our collection and implementation of the information on the countries and variables needed for the advance of these contracts.

In the period September 2009 – February 2010, based on the experience extracted from the previous objectives, **we will be able to finish the process of filling the database with information** for one or two indicators, country by country, until we complete the first field. **Recovering the information available in the SIRE database** will be our second objective for this period, in order to obtain and offer a functional and chronological coherent set of minimal indicators.

At the present moment the proposed time table is subject of revisions, because the finalization of an objective depends on some external factors such as the reception of an adequate base map, the calibration of the collecting process with the eventual changes at the level of the information from NSI or the reconfiguration of the administrative frame.

2.5 Challenge 5: Social / Environmental data



Coordinator: UAB (ETC-LUSI)

Combining socio-economic data measured for administrative zoning (Nuts level) and environmental data defined on a regular grid (like Corine Land cover or any spatiomap)

2.5.1 Objectives

Most of the socioeconomic variables or indicators are associated with administrative unit, i.e. NUTS regions, whereas the environmental data is usually not following those boundaries, but given by natural units or regular grid cells. The ESPON 2006 program developed some indicators in which the environmental data was transposed to NUTS division by means of GIS tools, in order to make them comparable to socioeconomic data. This solution introduces some problems revealed by the MAUP study (ESPON 3.4.3) and it seems better to find other solutions for data harmonization.

Therefore, this challenge is aimed at defining a suitable methodology for integrating and making comparable data coming from statistical sources (e.g. EUROSTAT) and measured by administrative unit, together with environmental data stored by natural unit or regular grid structure (e.g. Corine Land Cover).

2.5.2 Work done

We have splitted the work done into three separate sections, the first one regarding the background analysis, a second one about the methodology definition, and a third and last one listing the main conclusions after the results obtained.

Background review

The first step we have faced this challenge was the review of several methodological approaches made by other institutions or researchers.

Special attention has been paid to the work done by JRC in population: “Downscaling population density in the European Union with a land cover map and a point survey” by Francisco Javier Gallego[9], or the G-Econ Research project developed by the University

[9] http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP_RESEARCH/PGE_RESEARCH_NTTS/S14P3%20-%20JAVIER%20GALLEGO%20-%20%20DOWNSCALED%20POPULATION%20DENSITY.PDF

of Yale on Gross Cell Product (GCP): "New Metrics for Environmental Economics: Gridded Economic Data" by William D.Nordhaus[10].

Other methodologies were also explored, such as the one applied by the FARO-EU on the GDP at 1km grid, and the work done by the University of Columbia by Deborah Balk and Greg Yetman: "Transforming Population Data for Interdisciplinary Usages: From Census to grid"[11].

The main conclusion after this research has been that the way proposed by most of the studies revised, in order to downscale socioeconomic data and make it comparable to other kind of data, is using a regular grid structure, in which each cell takes a figure of the indicator or variable. It is also remarkable that each type of variable or indicator requires a different type of integration method into the regular grid. This is discussed in the next section.

Methodology definition

After reviewing several studies and taking into account our experience at the UAB (ETC-LUSI) and the EEA, we propose to integrate socioeconomic data in the 1 km European Reference Grid (figure 13).

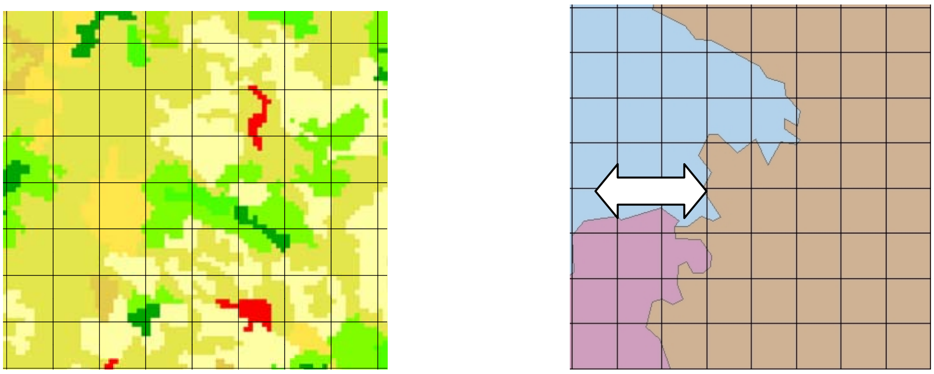


Figure 13 -The 1 km European Reference Grid will hold both environmental and socioeconomic information.

Therefore, the first step to be undertaken should be the intersection between the 1 km European Reference Grid and the administrative units by which the indicator is given.

Furthermore, we have realised that depending on the nature of each indicator, a different kind of integration procedure should be defined. In this regard, we define three general integration methodologies:

Maximum area criteria: the cell takes the value of the unit which covers most of the cell area. It should be a good option for uncountable variables (figures 14, 15 and 16).

[10] http://www.oecd.org/dataoecd/44/7/37117455.pdf

[11] http://sedac.ciesin.columbia.edu/gpw-v2/GPWdocumentation.pdf

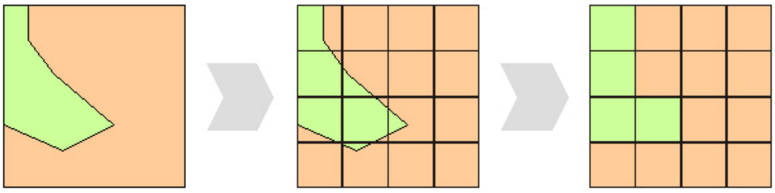
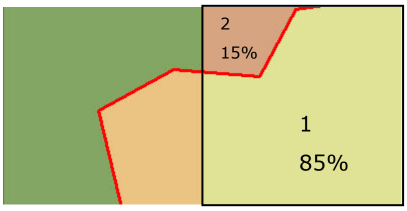


Figure 14 – maximum area criteria

Proportional calculation: the cell takes a calculated value depending on the values of the units falling inside and their share within the cell. This method seems very appropriate for countable variables.



Cell value = Σ (V_i * $Share_i$)

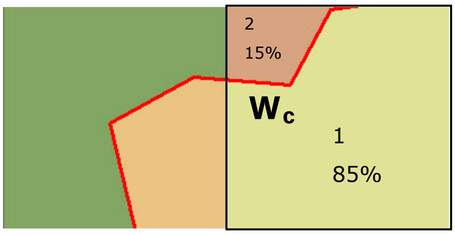
Where: V_i = Value of unit i

$Share_i$ = Share of unit i within the cell

In the example: V_1 * 0.85 + V_2 * 0.15

Figure 15 – proportional calulation

Proportional and weighted calculation: the cell takes also a proportionally calculated value, but this value is weighted for each cell, according to an external variable (e.g. population). This method can be applied to improve the territorial distribution of a socioeconomic indicator. For instance, a GDP indicator can be redistributed by 1 km grid and weighted by the population figures of each cell (coming from the 1 km population density dataset produced by JRC).



Cell value = W_c Σ (V_i * $Share_i$)

Where: V_i = Value of unit i

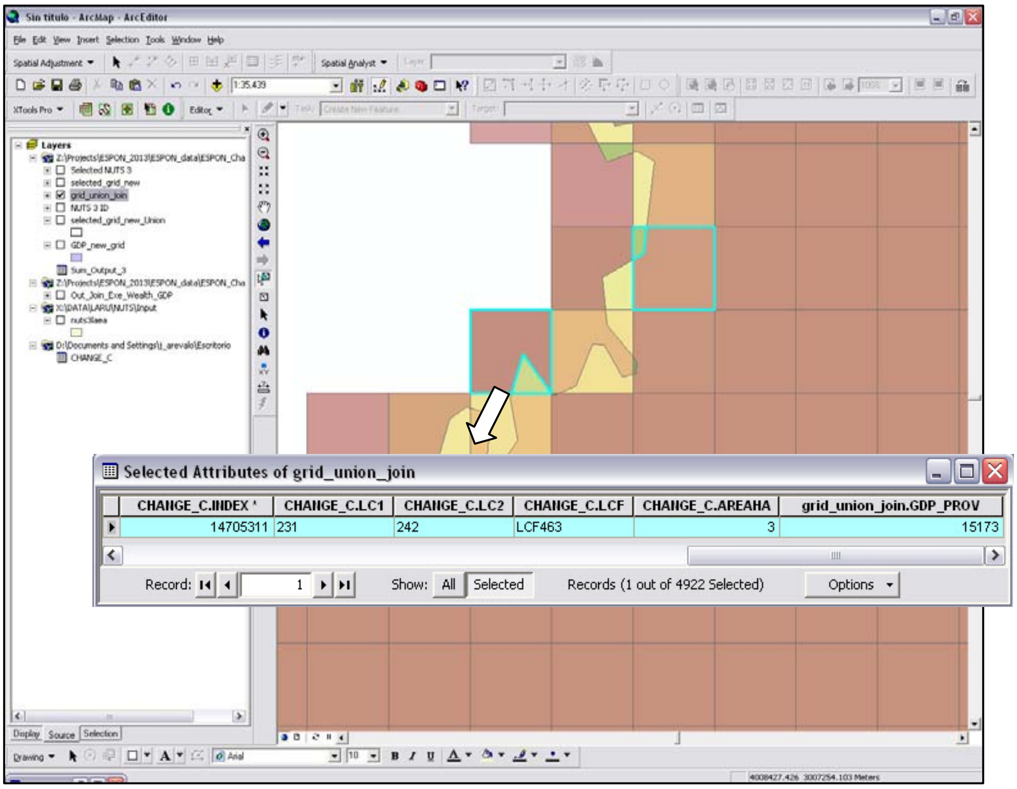
$Share_i$ = Share of unit i within the cell

W_c = weight assigned to cell c

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

Figure 16 - proportional and weighted calulation

Depending on each type of indicator or variable to be integrated within the reference grid, a different type of integration should be decided and tested. Besides the method finally chosen to integrate, it is important to highlight that indicator figures given by area unit, e.g. by square kilometre, should be converted considering that each cell has a total area of 1 km^2. (Figure 17)



Selected Attributes of grid_union_join

CHANGE_C.INDEX *	CHANGE_C.LC1	CHANGE_C.LC2	CHANGE_C.LCF	CHANGE_C.AREAHA	grid_union_join.GDP_PROV
14705311	231	242	LCF463	3	15173

Record: 1 Show: All Selected Records (1 out of 4922 Selected) Options

Figure 17 – Selected attributes of grid

Once the variable has been distributed by 1 km cell, it can be compared to other variables or indicators on a cell-by-cell basis, or it can be integrated into the EEA's Land and Ecosystem Accounting (LEAC system)[12].

In this example, we have been able to put together a "GDP in purchasing power" value, originally measured by NUTS3 region, together with the land cover flows between 1990 and 2000, coming by the Corine Land Cover changes:

Conclusions

After the bibliographic review and tests undertaken, we can draw some few conclusions about the work done so far in this challenge:

Integration of socio-economic and environmental data by 1 km grid is a good solution, since data can be harmonised without losing resolution.

We propose the use of the 1 km European Reference Grid as reference grid of data integration.

Depending on the nature of each variable, a different integration approach should be decided and implemented.

The method can indistinctly be applied to both vector and raster data.

[12] http://etc-lusi.eionet.europa.eu/LEAC

This approach facilitates the compatibility between ESPON databse and the EEA's LEAC assessment system.

2.5.3 Identified difficulties

Some of the difficulties encountered become challenges for the future work and are included as well in the work plan for the following months.

The first challenge is to make the decision about the integration method to use regarding a given variable or indicator. This point can be easy in some cases and more complicated in other cases, and will be studied in the following months.

A second problem or challenge is the feasibility of integrating such data into the EEA's LEAC System, in a way that can be easily compared to environmental data and queried online.

The processing of huge volumes of data might become also a problem. Partial or total automation of processes will be tested and applied to the methodology in order to verify the feasibility.

2.5.4 Work plan

Having set up the general terms of the methodology and having done the first tests, in the next months we should undertake the following tasks:

Test different integration methods for different socioeconomic indicators and variables.

Develop some automatic tools to improve or speed up some of the procedural steps.

Test the compatibility of some of the studied variables or indicators with the EEA's LEAC System.

Milestones

June 2009: A sufficient number of tests done for different variables or indicators, using all integration methods. Technical report about the conclusions derived from those tests.

December 2009: Integration of some variables or indicators into the EEA's LEAC System and assessment of the results.

2.6 Challenge 6: Urban data

Coordinator: Géographie-cités

Constructing complex geographical objects of higher level such as cities, resulting from an aggregation of elementary objects according to a measure of relation in space (proximity, links and flows…).

2.6.1 Objectives

Urban objects are complex geographical objects, which result from an aggregation of elementary units according to different possible measures of relation in space (distance, contiguity, flows, or other links). Another source of complexity comes from the international dimension of the European data sets and from the large number of urban data bases that have been built for the 10 last years, at different scale levels and according to different approaches: sub-cities, central municipalities, morphological agglomerations, functional areas, regional areas... In addition to classical tasks realized by the Espon DB (storing the urban data and metadata, updating the geometrical and statistical sources when possible, working on attributes), we conduct a semantic and empirical expertise in order to insure compatibility between the different definitions of cities and urban areas currently available.

2.6.2 Work done

Three different directions have been followed since the beginning of the project.

a) Gathering data bases and their documentation

The first step of the work consisted in enumerating and collecting the different urban data bases that could be of interest for the Espon Projects at the different levels of definitions. We obtained 12 databases, created by Urban Audit (3 databases for 2 reference years, 2001 and 2004, and a Proxy LUZ/Nuts3 for 2000), by EEA (UMZ 1990 and 2000), and by previous Espon Projects (MUAS: Espon 1.4.3, reference year 2000; FUAS 2000: Espon 1.1.1 and 1.4.3, reference year 2000)[13]. Obviously, all these data

[13] LUZ: Larger Urban Zone; UMZ: Urban Morphological Zone; MUA: Morphological Urban Area ; FUA: Functional Urban Area.

bases do not have the same geographical coverage in terms of sets of European countries, as illustrated in Annex 4.

The databases have been collected with their documentation when available in reports, websites and publications, and fulfilled by contacting some of the authors (IGEAT, NordRegio). Some databases or documentations still remain uncomplete (figure 18).

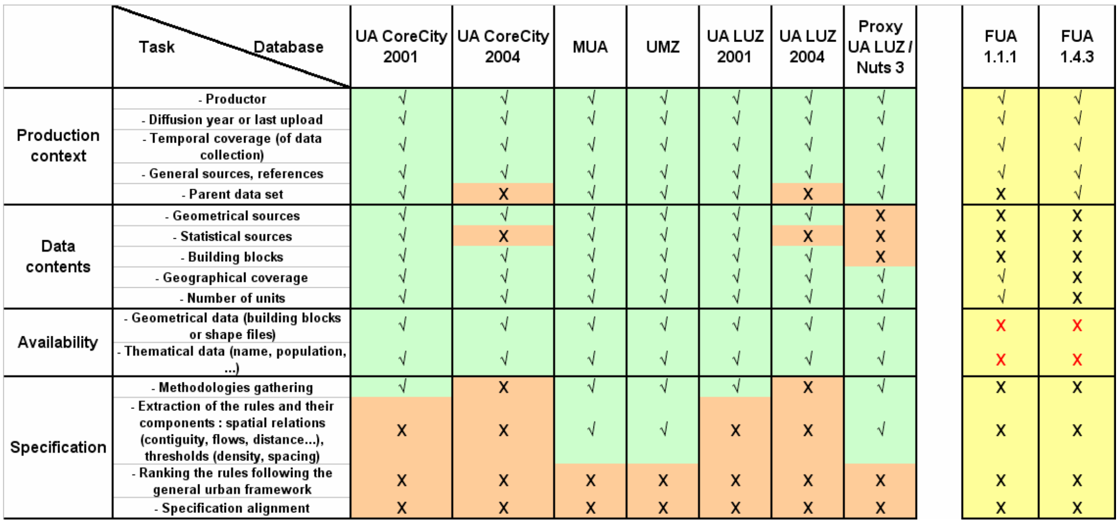
b) Semantic expertise

The aim of the semantic expertise is to produce databases integration, i.e. to precise the relationships between two different databases, to compare them and to be able to explain the differences. The first step is the extraction of the rules used to build urban objects (spatial relations, population or density thresholds etc.) in order to align the specifications and to be able to evaluate qualitatively the quantitative differences between data bases. First results have been obtained for the two databases using morpho-statistical criteria, MUAS and UMZ and will be provided through a technical report.

c) Delivering urban databases and metadata for the Espon Data Base

We have prepared a new version of the UMZ database (coming from CLC2000), which improves in two different ways the current one that can be loaded on the EEA website Using automatic methods, we have added a statistical variable (population 2000 from the last version of the data-grid built by Joint Research Center[14]) for the 112 168 UMZ of Europe. Next steps will be devoted to the preparation of national files (we still need LAU2 version 2006) and to the application of different possible methods for naming the UMZ. For practical purposes, we will test these methods using a minimum population threshold (10000 inhabitants, i.e. 4400 UMZ).

Green mark: work done; black cross: work in process; red cross: no data available



	Task / Database	UA CoreCity 2001	UA CoreCity 2004	MUA	UMZ	UA LUZ 2001	UA LUZ 2004	Proxy UA LUZ / Nuts 3	FUA 1.1.1	FUA 1.4.3
Production context	- Productor	√	√	√	√	√	√	√	√	√
	- Diffusion year or last upload	√	√	√	√	√	√	√	√	√
	- Temporal coverage (of data collection)	√	√	√	√	√	√	√	√	√
	- General sources, references	√	√	√	√	√	√	√	√	√
	- Parent data set	√	X	√	√	√	X	√	X	√
Data contents	- Geometrical sources	√	√	√	√	√	√	X	X	X
	- Statistical sources	√	X	√	√	√	X	X	X	X
	- Building blocks	√	√	√	√	√	√	X	X	X
	- Geographical coverage	√	√	√	√	√	√	√	√	X
	- Number of units	√	√	√	√	√	√	√	√	X
Availability	- Geometrical data (building blocks or shape files)	√	√	√	√	√	√	√	X	X
	- Thematical data (name, population, ...)	√	√	√	√	√	√	√	X	X
Specification	- Methodologies gathering	√	X	√	√	√	X	√	X	X
	- Extraction of the rules and their components : spatial relations (contiguity, flows, distance...), thresholds (density, spacing)	X	X	√	√	X	X	√	X	X
	- Ranking the rules following the general urban framework	X	X	X	X	X	X	X	X	X
	- Specification alignment	X	X	X	X	X	X	X	X	X

Figure 108 - Urban data bases collected in the 2013 Espon DB, February 2009

[14] Gallego J., 2007; *Downscaling population density in the European Union with a land cover map and a point survey*, http://dataservice.eea.europa.eu/dataservice.

2.6.3 Identified difficulties

The main difficulties provided by urban data can be illustrated by two geographical examples. The first one is the Bordeaux case (Figure 19), which enlightens the inadequacy between the semantic approach (leading to regular nested urban perimeters, from the city core definition to the larger urban zone one) and the reality given by the data (city core larger than agglomerations, proxy Luz highly different from Luz).

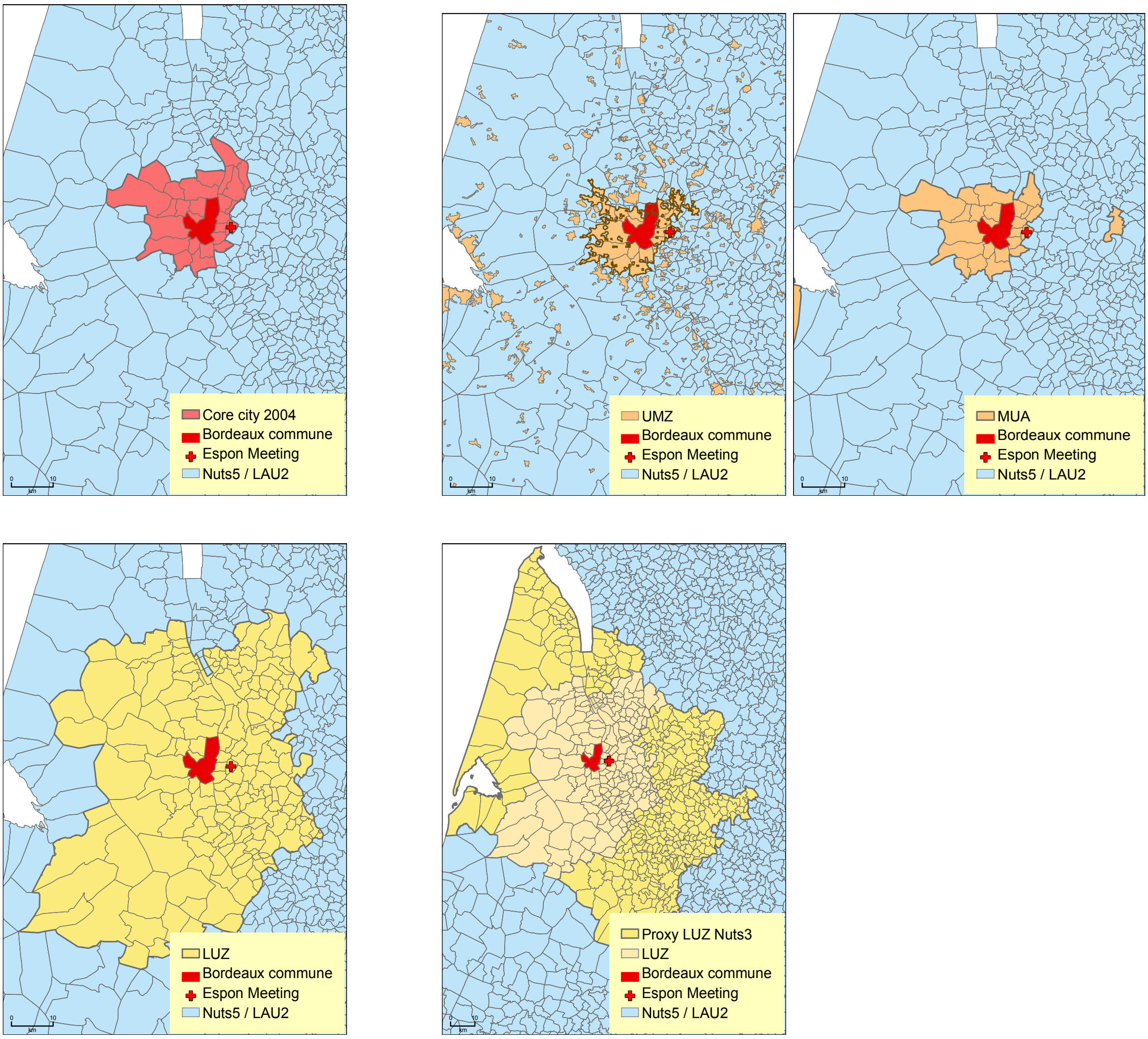
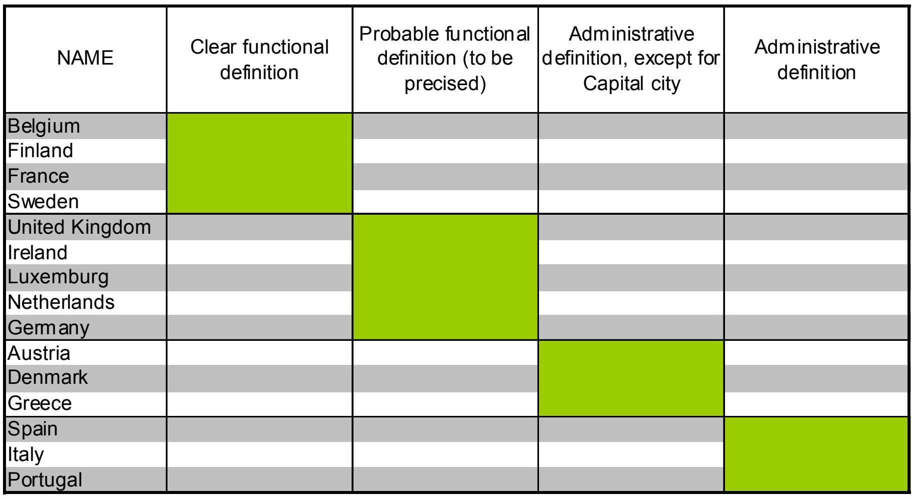


Figure 19 - Bordeaux, according to 5 different urban databases

We raise here two different problems, the heterogeneity of building blocks from one country to another in Europe and the heterogeneity of national definitions used by Urban Audit (figure 20).



NAME	Clear functional definition	Probable functional definition (to be precised)	Administrative definition, except for Capital city	Administrative definition
Belgium	🟩			
Finland	🟩			
France	🟩			
Sweden	🟩			
United Kingdom		🟩		
Ireland		🟩		
Luxemburg		🟩		
Netherlands		🟩		
Germany		🟩		
Austria			🟩	
Denmark			🟩	
Greece			🟩	
Spain				🟩
Italy				🟩
Portugal				🟩

Figure 20 - The heterogeneity of LUZ, between functional and administrative definitions

(Source: Urban Audit 2001, in Methodological Handbook 2004)

The second geographical case is the Saarbrücken one, represented through UMZ and MUAS databases (figure 21). The result enlightens the gap between two representations of a same semantic object, the "urban agglomeration". Whereas Saarbrücken is clearly separated from Forbach and represents a population of 552000 inhabitants, in the MUA database, the main UMZ covers both the two cities in a polycentric way but contains only 357000 inhabitants. We have here an empirical illustration of the result obtained by the semantic expertise, i.e. a major incompatibility between the two databases.

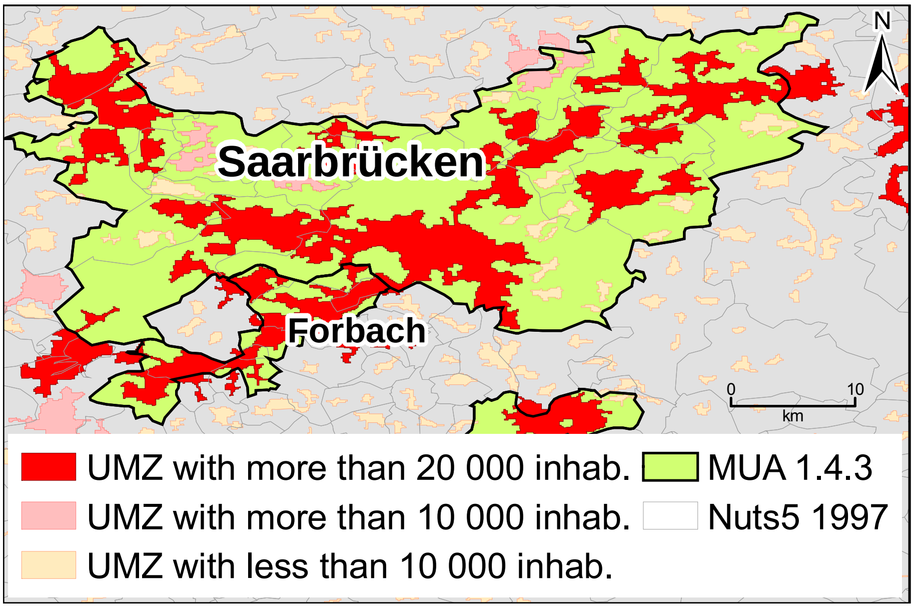

N
Saarbrücken
Forbach
0
10
km
UMZ with more than 20 000 inhab.
UMZ with more than 10 000 inhab.
UMZ with less than 10 000 inhab.
MUA 1.4.3
Nuts5 1997

Figure 21 - Saarbrücken, defined as a MUA and as a UMZ

2.6.4 Work plan

For June 2009, the work will progress following different directions.

- Integrating databases and metadata into the Espon DB: among the different tasks devoted to storage (urban delineations, socio-economic data), we will develop more specifically three different points.

  ➔ For UMZ, as mentioned above, we will deliver in the Espon DB the metadata and the national database files, with population 2000 version V4.a, names and the other current attributes available (exchanges with Challenge 5), for UMZ larger than 10 000 or 2000 inhabitants.

  ➔ For MUAS, we will store the data and try to adapt the metadata that have been gathered to the EsponDB2013 template (exchanges with Challenge 9).

  ➔ For Urban Audit, we will develop further the work on 2001 metadata and will enlarge it to 2004 if the documentation is available (Methodological Handbook, national reports). This last point will be delivered through a technical report.

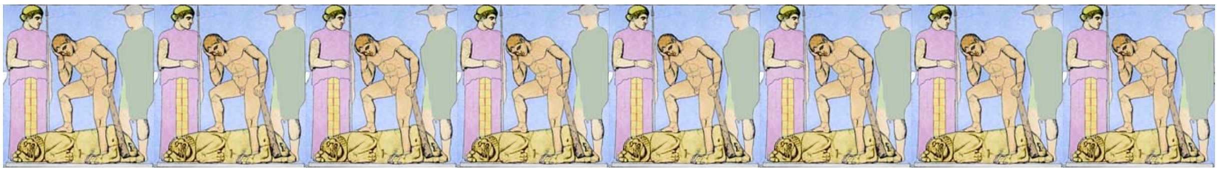
- Semantic expertise and database comparison: two points will be developed.

  ➔ First, in order to fulfill the semantic expertise of MUAS and UMZ, a comparison will be processed with LAU2 2006 through some statistical local data of reference (exchanges with Challenge 1 and 4), in order to estimate the ranges of deviations at different geographical scales.

  ➔ Secondly, a comparison between UMZ and the Swedish morpho-statistic urban database (surfaces, perimeters, populations) will be developed in collaboration with Challenge 12. As it was the case with the semantic comparison MUAS/UMZ, this analyze will improve our knowledge on the use of UMZ using for urban studies. Another aspect, leaded mostly by Challenge 12, will concern an estimation of the accuracy of the data grid Version V4.1 for estimating urban populations, through a comparison with Swedish individual data.

The results of these 2 points will be delivered through technical reports.

2.7 Challenge 7: Extra-ESPON data exchange



Coordinator: University of Luxembourg

Constructing complex geographical objects of higher level such as cities, resulting from an aggregation of elementary objects according to a measure of relation in space (proximity, links and flows…).

2.7.1 Objectives

The objective of the seventh challenge is to foster the exchange of data and expertise with main data providers and potential external (non-ESPON project) users of the ESPON database. It therefore regroups the set of networking activities undertaken with external organisations, primarily EUROSTAT (linkages with EEA are more specifically developped in challenge 5 and linkages with UN in challenge 3).

Along the time frame of the project, we seek to develop with several institutions two-way flows of data (statistical, urban zones, grids, networks, …) but also of expertise and knowledge in spatial data analysis, data integration, data management (metadata, data publication (web,…). We also aim at identifying longer-run synergies and collaboration between the ESPON database and those institutions.

This challenge also includes the promotion of the ESPON database via participations to international workshop and conferences, as well as regular contacts with the Coordination Unit for providing particular pieces of work to promote the program.

2.7.2 Method

The challenge is coordinated by University of Luxembourg (UL) who is responsible to draw the broad picture of networking for the Database project and record all external information and data flows. Contacts with EU institutions are made in agreement and with the help of ESPON Coordination Unit, particularly on licensing issues. Each ESPON DB partner however maintains direct contacts on more technical issues with resource persons within each institution of interest:

EUROSTAT: UL, RIATE, LIG, IGEAT

EEA: UAB, RIATE, LIG

OECD: IGEAT

URBAN AUDIT: Géographie-cités

National statistical offices: TIGRIS & UL

In addition, the Lead Partner (RIATE) is responsible for specific work requested by ESPON coordination unit for the program promotion.

2.7.3 Work done

ESPON DB – Institutions' contacts

So far, the ESPON DB has participated in several meetings with external institutions in order topresent the objectives and structure of the ESPON DB project, thus promoting the future database,inform mutually on data availabilities, spatial data management (including database structure and metadata), and cartographic issues, and identify a strategy to inform each other during the course of the project and possible synergies.

The first and main external institution contacted is obviously DG EUROSTAT. Obviously EUROSTAT is the main source of homogenised regional data at different scales from NUTS 0 to LAU 2. It is also responsible for Urban Audit and involved in the preparation and dissemination of geographical data through its GISCO service. In addition it is involved in the development and implementation of geographical and statistical standards (INSPIRE and SDMX). All these issues are crucial for the ESPON Database project as well.

Meeting was held on November 24th 2008 with the Coordination Unit and EUROSTAT GISCO, REGIO and Urban statics. The meeting resulted in an action plan (details, see annexe 4) with short, medium-term and re-occurring actions. It was decided to have regular meeting with EUROSTAT to update on data and indicator availabilities both way, and the evolution of the Database project and EUROSTAT projects, as well as to debate on methodological and technical issues. So far technical follow-up of this meeting has been made through direct contact between persons in charge of specific issues within the ESPON DB and within EUROSTAT, particularly on Urban Audit variables and local level data (SIRE database)

The second institution contacted is DG-REGIO. We find that it is particularly important to develop information exchange between DG-REGIO and the ESPON Database, since DG-REGIO has gained a lot of expertise on mapping and spatial data analysis to serve regional policies. They are also a potential user of the database and output maps. A first meeting was held on January 16th 2009 in order to inform on ongoing projects, identify available data, and discuss data integration, database structures and expertise in metadata. Decision has been made to exchange some datasets and provide regular updates about the ESPON Database and DG-Regio projects (See minutes from the meeting attached, annexe 7)

Third, regular informal contacts have been maintained between the European Environmental Agency (EEA) and the Joint Research Centre of the Commission (JRC) particularly on topics related to environmental data, land use grid data, and the INSPIRE metadata initiative.

General networking

Beyond particular contacts with Institutions, the ESPON DB also participated in different activities where the project and the database were promoted and where ESPON DB partners could develop their network with key resource persons and institutions. Particularly:

An extensive presentation of the project at the ESPON seminar in Bordeaux (December 2008) by all project teams.

Presentation of the project at the ESPON workshop on "Monitoring Spatial Dynamics" (ESch-sur Alzette, November 2008) by University of Luxembourg

Attendance of LIG-STEAMER researchers at the SDMX ***Global Conference organised by OECD the 19-21 January 2009 in Paris.***

Attendance of University Autonomous Barcelona (and presentation of Land Use Data Centre prototype based on Geonetwork) at the WISE (Water Information System of Europe)), related to the INSPIRE initiative.

2.7.4 Work plan

Follow-up on EUROSTAT-ESPON action plan:

Following the action plan (see annexe 5), the ESPON DB will participate in 2-3 meetings a year with EUROSTAT on general cooperation and more specific/technical issues. The action plan also identified a set of short and mid-term actions which are currently undertaken.

Among the short tem actions identified, the ESPON DB needs to further the situation of accessing archived regional data in order to develop time series (to be sorted out in April 2009). The other short-term actions identified have been done or are underway.

The ESPON DB has also now progressed on the three issues mentioned in the medium term plan: grid data, missing value estimates and INSPIRE compliant metadata (see respective challenges in this First Interim Report). The solutions developed within the ESPON DB project can be the object of a future routing meeting (e.g. to be taken in September, October after first implementation of a metadata tool) so that EUROSTAT is informed on and can comment on these progresses.

Information exchange with Commission DG's:

The project will go on with contacts with DG-Regio for specific data issues and mutual update about projects. In the medium term, it is also important to identify a series of ESPON database indicators and maps that could directly be useful to the preparation of the next Cohesion Report (End of 2009).

The project also aims at setting-up deeper contacts with other DG's, particularly DG ENVI to identify whether other resources (not EEA-based indicators) can be accessed, and DG AGRI for agricultural land data as well as to discuss the organisation and estimates within their own regional database. (Autumn 2009)

Wider networking:

The project will continue to develop its contacts with other organisations and expert groups both at the European scale (INSPIRE, EEA, etc..) but also at more global (UN, OECD,…) and local (national statistics institutes) scales. For the wider perspective, we can from now on benefit from the experience of an expert partner (see new challenges) on world data. For the local level (see challenge 4), stock of information has been automatically taken (web based queries) from national statistics offices. It will be complemented where necessary by direct contacts with resource persons within these institutions.

As part of its networking, the project will continue to attend meetings on spatial data management at the EU level. The ESPON DB project will for example be represented at the next meeting with EUROSTAT and EU statistical and cartographic institutes (Luxembourg, March 2009).

Finally, the ESPON DB project will be present in most ESPON workshop and activities (including next seminar in June 2009) in order to present the originalities and evolution of the database and promote its usage.

2.8 Challenge 8: Intra-ESPON data exchange

Coordinator: RIATE & LIG

Internal networking (other ESPON Projects)

2.8.1 Objectives

The project ESPON 2013 DB should develop regular contacts with other ESPON Projects from Priority 1, Priority 2 and Priority 3. This challenge is crucial as it is related to the circulation of data inside the whole program. Its main high level objectives are:

To provide an efficient framework for communication and collaboration between ESPON projects (Priority 1 and Priority 2) and statistical or cartographic institutions (Eurostat, EEA, Eurogeographics…)

To stimulate the launching of future ESPON Projects by incorporating indicators from neighbouring countries (Balkans, Ukraine, Belarus, Moldova, Maghreb, Turkey…) or elaborating long term (1960-1980-2000) series for a selected number of indicators

To reach these objectives, the definition of both files formats and a protocol for data exchange, shared by all the ESPON Projects, is a prerequisite to an efficient circulation of information and knowledge inside the project. We give here an overview of the actions we have undertaken in this direction. In particular,we describe how information transits between ESPON Projects, through the software architecture designed and implemented within the ESPON 2013 DataBase Project.

2.8.2 Work done

Contact with ESPON Priority 1 and Priority 2 project

The first action undertaken under this priority was to define contact points between each ESPON Project under priority 1 and priority 2 and the ESPON DB project (see figure 22) since the beginning for Priority 1 projects and was completed at the Bordeaux Meeting for the new priority 2 projects, where a special session was organised. The table of contact is reproduced below, and will be extended in the future when new projects will be appointed under priority 1, 2

ESPON projects - Priority 1

Project title	Acronym	Contact person – ESPON Database Project	E-mail	Phone number
Future Orientation for Cities	FOCI	**Géographie-cités – Paris** Anne Bretagnolle	Anne.bretagnolle@parisgeo.cnrs.fr	+33 1 40 46 40 00
European Development Opportunities in Rural Areas	EDORA	**University Alexandru Ion Cuza – Iasi** Octavian Groza	Grozaoctavian@yahoo.fr	+40 23 22 01 487
Demographic and Migratory Flows affecting European Regions and Cities	DEMIFER	**UMS RIATE – Paris** Maher Ben Rebah/Ronan Ysebaert Claude Grasland	manager@espondb.eu grasland@parisgeo.cnrs.fr	+33 1 57 27 65 32 +33 1 57 27 65 33
Regions at Risk of Energy Poverty	ReRisk	**UAB – Barcelona** Andreas Littkopf Alejandro Iglesias	andreas.littkopf@uab.es alejandro.iglesias@uab.es	+34 9358 13 519 +34 9358 13 866
Territorial Impact Package for Transport and Agricultural Policies	TIPTAP	**IGEAT – Brussels** Moritz Lennert Gilles Van Hamme Didier Peeters	moritz.lennert@ulb.ac.be gvhamme@ulb.ac.be dpeeter1@ulb.ac.be	+32 2 650 50 78 +32 2 650 50 74 +32 2 650 50 77

Figure 22 – Table of contacts

ESPON projects - Priority 2

Project title (Acronym)	**Contact person – ESPON Database Project**	**E-mail**	**Phone number**
The Case for Agglomeration Economies in Europe (CAEE)	**UMS RIATE – Paris**, Maher Ben Rebah, Ronan Ysebaert, Claude Grasland	manager@espondb.eu	+33 1 5727 6532
The development of the Islands-European Islands and Cohesion Policy (EUROISLANDS)	**UMS RIATE – Paris**, Maher Ben Rebah, Ronan Ysebaert, Claude Grasland	manager@espondb.eu	+33 1 5727 6532
Cross-border Polycentric Metropolitan Regions (METROBORDER)	**UMS RIATE – Paris**, Maher Ben Rebah, Ronan Ysebaert, Claude Grasland	manager@espondb.eu	+33 1 5727 6532
SUccess for convergence Regions' Economies (SURE)	**UMS RIATE – Paris**, Maher Ben Rebah, Ronan Ysebaert, Claude Grasland	manager@espondb.eu	+33 1 5727 6532
Spatial scenarios: New tools for Local-regional Territories (SS-LR)	**UMS RIATE – Paris**, Maher Ben Rebah, Ronan Ysebaert, Claude Grasland	manager@espondb.eu	+33 1 5727 6532
Territorial Diversity (TD)	**UMS RIATE – Paris**, Maher Ben Rebah, Ronan Ysebaert, Claude Grasland	manager@espondb.eu	+33 1 5727 6532

Delivery and update of mapkit tool

In the frame of the Intra-ESPON data exchange, one of the tasks of the project was to develop a Map Kit that could be used by all other TPGs of the ESPON Programme.
Based on an ArcMap document (*.mxd) and a collection of shapfiles (*.shp), the aim of this map Kit is to have a Common ESPON map layout (same design).
The covering of the cartographic elements is the EU31 territory (EU27 + Switzerland + Norway + Liechtenstein + Island) on different regional levels (NUTS0, NUTS1, NUTS2,

NUTS3). The administrative units are available for the NUTS2003 and NUTS2006 versions

Gradually, the map kit will contain also vector data for the municipalities of the ESPON countries, a 1km European reference grid and other elements that can be useful for TPGs.

For a more precise description of the map kit, see chapter 3.1.

Answer to specific requests of ESPON CU

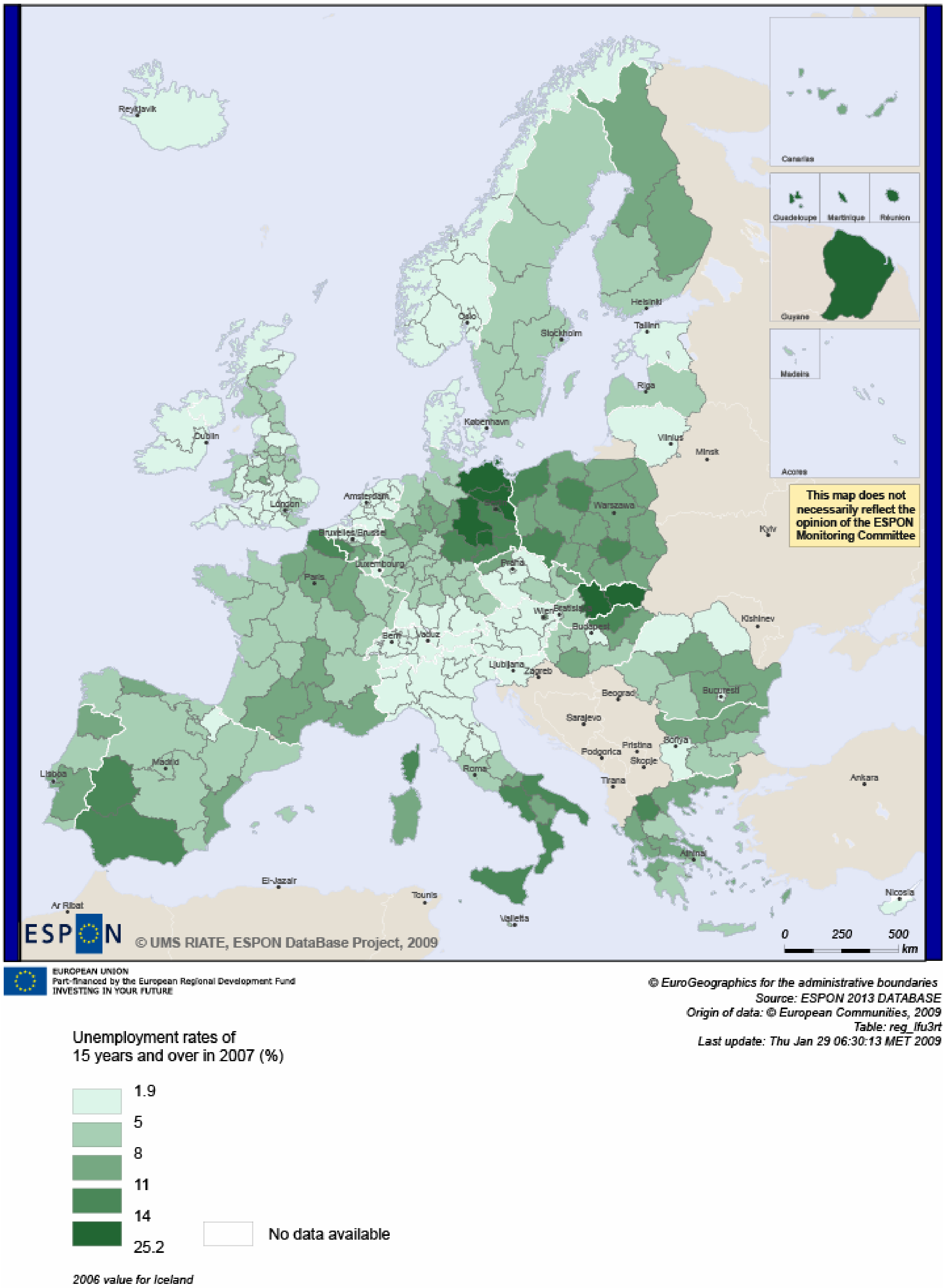
A specific situation was created by a direct action of ESPON CU that launched the production of "Territorial Observation Projects n°1" which was an update of the demographic typology. In this case, the ESPON DB project had not been associated with the action and could not be in contact with the researcher in charge of the project before the publication of the report (presented in Bordeaux in Dec. 2008). The data created by the researcher in charge of the publication was sent to ESPON DB project for inclusion into the database and also for quality control in order to order the payment of the study. This situation was at the same time positive and negative :

- ➔ The positive aspect is that ESPON DB received a first important delivery of new data that was a good opportunity to test the procedures of data quality control, metadata specifications, etc. Many experiences was made by RIATE and LIG on this datasets, providing concrete examples of problems and difficulties that have to be solved in the dataflows.
- ➔ The negative aspect is that the publication "ESPON Territorial Observation Projects n°1" was presented in Bordeaux in December 2008 without quality control of data but mentioning "ESPON Database" at a moment where data was not included. It appeared also (see. Annex 1) that data received from ESPON CU about this project did not fit with the maps presented in the publication and that some mistakes should be corrected.

In February, the ESPON CU sent another request of data to ESPON DB project for another project of Territorial Observation to be presented in June 2009, but one more time the request arrived very late and the discussion between ESPON DB project and the research team in charge of the study was rather chaotic and not well planified (see conclusion 4.3).

Support to ESPON CU for mapping

RIATE has provide some help to ESPON CU for mapping, out of the makit tool. For example, the realisation of a map of unemployment based on a recent Eurostat release, with estimation of missiong values for non EU members (see figure 23).



This map does not necessarily reflect the opinion of the ESPON Monitoring Committee

© UMS RIATE, ESPON DataBase Project, 2009

© EuroGeographics for the administrative boundaries
Source: ESPON 2013 DATABASE
Origin of data: © European Communities, 2009
Table: reg_lfu3rt
Last update: Thu Jan 29 06:30:13 MET 2009

2006 value for Iceland

Fig 23 : Regional unemployment rates in 2007

Elaboration of a provisional pragmatic scheme for the data flow

In order to avoid the problem created by "uncontrolled" delivery of data, it was necessary to provide precise guidelines to the ESPON DB project and to build a general scheme of the data flows. A first attempt was presented at the ESPON Bordeaux Seminar in December 2008, that appears to be rather a pragmatic provisional solution than a definitive answer.

The data flow considered here defines the itinerary followed by data coming from external sources, circulating between some of the ESPON 2013 Database Project partners for various processing, and being then made available to the other ESPON projects as well as, later and under certain conditions, to the external world.
A first data flow has been elaborated within the ESPON 2013 Database Project. This first version of the data flow has been discussed between the partners who have experienced the data integration process using the proposed file formats. The feedbacks allowed us to highlight some shortcomings in the formats initially proposed for the management of the whole knowledge required by the ESPON 2013 DataBase. This observation has led to the proposal and implementation of a second version of the data flow which consists of a refining of the files formats, the definition of the data acquisition and validation protocols, and modularity of the database itself as it was initially envisioned. The second version of the data flow is depicted on Figure .24

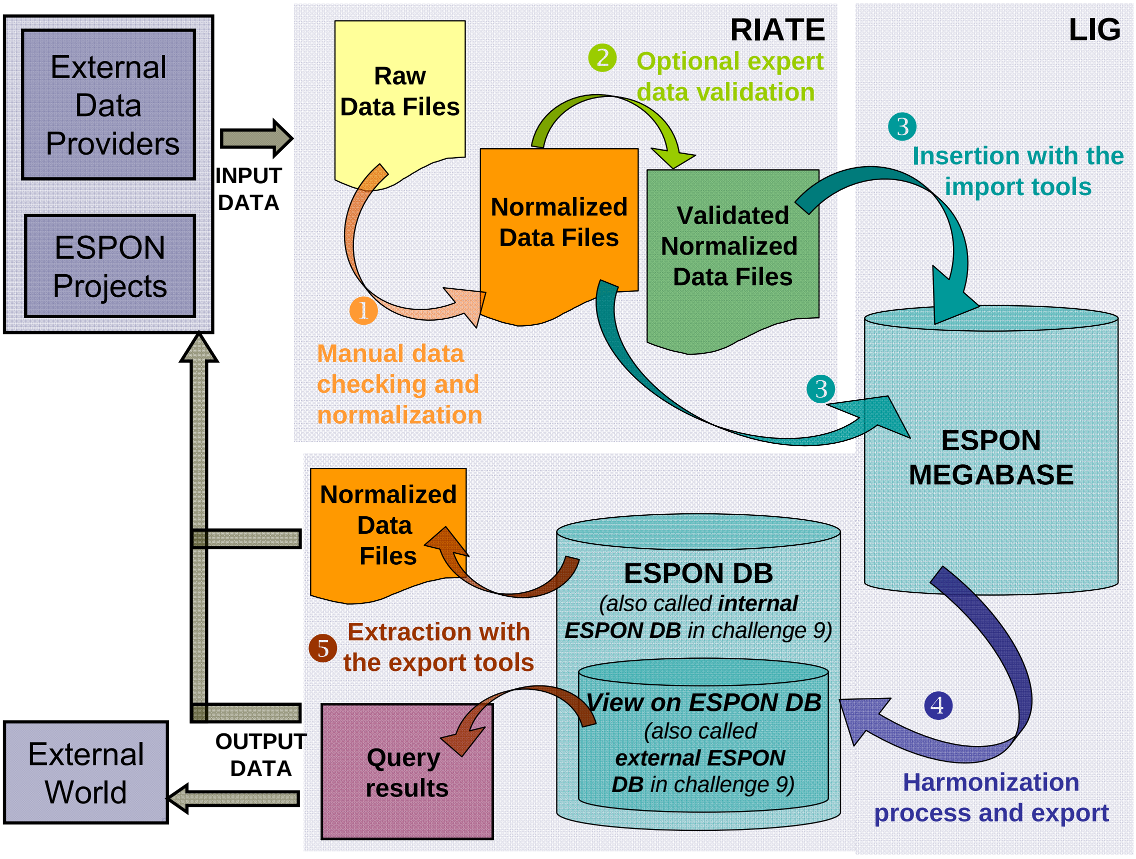

External Data Providers
ESPON Projects
INPUT DATA
RIATE
Raw Data Files
2 Optional expert data validation
Normalized Data Files
Validated Normalized Data Files
1
Manual data checking and normalization
LIG
3 Insertion with the import tools
3
ESPON MEGABASE
Normalized Data Files
ESPON DB
(also called internal ESPON DB in challenge 9)
5 Extraction with the export tools
View on ESPON DB
(also called external ESPON DB in challenge 9)
4
Harmonization process and export
Query results
OUTPUT DATA
External World

Figure 24 - Second version of the Data flow in the ESPON 2013 Database Project.

In this second version, external data are considered as input data in the flow. They can be provided in the shape of a statistical table downloaded from the Eurostat Web site, or a data file transmitted by any ESPON project, or a database received from a non ESPON project (e.g. a database resulting from a study of European Parliament on Shrinking Regions), or from different EC services (e.g. a database provided by the DG AGRI). External data can as well refer to national statistical figures provided by non EU countries (e.g. a database received from the ministry of employment of Switzerland, or a publication named « Liechtenstein in Figures »…). External data, whatever their format, are collected and centralized as so-called "raw data files" by the lead partner RIATE. Usually, when they enter the import pool of the ESPON DataBase (see challenge 9), raw data files respect some features: *i)* they are free from any previous transformation, *ii) they* have not been validated by any quality control yet, *iii) they* are not necessarily compatible with the standards used in the ESPON 2013 DB Project and, in particular with the cartographic standards.
Therefore, raw data undergo a primary, manual data checking process performed by the lead partner RIATE (see step 1 on Figure), which leads to the production of "normalized data files" as defined in Challenge 9. A normalized file is an Excel file compliant with the structure shown in table 1 (challenge 9). This step makes easier their importation into the DataBase (since the structure of the file is standardized). Optionally, when they seem doubtful or incomplete, normalized data files can also be validated by thematician experts (partner RIATE). If necessary, communication with the external data provider will be initiated in order to eliminate doubts or to correct erroneous values. After this step, files become "validated normalized data files" (see step 2 on the Figure).

Datasets passing the quality and consistency tests will be inserted in the MegaBase using the importation tools developed by the lead partner LIG (see step 3 on Figure). The information stored in the MegaBase can undergo further harmonisation processes before being exported towards the ESPON 2013 DataBase (see step 4 on Figure). Views on ESPON 2013 DataBase are updated to reflect the MegaBase content, if needed. Although described in Challenge 9, we recall here some precisions about the different databases to be considered.

The MegaBase is a complex structure, allowing data harmonization. The ESPON DataBase presents a simpler structure, optimized for easy and fast querying. We distinguish between the *internal* ESPON DataBase that contains data to be diffused only *from* and *towards* ESPON Projects and the *external* DataBase that contains data free of copyright, which can be diffused towards a wider public. We should point out that, from a content point of view, there are two databases – internal and external – we refer to by convenience, while, from a structural point of view, there is only one database – the ESPON 2013 DataBase. The external Database can be considered as a set of views on the ESPON 2013 DataBase.

Data outputs (see step 5 on Figure) for the other ESPON projects will be made available either in the shape of normalized data files diffused via the ESPON site and/or directly by the lead partner (similar to the ones used to populate the DataBase), or as results of queries formulated on the ESPON DataBase (via a server) through a simple Web interface. The Espon DataBase, installed on a Web server hosted by ESPON, will grant different access rights for different groups of users (ESPON projects in priority, EC members, general public, etc.), in order to cope with data rights issues.

As a side note, it is important to stress that the process of formatting heterogeneous file structures dynamically could prove to be highly time consuming and could slow the overall progress of the ESPON 2013 Database Project. Therefore, in order to fully automate the data validation and importation process, it is desirable that data inputs are provided directly by other ESPON projects using the normalized format we have defined. As a step in this direction, the two works mentioned below have been initiated.

Definition of a first data and metadata profile for socio-economic data

The notion of profile aims at offering external data providers some guidelines to help them structuring the datasets they want to share with the ESPON 2013 DataBase Project. These guidelines essentially take the form of documented file formats and high level tools for producing datasets files in compliance with these formats. In challenge 9, a full overview is given concerning the format of the files that have been established to integrate socio-economical data (the ones available on various partitions) with their metadata. A first template for filling in metadata is described in challenge 9 and available in chapter 3.2.

Implementation of an interface on Espon DataBase

The interface on ESPON Database aims at increasing the usability of the database by a wide public (ESPON projects in priority, EC members, general public, etc.). In Chapter 9, is given a full overview of the interface to be used for the extraction of data from the ESPON DataBase.

2.8.3 Identified difficulties

ESPON DB Project should have started 6 months before the other ESPON Projects

This challenge has revealed to be one of the most difficult. One of the main reason of the difficulties (mentioned in the Inception Report) is the fact that the ESPON DB project started at the same moment than other ESPON Projects under Priority 1 when it would have been much more convenient to introduce a delay of minimum 6 months before the beginning of ESPON DB and the other projects, making possible to work without pressure on the elaboration of clear recommendation for other ESPON projects in terms of dataflows and metadata. The problem was made increasingly difficult by the delay in the signature of ESPON DB contracts[15]. Finally, there was a lack of coordination between ESPON DB and ESPON CU concerning the specific contract related to "data updates" and "territorial observation" publications. And the consequence was a huge amount of work for the integration ex-post of the data that

[15] From a legal point of view, it is true that the ESPON DB project was supposed to start in July 2008 because expenses are *theoretically eligible* since this period. But *in practice*, the universities that are the legal authorities of most of the pattern of ESPON DB project never accept this practice, especially concerning staff costs. It is only when the signature of the contract was fully achieved (in November 2008) that it was really possible to engage sufficient work force on the project.

was not delivered in right form and was not consistent initially with the specification of ESPON DB (that was not achieved at the moment where this studies was launched).

The metadata question has revealed to be much more difficult than expected

Whatever the previous difficulties, ESPON DB has also to face a moajor difficulty concerning the choice of the right model of metadata, both for input of information from statistical organisation (Eurostat, EEA, UNEP, …) and project partners (Priority 1, Priority 2, …) or from output of information toward ESPON community and external world. Contrary to our initial expectations, there was no clear solutions available at European level, especially considering the diversity of geographical objects (grid, NUTS, States, Cities) and the diversity of thematics (social, economical, environmental, transport, …) that should be included in the ESPON database and that was requested by the Projects under priority 1. This topic of metadata was a major concern of the ESPON DB internal meeting of Februrary and implies all project partners in a common reflection that is not fully achieved but has made substantial progress (see. 3.2).

Also, experts from RIATE have reported some integration problems they had to cope with when integrating data. Such a feedback has been used to improve both the data flow and the ESPON 2013 DataBases. Such a situation may happen again when other external data providers will use our profiles, templates and interfaces for submitting their data. For instance, present metadata template may not be complete or adapted to the various data formats we may have to integrate. This should be the case, for non socio-economical indicators, as well as for non NUTS units (global, local, urban). Anyway, we are convinced that the first experience we had with socio-economical data on NUTS territorial units will be highly beneficial for those new configurations.

The identification of data to be collected by ESPON DB or by other ESPON Projects

Related to the previous debates, appears the important question to decide on which data has to be collected by ESPON DB project and which data has to be elaborated by other ESPON projects and simply stored by ESPON DB after quality control. This questions is difficult, as demonstrated by the experience of the integration of demographic data produced by the "Territorial Observation N°1" publication. On the one hand, it could have been possible to store all the data received by this project "as they was", just mentioning the name of the author. On the other hand, it was difficult to put in the ESPON DB some data collected by this project (population, births, deaths) that was contradictory with the same data collected by ESPON DB project from Eurostat. When data are of general use (as population) it is better to use data collected by ESPON DB project that are the same for all other projects. But what is the limit between data of general use and data specifically related to one thematic project under priority 1 or priority 2.

2.8.4 Work plan

We have drawn the conclusion that, in the future, the questions of dataflows (that has been discussed in Challenge 8) should not be separated from the question of data model (that are discussed in Challenge 9). We propose therefore to transform this challenges as follow :

New Challenge 8 : Mapkit tool and support to CU for map publication

New Challenge 9 : Data flow and data model integration.

As a consequence, we describe here only the workplan related to the new challenge 8 and we transfer to challenge 9 the work plan related to data flows.

June 2009

Updated version of map kit tool

Technical report on cartographic principle to be followed in ESPON

Cartographic support to ESPON CU

December 2009

Updated version of map kit tool

Proposal of one day formation in ESPON cartography

Cartographic support to ESPON CU

2.9 Challenge 9: Data model and integration

Coordinator: LIG & ULG

Internal networking (other ESPON Projects)

2.9.1 Objectives

The general objective through this challenge is to design and implement an operational software architecture that ensures the importation, the integration, and the exportation of sets of statistical indicators linked to geographical units, mainly NUTS but not only units. That is to say that the ESPON 2013 Database can be seen as a Spatio-Temporal Information System (or STIS), relying on several software components: databases (for data and metadata storage), ontologies (for data and metadata structuring) and programming routines (implementing procedures in charge of data harmonization, verification and estimation).

As a consequence, Challenge 9 in itself comprises several sub-challenges, listed below:

First, a sound and flexible data model capable of storing most types of data provided by statistical organisms must be built. Not only should this model be able to manage indicators provided for evolving (in time, in space) NUTS, but also for more global (WUTS) and more local (LA) spatial scales, as well as for urban areas (defined as UA, MUAS, FUAS), and also for grids as far as environmental data are concerned.

Second, the ESPON 2013 DataBase must handle the storage of metadata at the finest levels of detail available. However, prior to this, an extensible ESPON format for metadata must be defined, relying on the existing Inspire (ISO 19115) standard for the geographical part of data description, and on the SDMX recommendations for the statistical part of data description.

Third, the ESPON 2013 DataBase must integrate data harmonization algorithms, based on the knowledge of thematic experts, in order to reconstitute complete past temporal series and to allow predictions about future trends and evolutions. These algorithms – whose goal is to produce complete medium and long-term time series –, and the expert knowledge that supports the system to determine which evaluation method to activate, will be encapsulated in a problem-solving environment based on ontologies.

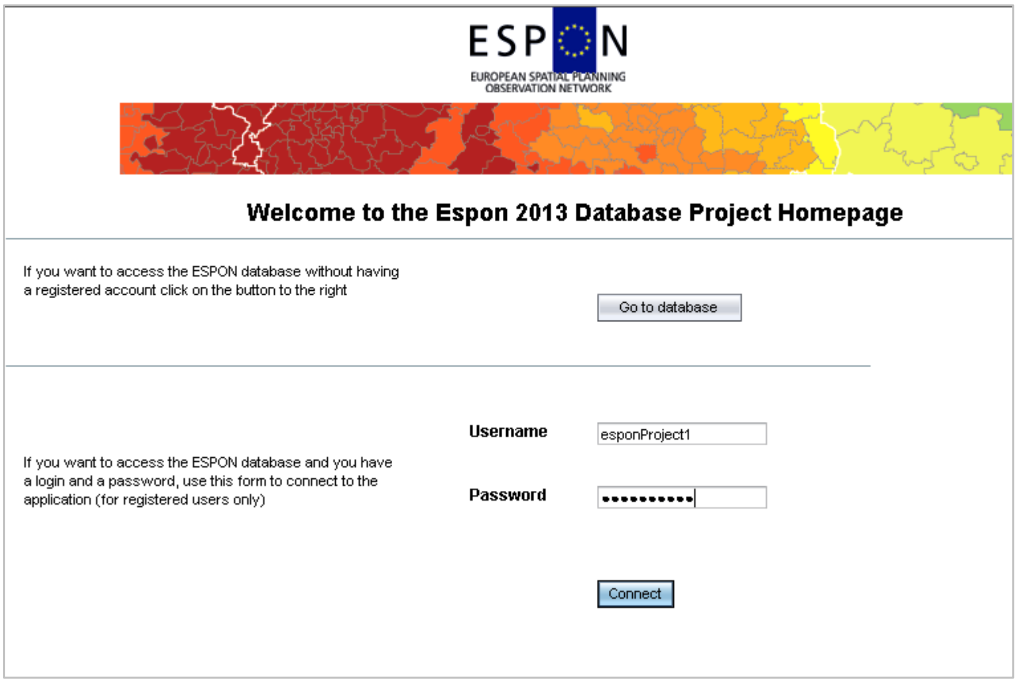
Finally, besides these three main sub-challenges, the ESPON 2013 DataBase has also two challenges to face in terms of: 1) usability (for experts as well as for non experts) attainable through user-friendly interfaces that authorize users to provide and/or extract data together with their metadata, and 2) performance (in terms of CPU time and memory size) through a Web server.

2.9.2 Work done

We describe in this section the work done since the Inception Report.

The interaction between users and the ESPON 2013 DataBase is performed through an application (called the ESPON 2013 DataBase Application), which has been implemented and offers a Web-based interface (see figure 25). Being Web-based, no particular tool or software is required to access this interface; any Web browser can be used. The ESPON 2013 DataBase Application will be installed on the ESPON Web server and accessible here at a given url.

The application is developed in Java Server Pages (JSP), a Java Web development framework combining Java Servlets, Java Server Pages and Java classes in a component-oriented architecture. The databases are implemented using the PostgreSQL/PostGIS database management system. Free and open source libraries are used for database connectivity and linking with the different data formats (.xls, mif/mid, etc.). For data transformation and manipulation purposes, the open source Kettle Pentaho Data Integration tool is used.



Welcome to the Espon 2013 Database Project Homepage

If you want to access the ESPON database without having a registered account click on the button to the right

If you want to access the ESPON database and you have a login and a password, use this form to connect to the application (for registered users only)

Username: esponProject1

Password: ●●●●●●●●●●●

Figure 25 - The first screen of the ESPON DB Application Web interface.

The first screen of this interface is shown on Figure It allows two types of access:

One registered access, for users to whom a login and a password have been given in order to access both the data upload and download interfaces. Clearly, this controlled access is meant for other members of ESPON Projects who will be able through it to upload and download their own data and metadata. In the future, access to different partitions or views of the internal part the ESPON DataBase will be given, from access rights associated with logins, depending on the category and ESPON projects to which the users belong.

One anonymous access (by directly clicking on the “Go to database” button), meant for a wider public, which only permits to visualize and download data from the external part of the ESPON DataBase (i.e. data free from copyright rights).

In the following sub-sections, the presentation of the work done since the Inception Report describes the three main components of the ESPON DataBase Application architecture: the import pool, the integration pool and the export pool.

The import pool

We describe here the Web interface, which has been implemented for uploading data. This upload interface ideally allows data providers, external to the ESPON DB Projects, to send their own data so that they can be integrated in the ESPON DataBase (see figure 26). So far, accordingly to the data flow presented in challenge 8, this integration phase has been performed under the control of the partner RIATE. This means that this simple interface is accessible only to registered users and provides 4 buttons for data upload purposes.

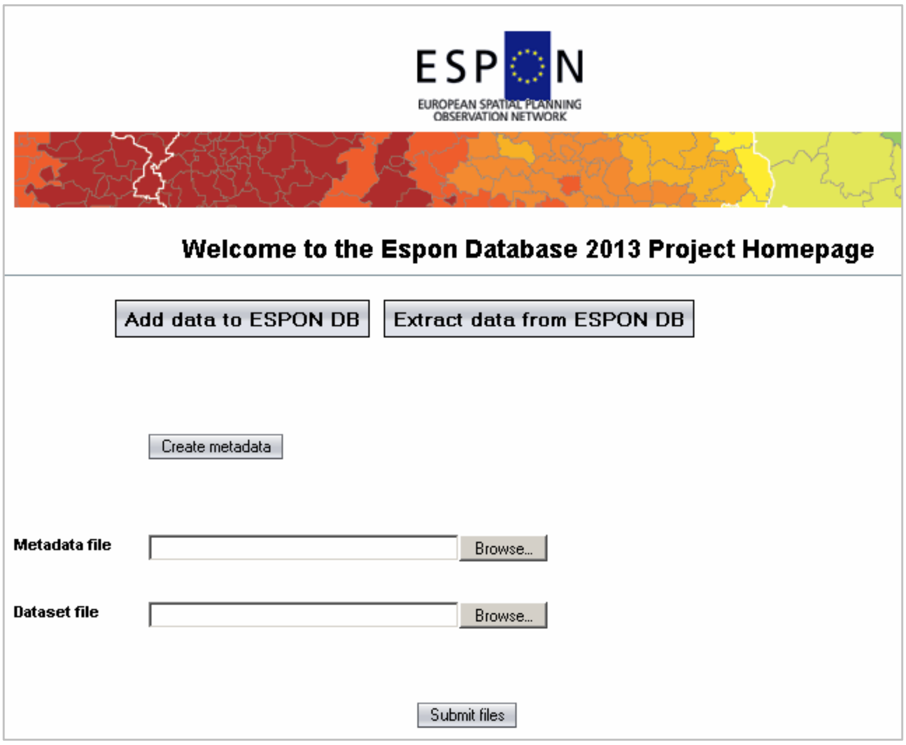

ESPON
EUROPEAN SPATIAL PLANNING OBSERVATION NETWORK
Welcome to the Espon Database 2013 Project Homepage
Add data to ESPON DB
Extract data from ESPON DB
Create metadata
Metadata file
Browse...
Dataset file
Browse...
Submit files

Figure 26 - The data upload Web interface of the ESPON DB 2013 Application.

At the top of the data upload interface, the two buttons “Add data to ESPON DB” and “Extract data from ESPON DB” allow registered users to navigate between the data import and export pages.

The “Create metadata” button allows data providers to open the ESPON metadata editor in order to fill and upload the metadata that describe the dataset they upload. Up to now, the metadata editor is a simple Excel form, containing the fields required by the ESPON DB application metadata processing. Once edited and filled, the metadata file can be saved on the computer of the user (here, authorized provider).

The two browsing buttons for metadata and, respectively, for data files, allow users to upload files saved on their hard disks and to send them to the ESPON Web server. Files will be sent by pairs (data file + metadata file). The metadata file is a file that has been edited using the “Create metadata” button above.

The “Submit files” button confirms the sending of a pair of files (data and corresponding metadata) to the server.

Towards Input Standard Formats

The ESPON DB 2013 Project will collect a minimal set of metadata so that data with their corresponding metadata (according to INSPIRE directive, and the included ISO 19115 recommendation) are delivered. It can be noticed that a subset of metadata information can be computed from the data files themselves. For example, the geographical extent of an indicator can be computed from the set of units code of the NUTS nomenclature.

In order to ease the capture of metadata, this process should be coupled with the data acquisition process (see Figure 27). This way, only one metadata file should be associated with one data file. However, a data file can contain various columns of indicators, concerning many themes and having their own lineage, quality level, temporal extent, and so on. Then, the only thing that is common to all data of a data file is the geographic extent. As shown on Figure , data should be first parsed to compute as much as possible metadata information automatically before providing the user with forms in order to complete information.

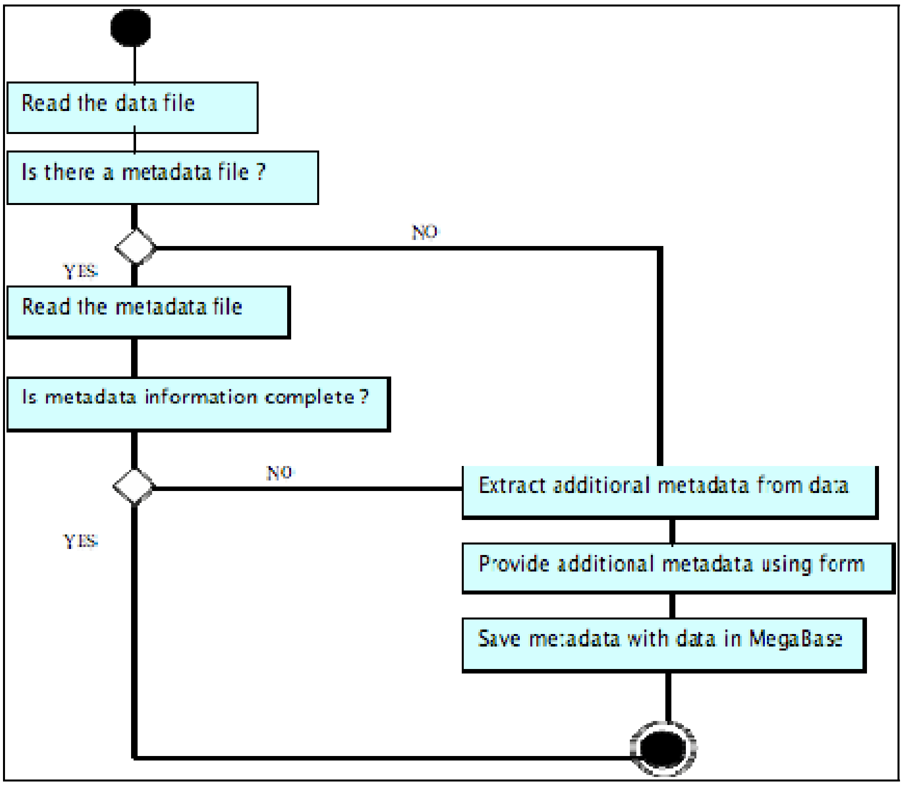

Read the data file
Is there a metadata file ?
NO
YES
Read the metadata file
Is metadata information complete ?
NO
YES
Extract additional metadata from data
Provide additional metadata using form
Save metadata with data in MegaBase

Figure 27 - Metadata and data acquisition process.

In order to ease the automatic processing and integration in the ESPON MegaBase (see for its definition below), a standard structure for the input Excel data files has been defined. This structure is the result of discussions led in close collaboration with the thematic experts involved in the ESPON 2013 DB Project, mainly partners from RIATE. It represents a good compromise between computer requirements (for easier automation and integration) and human requirements (for easier production and manipulation by users without computer science expertise).

Basically, each data file is provided in the shape of one Excel sheet. The main idea of this format is to capture all the necessary elements defining an indicator value. This way, information is made explicit and not hidden within the file path, the file name or the sheet name. Some of these elements are valid for an entire column (the indicator code and the data production year) and as such, they can be put into the header of the column containing the values of the indicators (see Table 1).

id	level	Area	category1	pop_t	category2	gdp_eur	Category3
		2003		2003		2003	
AT111	NUTS3	701,5	Eurostat,2007	37,5	Eurostat,2007	677	Eurostat,2007
AT112	NUTS3	1792,6	Eurostat,2007	141,7	Eurostat,2007	3108,3	Eurostat,2007
AT113	NUTS3	1471,4	Eurostat,2007	97,4	Eurostat,2007	1571,9	Eurostat,2007
AT121	NUTS3	3356,7	Eurostat,2007	237,9	Eurostat,2007	4585,8	Eurostat,2007
AT122	NUTS3	3367,1	Eurostat,2007	247,8	Eurostat,2007	5352,8	Eurostat,2007

Table 1 - Example of fragment from a standard input data file

For the codes of the indicator, the EUROSTAT naming conventions should be used, since EUROSTAT already provides a kind of standard, recognized and used by other important data providers as well (e.g. DG Regio).

The identifiers of the territorial units are valid for a whole row, so they are reported only once in the first column (named "id").

As an additional means of identification, the level of the territorial NUTS can be specified (see column "level" in Figure 1.9.4.). This is mandatory information only for older nomenclatures, because starting from NUTS 1999, the unit codes are normalized: all territorial units have a unique code, which also accounts for their level in the hierarchy. For some older nomenclatures, the same code was used for territorial units situated on two different levels. For instance, if some big region was composed of only one region (they had the same geographical extent), the codes were sometimes the same, although the indicator values were different, as semantically the 2 entities were different.

The only element that requires one field each time is the actual value of the indicator. It is then associated with a metadata label referencing some quality information that can be found in the metadata file that is delivered with the data file. Unknown indicator values should be simply left blank.

A first metadata profile for ESPON DB 2013

A full description of the template for socio-economical data is provided in the section 3.2.3.2. It has been agreed with the RIATE partner that this information should be provided at indicator level. The crucial point to understand is that metadata for each indicator may have various scopes (a spatial and temporal context identified by a label) and this scope is referenced in the Excel data file, in the column 'category' next to the column 'value'.

The integration pool

The integration pool of the ESPON 2013 DB Project is based on two main databases: the MegaBase, which is a complex structure, allowing data harmonization, and the ESPON DataBase, which is a simpler structure, optimized for easy and fast querying. In the ESPON DataBase, one should distinguish between the internal ESPON DataBase and the external ESPON DataBase, on the basis of the manipulated content. The internal ESPON DataBase is to contain data that can be diffused only from and towards ESPON Projects, while the external database is to contain data free of copyright, which can be diffused towards a wider public. We should point out that, from a content point of view, there are two databases – internal and external – we refer to by convenience, while, from a structural point of view, there is only one database – the ESPON DataBase. The explicit distinction will be made in the remainder only when content issues are addressed.

Model of the ESPON DataBase

Since the Inception Report, the model of the ESPON DataBase has been subject to substantial modifications we report here. The stable version of this model now accepts and integrates formatted set of data as defined above and provided by the thematic experts from RIATE.

The latest version of the model is presented in Figure . The new improved model contains the following tables:

Territorial_unit contains every territorial unit defined by the NUTS nomenclatures (e.g. Luxembourg, Île-de-France, etc.);

Study_area contains the predefined study areas that might be required by the end-users as a whole (EU27, EU15, etc.). They are provided to make data retrieval easier and faster for the end-users;

Custom_TU_set allows storing arbitrarily created spatial divisions (like the "standard" NUTS23 division or the modified NUTS23 mixture used in the Mats Johanson database);

Visibility allows managing the copyright issues and making the difference between the content accessible in both the internal and the external parts of the database;

Category describes the lineage of the data at the level of the indicator value. It allows the dynamic exportation of metadata simultaneously with data exportation;

Footprint stores the spatial boundaries of all the territorial units;

Indicator stores the description of all the indicators in the database, allowing the association of a standard code (like those used in the EUROSTAT naming conventions) with an indicator meaningful name;

Indicator_value holds the actual indicator values for the different years, different sources and for different territorial units.

The ESPON DataBase has been filled with the values for the basic indicators for NUTS2003 and NUTS2006, which were also delivered in Excel file format by the thematic experts from the partner RIATE, and can be used for data exportation. It will be made available on the ESPON site.

Although most development efforts were until now directed towards the proposal of a first version of the internal/external ESPON DataBase, in order to answer as quickly as possible the demand coming from other ESPON Projects concerning data, some high attention was also devoted to the finalisation of the data model and in the implementation of the MegaBase.

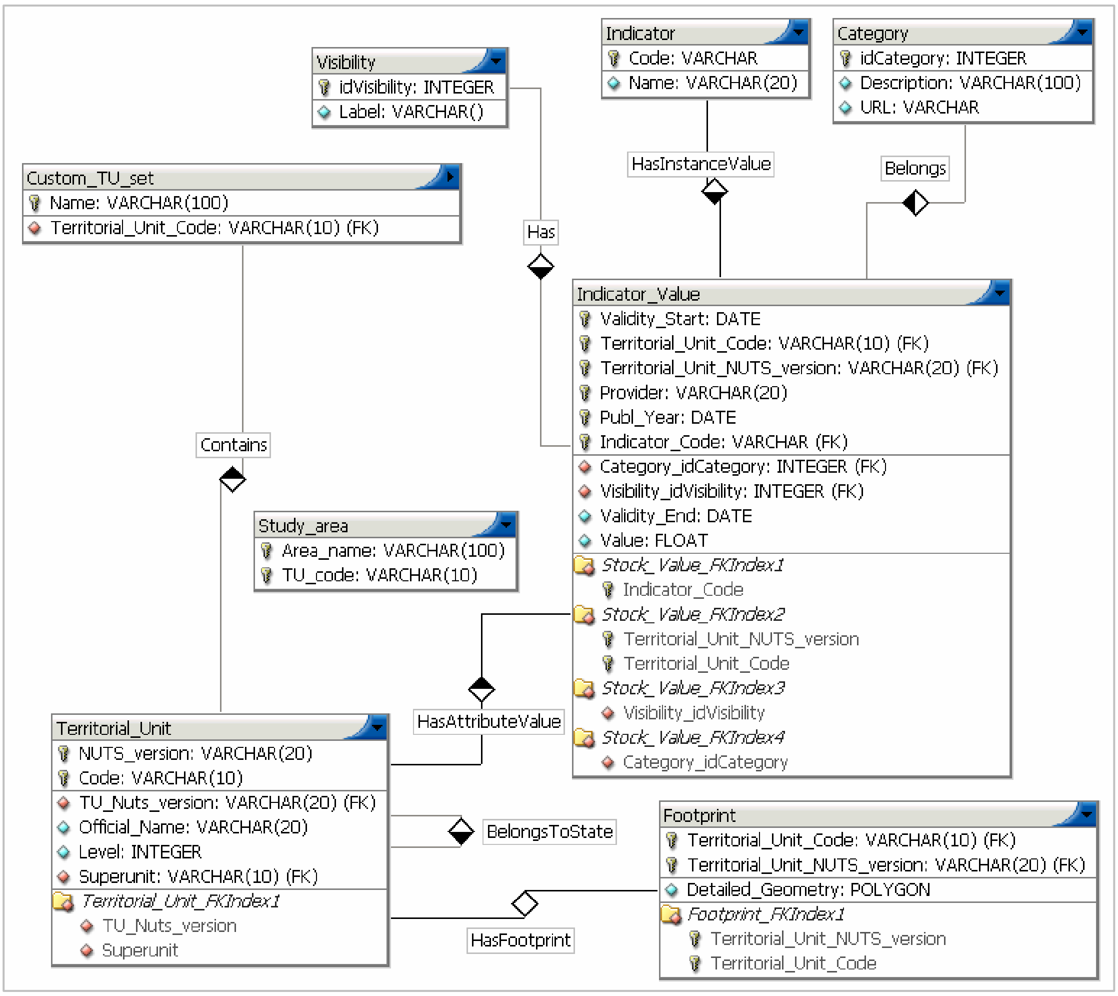


Figure 28 - Data model of the internal ESPON database

Model of the MegaBase

A first version of the operational model (see Figure 28 and 29) has been implemented and the data acquisition process has started with the integration of territorial units in the NUTS2003 and NUTS2006 nomenclatures. The first version of the MegaBase is designed in order to fulfill the requirements for spatial tessellation data (like NUTS0-NUTS5 and WUTS). It contains the following tables:

Territorial_unit holds the internal identifier of the territorial unit and its temporal extent;

Name describes all the official names of the territorial unit (they can change in time);

Code describes the conventional codes identifying the territorial unit within different nomenclatures or different nomenclature versions;

Nomenclature stores the different nomenclatures and nomenclature versions (e.g. NUTS1999, NUTS2006, WUTS). Nomenclatures have different temporal extents;

Aggregation is the relation describing the hierarchical composition relations between territorial units. An aggregation relation is valid only within the scope of a certain nomenclature version;

Geometry contains the spatial representations of a territorial unit. One territorial unit can have several corresponding spatial representations, for different scales and usages;

Genealogy describes the historical horizontal relations existing between territorial units (e.g. Germany today is strongly related to the ex-GDR);

Provider allows describing the different data providers (statistical organisms, national institutes, etc.);

Source describes databases, which can provide indicator values and are created and maintained by data providers;

Indicator describes all the variable values (may they be stocks, ratios, or more complex indicators, in different terminologies);

Indicator_value holds the actual variable values for one territorial unit, one period of time (e.g. one year) one data source, and one type of indicator.

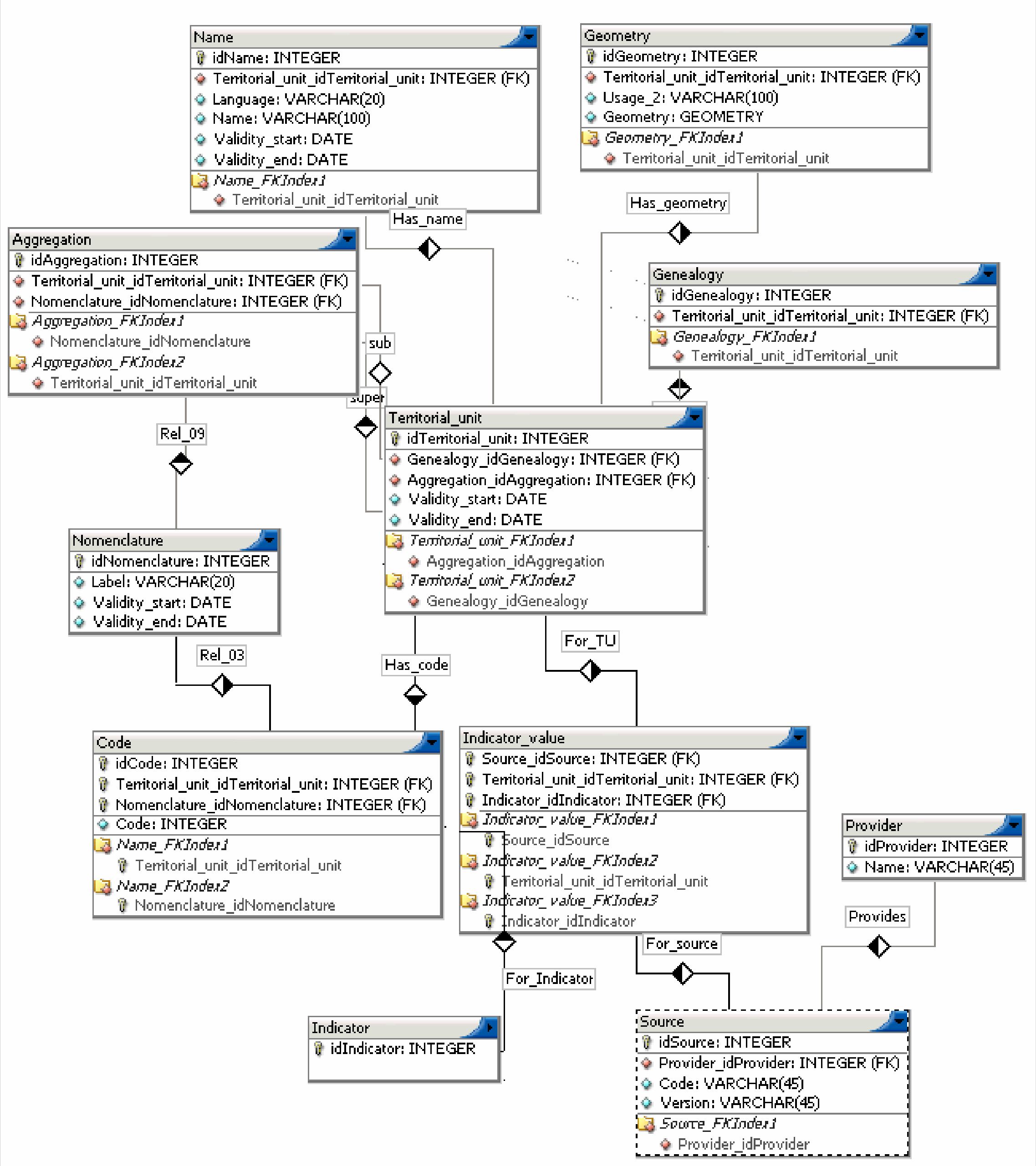

Name
idName: INTEGER
Territorial_unit_idTerritorial_unit: INTEGER (FK)
Language: VARCHAR(20)
Name: VARCHAR(100)
Validity_start: DATE
Validity_end: DATE
Name_FKIndex1
Territorial_unit_idTerritorial_unit
Geometry
idGeometry: INTEGER
Territorial_unit_idTerritorial_unit: INTEGER (FK)
Usage_2: VARCHAR(100)
Geometry: GEOMETRY
Geometry_FKIndex1
Territorial_unit_idTerritorial_unit
Has_name
Has_geometry
Aggregation
idAggregation: INTEGER
Territorial_unit_idTerritorial_unit: INTEGER (FK)
Nomenclature_idNomenclature: INTEGER (FK)
Aggregation_FKIndex1
Nomenclature_idNomenclature
Aggregation_FKIndex2
Territorial_unit_idTerritorial_unit
Genealogy
idGenealogy: INTEGER
Territorial_unit_idTerritorial_unit: INTEGER (FK)
Genealogy_FKIndex1
Territorial_unit_idTerritorial_unit
sub
super
Rel_09
Territorial_unit
idTerritorial_unit: INTEGER
Genealogy_idGenealogy: INTEGER (FK)
Aggregation_idAggregation: INTEGER (FK)
Validity_start: DATE
Validity_end: DATE
Territorial_unit_FKIndex1
Aggregation_idAggregation
Territorial_unit_FKIndex2
Genealogy_idGenealogy
Nomenclature
idNomenclature: INTEGER
Label: VARCHAR(20)
Validity_start: DATE
Validity_end: DATE
Rel_03
Has_code
For_TU
Code
idCode: INTEGER
Territorial_unit_idTerritorial_unit: INTEGER (FK)
Nomenclature_idNomenclature: INTEGER (FK)
Code: INTEGER
Name_FKIndex1
Territorial_unit_idTerritorial_unit
Name_FKIndex2
Nomenclature_idNomenclature
Indicator_value
Source_idSource: INTEGER (FK)
Territorial_unit_idTerritorial_unit: INTEGER (FK)
Indicator_idIndicator: INTEGER (FK)
Indicator_value_FKIndex1
Source_idSource
Indicator_value_FKIndex2
Territorial_unit_idTerritorial_unit
Indicator_value_FKIndex3
Indicator_idIndicator
Provider
idProvider: INTEGER
Name: VARCHAR(45)
Provides
For_source
For_Indicator
Indicator
idIndicator: INTEGER
Source
idSource: INTEGER
Provider_idProvider: INTEGER (FK)
Code: VARCHAR(45)
Version: VARCHAR(45)
Source_FKIndex1
Provider_idProvider

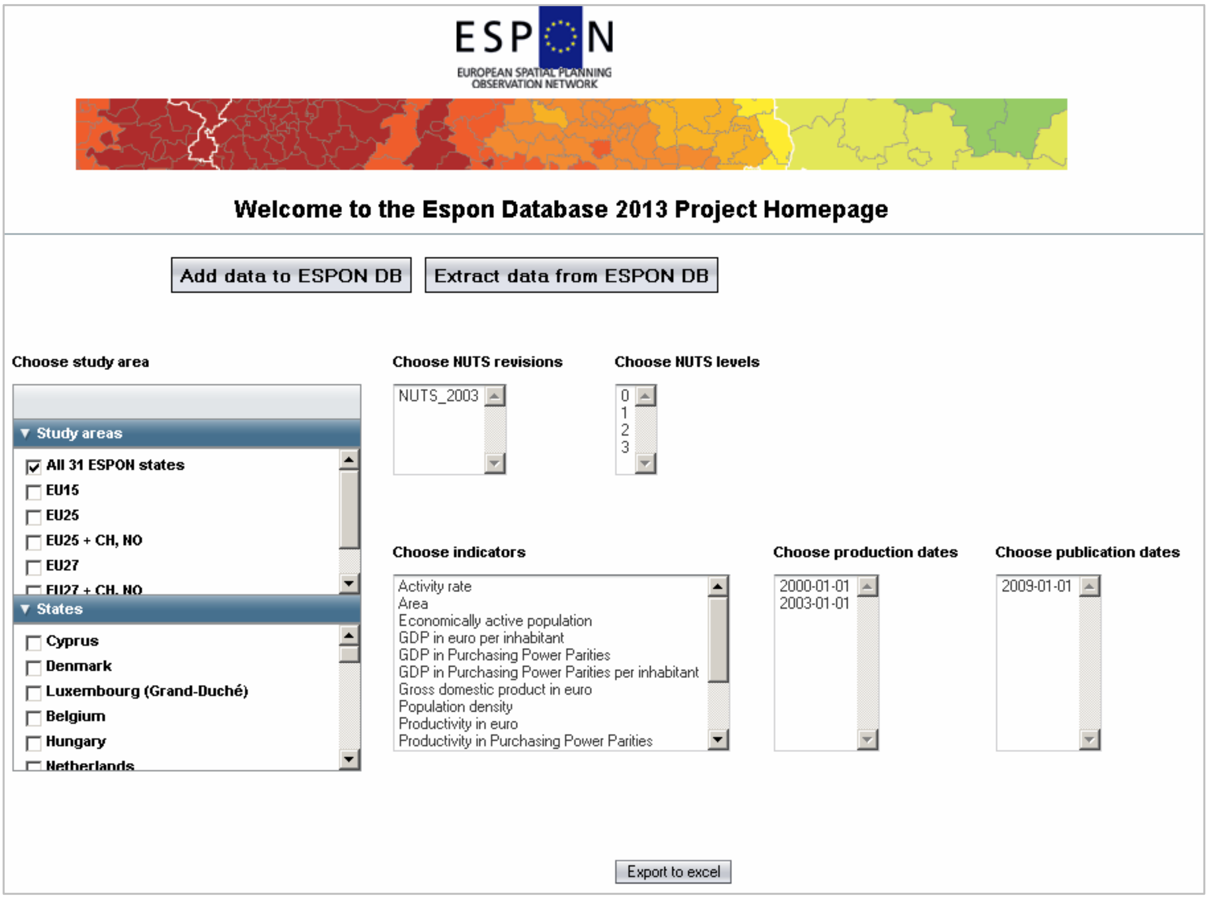
Figure 29 -Data model of the ESPON MegaBase.

The data acquisition process will be pursued in the short term, while the MegaBase data model will also be gradually extended and updated in order to integrate conceptual progresses made in the definition of non-NUTS territorial units (WUTS, LA2, LA1…) and cities representation (UA, MUAS, FUAS…).

The export pool

Within the export pool, a Web interface for querying and exporting data has been implemented (see Figure 30). This interface allows users to search for data in the database by choosing a series of parameters (study area, NUTS version, NUTS level, production date and possibly others, like the data source and publication date). Each time a user selects a parameter, the lists of available parameters are dynamically updated, according to the kind of data available in the database. For instance, if a user selects the "NUTS2003" version, the indicator list available is limited to the indicators available for the territorial units in the NUTS2003 revision.

After selecting all the needed parameters, the user can export the data as an Excel file containing the result of the query. The resulting Excel file has exactly the same structure as the standard input format, allowing users to retrieve data for one or more NUTS levels, indicators and production years, and for exactly one NUTS revision and publication date.



ESPON
EUROPEAN SPATIAL PLANNING OBSERVATION NETWORK
Add data to ESPON DB
Extract data from ESPON DB
Export to excel

Welcome to the Espon Database 2013 Project Homepage

Choose study area

▼ Study areas
☑ All 31 ESPON states
□ EU15
□ EU25
□ EU25 + CH, NO
□ EU27
□ EU27 + CH, NO

▼ States
□ Cyprus
□ Denmark
□ Luxembourg (Grand-Duché)
□ Belgium
□ Hungary
□ Netherlands

Choose NUTS revisions
NUTS_2003

Choose NUTS levels
0
1
2
3

Choose indicators
Activity rate
Area
Economically active population
GDP in euro per inhabitant
GDP in Purchasing Power Parities
GDP in Purchasing Power Parities per inhabitant
Gross domestic product in euro
Population density
Productivity in euro
Productivity in Purchasing Power Parities

Choose production dates
2000-01-01
2003-01-01

Choose publication dates
2009-01-01

Figure 30 - The data download Web interface.

2.9.3 Identified difficulties

During this first phase of the project, we have encountered some difficulties mainly due to the short period of development granted. Thus, since other ESPON Projects have started in parallel, the ESPON 2013 DataBase Project has to be able to deliver data sets at the same time as the design and implementation of the ESPON DataBase were launched. As a consequence, there was a short allocated time to design and develop the internal/external ESPON DataBase (in order to be able to provide other ESPON projects with valuable data to work with), whereas, it should have been more profitable to have time left to design and implement the MegaBase (required for data harmonization) before the ESPON DataBase which is supposed to contain harmonized data. This situation has led us to readjust the implementation frequently according to the progress we made in the conception activity. Normally, the ideal situation would have been to have a solid conception before getting to the implementation.

Also, apart from the fact that one of the objectives of the ESPON 2013 DataBase Project is to provide Inspire compliant metadata together with sets of indicators, it seems that the difficulty of this task must not be underestimated. That is to say, as shown by the study of existing metadata standards and approaches (INSPIRE, SDMX, and the way they are used within the European and global data infrastructures), the needs of the ESPON 2013 DataBase Project go well beyond what is presently specified in these standards and what is done in the current applications. In the ESPON 2013 DataBase Project, the objective is to be able to integrate automatically data and metadata into one model/database and also to dynamically generate data and metadata. This will be the main difficulty to solve in the mid term.

2.9.4 Work plan

We present in this section the tasks (described here in terms of activities of both conception and development) that we will lead in the forthcoming year, up to the Second Interim Report (planned for February 2010), including the two milestones that correspond to the two ESPON seminars to be held in 2009.

Work in the Challenge 9 is to be pursued as follows:

June 2009

Partners involved in the definition of an ESPON format for metadata (namely, UL, AUB, RIATE, LIG) will meet in March in Barcelona. As a result, one template (or profile), compliant to the standards Inspire ISO 19115 and SDMX, will be then defined and implemented.

Write documentation on the data flow for the attention of external providers (among which other ESPON projects): "guidelines for datasets submission" including an extensive guide concerning metadata information to provide.

Design and implementation of the first version of an editor of metadata based on the template defined. This editor will mainly allow the association of metadata with socio-economical indicators.

Extension of the existing DataBase schema in order to integrate metadata items as described by the defined template.

Propose an update of the data flow after it has been more largely put to the test by the lead partners RIATE and LIG by integrating wider datasets supplied by different external providers.

Specification of the automation of the data checking process on import side, that is to say how to reduce the amount of manual checking for RIATE.

Results of the first discussions held on the extension of the ESPON DataBase towards the integration of global, local and urban units.

The ESPON DataBase will be filled periodically with new sets of indicators.

December 2009

More evolved version of the metadata editor, according to the evolution of the metadata template.

Update the documentation on the data flow for the attention of external providers (among which other ESPON projects): "guidelines for datasets submission" including an extensive guide concerning metadata information to provide.

New version of the ESPON DataBase schema: extension of the MegaBase schema towards the integration of quality information.

First conceptual extension of the ESPON DataBase schema towards the integration of socio-economical indicators for global, local and urban units.

Design of the first version of an ontology in charge of describing the genealogy of the geographical units.

Design of the first version of an ontology in charge of describing the evolution of the definition of indicators in time, country, providers, etc.

Development for the automation of the data checking process on import side.

Development for the automation of ready to use harmonized datasets extraction from the MegaBase towards the DataBase. Datasets will be made available to other ESPON projects and to the external world through a Web interface.

The ESPON DataBase will be filled periodically with new sets of indicators.

February 2010

First implementation of the ESPON DataBase schema integrating socio-economical indicators for global, local and urban units.

First considerations about the Problem Solving Environment dedicated to the estimation of missing values.

Implementation of the first version of an ontology in charge of describing the genealogy of the geographical units.

Implementation of the first version of an ontology in charge of describing the evolution of the definition of indicators in time, country, providers, etc.

The ESPON DataBase will be filled periodically with new sets of indicators.

Second Interim Report

2.10 Challenge 10: Spatial analysis for quality control

Coordinator: LIG & NCG

Objectives: To develop spatial analysis and datamining methods in order to identify exceptional values

The National Centre for Geocomputation (NCG) at the National University of Ireland, Maynooth are acting as experts for the ESPON 2013 Database project. They will contribute a case study on statistical analysis tools which will be applied to quality control procedures in the database. This case study will be realised in two 12 month steps, each of which will be terminated with a report. The same expert team will be used for both steps.

The first step will run from December 2008 to December 2009. The key aim of this work is to examine how statistical analysis tools can be applied to the ESPON 2013 Database in order to find 'outliers', i.e. data that are exceptional as compared to neighbouring data with respect to: (a) attribute-space, (b) geographical-space and/or (c) temporal-space (and where each dimension depends on the scale it is viewed at). The outcome will be a review of existing tools in the field of statistics, data mining, GIS, and spatial analysis and an examination of how these tools can assist in the improved detection of errors (i.e. outliers) and quality control in the ESPON 2013 Database. A second and related aim of the study is to investigate different solutions to the Modifiable Areal Unit Problem (MAUP). It is important to investigate the MAUP, as the correct detection of outliers depends on the level of aggregation used (i.e. a scale issue).

The second step will run from December 2009 to December 2010. Work in this stage will continue to examine how statistical tools can be used, not only for quality control but also for research purposes in the ESPON program. Work will concentrate on: (a) methods for measuring local changes in spatial autocorrelation (via local indicators of spatial association, LISA) and (b) methods for investigating local changes in spatial relationships (via Geographically Weighted Regression, GWR), where both methods are applied in the context of outlier detection. Work will demonstrate the utility of these analytical tools for the database project using a concrete example. Again with respect to the MAUP, appropriate levels of aggregation need to be found for these tools to be applied. The ultimate aim is to improve data quality control in the ESPON 2013 Database by the implementation of procedures for the *automatic* detection of unusual values.

Aspatial outliers: univariate to multivariate data forms

Outliers in statistical distributions may arise for a number of reasons. One reason is that the data may have been incorrectly entered, for example 29 may be inadvertently entered as 92. In these cases, it is hoped that compared with the rest of the distribution, this entered value is unusual and can be identified. Alternatively, the observation itself is unusual, for example its value may be unusually high or unusually low. Determining what values are 'unusual' and how cases with 'unusual' values' should be treated is one of the objectives of this research.

A simple tool for the detection of outliers in univariate data is the boxplot, which can be extended to the bivariate case with the bagplot. Central to the creation of the boxplot is the inter-quartile range (Q3-Q1) around the median value Q2. At the upper end of the distribution, the inner fence is defined as the value given by Q2+1.5(Q3-Q1) and the outer fence as the value given by Q2+3(Q3-Q1); and there are corresponding values for the lower end of the distribution. Observations whose value lies between the inner and outer fences are referred to as 'outside' and those whose value lies beyond the outer fence are referred to as 'far out'. These are usually shown graphically, but the identification of these values is convenient and can be used as a simple, initial data filter. The bivariate bagplot extends the concepts of the boxplot to: (a) a bag which contains 50% of the data points, (b) a fence which defines potential outliers and (c) a loop which lies outside the bag and inside the fence.

Higher dimensional bagplots for multivariate data sets require larger amounts of data for the reliable detection of outliers. However, principal components can be used to reduce the dimensions of the dataset and outliers are often readily observable in this transformed space. Recent work by Filzmoser and colleagues (2005, 2007) has investigated the use of principal components for the detection of outliers in such data cases. It is expected that many inputs to the ESPON 2013 Database will be characterised by such high dimensional data.

Spatial outliers: univariate to multivariate data forms

Commonly methods ignore any spatial element to the data. Data not observed as an outlier when an aspatial method is used, may very well be a spatial outlier. Therefore it is important to consider spatial aspects if false negatives (i.e. outliers undetected by an aspatial method) are to be avoided. Likewise a group of observations identified as outliers may actually be spatially clustered with a substantive reason for their 'unusualness' (i.e. false positives are to be avoided as well).

Spatial methods for univariate data include the use of local versions of Moran's I (spatial autocorrelation) and the Getis G* statistic (spatial clustering). Positive spatial autocorrelation exists when neighbouring spatial units tend to have similar values of a variable; negative spatial autocorrelation exists when they do not. Related methods include the deconstruction of the empirical variogram, where now data pairs (that are separated by some pre-defined distance interval) can be identified as unusual.

Spatial methods for multivariate data include the use of GWR, a technique in which the parameters in a regression model are allowed to vary spatially. Large residuals from a conventional regression tend to be influenced by spatial structure in the data. As GWR takes this spatial structure into account, large residuals from GWR tend to reflect a poor model fit due to a particular observation (i.e. a likely spatial outlier). The down-weighting of such data can lead to an outlier-resistant (or robust) version of the GWR technique. If a data reduction transform is required, then GW principal components can also be found and utilised.

Summary

In summary, this leads us to a typology of methods where variables may be analysed singly or in combination; and aspatially or spatially (where any temporal investigation can be viewed as a simpler sub-class of a spatial investigation). Underlying all of these methods is the spatial structure of the reporting units, where results can be influenced not only by the level of spatial aggregation used but also by the spatial configuration of the reporting units (i.e. MAUP). A previous report for the ESPON 2006 Database from the NCG team examined the influence of the MAUP on spatial models, where its findings will help inform on the research to be undertaken here.

2.11 Challenge 11: Enlargement to neighbourhood

Coordinator: RIATE & NTUA

Objectives: To collect data at regional and local level for neighbouring countries of ESPON territory

Situation:

For the purposes of the Project, we should distinguish between Candidate Countries / CC -the Western Balkans countries and Turkey-, other Eastern Neighbouring countries (ENC) like Russia, Ukraine and the Republic of Moldavia and other Southern Mediterranean Neighbouring countries (MNC).

A first attempt to include data from these countries in the ESPON 2006 Program was made at the end of the Program –see in the Annex of DG Regio paper on territorial cohesion; therefore, the inclusion of the relevant data in the ESPON Database 2013 should be more or less made from scratch.

Strategy and Work plan:

Consequently, it is important to formulate and apply a specific strategy of inclusion of data from the CC, ENC and MNC in the ESPON Database 2013.

This strategy should rely on the networking with other ESPON priorities 1 and 2 projects, as exchange (inputs / outputs) of data between the ESPON Database 2013 and these projects will be made. It should also rely on the cooperation with relevant European (Eurostat, EEA, JRC, Eurogeographics etc) and World Organisations as the former have already created useful datasets concerning the CC, ENC and MNC (some of them will enhance in the near future their CC and ENC database -see for example for the case of CLC 2000).

The strategy should be implemented through several methodological steps / working phases:

(1) The general aim of the **first step** is to provide data collection at regional level -equivalent of NUTS2 and NUTS3- for countries located in the Western Balkans and Turkey (see figure 31).

This working phase contains three tasks: (a) Evaluation of the situation of data available in these countries, following the relevant methodology and preliminary studies elaborated in ESPON 2006; this evaluation has already begun. A corresponding technical report will be submitted. (b) Establishment of contacts with national offices of these countries and assessment how it is possible to establish regular dataflow with ESPON 2013 Database Project. The work on tasks (a) and (b) will last until December

2009. A short paper describing the inclusion of regional data of these neighbouring countries in ESPON 2013 Database will be included in the 2nd Interim Report (February 1010) (c) Insurance of a regular flow of data at regional level for W. Balkans and Turkey and exploration of the possibility to collect new statistics related to different geographical objects. This task will focus in particular on the urban data, making it possible to enlarge the urban database elaborated by ESPON 2006 project and further developed by ESPON 2013. The work on this task will last from December 2009 to December 2010. Its results will be incorporated in the Final Report.

(2) A **second step** aims at enlarging the scope of ESPON Database 2013 to cover regional data (NUTS2, NUTS3): (a) for the Eastern Neighbouring countries (ENC) and (b) for the Southern Mediterranean Neighbouring countries (MNC). This step will begin on December 2009. The results of a preliminary assessment will be included in the 2nd Interim Report (February 2010) while the conclusions will feed the Final Report.

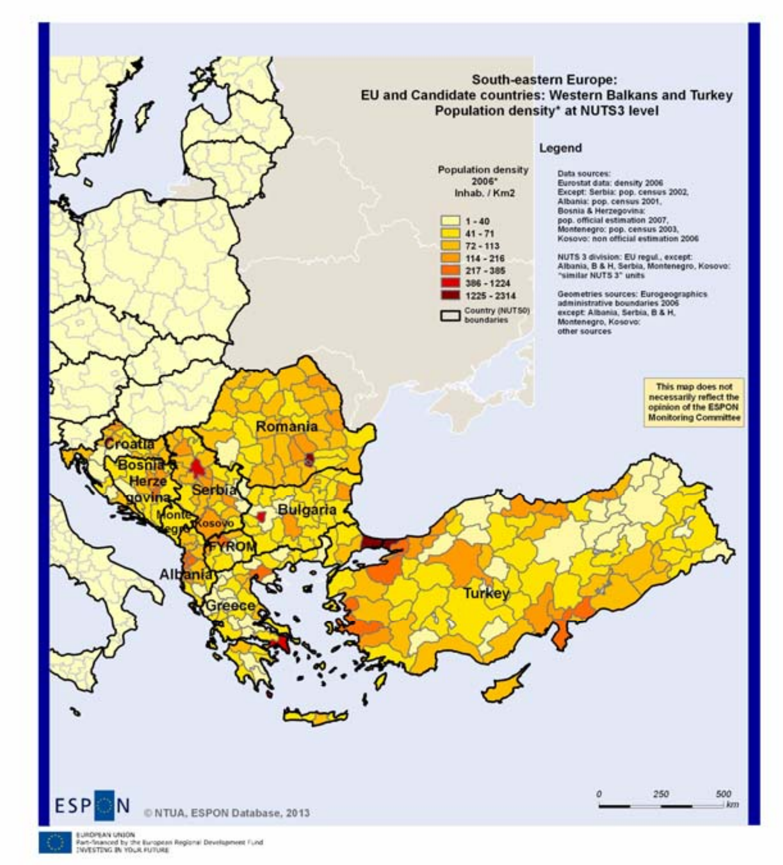

South-eastern Europe:
EU and Candidate countries: Western Balkans and Turkey
Population density* at NUTS3 level
Legend
Population density 2006* Inhab. / Km2
1 - 40
41 - 71
72 - 113
114 - 216
217 - 385
386 - 1224
1225 - 2314
Country (NUTS0) boundaries
Data sources:
Eurostat data: density 2006
Except: Serbia: pop. census 2002,
Albania: pop. census 2001,
Bosnia & Herzegovina:
pop. official estimation 2007,
Montenegro: pop. census 2003,
Kosovo: non official estimation 2006
NUTS 3 division: EU regul., except:
Albania, B & H, Serbia, Montenegro, Kosovo:
"similar NUTS 3" units
Geometries sources: Eurogeographics
administrative boundaries 2006
except: Albania, Serbia, B & H,
Montenegro, Kosovo:
other sources
This map does not necessarily reflect the opinion of the ESPON Monitoring Committee
Croatia
Bosnia Herzegovina
Serbia
Romania
Bulgaria
Monte negro
Kosovo
FYROM
Albania
Greece
Turkey
0
250
500
km
ESPON
© NTUA, ESPON Database, 2013
EUROPEAN UNION
Part-financed by the European Regional Development Fund
INVESTING IN YOUR FUTURE

Figure 31 – Population density in South-easter Europe, 2006

2.12 Challenge 12: individual data and surveys

Coordinator: RIATE & U.UMEA

Objectives: Examine how to integrate individual data based on census or surveys in the ESPON database.

Background

The Department of Social and Economic Geography, Umeå University has access to and thorough experience in using individual, longitudinal population data. The department manages the database ASTRID, which not only covers the entire population of Sweden for a substantial time period, but also has a high degree of geographic resolution. This database has been utilized to carry out novel geographical research concerning a wide range of demographic and socio-economic phenomena. The availability of and access to comprehensive microdata for purposes of research is neither unique for Umeå University, nor Sweden as a whole. Indeed, in other Nordic countries, for example, register data of the population has also been made available for research purposes. However, at the European level, individual data is primarily available from various surveys. In the context of Challenge 12, Swedish data will be used for purposes of exploratory studies, comparisons and methodological development, which otherwise would be difficult to conduct. The main aim is to examine how microdata could improve the ESPON database, as well as research carried out by ESPON projects.

The planned research activity in the project, which will have a focus on individual data and surveys, can be divided into four main themes/activities:

1. Availability and usefulness of survey data.
2. The modifiable areal unit problem (MAUP).
3. Comparisons between Swedish and European-wide data.
4. Integration of survey data in the ESPON database.

1) Availability and Usefulness of Survey Data

Initial work will be devoted to an overview of available surveys that might be of interest for ESPON, in terms of aspects such as focus, content, temporal and geographical scope, current utilization, etc. Surveys that will be looked at include available Eurostat datasets, such as the Labour Force Survey and the European Community Household Panel, but also surveys from other sources, e.g. the European Values Study and surveys originating from organizations such as UN and OECD.

2) The Modifiable Areal Unit Problem (MAUP)

Previously, as part of the Swedish case study in ESPON 3.4.3, the MAUP was examined using Swedish register data. In the study, three variables—population density, population change and disposable income—were aggregated to administrative subdivisions and grids with different resolution/number of units, followed by an analysis of the impact of map type and scale on visual map appearance and statistical measures.

In the context of the present project, further studies of the MAUP will be undertaken, based on 2005 Swedish register data. A selection of individual attributes will be presented in aggregate form in different maps at various levels of resolution. For each level of resolution, the maps will have an equal (or approximately the same) number of units, but the characteristics of the zoning will different. This represents a similar—but somewhat more structured and systematic—approach compared to the ESPON 3.4.3 case study. In this context, the way the MAUP relates to characteristics of space and place will be given attention. Furthermore, the impact of map type and scale on statistical analyses, for instance OLS regression, will be examined.

3) Comparisons between Swedish and European-wide Data

Swedish data will be leveraged for comparisons with existing European-wide datasets. The Joint Research Centre dataset "Population density disaggregated with Corine land cover 2000" is a European 100 meter grid of population density, mainly created by distributing the 2001 local administrative population based on Corine land cover data. In cooperation with Challenge 6 ("Urban data"), the 4.1 version of the grid will be compared with Swedish register data of the population. Differences between the datasets will be calculated, and described and analyzed in various respects. In addition to a summary of deviations for each LAU 2 unit, spatial statistics, such as the Getis-Ord Gi* method for hot spot analysis, will be used in order to identify statistically significant clusters of positive and negative deviations. Furthermore, the comparison will examine local discrepancies within specific urban areas. This analysis, in particular, is closely related to the Challenge 6 comparison between Urban Morphological Zones and the Swedish delimitation of urban localities.

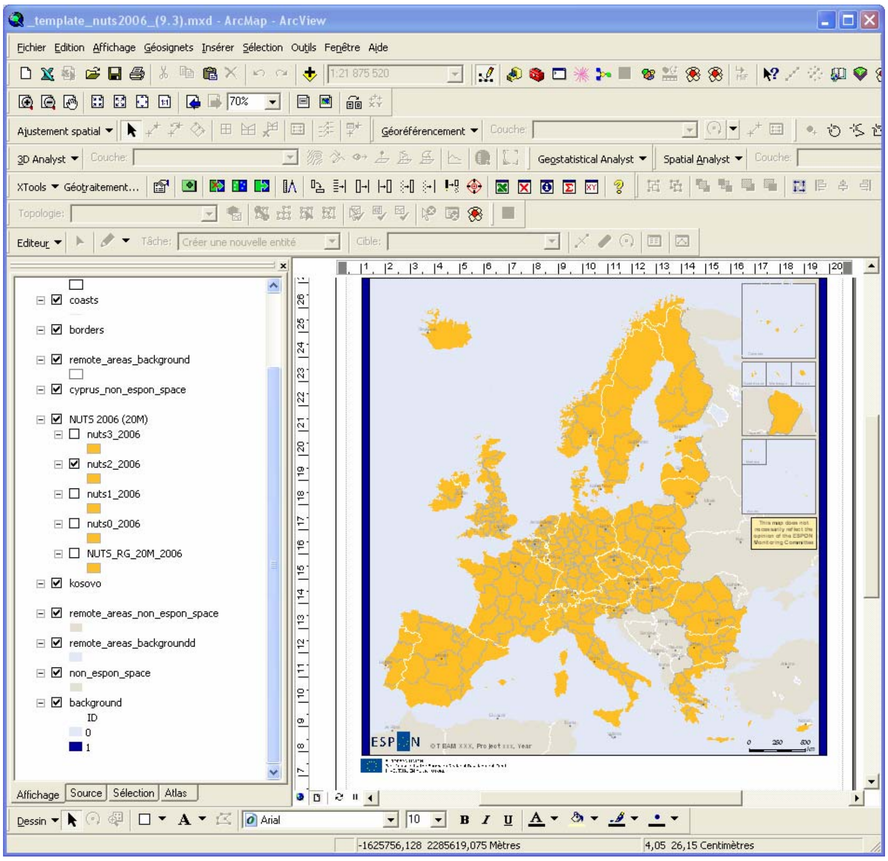
4) Integration of Survey Data in the ESPON Database

This part of the study will examine the possibilities of integrating survey data in the ESPON database. While Eurostat already employs surveys to produce certain statistics, variables tends to be presented with a low degree of geographical resolution, typically in the form of national statistics. Various strategies for presenting survey data with higher geographical resolution will be explored, discussed, and (depending on viability) tested. A crude way to achieve such a disaggregation would be to "rescale" survey data, based on regional variations in population composition (sex, age, etc.). A more interesting approach might be to utilize eventual regional dimensions that are captured in existing surveys. The extent to which information with spatial meaning or connotations is collected is interesting as such, by indicating for which subject areas and data the regional dimension is considered of special interest. Possibly, such information could also be utilized to facilitate or enhance geographical disaggregation of survey data. The production of synthetic samples of individuals is another topic that will be given attention. For certain research purposes, such as microsimulation, the availability of individual datasets constitutes an important methodological precondition. In the absence of comprehensive register data, a synthetic population that is fairly consistent with available and disaggregated regional statistics would be a possible way to enable the use of spatial microsimulation methods.

3 Transversal questions

3.1 New version of the Map Kit Tool

Based on the fact that a lot of data are not yet available for the NUTS2006 version, this consolidated version of the ESPON map-kit tool is divided in two folders. The first folder is a based on the NUTS 2003 version and the second one is based on the NUTS 2006 version. Each folder contains geometries from NUTS0 to NUTS3 of the ESPON territory (EU27 + Switzerland + Norway + Iceland + Liechtenstein). Theses geometries are an extraction of the geometry of EBM of Eurogeographics with a scale of 1:20 Million downloaded on the Eurostat website (GISCO). To finalise the cartographical template, other elements are also available (e.g. Coast lines, North part of Cyprus, The delimitation of the Kosovo, remote territories, capital cities). Compatible with the EPSON 2013 database, all theses elements are included in an ARCGIS mxd document (see figure 32), that is an easy way to make harmonized maps. For the TPGs that don't use the ARCGIS Software, a template folder is also available to show directly how to build the layout of the map.



_template_nuts2006_(9.3).mxd - ArcMap - ArcView

- ☑ coasts
- ☑ borders
- ☑ remote_areas_background
- ☑ cyprus_non_espon_space
- ☑ NUTS 2006 (20M)
  - ☐ nuts3_2006
  - ☑ nuts2_2006
  - ☐ nuts1_2006
  - ☐ nuts0_2006
  - ☐ NUTS_RG_20M_2006
- ☑ kosovo
- ☑ remote_areas_non_espon_space
- ☑ remote_areas_backgroundd
- ☑ non_espon_space
- ☑ background
  - ID
  - 0
  - 1

1:21 875 520

70%

Arial 10

-1625756,128 2285619,075 Mètres

4,05 26,15 Centimètres

Figure 32 – Map Kit tool

3.1.1 Technical elements

Data type: geometries

Format: shp + mxd

Spheroid: ETRS89

Projection: Lambert azimutal equal area. 15°E50°N

Resolution: 1/20 000 000

Coverage: 31 Countries (EU27 + Norway + Switzerland + Island + Liechtenstein)

Copyright: EuroGeographics for the administrative boundaries

3.1.2 Elements for ESPON Cartography

Capital cities (fig 33)

45 capital cities are represented in the Map Kit.

Vilnius, Minsk, Dublin, Berlin, Amsterdam, Warszawa, London, Bruxelles/Brussel, Kyiv, Praha, Paris, Wien, Budapest, Bern, Beograd, Bucuresti, Sofiya, Tirana, Madrid, Ankara, Helsinki, Zagreb, Nicosia, Luxembourg, Bratislava, Tallinn, Sarajevo, Skopje, Athinai, Kishinev, Kobenhavn, Lisboa, Oslo, Reykjavik, Riga, Roma, Stockholm, Valletta, Ljubljana, El-Jazair, Tounis, Ar Ribat, Podgorica, Vaduz, Pristina

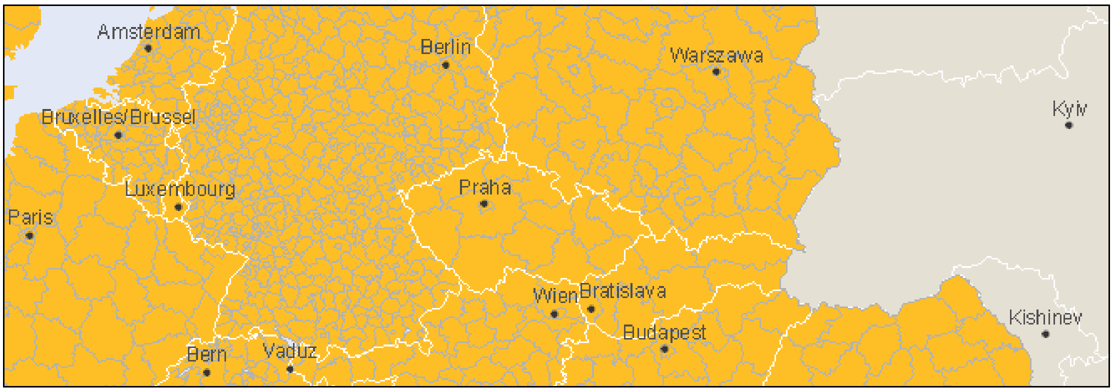


Figure 33 – Capital cities

Western Balkans / Kosovo (fig 34)

Design of borders and some countries must follow precise rules for political reason. In general, ESPON follows the rules established by European Commission. When these rules do not exist at EU level (for example because of lack of consensus) the rules of

UN are used as reference. According to these considerations, the borders of Kosovo are thinner (0.15 pt) than the other boundaries (0.30 pt).

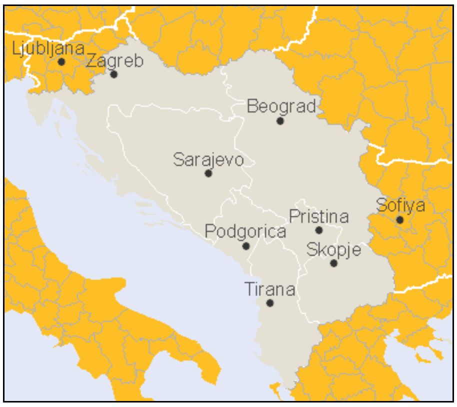


Figure 34 – Western balkans

Remote territories (fig 35)

Remote territories of France (Martinique, Guadeloupe, Guyane française, Réunion), Spain (Canarias) and Portugal (Acores, Madeira) are territories members of the European Union that should be represented on maps even when data are not available.

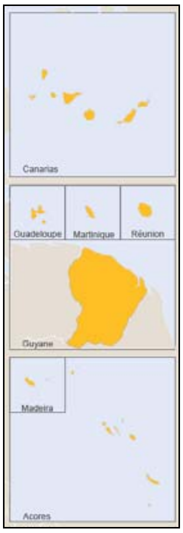


Figure 35 – Remote territories

Cyprus (fig 36)

Cyprus is represented in two different colours. The two parts are separated with a NUTS line (without border line). The area not controlled by the government appears in white as "No data available".

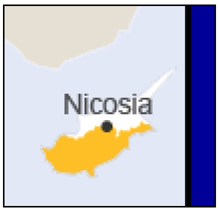


Figure 36 – Cyprus

Malta (fig 37)

To be sure that Malta is always visible on the maps, we do not use light blue coast-line that could be reduce on the map the size of this country.

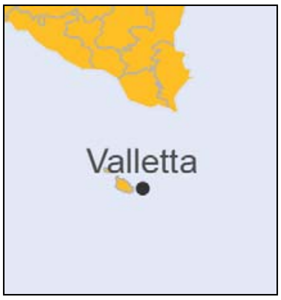


Figure 37 – Malta

Logos, disclaimer, layout, … (fig 38)

The map template contains also different elements that the partners have always to put on the map

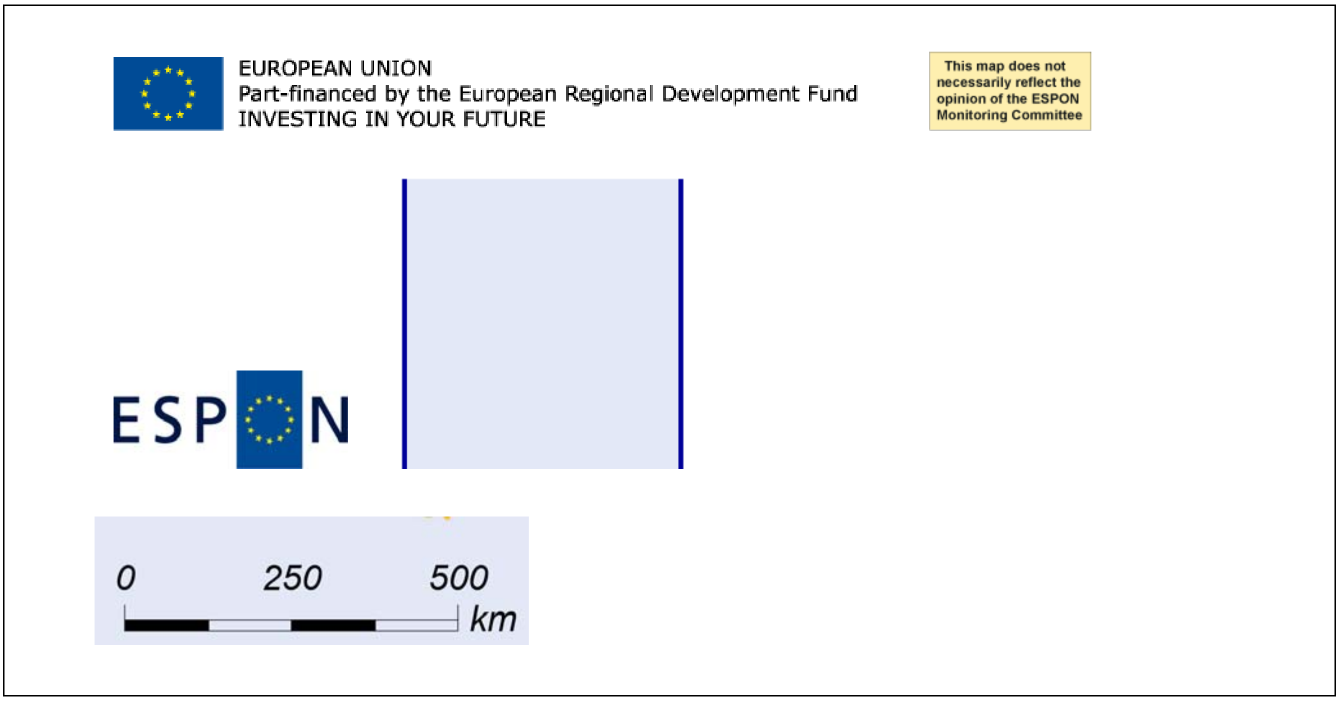

EUROPEAN UNION
Part-financed by the European Regional Development Fund
INVESTING IN YOUR FUTURE
This map does not necessarily reflect the opinion of the ESPON Monitoring Committee
ESPON
0
250
500
km

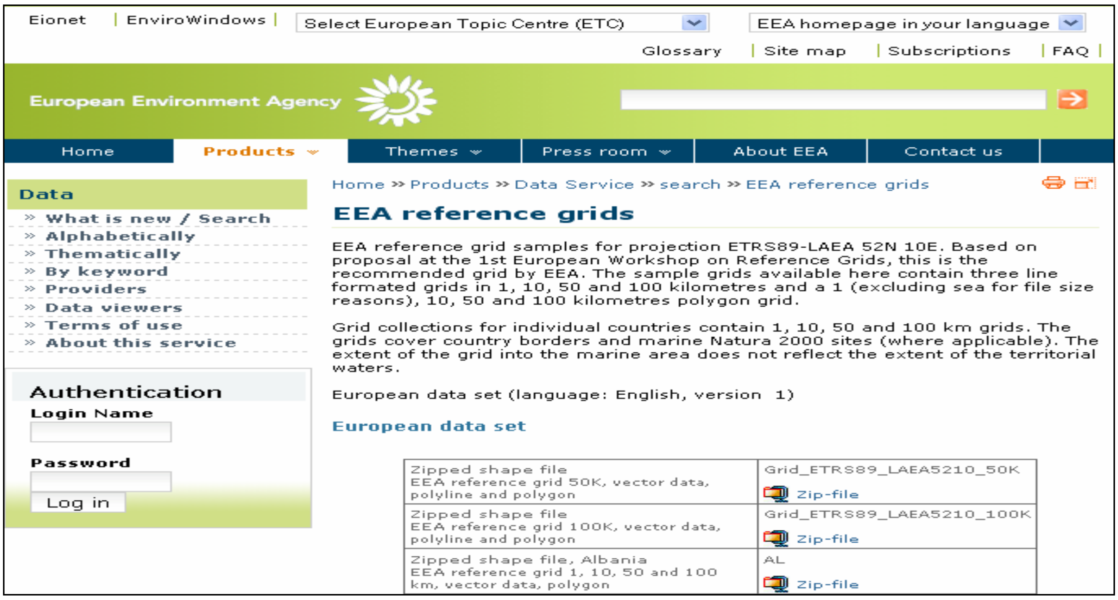
Figure 38 – Logos, disclamer, layout, …

3.1.3 Other elements

Gradually, the map kit will contain also vector data for the municipalities of the ESPON countries, a 1km European reference grid and other elements that can be useful for TPGs.

- **EEA reference GRID**

See http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=760



Home » Products » Data Service » search » EEA reference grids

EEA reference grids

EEA reference grid samples for projection ETRS89-LAEA 52N 10E. Based on proposal at the 1st European Workshop on Reference Grids, this is the recommended grid by EEA. The sample grids available here contain three line formated grids in 1, 10, 50 and 100 kilometres and a 1 (excluding sea for file size reasons), 10, 50 and 100 kilometres polygon grid.

Grid collections for individual countries contain 1, 10, 50 and 100 km grids. The grids cover country borders and marine Natura 2000 sites (where applicable). The extent of the grid into the marine area does not reflect the extent of the territorial waters.

European data set (language: English, version 1)

European data set

Zipped shape file EEA reference grid 50K, vector data, polyline and polygon	Grid_ETRS89_LAEA5210_50K Zip-file
Zipped shape file EEA reference grid 100K, vector data, polyline and polygon	Grid_ETRS89_LAEA5210_100K Zip-file
Zipped shape file, Albania EEA reference grid 1, 10, 50 and 100 km, vector data, polygon	AL Zip-file

Figure 40 – EEA dataservice website

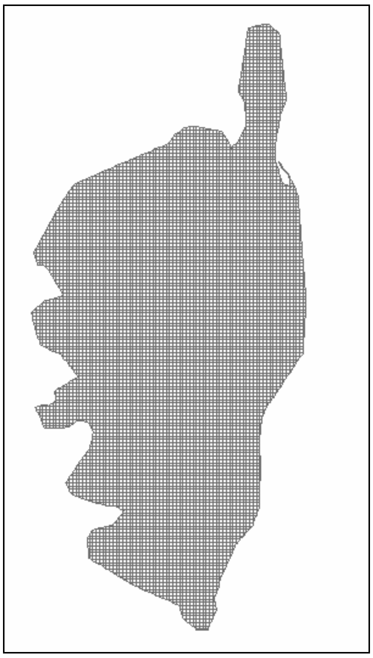


Figure 41 – EEA reference grid: example of Corse

- Local administrative boundaries

Extracted from EuroBoundaryMap V3, these boundaries are defined according to the administrative situation as it was on 1st January 2008 for an application scale of 1:100 000.

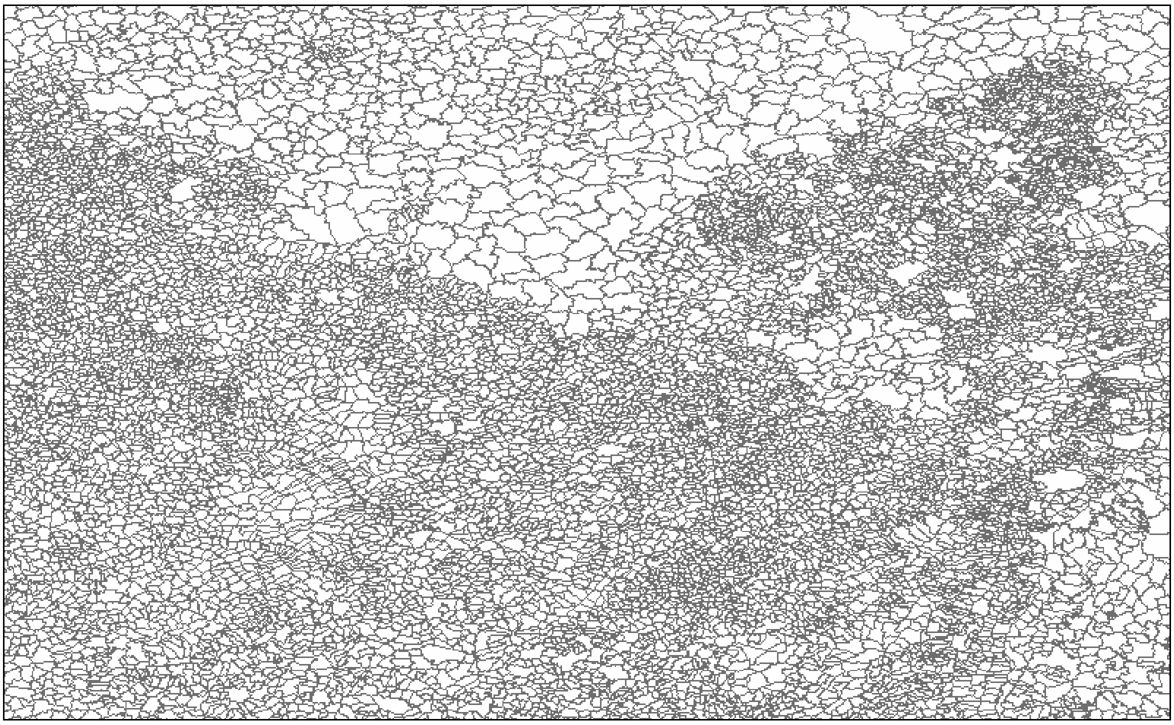
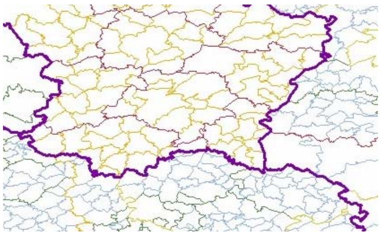


Figure 42 – LAU-2

- Others

More generally, the ESPON DB project can give access on request to the TPGs, to different products from EurogeoGraphics[16]:

EuroBoundarymap



EuroGlobaMap

EuroRegionalMap

[16] Due to the size and complexity of these products, they have not been involved in the map kit tool.

Figure 43 –Eurogeographics products

3.2 Data and metadata

…about things given !

The word data is originally Latin for "<u>things given or granted</u>".
Data is Latin translation of a work by Euclid entitled Dedomena (Greek: Δεδομένα).
Dedomena […] are pure data […], i.e. <u>data before they are interpreted</u>.
(source: article "Data" International Encyclopedia of the Social Sciences, 2nd ed. (Darity ed., 2008)

In epistemology, the prefix meta (from Greek: μετά) is used to mean "<u>about</u>".
(source: Wikipedia: Data (Euclid))

Euclid, detail from "The School of Athens" painting by Raphael.

3.2.1 The metadata challenge

Scarcity of data documentation within the previous ESPON program has been seen as an important impediment to the building and use of the ESPON database. Difficulties arised from uncertainties about legal constraints, sources, units, etc…

In the ESPON Database 2013 project, the database will be enriched and expanded in the time, spatial and thematic dimensions (see our respective challenges). Information about the data made available is thus even more crucial. Building a rich database would be useless without a strong effort to inform about the things that have been gathered and integrated within the database. This information about data is known as **metadata**. **Creating and organising metadata is therefore an additional, important, and transversal challenge** for our project.

To be useful for ESPON projects and other end-users, data should always be accompanied with metadata, including information about quality, sources, and lineage. It is also particularly important that metadata are created in a manner that is consistent with international standards so as to ensure the use of the database in the longer-run and compatibility with other national and international database initiatives.

In the next section (3.2.2), we present the main **components of metadata** and how they are implemented within international standards, particularly based on the **INSPIRE** initiative and following the ISO **standards** for geographical data, and the SDMX (*Statistical Data and Metadata Exchange*) standard proposed for statistical data. Beyond the existing standards, it is worth mentioning however that we found from our various external contacts that "being INSPIRE compliant" today is still currently a process in most European institutions and services. We are not yet in a situation where all external data providers fulfil a clear set of binding regulations. There are some degrees of freedom therefore which the ESPON Database can use to develop metadata that are well tuned to its users.

In the third part of this chapter (3.2.3), we remind the structure of **data and metadata flows** within the project and propose a short-term strategy for creating and maintaining metadata all along these flows.

The **short-term strategy** is aimed at proposing a quick and easy metadata solution for ongoing ESPON projects (Priority 1 and 2) when exchanging data with the ESPON Database. This solution is needed as long as our project is furthering the structures and tools that are necessary for good metadata. Creating well-thought metadata structures is a research in itself and should be made carefully because it conditions future data quality and access. The short-term metadata solution is based on a simple **spreadsheet** format to accompany data provided to the database by the different projects. The spreadsheet is structured in a way that metadata will then be easily integrated in the database when the mid-term solution will be in place.

The **mid-term solution** will be based on a **web-editor** and linked to the Database import pool. We have explored a few metadata editors in order to find out the most suitable one for the ESPON community (Inspire Metadata Editor, CatMEdit, Geonetwork,…). We found that a web metadata editor (rather than a desktop one) should be preferred in the longer-run because it requires no install (and updates) from the users but also because it can more easily provide help to the users when filling in the different necessary fields.

We have found no existing web-based metadata editor that is well tuned to ESPON specificities, particularly because most of the data to be exchanged are likely to be statistical tables rather than geographical features. We therefore aim at developing a new metadata editor. Of course we will not start from scratch, but capitalise on the **Geonetwork** tools as used at the University of Barcelona for the European Environment Agency. We present in section (3.2.4.1) the features that are to be included in the mid-term solution as well as current Geonetwork concepts.

The steps to be undertaken in the next months to achieve this mid-term solution are presented in the last section of the present chapter (3.2.5).

3.2.2 About metadata and standards

Quality information and lineage

In [Servigne et al., 2006], a comprehensive list of items that compose **quality information** (see table 2) is established. Accordingly, the ESPON DB 2013 Project needs to establish a way to collect, compute, and then deliver all 7 items that make this quality information.

Item	Designation	description	example
1	geometric/positional accuracy	Defines the deviation in the values of the respective positions between data and the nominal ground.	
2	thematic/attributes accuracy	Defines the deviation of measurements of qualitative attributes or quantitative attributes (classification) to the real values.	Percentage of "true" outliers. Level of confidence in a value. Sampling errors (mathematical expression)
3	Completeness	Evaluates the ratio of omission (abnormal absence) or commission (abnormal presence) of certain information. It concerns also the data model (whether it fulfills the application requirements or not) : "fitness for use" concept	Percentage of missing values
4	logical consistency	Describes the matching between the dataset and the structure of the model used (respecting	Deviation between the sum of indicator

		specified integrity constraints)	values on sub-units and their super-unit's indicator value.
5	temporal accuracy	Provides information about the dates of data observation (origin), types and frequency of updates, the period of validity of data, and timeliness (which is length of time between data availability and the event or phenomenon they describe	Frequency : 5 years Validity period : [2005-2010[[17] Observation date : 2005
6	Semantic consistency	Describes the number of objects, of relationships, and attributes correctly encoded with the set of rules and specifications.	Number of values correctly encoded with the required precision (speaking about currency rate for example).
7	specific quality	What can not be classified in other criteria.	comparability (over time or space): level of conformance to international standards concerning the methodology, definition, etc.

Table 2. List of items composing quality information.

In addition to those items, the **lineage** contributes also to give some information about the quality of the dataset: it describes the procedures of acquisition, the sources, and the methods used for deriving and transforming data and is used to rebuild the history of a dataset by indicating for instance [Servigne, 2006]:

- the source of data (clearly, the organization's reputation should be taken into account), the origin, the reference domain, characteristics of spatial data, coordinate and projection systems
- the acquisition, compilation and derivation processes used: this consists of the fundamental hypotheses of observation, calibration and correction, as well as the methods used for interpreting, interpolating or aggregating data. This is the kind of

[17] The notation [2005-2010[means that the validy period ends the 31 december 2009, at midnight. This unusual notation allows for some algebraic operations on time intervals (union, intersection, etc.) in the sense of Allen's grammar (1983).

information that can be found inside the methodological guides that are delivered on providers (INSEE, EUROSTAT, OECD, etc.) web sites.

- the data conversion process that leads to a transformation of data. For example, define how the GDP values have been converted from national currency (pounds sterling for example) to Euros.

- the dates of the different stages of processing

- transformation or analyses of data: transformation of coordinates, generalization, translation, reclassification, all defined, as far as possible in precise mathematical terms. A simple example is: source = Eurostat, transformation: aggregation from NUTS3 level, formula: density = population(inh.)/area(km2).

One important thing about the metadata process is that each new provider enriches metadata's lineage information with its own transformations process descriptions, by mentioning only the source and a reference URL. The user can then reconstitute the whole lineage chain backwards from a metadata to the previous one.

Metadata standards

A metadata standard is a schema to describe a dataset. It does not contain data values themselves. A standard describes structures and attribute names ("rubriques", fields) in order to ensure **"syntaxic interoperability"**. A standard may also implement a set of controls on values so as to ensure **"semantic interoperability"**, e.g. through thesaurus, glossary

Beyond interoperability, it is also important to have some flexibility in order to fit the requirements of particular users. Actually, within a standard, every field is not constrained or mandatory, and fields can also be added. By varying those flexible elements, a metada profile is created and allows institutions to fit the needs of their users while ensuring interoperability with others. This way, we aim at creating an **ESPON metadata profile**.

The **INSPIRE directive** about data dissemination clearly recommends the use of the standard ISO 19115 or Dublin Core for data with geographic references, which is the case for all datasets (environmental or socio-economic) within ESPON DB 2013. Knowing that the 15 mandatory fields of the Dublin Core standard do have equivalent fields in the ISO 19115 standard [Barde, 2005], the ESPON DB 2013 Project can, without loss of metainformation, make the choice of the ISO 19115 standard. This standard (together with ISO 19139 for implementation) is more and more extensively used within different institutions such as the EEA and JRC among others. We take those experiences into account when defining the ESPON Metadata Profile.

For statistical data, the main standard is named **SDMX** (Statistical Data and Metadata Exchange).[18] We aim at using this standard as well in this project since it is supported by Eurostat, which is the main statistical data provider for the ESPON DB.

The ISO 19115 and SDMX standards are further detailed below.

[18] *Note: The Statistical Data and Metadata Exchange web: http://sdmx.org, contains oriented guidelines and recommended practices for creating data and metadata for statistical domains.*

Geography - ISO19115

The **standard ISO 19115** offers a way to structure metadata, through various topics ("rubriques"), each of them corresponding to specific information (see Figure 44).

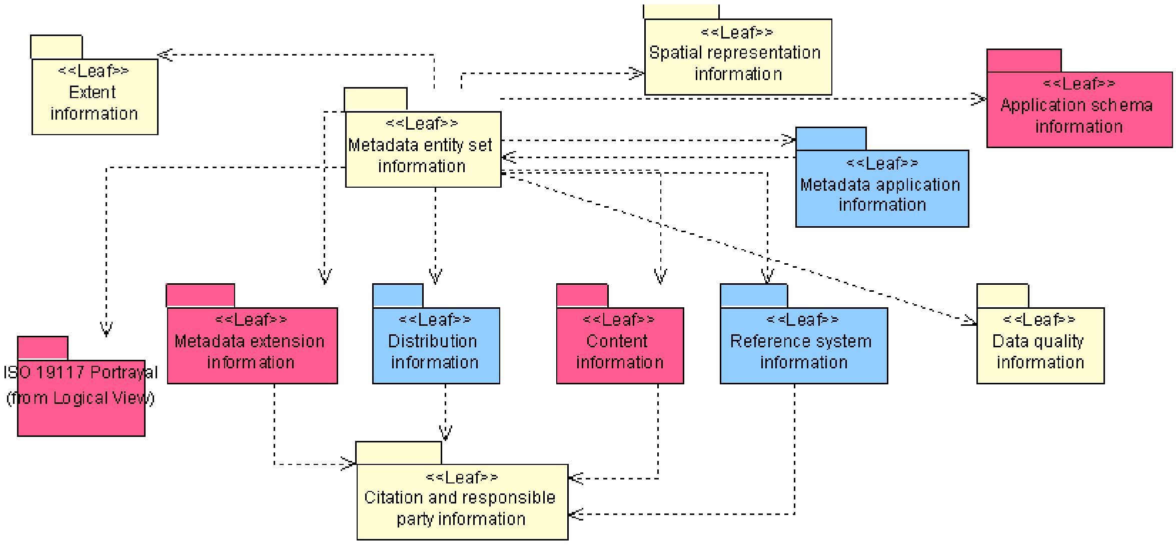

<<Leaf>>
Extent information
<<Leaf>>
Spatial representation information
<<Leaf>>
Application schema information
<<Leaf>>
Metadata entity set information
<<Leaf>>
Metadata application information
ISO 19117 Portrayal (from Logical View)
<<Leaf>>
Metadata extension information
<<Leaf>>
Distribution information
<<Leaf>>
Content information
<<Leaf>>
Reference system information
<<Leaf>>
Data quality information
<<Leaf>>
Citation and responsible party information

Figure 44 - UML representation of ISO 19115 components. Source :

http://www.isotc211.org/hmmg/HTML/root.html

In detail, theses topics are:

Identification Information: identifies the main characteristics of the datum: a title, an abstract in free text, the language, the set of characters (UTF-8, ISO 8859-1, etc.), themes which should help for further classification, and some keywords, both extracted from thesaurus, and the name of the data file.

Metadata entity Set Information: information about metadata themselves: the language in use (English for ESPON projects), the set of characters (UTF-8, ISO 8859-1, etc.), contact (organism or person responsible for this metadata: ESPON 2013 project), the last update date, name of the metadata standard used and its version, identification of the data file, and identification of the associated metadata file.

Data Quality Information: information about the quality of data, and explains how data have been produced: their lineage (the methodology and transformation process) as well as the quality level of the result expressed through spatial and temporal accuracy for example, or one of the items listed in the table 1. Usually, providers establish various reports relating the quality level, in separated documents having heterogeneous format. For the end-user, we could store those files or give a URL to access to these documents, yet, we need to extract and format in standard attributes as much information as possible. Furthermore, we may attach as many quality topics as required for each data sub-set, since a mandatory field (scope) should indicate the set (or sub-set) of data that are concerned by this information.

Constraints Information: is used to indicate if data can be downloaded for free, or if there are some restrictions (limited to a certain group of users, or organisations). It gives any legal information concerning the copyright, use and access rights on data. It gives also indication about the access constraints for the metadata file itself.

Maintenance Information: information about the frequency of data updates (regular or not) and the periodicity of these updates. The scope of updates is also given, indicating which parts of the dataset may be updated.

Spatial Representation Information: following the kind of resource described, this topic gives detail about the representation mode of spatial attributes: if it is a grid support, attributes such as resolution, number of dimensions, number of pixels by dimension, and pixel resolution cell geometry, geo-referenced parameters, etc., should be provided. For a vector support, the topology level, the geometric object type should be provided.

Reference System Information: information about spatial reference system (the Reference System Identifier, its name and code, the coordinate system with ellipsoid parameters and projection parameters) and temporal reference system.

Extent Information: information about the spatial and temporal coverage of the dataset. The geographic extent is given by a geographic bounding box (longitude and latitude, minimal and maximal) and a geographic description with an identifier of the studied territory. The temporal extent is bounded by the dates of oldest data and more recent data.

Content Information: more technical specifications about data. It has first been thought to describe images coming from remote sensing: sensor characteristics, minimal and maximal band wave length, units, tone gradation, etc. As a matter of fact, this topic is well suited for environnemental data, but as far as socio-economic statistical indicators are concerned, a new nomenclature needs to be established, giving for example some information about the survey, the sampling mode, or the list of category in use (age stratification or socio-professional categories).

Distribution Information: this topic describes the distribution process of data: who has to be contacted (the distributor) to get the data: name, address, phone… and the digital transfer options: on line, or off line, the medium description (a CD, its name and size for example), and the standard order process that is to say the fees, the ordering instructions, etc.

Citation and Responsible party Information: describes how to cite the resource as a bibliographic reference (title, reference date, linkage (URL)) and people to be contacted for further information about the dataset: name, organization, position, address, email, phone. This part is included within the identification topic.

Only the first two topics, **Identification (including citation) and Metadata, are mandatory**. All other topics can be extended to build a specific profile. It is under the ESPON DB 2013 Project responsibility to decide which topic should be made mandatory, by producing its own profile. For example, we think the quality topic should be mandatory in the ESPON DB 2013 profile. However, in most cases, the quality information topic is described using free text, which makes the processing of this metadata value a difficult task. So, we might decide to further specify this topic by adding specific attributes and valuing them using a controlled vocabulary, authorising later automatic checking, and global quality evaluation.

Statistics - SDMX

SDMX is another standard for structuring information that mixes metadata within data, and is well suited for statistical data. It is a standard *"de jure"*, since it is developed and supported by the main providers of statistical datasets: OECD, UN, Eurostat. This model will be soon used by Eurostat to disseminate its data.

SDMX eases the transfer of data, but, as successor of GESMES, it is mainly a logistic metadata [Kent et Schuerhoff, 1997]. That is to say that SDMX is based on the use of schemas (i.e. grammars) that structure information, information being encoded inside those schemas using some keywords and controlled vocabulary. A schema uses metadata tags whose semantic is shared by users of statistical data. SDMX is a logistic metadata standard allowing the exchange of data between computers, and the automatic checking of the quality of information.

For these reasons, the ESPON **DB 2013 Project should certainly provide ways to export data in SDMX format** (but not only). The tags used for structuring the information are called the "metadata tags", and have been well documented in the "Metadata Common Vocabulary" glossary [SDMX MCV]. Moreover, they have been thought to cover the various items listed in the table 1 above concerning the quality information. Furthermore, they also allow for the description of maintenance information, and various other topics that have been listed in the ISO 19115 standard (though the terms used are not always the same, e.g. "Release policy" is used for "Maintenance" by example).

The user could give a list of codes to valuate the information designed by theses tags, that have been internationally standardized, and documented in the SDMX "Cross Domain Code Lists" [SDMX CL]. For example, the frequency can be valuated by following codes defined in CL_FREQ: A (Annual), S (Semester), Q (Quarterly), M (Monthly), W (Weekly), D (Daily), B (business week), N (Minutes).

Unfortunately, this format being very verbose, it is not well-suited for grid information (as shown in figures 9a and b, end of this chapter). In addition its structure is rather complicated for users who might be more used to collect statistical data using spreadsheets (Excel or other). Using this standard within ESPON will require the development a set of tools, templates and guidelines for filling in quality information. Whenever data are provided in SDMX format, we also have to be able to parse them for acquisition process and storage in the database. Moreover, we should think of a good integration with ISO 19115.

3.2.3 ESPON DB metadata flows

Data and Metadata flows

The **data and metadata flow** to and from the ESPON Database can be defined as follows: first, data are imported from ESPON internal projects or external providers, through the import pool. Second, the ESPON DB 2013 Project is in charge of enriching metadata description, completing datasets, and compiling new datasets gathering many sources, through the integration pool. Third, those data are then redistributed together with their associated metadata, through the export pool. Figure 45 gives an overview of the whole data and metadata flow related to the ESPON 2013 DB project.

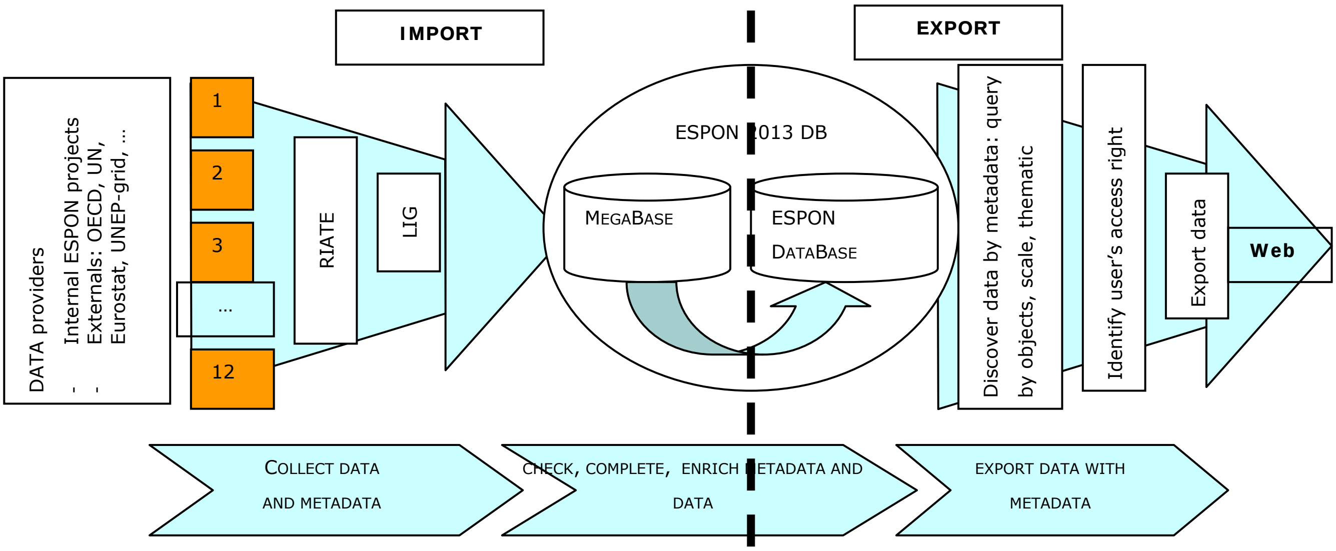

IMPORT
EXPORT
DATA providers
- Internal ESPON projects
- Externals: OECD, UN, Eurostat, UNEP-grid, …
1
2
3
…
12
RIATE
LIG
ESPON 2013 DB
MEGABASE
ESPON DATABASE
Discover data by metadata: query by objects, scale, thematic
Identify user's access right
Export data
Web
COLLECT DATA AND METADATA
CHECK, COMPLETE, ENRICH METADATA AND DATA
EXPORT DATA WITH METADATA

Figure 45. Data and metadata flow in ESPON 2013 DB project.

Following the lineage concept (i.e. backward metadata chain), the ESPON DB 2013 Project does not need to re-import metadata that are provided in the lineage part of its **external data providers** like EUROSTAT, UN, or EEA for example. The ESPON DB 2013 Project just needs to add a description of its own data transformation process (for instance, when computing missing values or re-adjusting datasets), and to provide a precise reference towards the data provider.

However, for other **internal ESPON Projects**, whenever they produce a new variable (indicator), they should provide the ESPON DB 2013 Project with a metadata document comprising all the items listed above. If their own sources also provided a consistent metadata, then the projects should deliver a description of transformation plus reference to the source metadata. The metadata provided by the projects should then be structured and stored within the ESPON DB 2013, since ESPON is the publisher of those data.

A short-run solution for internal ESPON flows

Further research should be accomplished to achieve an ESPON metadata profile encompassing geographical and statistical data and relying on existing standards. Efforts will also be put on the developement of a dedicated metadata editor solution. Meanwhile, we propose to use a spreadsheet-based metadata to exchange statistical data between the ESPON Database and the other ESPON projects.

Note that for the (presumably rare) cases of exchange of geographical objects from ESPON projects to the Database, the former should provide ISO19115 metadata files from their own source or resources (e.g. using a GIS editor, INSPIRE web-based facilities, etc.) or contact the Database project to find out a metadata solution on a case by case basis.

We should raise awareness to the fact that there is no legal obligation to respect any standard template for the import process of data. This means that the structuring of metadata fields itself has no importance when collecting metadata, from the legal point of view. The emphasis should rather be put on the content of collected metadata information. The ESPON DB 2013 Project should collect as much as possible information, whatever the format. But for practical reasons, we need a format, i.e. templates and guide lines to ease the metadata acquisition process and its integration into the database. Indeed we would like to begin with an automate acquisition of the metadata as soon as possible, and if every body provide us metadata in free text, on various supports (Excel files, PDF, and so on) we won't be able to automate the acquisition. This is why we try to propose a structure (an Excel file with specific fields) to ease the collect of metadata information, and that is a short-term solution.

In order to comply with INSPIRE directive, the proposed spreadsheet metadata for statistical information contains all ISO 19115 mandatory information but also the other fields that make good quality information. Every exchange of a dataset should be accompanied by this spreadsheet (see challenge 9 for internal ESPON exchanges and data sheet structures).

An example of metadata spreadsheet is provided below. Though simple, this solution allows for example for a three tiers characterisation: the whole dataset level, the variable level, and the record level.

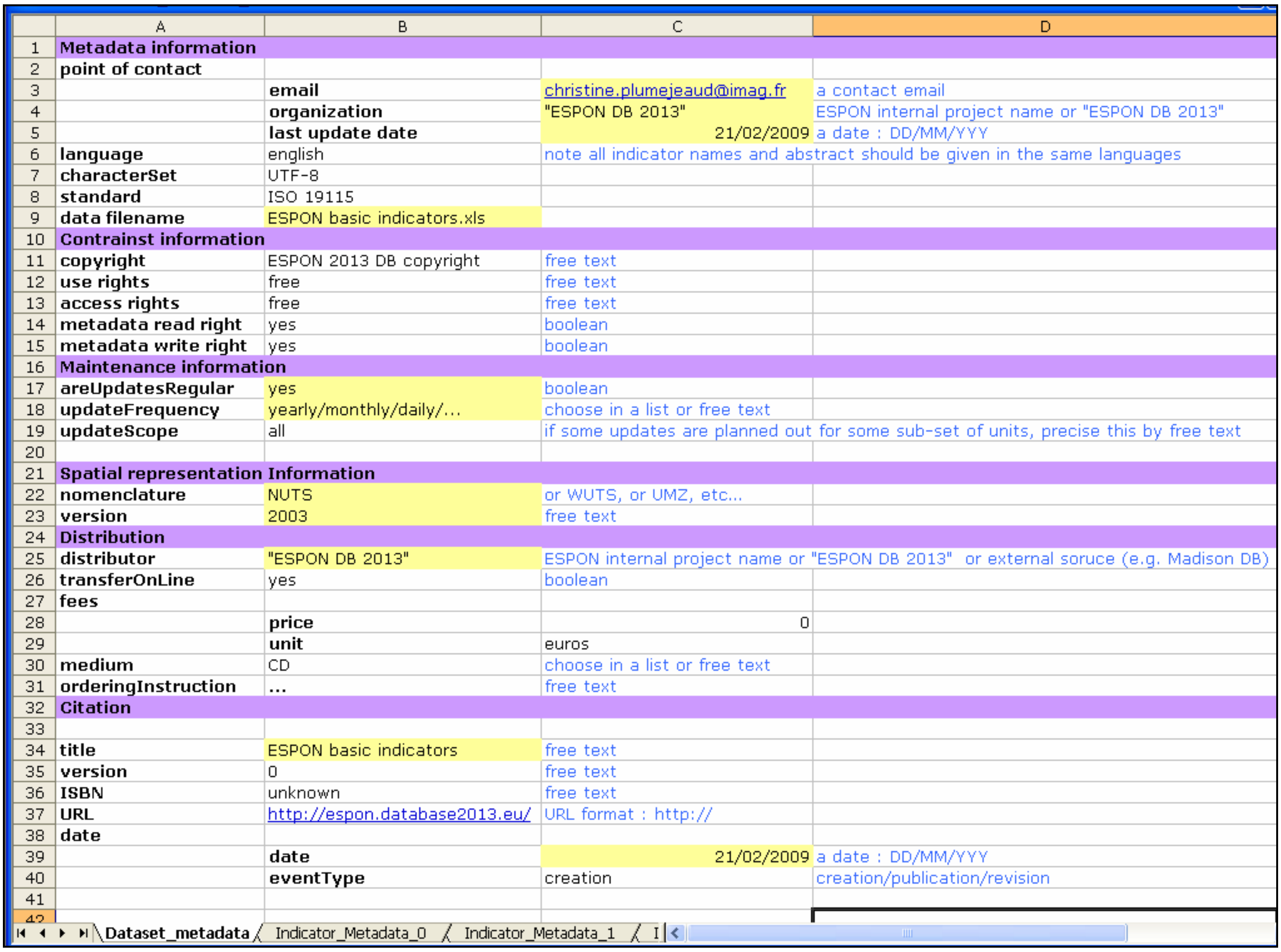
In fact, the metadata spreadsheet is composed of a first sheet: the **"dataset_metadata"** (Figure 3) which refers to information for the whole dataset, and then additional sheets, **"indicator_metadata_X"**, for each different variable in the dataset (Figure 4).

In the sheet "dataset_metadata", the distribution topic is not mandatory. It should be filled only if data are not available for free use and access. It is dedicated to data under the protection of a copyright, like the Madison database for example.

The description conventions are the following:

In blue are given some comments and precisions about the data format.
In **bold**, the designation of the fields is given
Cells filled in yellow indicate the mandatory fields.
In black, appear the default values of the fields.
* means all
The fields that are marked with an * can be multi-valued, and the separator is the common ','.

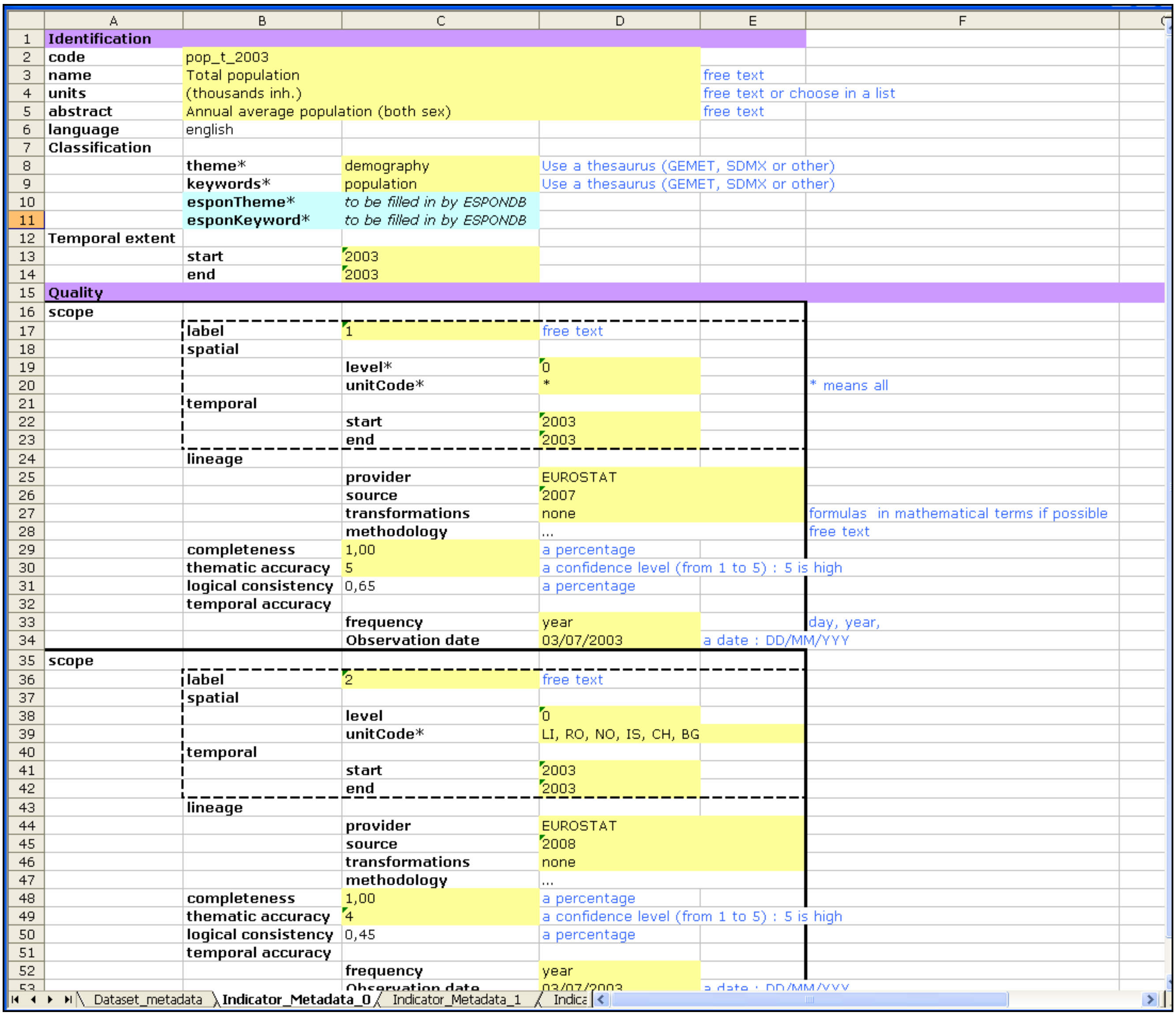
For the sheets "indicator_metadata_X**"**, the user has to fill Identification, Temporal extent, Quality and Lineage information, and information about the collected indicators. It has been agreed with the RIATE partner that this information should be provided at indicator level (that is to say in a column in an Excel file) for the minimum. And at the maximum, we can go down to the unit level to provide quality and lineage information (that is to say on a line in an Excel file). There is as many **indicator_metadata_X** as indicators to be described: one can copy paste the first sheet (X==0) and modify the *identification* part (see Figure 46). Thus, for *n* indicators, *n* indicator_metadata_X sheets should be filled.



	A	B	C	D
1	**Metadata information**			
2	**point of contact**			
3		**email**	christine.plumejeaud@imag.fr	a contact email
4		**organization**	"ESPON DB 2013"	ESPON internal project name or "ESPON DB 2013"
5		**last update date**	21/02/2009	a date : DD/MM/YYY
6	**language**	english	note all indicator names and abstract should be given in the same languages	
7	**characterSet**	UTF-8		
8	**standard**	ISO 19115		
9	**data filename**	ESPON basic indicators.xls		
10	**Contrainst information**			
11	**copyright**	ESPON 2013 DB copyright	free text	
12	**use rights**	free	free text	
13	**access rights**	free	free text	
14	**metadata read right**	yes	boolean	
15	**metadata write right**	yes	boolean	
16	**Maintenance information**			
17	**areUpdatesRegular**	yes	boolean	
18	**updateFrequency**	yearly/monthly/daily/...	choose in a list or free text	
19	**updateScope**	all	if some updates are planned out for some sub-set of units, precise this by free text	
20				
21	**Spatial representation Information**			
22	**nomenclature**	NUTS	or WUTS, or UMZ, etc...	
23	**version**	2003	free text	
24	**Distribution**			
25	**distributor**	"ESPON DB 2013"	ESPON internal project name or "ESPON DB 2013" or external soruce (e.g. Madison DB)	
26	**transferOnLine**	yes	boolean	
27	**fees**			
28		**price**	0	
29		**unit**	euros	
30	**medium**	CD	choose in a list or free text	
31	**orderingInstruction**	...	free text	
32	**Citation**			
33				
34	**title**	ESPON basic indicators	free text	
35	**version**	0	free text	
36	**ISBN**	unknown	free text	
37	**URL**	http://espon.database2013.eu/	URL format : http://	
38	**date**			
39		**date**	21/02/2009	a date : DD/MM/YYY
40		**eventType**	creation	creation/publication/revision
41				

Figure 46. Spreadsheet solution: Dataset metadata sheet

The quality and lineage information should be as detailed as possible, and can have various scopes: a whole temporal series, a specific date, a full geographic area, a sub-area, or even a specific unit (like for example the Lichtenstein). A mechanism of specialization of metadata information will be implemented for the metadata import: information concerning a specific unit prevails over information concerning its enclosing unit, or even the whole geographic space, or even the whole covering time. Concretely, data providers can customize, as needed, each *quality scope*, and add as many scopes as needed following the exceptions appearing in the dataset (Figure 47). The parts that are surrounded with a dash line define a scope by its temporal and spatial validity. The scope having the finest validity extent (concerning the units, or the period) prevails above all for the concerned units. Each scope of an indicator is uniquely identified by a label, and this label will be re-used to identify the *category* of each value in the dataset.



	A	B	C	D	E	F
1	Identification					
2	code	pop_t_2003				
3	name	Total population			free text	
4	units	(thousands inh.)			free text or choose in a list	
5	abstract	Annual average population (both sex)			free text	
6	language	english				
7	Classification					
8		theme*	demography	Use a thesaurus (GEMET, SDMX or other)		
9		keywords*	population	Use a thesaurus (GEMET, SDMX or other)		
10		esponTheme*	*to be filled in by ESPONDB*			
11		esponKeyword*	*to be filled in by ESPONDB*			
12	Temporal extent					
13		start	2003			
14		end	2003			
15	Quality					
16	scope					
17		label	1	free text		
18		spatial				
19			level*	0		
20			unitCode*	*		* means all
21		temporal				
22			start	2003		
23			end	2003		
24		lineage				
25			provider	EUROSTAT		
26			source	2007		
27			transformations	none		formulas in mathematical terms if possible
28			methodology	...		free text
29		completeness	1,00	a percentage		
30		thematic accuracy	5	a confidence level (from 1 to 5) : 5 is high		
31		logical consistency	0,65	a percentage		
32		temporal accuracy				
33			frequency	year		day, year,
34			Observation date	03/07/2003	a date : DD/MM/YYY	
35	scope					
36		label	2	free text		
37		spatial				
38			level	0		
39			unitCode*	LI, RO, NO, IS, CH, BG		
40		temporal				
41			start	2003		
42			end	2003		
43		lineage				
44			provider	EUROSTAT		
45			source	2008		
46			transformations	none		
47			methodology	...		
48		completeness	1,00	a percentage		
49		thematic accuracy	4	a confidence level (from 1 to 5) : 5 is high		
50		logical consistency	0,45	a percentage		
51		temporal accuracy				
52			frequency	year		
53			Observation date	03/07/2003	a date : DD/MM/YYY	

Figure 47. Spreadsheet solution: Indicator metadata sheet

More examples will be provided to each ESPON Projects explaining how this metadata sheet is to be linked with the actual data files.

Nevertheless, it is worth mentioning the thematic characterisation of the dataset based on themes and keywords for two reasons. First, these keywords and themes can be taken from thesaurus and glossaries. INSPIRE (GEMET) and SDMX propose such themes. It is important to fit with these initiatives. One can see obviously here that a web-based application would enhance the way one can fill-in these forms by given direct access to glossaries. Second, the fields espontheme and esponkeywords have been introduced to allow for a thematic characterisation that is more focussed on ESPON needs and objectives. Those two fields will be used to structure the database into thematics that are proper to the ESPON community (cohesion, polycentricity,… or any policy driven concept). They will constitute further entry points to the ESPON users when accessing the database. The ESPON DB project will define these values.

3.2.4 ESPON DB metadata mid-term needs and strategy

The previously described solution is not adequate in the longer-run because of its low-automatisation and desktop approach. A Web-based metadata editor is to be developed and integrated together with a data import interface. We will tackle this tool issue and refine the ESPON metadata profile in the next stages of the project.

We detail below the desired features both on the import and export sides of the database (section 5.1.4.1) and then shortly present a web-based solution: Geonetwork on which we are going to encroach further tool developments for metadata.

Import features

Here is a list of basic functionalities that should be provided in this tool for importing metadata:

A thesaurus management tool, an independent component that we can plug into the architecture :

creation, deletion, modification of a thesaurus,

edition/visualization of terms in a hierarchical and alphabetical structure,

import/export from/to text files in different formats.

This thesaurus should help the classification and description of indicators using keywords in addition to ESPON specific ones. For example, the GEMET thesaurus or the SMDX "COG".

An XML import/export tool that enables the exchange of metadata records in XML format conforming to various formats (SDMX, ISO 19115, CSDGM). Using XSL transformations, this tool should enable the production of various HTML presentation of metadata, according to various user profiles (or "preferences").

Automatic metadata generation for some data file formats. When these formats correspond to raster formats (Shapefile, DGN, ECW, FICC, GeoTiff, GIF/GFW, JPG/JGW, PNG/PGW), the tool could just interface with software that can read them. [Diaz et al., 2007] If parsing Excel data files, it is possible to extract some information from the data file itself if it is well organized.

The management of group of metadata: that is to say that the tool should handle various levels of descriptions for data collection. Sometimes, a set of metadata is common to a whole set of data (for example, the source, and the temporal extent), and then the lineage should be refined only for a sub-set of data. Generally, this refinement is done according geographic criteria, and can be applied downwards to the finest geographic unit.

A tool for the definition of the geographic reference area that is covered by metadata: some graphical interface allowing user to select automatically the region of interest, either by its selection on a map, or by the drawing of a bounding box using the mouse on a map. This could also be done by providing the name/code of the region. This should lead to the automatic computation of bounding box coordinates of the region of interest that should be stored with metadata.

A validation tool for checking whether all mandatory fields are present, and reminding the user to provide them. The verification can be done according various standards

(FGDC, Dublin Core, ISO 19115, SDMX). Note the SDMX infrastructure provides some small validator tools checking the validity of data files according a Data Structure Definition.

A contact directory (name, address, mail, telephone) for managing the various contacts (official providers, internal ESPON project's point of contact, ESPON 2013 DB contact) to ease the filling of fields such as "data provider", "metadata publisher" and so on.

A user access control system to restrict the set of users allowed to edit metadata: the members of ESPON 2013 DB project, or internal ESPON's projects. A session identification using a login/password for example could be convenient.

Support for internationalization that is to say to allow the translation in various languages of the metadata.

On-line help about the metadata elements defined in a specific metadata profile: definitions, maximum occurrence, examples

Finally, metadata should also be collected even if they are provided in other templates, using other standards. The ESPON DB 2013 Project may have to import in ESPON DB data that have already been fully described using standards (FGDC, ISO 19115 or SDMX). The future import tool should thus be able to handle those metadata structures to avoid any redundant acquisition for the user.

Export features

The EXPORT side of the ESPON 2013 DB should allow for

The discovery of data by querying metadata: that is to say to get some information about the completeness of a dataset, the quality level, the lineage, the theme (and so on) of a particular dataset before to decide to export it.

Some Access Control Level (ACL) should also be put on metadata concerning the copyrights of data to inform the user about the possibility to get (or not) these data. If data are for internal use inside ESPON projects only, a password and login should be required to download them from ESPON web site.

Export data using the standards, and associate them with the corresponding metadata. That is to say that metadata information (even if collected in one single file for various indicators) should be delivered "on demand" for each specific indicator.

There are two further points that derive from those expected features:

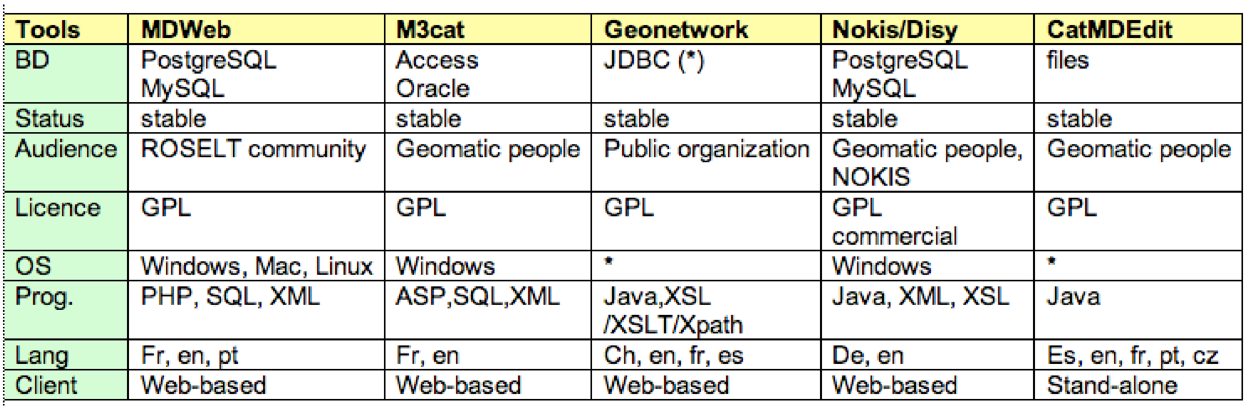
The first point means that we need to offer a web-based component for searching data using various criteria (scale level, geographic location, and thematic fields), thus showing the metadata to the user BEFORE he/she confirms the download. For example, the user could be aware of the level of completeness of specific dataset before to decide to download (if he/she is granted, according to its ACL). On this point, INSPIRE gives some precision concerning the query criterias:
"the Discovery service shall implement as a minimum the following combination of search criteria:
(a) keywords;
(b) classification of spatial data and services;
(c) the quality and validity of spatial data sets;
(d) degree of conformity with the implementing rules provided for in Article 7(1);
(e) geographical location;
(f) conditions applying to the access to and use of spatial data sets and services;

(g) the public authorities responsible for the establishment, management, maintenance and distribution of spatial data sets and services.”

The second point concerns the model used by the ESPON DB 2013 Project for exportation: if the SDMX standard is used, data will be mixed with their metadata inside XML files. If we use a more classic way, the data will come in a spreadsheet format, but associated with another file, the metadata file, that should be formatted according ISO 19115 standard like an XML file. Ideally, our interface for exportation of data should offer the two choices. Even better, would be a geographic web service based on OGC standards for cataloging (WCS).

Towards a web-based tool

There are several metadata tools adapted to geographical information (see Figure 48). We have reviewed some of them plus the online INSPIRE Metadata Editor.



Tools	MDWeb	M3cat	Geonetwork	Nokis/Disy	CatMDEdit
BD	PostgreSQL MySQL	Access Oracle	JDBC (*)	PostgreSQL MySQL	files
Status	stable	stable	stable	stable	stable
Audience	ROSELT community	Geomatic people	Public organization	Geomatic people, NOKIS	Geomatic people
Licence	GPL	GPL	GPL	GPL commercial	GPL
OS	Windows, Mac, Linux	Windows	*	Windows	*
Prog.	PHP, SQL, XML	ASP,SQL,XML	Java,XSL /XSLT/Xpath	Java, XML, XSL	Java
Lang	Fr, en, pt	Fr, en	Ch, en, fr, es	De, en	Es, en, fr, pt, cz
Client	Web-based	Web-based	Web-based	Web-based	Stand-alone

Figure 48: Metadata tools comparison

We have rejected the use of a desktop tool for (e.g. CatMEdit, [Zarazaga-Soria et al., 2003]) for the reasons mentioned above.

We also rejected the use of the INSPIRE web-tool because of several shortcomings, particularly the fact that it is not tuned to statistical data. Moreover, we found that several fields could be pre-filled either because they correspond to ESPON related features (ESPON projects name, repository, ESPON URL, a dedicated set of themes and keywords, etc.), or because they could be derived from the data if integrated into our import pool.

Given the list of needs and requirements detailed above, the most promising tool seems to be the Geonetwork metadata editor, which is open and for which we can build on the expertise of the University of Barcelona.

Geonetwork **(see http://geonetwork-opensource.org/) is a complete open source package for setting up a web geoportal.**

The portal provides a catalog application to manage spatially referenced resources through the web. The metadata catalogue is fully compliant with the Catalog Web Service standard (CSW 2.0) allowing the connection from

different metadata catalogues such as: FAO, UNEP, ESA, INSPIRE, ETC-LUSI, ETC-WATER, EEA, etc.

It also provides a powerful web metadata editor which has already implemented different metadata standards: FGDC, Dublin Core, ISO 19115/ ISO 19139. Other templates could be easily added to create metadata according to different standards such as the Statistical Data and Metadata Exchange (SDMX).

Once metadata are created, XML files are produced out of the Web form and stored in a database. With such a solution the metadata files provided by different ESPON projects could for example be validated and uploaded into the database automatically.

In addition to those features, search functions are also implemented and an interactive web map viewer is embedded.

Examples of such facilities are shown below in Figures 49 to 51.

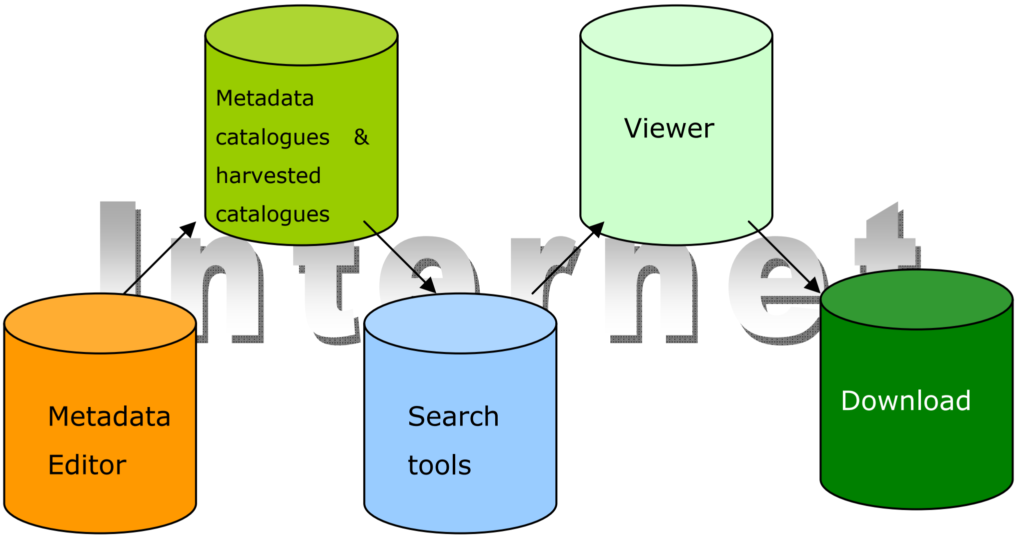
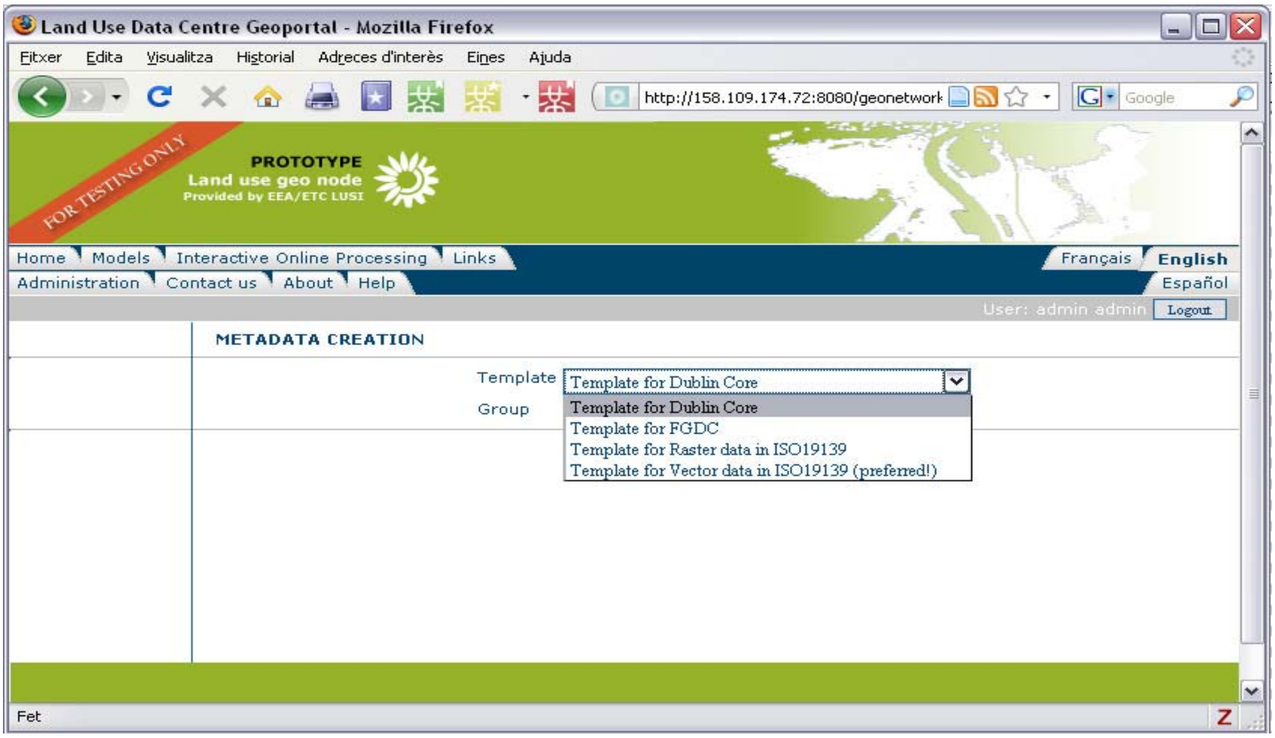


Figure 49: Schema of different functionalities within the Geonetwork platform.



Land Use Data Centre Geoportal - Mozilla Firefox
Fitxer Edita Visualitza Historial Adreces d'interès Eines Ajuda
http://158.109.174.72:8080/geonetwork
FOR TESTING ONLY
PROTOTYPE
Land use geo node
Provided by EEA/ETC LUSI
Home Models Interactive Online Processing Links
Administration Contact us About Help
Français English Español
User: admin admin Logout
Fet

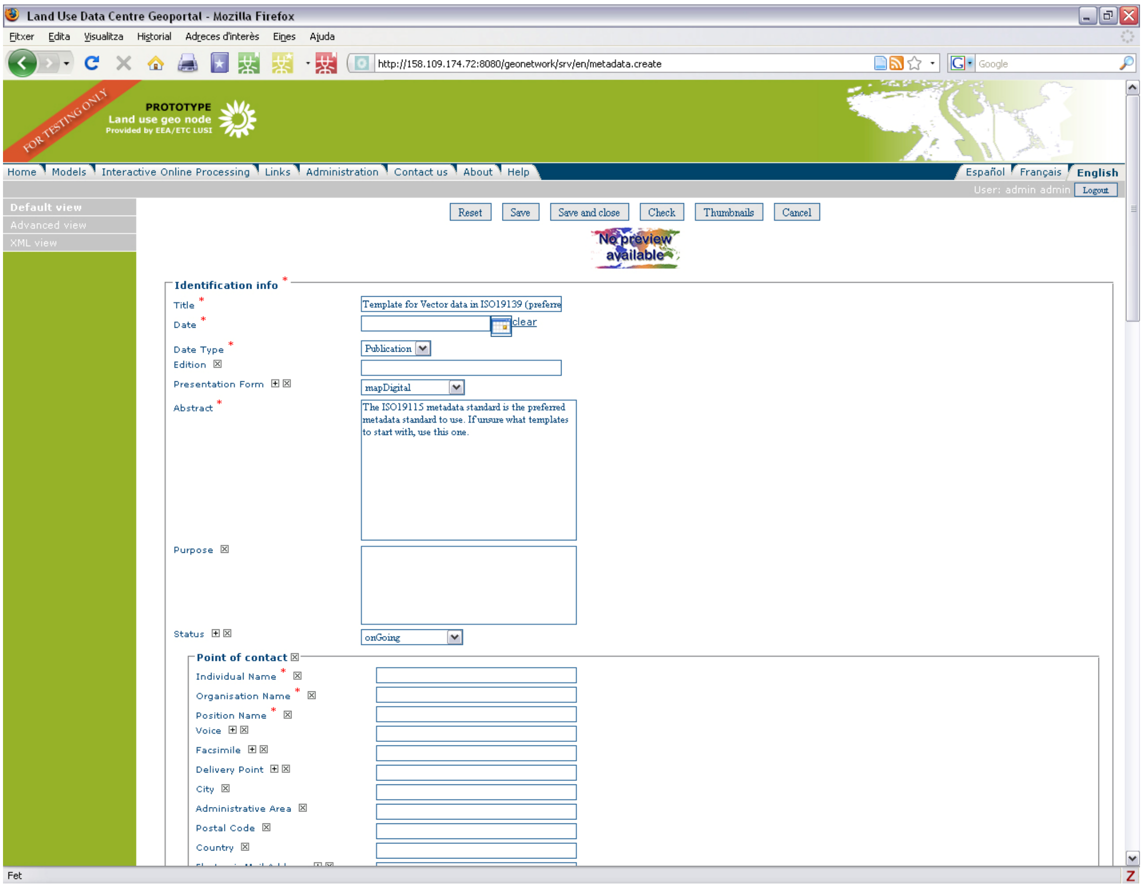
METADATA CREATION

Template: Template for Dublin Core

Group

- Template for Dublin Core
- Template for FGDC
- Template for Raster data in ISO19139
- Template for Vector data in ISO19139 (preferred!)

Figure 50: The metadata creation tool allows the user to generate a new metadata file based on a template according to a specific metadata standard.



Land Use Data Centre Geoportal - Mozilla Firefox
Fitxer Edita Visualitza Historial Adreces d'interès Eines Ajuda
http://158.109.174.72:8080/geonetwork/srv/en/metadata.create
FOR TESTING ONLY
PROTOTYPE
Land use geo node
Provided by EEA/ETC LUSI
Home Models Interactive Online Processing Links Administration Contact us About Help
Español Français English
User: admin admin Logout
Default view
Advanced view
XML view
Reset Save Save and close Check Thumbnails Cancel
No preview available
Fet

Identification info *

Title *: Template for Vector data in ISO19139 (preferre

Date *: clear

Date Type *: Publication

Edition ☒

Presentation Form ⊞ ☒: mapDigital

Abstract *: The ISO19115 metadata standard is the preferred metadata standard to use. If unsure what templates to start with, use this one.

Purpose ☒

Status ⊞ ☒: onGoing

Point of contact ☒

Individual Name * ☒

Organisation Name * ☒

Position Name * ☒

Voice ⊞ ☒

Facsimile ⊞ ☒

Delivery Point ⊞ ☒

City ☒

Administrative Area ☒

Postal Code ☒

Country ☒

Figure 51: Overview of Geonetwork metadata editor for the Land Use Data Centre Prototype.

3.2.5 Next steps

In the next months we will further our research in order to pass from the short to the mid-term solution for metadata exchange.

First, we will start developing the Geoportal metadata editor that can integrate the ISO 19115 and SDMX standards, as well as other specificities of our ESPON profile, and link this development to the Database import pool. Second, we should refine the ESPON Metadata profile based on several examples and experiences of users. A review of metadata fields used in other similar databases (e.g. UN, OECD, FAO) will also help in preparing an ESPON specific thematic structure of the database.

A metadata specific meeting will be held in Barcelona in March 2009 to discuss the metadata profile and the computer implementation.

We also intend to circulate a Metadata guideline for the next seminar in June.

Additional figures

SDMX example for currency exchange

First the data structure definition provides the metadata tags to use, the code lists to use for the valuation, and their organization inside a hierarchical structure (Dataset, Group, Serie, Observation).

```
<!DOCTYPE root>
<root>
        <Structure xmlns="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message"
                xmlns:message="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message"
                xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                xsi:schemaLocation="
        http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message
                SDMXMessage.xsd
        http://www.SDMX.org/resources/SDMXML/schemas/v2_0/structure
        SDMXStructure.xsd">
                <Header>
                        <ID>IREF000506</ID>
                        <Test>false</Test>
                        <Name>ECB structural definitions</Name>
                        <Prepared>2006-10-25T14:26:00</Prepared>
                        <Sender id="4F0"/>
                </Header>
<!-- The Concepts element contains a list of concepts used to identify and describe the data.
-->
                <Concept agencyID="ECB" id="COLLECTION">
                        <Name xml:lang="en">Collection indicator</Name>
                </Concept>
                <!-- The CodeLists element contains a list of CodeList elements.
                Each CodeList element contains 2 attributes:
                the ID of the Agency responsible for the code list ("ECB")
                and the code list ID (for example "CL_EXR_SUFFIX").
                -->
                <CodeList agencyID="ECB" id="CL_EXR_SUFFIX">
                        <Name xml:lang="en">Exch. rate series variation code list</Name>
                        <Code value="A">
                                <Description xml:lang="en">
                                 Average or standardised measure for given frequency
                                 </Description>
                        </Code>
```

```
            <Code value="E">
                <Description xml:lang="en">End-of-period</Description>
            </Code>
            <!-- … -->
        </CodeList>
        <KeyFamily agencyID="ECB" id="ECB_EXR1"
            uri="http://www.ecb.int/vocabulary/stats/exr/1">
            <Name xml:lang="en">Exchange Rates</Name>
            <Components>
            <!-- list the dimensions used to describe the observated value -->
                <Dimension        conceptRef="FREQ"        codelist="CL_FREQ"
isFrequencyDimension="true"/>
                <Dimension conceptRef="CURRENCY" codelist="CL_CURRENCY"/>
                <Dimension conceptRef="CURRENCY_DENOM" codelist="CL_CURRENCY"/>
                <Dimension conceptRef="EXR_TYPE" codelist="CL_EXR_TYPE"/>
                <Dimension conceptRef="EXR_SUFFIX" codelist="CL_EXR_SUFFIX"/>
                <TimeDimension conceptRef="TIME_PERIOD"/>

                <!-- define the attributes attached to the group level -->
                <Group id="Group">
                    <DimensionRef>CURRENCY_DENOM</DimensionRef>
                    <DimensionRef>EXR_TYPE</DimensionRef>
                    <DimensionRef>EXR_SUFFIX</DimensionRef>
                </Group>

<!-- Then, we indicate which attribute will contain the measured value.
Conventionally, it is associated with the OBS_VALUE concept.-->
                <PrimaryMeasure conceptRef="OBS_VALUE"/>

<!-- list the attributes
An Attribute element will contain information such as the concept used for the attribute,
the attachment level (i.e. "Observation", "Group", "Series", "DataSet")
and whether it is mandatory or not (i.e. "Mandatory" versus "Conditional").
-->
                <Attributes>
                <Attribute conceptRef="TIME_FORMAT" attachmentLevel="Series"
                        assignmentStatus="Mandatory" isTimeFormat="true">
                        <TextFormat textType="String" maxLength="3"/>
                </Attribute>
                <Attribute conceptRef="OBS_STATUS" attachmentLevel="Observation"
                        assignmentStatus="Mandatory"
                        codelist="CL_OBS_STATUS"/>
                <Attribute conceptRef="DECIMALS" attachmentLevel="Group"
```

```
                                              assignmentStatus="Mandatory"
                                               codelist="CL_DECIMALS">
                                              <AttachmentGroup>Group</AttachmentGroup>
                                </Attribute>
                                 <!-- … -->
                                </Attributes>

                        </Components>
                </KeyFamily>
        </Structure>
</root>
```

Figure 52a. The Data Structure Definition file (1ecb_exr1_compact.xsd file)

Second, the data file itself, using the tags previously choosen to describe the values

```
<!DOCTYPE root>
<root>
        <CompactData
                xmlns="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message"
                xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                xsi:schemaLocation="http://www.SDMX.org/resources/SDMXML/schemas/v2_0/message
SDMXMessage.xsd">
                <Header>
                        <ID>EXR-HIST_2006-11-29</ID>
                        <Test>false</Test>
                        <Name xml:lang="en">Euro foreign exchange reference rates</Name>
                        <Prepared>2006-11-23T08:26:29</Prepared>
                        <Sender id="4F0">
                                <Name xml:lang="en">European Central Bank</Name>
                                <Contact>
                                        <Department xml:lang="en">DG Statistics</Department>
                                        <URI>mailto:statistics@ecb.int</URI>
                                </Contact>
                        </Sender>
                </Header>

                <DataSet xmlns="http://www.ecb.int/vocabulary/stats/exr/1"
        xsi:schemaLocation="http://www.ecb.int/vocabulary/stats/exr/1ecb_exr1_compact.xsd"
                        datasetID="ECB_EXR1">
                        <Group CURRENCY="AUD" CURRENCY_DENOM="EUR" EXR_TYPE="SP00"
                                EXR_SUFFIX="A" DECIMALS="4" UNIT="AUD" UNIT_MULT="0"
                                TITLE_COMPL="ECB    reference    exchange    rate,    Australian
dollar/Euro"/>

                        <Series FREQ="D" CURRENCY="AUD" CURRENCY_DENOM="EUR" EXR_TYPE="SP00"
                                EXR_SUFFIX="A" TIME_FORMAT="P1D" COLLECTION="A">
                                <Obs TIME_PERIOD="1999-01-04" OBS_VALUE="1.9100" OBS_STATUS="A"
                                        OBS_CONF="F"/>
                        </Series>
                </DataSet>
        </CompactData>

</root>
```

The header is equivalent to the Identification topic of ISO 19113.

It is the reference to the "Data Structure Definition", that is to sav the schema

Data description is structured by various scopes:

- DataSet
- Group
- Series
- Obs : measure level

Value is 1.9100, is valid for 04 April 1991. The quality information OBS_STATUS indicates if it a forecast, an estimate, a normal value, etc. : normal in this case.

Figure 52b. The SDMX-ML data file

OBS_CONF gives the confidentiality level of this information: free in this case

References

[Barde, 2005] Mutualisation de Données et de Connaissances pour la Gestion Intégrée des Zones Côtières. Application au Projet SYSCOLAG. BARDE Julien - Université Montpellier II , 2005 http://papyrus.lirmm.fr/Auteur.htm?numrec=061967579914930

[Diaz et al., 2007] Laura Díaz, Cristian Martín , Michael Gould, Carlos Granell, Miguel Ángel Manso (2007) Semi-Automatic Metadata Extraction from Imagery and Cartographic Data. In proceedings of IGARRS'07

[Kent et Schuerhoff, 1997] Kent, J-P. and Schuerhoff, M. (1997). Some Thoughts About a Metadata Management System. In Proc. NinthInternational Conference on Scientific and Statistical Database Management, Olympia, Washington (pp. 174–185). http://portal.acm.org/citation.cfm?id=695474

[Servigne et al., 2006] Spatial data quality components, standards and metadata. S. Servigne, N. Lesage, T. Libourel. Fundamentals of Spatial Data Quality. International Scientific and Technical Encyclopedia. ISBN 1905209568. Pp. 179-210 2006.

[SDMX MCV]. Content-Oriented Guidelines Annex4 – Metadata Common Vocabulary http://sdmx.org/

[SDMX CL]. Content-Oriented Guidelines Annex2 – Cross-Domain Code Lists http://sdmx.org/

[Zarazaga-Soria et al., 2003] Zarazaga-Soria, F. J.; J. Lacasta; J. Nogueras-Iso et al. (2003): A Java Tool for Creating ISO/FGDC Geographic Metadata, in: Bernard, L.; A. Sliwinski und K. Senkler [eds.]: Geodaten- und Geodienste-Infrastrukturen - von der Forschung zur praktischen Anwendung. Beiträge zu den Munsteraner GI-Tagen. IfGIprints 18:17-30.

www.gi-tage.de/archive/2003/downloads/gitage2003/tagungsband/zarazaga_soria.pdf

ISO TC211 http://www.isotc211.org/hmmg/HTML/root.html

4 Conclusion

In this final section, the two scientific coordinators of the project, Jerome Gensel (LIG) and Claude Grasland (UMS RIATE) present some general observations on progress made and future workplan () and ask some questions to ESPON CU and ESPON MC.

4.1 Synthesis of progress made

About general objectives

The ESPON 2013 Database project is a central project in the ESPON 2013 program. Its mission is both to make available harmonized datasets, mostly coming from official statistical organisms, to the other ESPON program projects, and to collect datasets produced by these projects in order to promote data exchange between them, even to contribute to the launching of new projects. Contrary to the ESPON 2006 program, indicators handled here are not anymore only socio-economic statistical data attached to territorial units of NUTS type, but can also be wider thematic indicators (concerning environment for instance) associated with other division of the European space (regular grids, local units, urban zones…) or of world space (global units).

About the internal organisation by challenge

In order to carry out the project, 12 challenges (“so called” the 12 labours of Hercules) have been identified and assigned to the different partners and experts associated with the project. These 12 challenges cover the whole spectrum of scientific, technical and organizational problems linked to the global objective of the project.

Methodology

These 12 challenges are strongly interdependent and interconnected. Thus, if one coordinator exists for each challenge, other partners or experts, themselves coordinators of other challenges, are involved in each challenge. This promotes collaboration, dialog and reinforces the cohesion between the 12 challenges, which is indispensable for the success of the project as a whole.

The issues addressed

The issues addressed by the different challenges are recent scientific and technical problems which, each inside its own domain (for example in spatial analysis or in computer science), mobilise numerous research teams in Europe and further. It is worth noting that scientific and technological issues, but also administrative and legislative ones, that are tackled by challenges are not all settled, whether they concern the harmonisation of temporal series, the collection of global or local data, the merging of environmental or socio-economic data, the representation of urban zones, the elaboration of adapted metadata profiles, the conception and development of a system for managing evolutive spatio-temporal information, the elaboration of missing data estimation methods, the identification of outliers, the integration of data relative to Europe bordering countries or countries that request membership, or the management of data available at greater resolutions (at individual level for instance). This wide spectrum of present problems shows that partners and experts working together in the ESPON 2013 DataBase project tackle ambitious challenges to which they hope to answer with original and exploratory approaches, if not to bring solutions.

The central role of the Challenge 9

The challenge 9 is at the core of the ESPON 2013 DataBase project. It renders operational the circulation of indicators datasets since their collection up to their redistribution, taking also into account if needed their harmonization and their estimation. Moreover, for each dataset handled by the system, associated metadata are collected and can be used as criteria (i.e. queried) to select datasets.

The building process of such a spatio-temporal information system capable of supporting the diversity of indicators and metadata, of spatial divisions and of estimation and identification methods, gathers project partners and experts (in computer science and thematic fields) on every issue addressed.

Concerning the tools for the development, choice has been made clear about the selection of open-source technologies, those that are standardized and supported by a wide and active community. The motivations of this choice have been exposed during the ESPON seminar in Bordeaux, and validated by every actors of the project.

Moreover, the work on metadata has proven the attention paid by each ESPON 2013 DataBase partners to the standards (such as ISO 19115 for georeferenced data or SDMX for statistical data) and legal directives (INSPIRE), for at least the respect of them, and better, for a contribution of the project to their advancement.

First results

The description by each challenge of the work that has already been done shows that every one has a good understanding of its allocated task, is totally involved within, and can supply a work plan covering the whole project period, with many stakes marking the steps of the progress made. Furthermore, this report proves also that every partners or experts contributing to the project has coordinated themselves through an extensive usage of communication tools (such as an extranet web site that reveals to be a very efficient tool), so that they can plan a methodical and progressive advancement of the work, with various steps for the validation of the achievement of the objectives. The communication is a key point to deal with the huge task we must face to, in a coherent and homogeneous manner.

Update of objectives

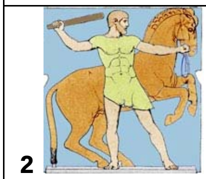
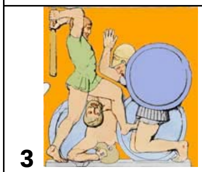
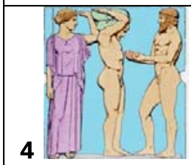
For each challenge, a list of identified problems has been provided, with the hope that the Monitoring Committee could consider the amount of efforts that have been put in order to pass over or work around the various difficulties, which can be of scientific, technologic nature, but also of administrative, financial or bureaucratic nature. For this survey of problems, we provide also a list of measures and decisions that have been, (or should be), acted to solve them.

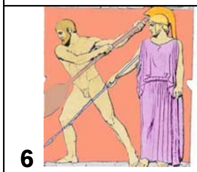
Whenever the difficulty to progress in one challenge or on a specific task becomes too high, one of the option might be to reduce the ambition of the challenge, or give up the task, or substitute it by another objective, close to the previous one and relevant for the global objective of the project.

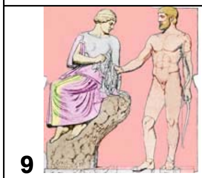
4.2 Workplan until SIR

The work plans provided by each challenge expose the possibilities of advancement that each partner or expert consider as reasonnable, and sustainable within the devoted time and resources, to reach the goal fixed for each challenge, and further, the ESPON 2013 Database project's objectives.

Even though the shedule provided within the Subsidy Contract was restricted to one next stake in February 2010, with the deliverance of the second interim report, we have decided to add two more intermediary stakes, each ones for the ESPON's seminars, in June and December 2009. This allows partners and experts to coordinate their various tasks on more reduced time periods. Furthermore, the Monitoring Committe could be re-assured about our progress more regularly, and could evaluate it in a more detailed manner. This could be endorsed by the establishement of a presentations planning, in concertation with all members of the project, for demonstrations and talks at the ESPON's seminars.

Challenges	**June 2009**	**December 2009**	**February 2010**
1	Continue to check and integrate dataset from other ESPON projects or expertises	Try to enlarge the integration of two basic data - Population and area - to other geographical objects and scales: World, cities, grids (exchanges with challenges 3, 5 and 6).	Try to define a methodology to detect spatial and statistical outlier in these basic datasets to point out extraordinary values (exchanges with challenge 10)
2 	Diagnostic of time series' availability in the ESPON area (technical report)	Elaboration of dictionary NUTS' changes (technical report)	Computing data models and automating some proceedings
3 	Partnership agreement ESPON-UNEP GEO TECHNICAL REPORT "ESPON World database (I) : Dictionary of units and regions" ESPON World Database version 1.0 (Data + Geometry) Networking with FP7-EuroBroadmap	Partnership agreement ESPON-UNEP GEO TECHNICAL REPORT "ESPON World database (II) : Integration of national and regional levels" ESPON World Database version 2.0 (Data + Geometry) Support to ESPON project Priority 1 / Globalization Networking with FP7-EuroBroadmap	Preparation of SIR Integration of results with other challenges, in particular C.1 (basic data), C.2 (time series), C.5 (Grid) and C.6 (Cities).
4 	Providing a finalized database with indicators for at least two neighboring countries and completing the database with available indicators at LAU1/2 level for the ESPON space	Finalizing the indicator database for most of the countries and deriving a short history of the modifications in the LAU1/2 units' geometry or in the official denominations	Finishing the process of filling the database with information for one or two indicators, country by country, until we complete the first field and recovering the information available in the SIRE database

5	Testing different socioeconomic variables or indicators, using all integration methods. Technical report about the conclusions derived from those tests.	Automating of calculation processes and integration of some socioeconomic variables or indicators into the EEA's LEAC System and comparison with environmental data.	Assessment of the results of the integration of data into the LEAC System. Technical report detailing the procedure, results and challenges.
6 	a) Storage (urban delineations, socio-economic data) b) delivering UMZ data base: metadata, names and other current attributes c) Technical Report: Naming the UMZ	a) Storage (urban delineations, socio-economic data) .b) Technical Report: comparison UMZ/MUAS	a) Storage (urban delineations, socio-economic data) b) Technical Report: comparison UMZ/Swedish localities
7	1) Eurostat: archived regional database (April) ; 2) Meeting Statistical and Cartographic offices (March); 3) ESPON seminar (June)	1) Eurostat-routing meeting: metadata, grids, missing values (e.g. September)	1) Meeting DG-Envi; 2) Meeting DG Agri; 3) Potential indicators/maps to DG Regio (Cohesion Report)
8-a	Updated version of map kit tool Technical report on cartographic principle to be followed in ESPON Cartographic support to ESPON CU	Updated version of map kit tool Proposal of one day formation in ESPON cartography Cartographic support to ESPON CU	

8-b	Partners involved in the definition of an ESPON format for metadata (namely, UL, AUB, RIATE, and LIG) will meet in March in Barcelona. As a result, one template (or profile), compliant to the standards Inspire ISO 19115 and SDMX, will be then defined and implemented.	Design and implementation of the first version of an editor of metadata based on the template defined. This editor will mainly allow the association of metadata with socio-economical indicators.	More evolved version of the metadata editor, according to the evolution of the metadata template. • New version of the ESPON Databases schema: extension of the MegaBase schema towards the integration of quality information.
9 	Definition of an INSPIRE compliant metadata format (including ISO 19115 and SDMX) and implementation of a first version of Web editor for this format (with the corresponding user manual); inclusion of the metadata into the ESPON DataBase schema;	Extension of the MegaBasse schema with the integration of data quality descriptors, of global, local and urban units; design of a first version of the spatial and indicator ontology	Implementation of global, local and urban units in the the ESPON DataBase and implementation of the spatial and indicator ontologies; design of a the Problem Solving Environment dedicated to the data harmonization and estimation

10	The first step will run from December 2008 to December 2009. The key aim of this work is to examine how statistical analysis tools can be applied to the ESPON 2013 Database in order to find 'outliers', i.e. data that are exceptional as compared to neighbouring data with respect to:
	a) attribute-space;
	b) geographical-space;
	c) temporal-space (and where each dimension depends on the scale it is viewed at).		

11	(a) Evaluation of the situation of data available in C. countries. Technical report. (b) Establishment of contacts with CC national statistical .offices and assessment how it is possible to establish regular dataflow with ESPON 2013 Database		Continue tasks (a) and (b) / paper on the iclusion of CC reg. data in ESPON DB for the 2nd Int. Rep. - In Enlarge the scope of ESPON DB to cover ENC and MNC regional data
12	Initial work will be devoted to an overview of available surveys that might be of interest for ESPON, in terms of aspects such as focus, content, temporal and geographical scope, current utilization, etc. Surveys that will be looked at include available Eurostat datasets, such as the Labour Force Survey and the European Community Household Panel, but also surveys from other sources, e.g. the European Values Study and surveys originating from organizations such as UN and OECD.		

4.3 ESPON DB and ESPON Project priorities

We take the opportunity of the FIR to point some questions to be addressed to the ESPON CU concerning the relation between ESPON DB and other ESPON projects

ESPON DB and priority 1 projects : division of work for data collection.

The basic rule for the division of work in terms of data collection is that ESPON DB should collect the "*data of general interest*", that can be defined as data that are of general use for all ESPON project. The other data should be collected by ESPON projects under priority 1, 2 or 3, with a simple validation by ESPON DB project. This principle is theoretically simple but, in practical term, it is not always easy to define the border between "data of general interest" and other data to be collected by ESPON projects. As an example, we have put in annex 6 the data request sent by ReRisk project in Feb. 2009 to ESPON DB through their contact team (UAB). Starting from this example, we can distinguish different types of situations:

1. Basic data that should be obviously provided by ESPON DB : population density, GDP per capita, unemployment rate, etc.
2. Data that are currently produced by other ESPON projects : urban sprawl or indicators of urban poverty (FOCI), % of elderly people (DEMIFER),
3. Specific data for which ESPON DB can provide expertise and recommendations : a typical example is average annual minimum and maximum temperature for which ESPON DB can suggest methodology (methods of aggregation toward Nuts or grid) but not do directly the job.
4. Data that are clearly based on the thematic of the project: energy consumption by sources and sectors, regional energy costs, …
5. Data that are not basic but are not actually covered by a project : for example, regional gross added value by sector.

It is probably difficult to propose a perfect solution to this problem of division of work, but one possible solution could be the development of a kind of **"Data forum"** where the different ESPON projects can exchange their ideas and initiatives.

ESPON DB and priority 2 projects: the challenge of local zoom on territories.

Actually we are just at the beginning of the cooperation between ESPON DB and priority 2 projects, but some difficulties can be predicted and, possibly, solve in advance.

The general focus of priority 2 projects on local zoom on territories implies specific requests of data toward ESPON DB (i.e. data at LAU1 and LAU2 levels, coupled with Eurogeographics geometries) that will be very difficult to satisfy. As it was explained in Challenge 4, the creation of a complete set of data at LAU level, even for basic indicators, is a very difficult task that will probably not be achieved before the end of the current ESPON DB project in 2011. The support to this priority 2 project should therefore be "taylor made", taking into account their great diversity of scales and focus. It can be a very time consuming task if some rules are not clearly defined concerning the support that can be offered by ESPON DB.

A reverse problem is the integration of data produced by priority 2 projects in the general ESPON database. According to the diversity of scopes, scales and methods of this priority 2 projects, it is not obvious that the data they will collect should be always integrated in the general corpus of ESPON database and it could be better to examine, case by case, if the integration is justified. If no added value is expected from this integration, it should be better to let the data produced by this priority 2 projects out of the general database as standalone deliverable.

ESPON DB and priority 3 projects: the need for more integration.

As explained in the discussion of challenge 1 and 8, the level of integration has not been sufficient between ESPON DB and other priority 3 projects, especially the projects related to data update on demography (achieved in Dec. 2008) and accessibility (to be achieved in June 2008). In both case, the contact between ESPON DB and this specific project was establish too late and with strong time pressure that could have been avoided if a better coordination of linkage had been elaborated by ESPON CU between the actual and future priority 3 projects. The only case of anticipation was the request of information from candidates to the project on typologies. But in this case, it was impossible to establish contact with teams that was in the tendering process as it could have been a conflict of interest …

Our recommendation to ESPON CU is to establish a more integrated workplan for Priority 3 projects and to associate early the ESPON DB project to the process of data update realised by other teams working under priority 3. ESPON DB can organise specific meeting with the other priority 3 projects or associate them to the data forum suggested for priority 1.

ESPON DB could also be consulted regularly on the choice of priority for data updates. For example, we are convinced that ***the inclusion in the ESPON DB of a time or cost distance matrix by road*** (and if possible for other modes of transport) ***is an absolute priority*** and is much more interesting than aggregated indicator of potential accessibility to population or GDP that are particular outputs derived from this distance matrix. Moreover, we are convinced that such a distance matrix should be computed not only for NUTSxNUTS territorial units but also for Cities x Cities and eventually for GridxGrid. In this last case, it is not possible to store all the distances and what is requested is a software application for the computation of time or cost distance between the different types of geographical objects used in ESPON project.

ESPON DB and priority 4 projects: ECP network as resource for data collection.

We suggest to ESPON to analyze how the network of ESPON Contact Points could be used in order to support the process of data collection, especially in the case of innovative or complex datasets. This could be realized by specific tenders under priority

4 if a consensus is obtained on the interest of the topics. We just give some examples in order to stimulate the debates:

- *Support for the elaboration of LAU1 and LAU2 database* : the task of elaboration of data at local level is very difficult and the support of ESPON contact point could be very helpful on many topics. Some of them very simple like the translation of information from native language of the country to English. Some other more complex like the review of changes in territorial division at local level or the exploration of resources available in an historical perspective.
- *Support for the harmonization of city definition and completion of Urban Audit data*. In the same spirit, Espon Contact Point could be very helpful for the related topic of harmonization of city definition, for example through a review of national definitions available.

We discuss three possible objections to this proposal.

1. *The support for data collection at national level is normally made by the MC contacts rather than ECP*. It is true and logical when the problem is to establish an official contact with national data provider. But what we expect from ECP is different and is rather oriented to a technical expertise and an help for establishing network connections with experts from the country.
2. *All ESPON countries has not appointed an ECP able to fulfill this task*. It is true but most of them are able to engage in such action and, in the countries where ECP are not able to fulfill this task, it could be possible to replace them by a research team suggested by the national MC member.
3. *Are this actions really eligible to priority 4?* We consider that the answer is positive according to the "Bible" of the ESPON Program. It is indeed written about priority 4 that "The transnational networking activities shall in particular ensure an operational approach that can lead to new initiatives within the area" and that "transnational activities could contribute to the scientific consistency the applied research actions by via the CU giving feed back to the Transnational Project Groups based on national information (p. 49)[19]." Moreover, it is typically an activity that fulfill the expectation of wide involvement of countries : "Transnational activities will be selected in order to ensure a complete coverage of all relevant actors within the European territory during the implementation of the programme. The transnational activities shall be well prepared and involve a broad range of stakeholders, including the mobilisation of potential national scientific networks (p. 50)"

[19] **ESPON 2013 PROGRAMME, 2007**, European observation network on territorial development and cohesion, Adopted by European Commission Decision C(2007) 5313of 7 November 2007, CCI 2007 CB 163 PO 022

5 ANNEXES

Annexe 1 (challenge 1)

1.1 Discontinuities in time series

The regional data that are provided by Eurostat are often provisional and revised. In the case of population, many time series are clearly characterized by singularities that can be explained by several facts:

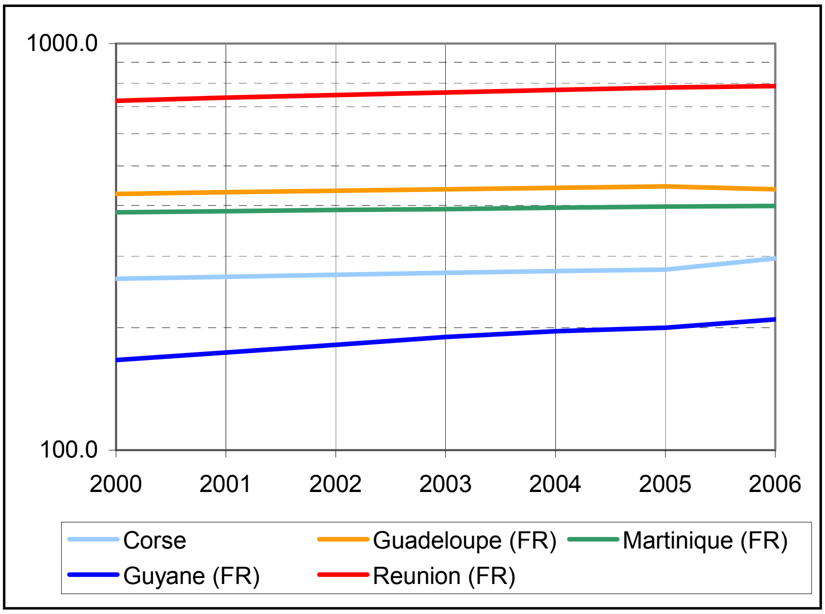
- A new census that provide "real" values after several years of estimation. In this case, the national offices decide generally to revised ex-post the previous estimations but it can takes some time.
- A revision of territorial limits that has not been registered. In this case, we observe generally a couple of neighboring units with opposite trends of increase or decrease.
- A pure mistake, related for example to an inversion of figures (698 instead of 968) or to an error of lines.

But we can never exclude the case of a real exception in reality, related for example to a crisis that is related to exceptional changes in trends. In this case, we do not have a mistake but a "false positive".In this short annex, we propose a preliminary solutions for the detection of breakdowns in tile series of population 2000-2006 at NUTS2 level, that could be further reproduced for other data or other geographical levels.

1.1.1 Visual inspection of population is not efficient

Looking directly at the data is time consuming and not very efficient. For example, the visualization of population evolution of 5 regions of France does not reveals obvious mistakes (Figure 1).

Figure 1: Analysis of population table

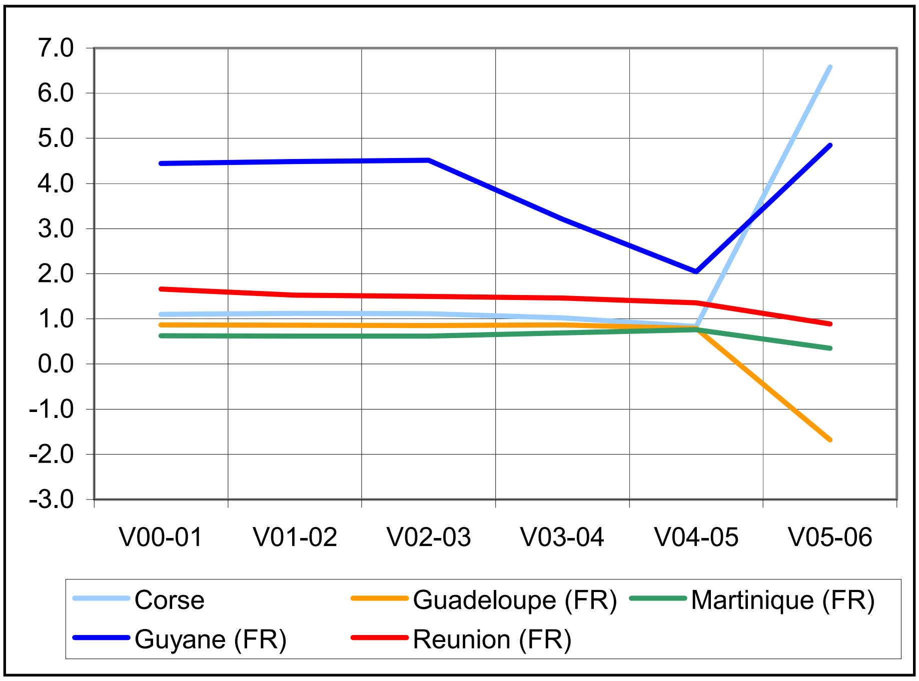


NUTS2	2000	2001	2002	2003	2004	2005	2006
Corse	264.0	266.9	269.9	272.9	275.7	278.0	296.3
Guadeloupe (FR)	427.1	430.8	434.5	438.2	442.0	445.5	438.0
Martinique (FR)	384.6	387.0	389.4	391.8	394.5	397.5	398.9
Guyane (FR)	166.5	173.9	181.7	189.9	196.0	200.0	209.7
Reunion (FR)	723.4	735.4	746.6	757.8	768.9	779.3	786.2

1.1.2 Analysis of population variation is more efficient

The transformation of the previous table into a table of population variation is a much more efficient way to observe singularities in time series. Looking at Figure 2 it is obvious that the evolution of population in Corse and Guadeloupe in 2005-2006 is not consistent with the evolution of the previous periods. The case of Guyane is less clear and we can not be sure that the fluctuations of the rate of population variation are mistakes.

Figure 2: Analysis of annual rate of population variation



NUTS2	V00-01	V01-02	V02-03	V03-04	V04-05	V05-06
Corse	1.1	1.1	1.1	1.0	0.8	6.6
Guadeloupe (FR)	0.9	0.9	0.9	0.9	0.8	-1.7
Martinique (FR)	0.6	0.6	0.6	0.7	0.8	0.4
Guyane (FR)	4.4	4.5	4.5	3.2	2.0	4.8
Reunion (FR)	1.7	1.5	1.5	1.5	1.4	0.9

We have now to define a general method for estimation of "singularities" in time series that could be realized automatically and where the human expert will only have to decide on selected situations where anomalies are suspected. It is indeed impossible to proceed by visual analysis for all cases, especially when we arrive at NUTS3 level …

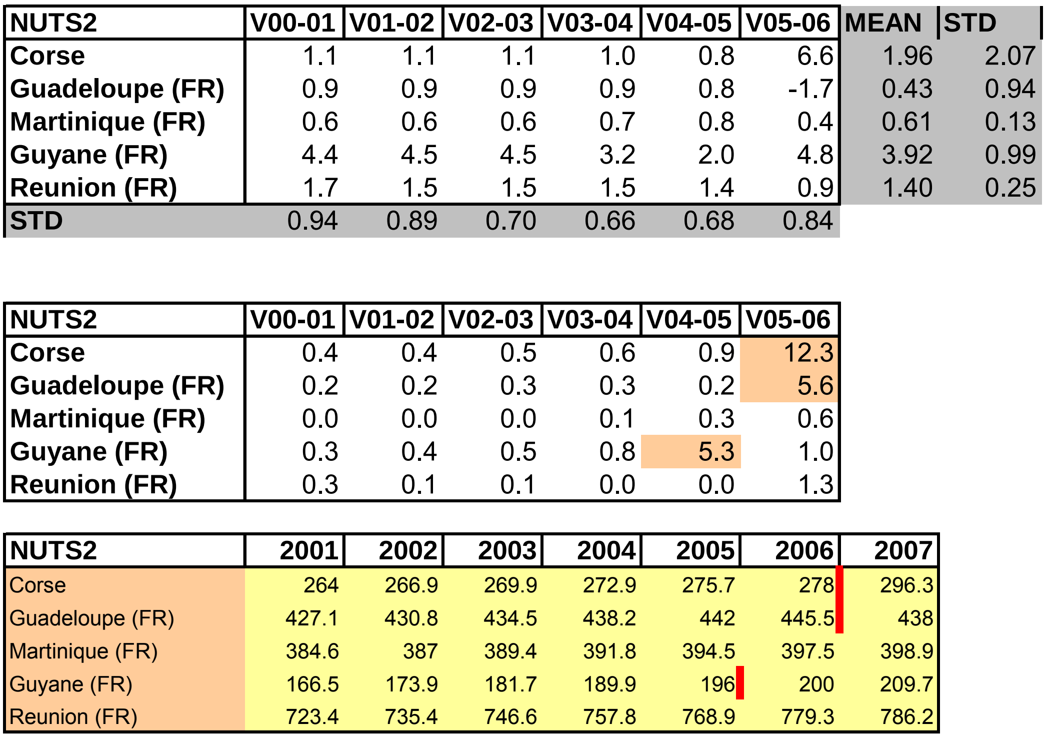
1.1.3 An index for the measure of discontinuities in time series

In order to evaluate the singularities of the time series, we propose to compute the following index :

$$H\ (i,t,t+1) = [VP(t,t+1) - mean(VP,\ t0..tn)\ ]^2 \ /\ [Std(VP,\ t0..tn)*Std(VP,1..i\ldots n)]$$

H is a standardised measure for a territorial unit (i) at a given time period(t,t+1). It combines the standard deviation of the different spatial units (at the same time period) and the standard deviation of the different time periods (for the same territorial unit). With this double standardisation, we take into account the fact that exceptional variations can depend both of the fact that a territorial unit OR a time period are subjected to more or less important variations. Under an assumption of gaussian repartition of the deviations, 95% of the standardised measure of H should be included between 0 and +2. It provide a simple way to identify the exceptional values and to display a visualisation of exceptional discontinuities in time series (Figure 3)

Figure 3: Determination of discontinuities in time series



NUTS2	V00-01	V01-02	V02-03	V03-04	V04-05	V05-06	MEAN	STD
Corse	1.1	1.1	1.1	1.0	0.8	6.6	1.96	2.07
Guadeloupe (FR)	0.9	0.9	0.9	0.9	0.8	-1.7	0.43	0.94
Martinique (FR)	0.6	0.6	0.6	0.7	0.8	0.4	0.61	0.13
Guyane (FR)	4.4	4.5	4.5	3.2	2.0	4.8	3.92	0.99
Reunion (FR)	1.7	1.5	1.5	1.5	1.4	0.9	1.40	0.25
STD	0.94	0.89	0.70	0.66	0.68	0.84		

NUTS2	V00-01	V01-02	V02-03	V03-04	V04-05	V05-06
Corse	0.4	0.4	0.5	0.6	0.9	12.3
Guadeloupe (FR)	0.2	0.2	0.3	0.3	0.2	5.6
Martinique (FR)	0.0	0.0	0.0	0.1	0.3	0.6
Guyane (FR)	0.3	0.4	0.5	0.8	5.3	1.0
Reunion (FR)	0.3	0.1	0.1	0.0	0.0	1.3

NUTS2	2001	2002	2003	2004	2005	2006	2007
Corse	264	266.9	269.9	272.9	275.7	278	296.3
Guadeloupe (FR)	427.1	430.8	434.5	438.2	442	445.5	438
Martinique (FR)	384.6	387	389.4	391.8	394.5	397.5	398.9
Guyane (FR)	166.5	173.9	181.7	189.9	196	200	209.7
Reunion (FR)	723.4	735.4	746.6	757.8	768.9	779.3	786.2

1.1.4 Application to the quality control of ESPON datasets

The method has been applied to a dataset elaborated by RIATE concerning the total population of NUTS2 region (NUTS2006 version) between the years 2000 and 2006. This data was a compilation of various sources (Eurostat, National institutes) and estimations of missing values. We have selected in Figure 4 the regions for which at less one discontinuity (H>2) has been discovered which is represented by a red line between two years. The major discontinuities (H>4) are represented with a thick line.

Figure 4: Extraction of suspect data in the datasets of population 2000-2006

NUTS2_	NAME	2000	2001	2002	2003	2004	2005	2006
bg31	Severozapaden	1074.2	1032.1	1016	999.5	982.9	966.3	950.8
bg32	Severen tsentralen	1043.9	991.4	980.7	971.8	962.9	954.1	945.3
bg34	Yugoiztochen	1206.5	1166	1158.8	1150.8	1143.3	1137.3	1132.3
bg41	Yugozapaden	2142.9	2097	2103.1	2107.1	2112.4	2116.8	2117.8
bg42	Yuzhen tsentralen	1680.7	1605.1	1595	1585.3	1575.8	1566.1	1557.6
ch01	Région lémanique	1295.3	1310.3	1325.5	1339.6	1339.6	1369.5	1383.1
ch04	Zürich	1201.2	1218.3	1235.6	1245.6	1245.6	1267.2	1278.3
ch06	Zentralschweiz	676.7	684.4	692.2	696.7	696.7	706.1	711.2
ch07	Ticino	310.2	311.7	313.2	316.3	316.3	321.1	323.6
cz01	Praha	1183.9	1164.7	1158.8	1163.8	1168.1	1176.1	1184.9
dee0	Sachsen-Anhalt	2633	2598.4	2564.8	2535.9	2508.7	2482.1	2427.1
fr83	Corse	264	266.9	269.9	272.9	275.7	278	296.3
fr91	Guadeloupe (FR)	427.1	430.8	434.5	438.2	442	445.5	438
fr93	Guyane (FR)	166.5	173.9	181.7	189.9	196	200	209.7
is00	Iceland	281	285	288	290	292	296.7	303.8
li00	Liechtenstein	33	33	34	34	34	34.8	35
nl23	Flevoland	323.1	335.3	346.7	355.8	362.9	368.3	372.5
nl31	Utrecht	1112.9	1123.7	1146.1	1157.2	1166.8	1175.7	1185.3
ro11	Nord-Vest	2847.2	2839.1	2756.5	2746.8	2743	2735.9	2729.2
ro12	Centru	2643.3	2640.2	2549.8	2545.9	2538.5	2533.9	2529.3
ro21	Nord-Est	3825.7	3835.5	3745.1	3744.6	3739.2	3735.2	3731.4
ro22	Sud-Est	2935.6	2935.2	2867.7	2859.2	2852.5	2846.8	2839
ro31	Sud - Muntenia	3469.3	3462.6	3376.1	3359.4	3344.2	3329.8	3313.1
ro32	Bucuresti - Ilfov	2279.3	2268.9	2211	2208.2	2209	2212.7	2223.9
ro41	Sud-Vest Oltenia	2401.5	2397.2	2342.2	2330.5	2319.5	2307.9	2293.8
ro42	Vest	2041.2	2032.2	1954.8	1947.3	1939.1	1932.1	1927.9
sk01	Bratislavský kraj	617.2	599.1	599	599.8	600.5	602.4	605.2
uki1	Inner London	2722.4	2853.9	2886.9	2908	2921.7	2942.1	2972.6
uki2	Outer London	4382	4459.3	4476.3	4486	4493.9	4511.3	4536.8

Generally speaking, the method appears to be efficient and all major discontinuities are related to problems in the data collections realized by Eurostat.

in the case of Bulgarian regions (discontinuity between 2000 and 2001) or the Romanian regions (discontinuity between 2001 and 2002) where systematic errors are probably due to the combination of estimation and census data that are not adjusted.

In the case of Switzerland, there is obviously a mistake for the year 2003 and 2004 where the figures of population are the same for all regions. It is good to see that the method had been able to capture this mistake.

In the case of Liechtenstein, the problem is related to the fact that values are rounded for a region with very small population. For small unit, it is necessary to introduce more precision.

Some isolated regions are also characterized by problems in France, CZ, SK, DE.

The only case of "false positive" appears to be Flevoland

1.2 Annexe 2: Some concrete problems and perspectives raised by the integration of datasets into the ESPON Database

Good news

- At short term, it is possible to use a kind of combination of NUTS2 and NUTS3 (called "NUTS23 B") which allows to transfer the almost data from NUTS 2003 version to the NUTS 2006 one.
- In most of the cases, it would be possible to check the data from ESPON with updating official sources (Eurostat, national offices)
- The errors generated in a dataset are generally recurring. With simple (but multiple) checks, they are easily pointed.

Bad news

- Some mistakes in the Excel sheet of ESPON projects are predictable in the future, because of the long and difficult process of indicator construction.
- Check manually these mistakes take a lot of time. It will not be the job of ESPON DB project. It will be difficult to make by ESPON projects also.

In the next steps of ESPON projects, some data will be developed by these projects and have to be integrated in the future ESPON database. The building of this kind of table is much risked as at each step of the long process of the establishing of the indicator a mistake can potentially appearing (downloading correctly the data, combining raw data and indicators; different sources; different level of NUTS in some cases). An error in the compilation of data can have a considerable impact on the results generated.

The aim of this paper is to show concretely what kind of problem can appear in this kind of table (here an Excel sheet). In this example, the data used by the project "Territorial dynamics in Europe: trends in population development"[20] for mapping has been compared by the raw data from Eurostat.

The idea is the following: Considering **the basis of an official source** of data, is it possible to obtain the **same results** with applying the **same methodology** as indicated in the metadata of a data file?

In deed, the operation can be compared as an operation of verification and validation of data.

[20] ESPON Territorial Observation No. 1, November 2008

In order to make this operation the most relevant and reproducible as possible, a clear protocol of experimentation has to be defined. The following is proposed for the analysis:

1- Identification of raw data and regional level used for the analysis.

2- Comparison of the value between the raw data from an official source (Eurostat) and the latter figuring in the Excel sheet from ESPON project.

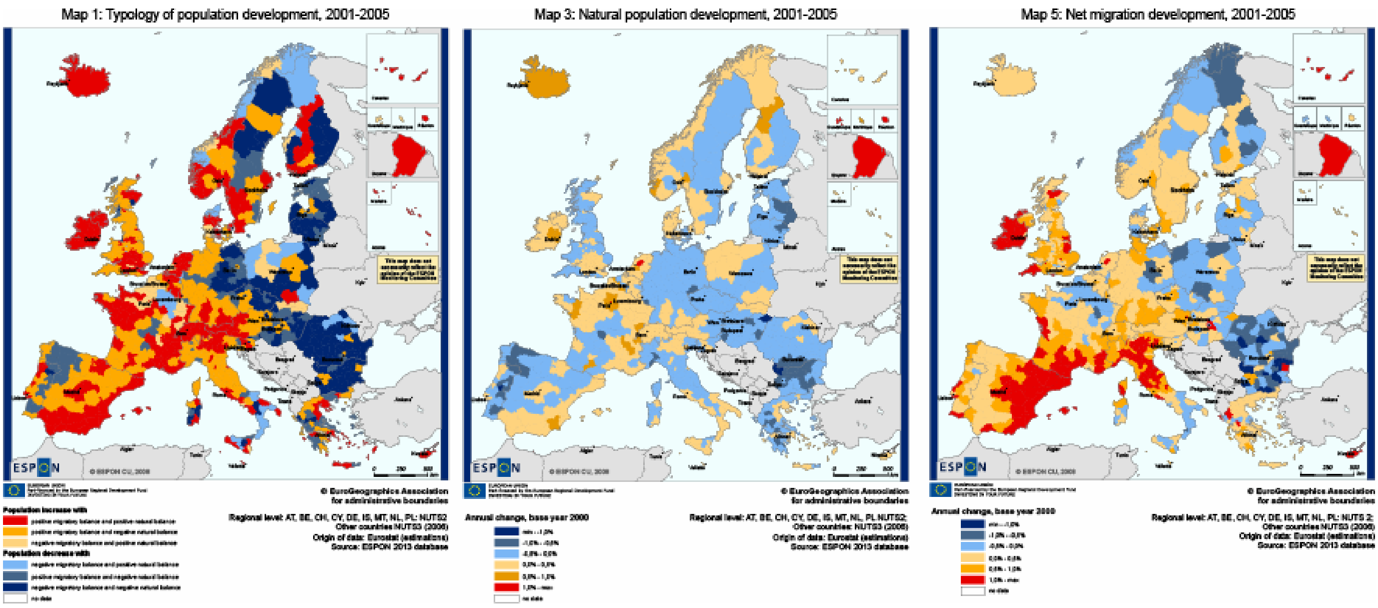
3- Identification of the methodology used to transform raw data in indicator and apply the same methodology.

1.2.1 Identification of raw data and regional level used for the analysis

The identification of the raw data used to establish the indicator is fundamental to begin the analysis. In the example taken (figure 1), these maps are created on the basis of three raw data: the total population (from 2000 to 2005), the number of births (from 2001 to 2005) and the number of deaths (from 2001 to 2005).

It would be interesting to **distinguish** systematically in metadata which column refers to a **raw data** (basic data) and which column refers to **indicator** (composite data) in order to make possible the re-building of an indicator.

Figure 5 – Results expected (maps of the report)



Map 1: Typology of population development, 2001-2005

Map 3: Natural population development, 2001-2005

Map 5: Net migration development, 2001-2005

ESPON
© ESPON CU, 2008
© EuroGeographics Association for administrative boundaries
Population increase with
positive migratory balance and positive natural balance
positive migratory balance and negative natural balance
negative migratory balance and positive natural balance
Population decrease with
negative migratory balance and positive natural balance
positive migratory balance and negative natural balance
negative migratory balance and negative natural balance
no data
Regional level: AT, BE, CH, CY, DE, IS, MT, NL, PL: NUTS2
Other countries NUTS3 (2006)
Origin of data: Eurostat (estimations)
Source: ESPON 2013 database
Annual change, base year 2000
min - -1,0%
-1,0% - -0,5%
-0,5% - 0,0%
0,0% - 0,5%
0,5% - 1,0%
1,0% - max
no data
Regional level: AT, BE, CH, CY, DE, IS, MT, NL, PL NUTS2;
Other countries: NUTS3 (2006)
Origin of data: Eurostat (estimations)
Source: ESPON 2013 database
Annual change, base year 2000
min - -1,0%
-1,0% - -0,5%
-0,5% - 0,0%
0,0% - 0,5%
0,5% - 1,0%
1,0% - max
no data
Regional level: AT, BE, CH, CY, DE, IS, MT, NL, PL: NUTS 2;
Other countries NUTS3 (2006)
Origin of data: Eurostat (estimations)
Source: ESPON 2013 database

At this step of the analysis it is also important to identify the regional level: here it is a combination of NUTS2/3 in the 2006 version. This **identification** can be made by the analysis of the **code** of the NUTS in the data file and/or by the analysis of the **geometries** on the map.

It is the occasion to precise what is behind this NUTS version and discuss about its interest. As compared to the NUTS2/3 (called "NUTS23 A") developed in the previous ESPON projects (ESPON 3.1., ESPON 3.4.1. …), the latter proposed by the project "territorial dynamics in Europe" is a new one (called "NUTS23 B", figure 2): in the previous version, 8 countries was defined in NUTS 2 (Belgium, Netherlands, Germany, United Kingdom, Switzerland, Austria, Portugal and Greece), the rest in NUTS 3. In the version proposed by this project, the NUTS2 is defined as level of reference for 7 countries (Belgium, Netherlands, Germany, Austria, Switzerland, Poland and Iceland).

The figure 3 illustrates the interest of the NUTS23 B. If we take separately the geometries from 2003 to 2006 at NUTS2 and NUTS3, one can notice that the new division does not imply so much modification that one can imagine. There is still a problem in Denmark but most of the cases the **transposition of data from NUTS 2003 version to NUTS 2006 version is conceivable.**

In deed, if the use of NUTS 2/3 A was relevant in theoretical point of view (cf. MAUP), the NUTS23 B is precious in concrete terms, to transpose data from NUTS 2003 to 2006.

Figure 6 – Comparison between the previous NUTS 2/3 (version = 2003) and the new one (version = 2006)

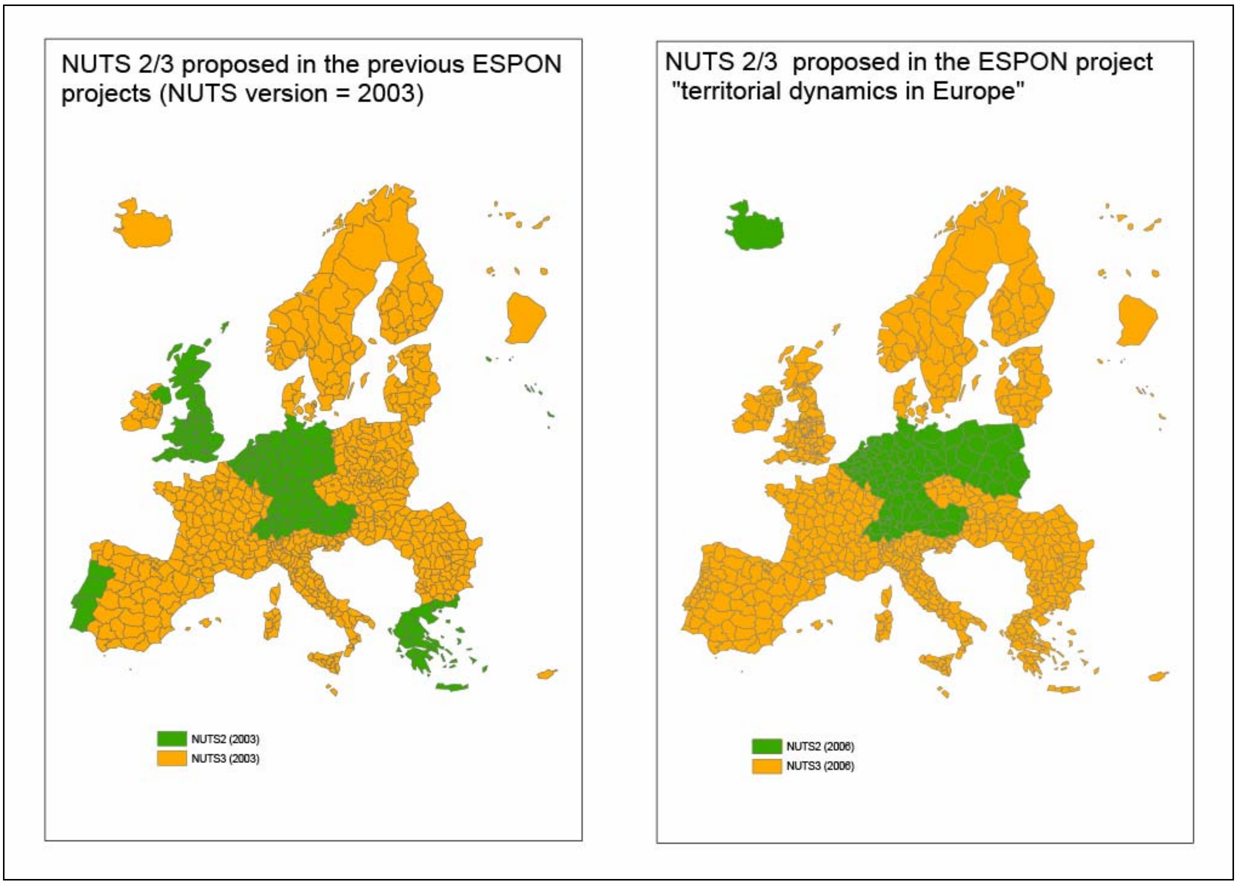
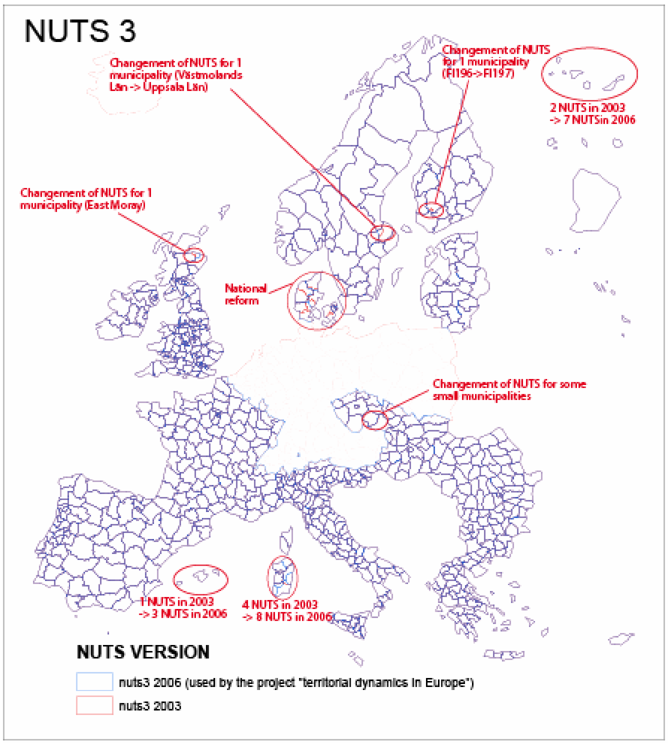
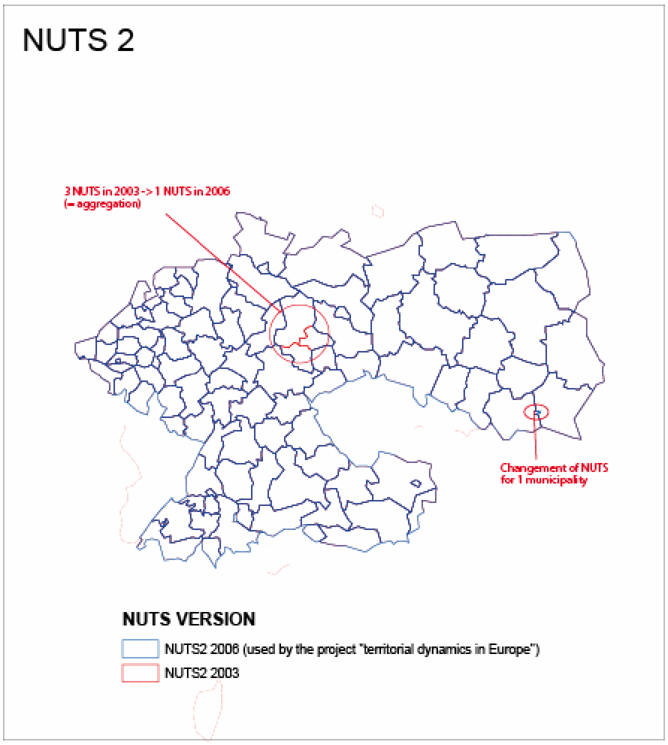

NUTS 2/3 proposed in the previous ESPON projects (NUTS version = 2003)
NUTS2 (2003)
NUTS3 (2003)
NUTS 2/3 proposed in the ESPON project "territorial dynamics in Europe"
NUTS2 (2006)
NUTS3 (2006)

Figure 7 – Modifications of NUTS from 2003 to 2006 in the new NUTS 23 at NUTS 2 and NUTS 3 level


NUTS 3
Changement of NUTS for 1 municipality (Västmolands Län -> Uppsala Län)
Changement of NUTS for 1 municipality (FI196->FI197)
2 NUTS in 2003 -> 7 NUTSin 2006
Changement of NUTS for 1 municipality (East Moray)
National reform
Changement of NUTS for some small municipalities
1 NUTS in 2003 -> 3 NUTS in 2006
4 NUTS in 2003 -> 8 NUTS in 2006
NUTS VERSION
nuts3 2006 (used by the project "territorial dynamics in Europe")
nuts3 2003


NUTS 2
3 NUTS in 2003 -> 1 NUTS in 2006 (= aggregation)
Changement of NUTS for 1 municipality
NUTS VERSION
NUTS2 2006 (used by the project "territorial dynamics in Europe")
NUTS2 2003

1.2.2 Comparison of the value between the raw data from an official source (Eurostat) and the latter figuring in the Excel sheet from ESPON project.

After taking into an account the raw data and the NUTS level used by the ESPON project for its analysis, it is possible to begin the process of comparison - ascertaining and validation of raw data - with data downloaded from official source. Theoretically, no differences between these two sources would have to be noticed. But the situation is more complex than expected.

The table below (figure 4) summarizes the different case of figure picked out by the comparison between data of population, death and birth when we take as a source Eurostat and the ESPON one:

Figure 8 – Typology characterising the coherence between Eurostat data and ESPON project data

→ No problem

🟩	Eurostat data = ESPON data
🟩	No data on Eurostat. The ESPON project has implemented a procedure of estimation validated and reproducible (implemented in the cell of the Excel sheet for example).
⬜	No data both on Eurostat and ESPON file

→ Problem which can be solved

🟫	Problem known on Eurostat data and corrected by ESPON project and specified in metadata (inversion of NUTS in Greece in this case).
🟧	Problem on ESPON data which can be corrected with the Eurostat one.
▨	No data on ESPON file. However, the data exists in Eurostat (update not known?)

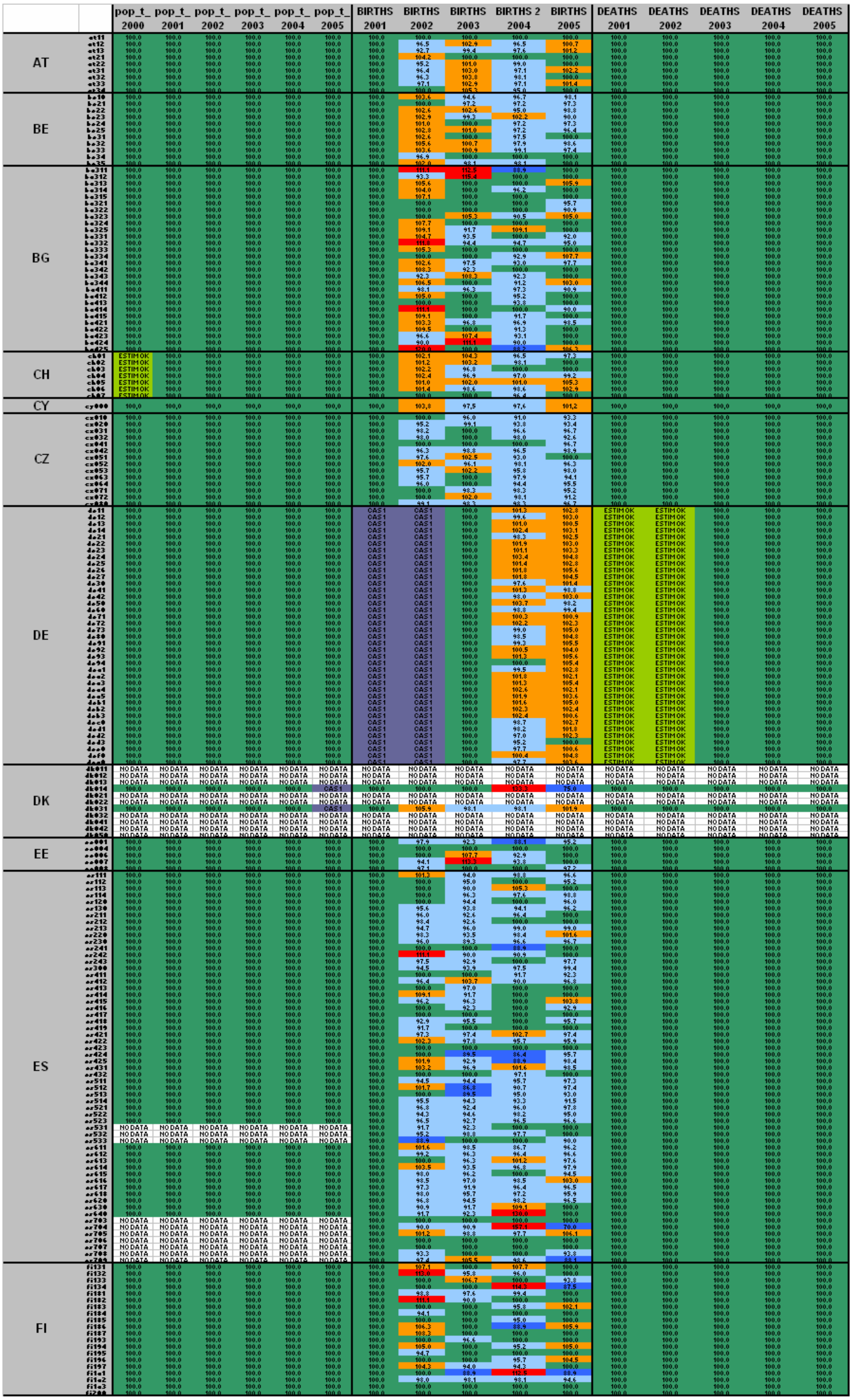
→ Problem which can not be solved without making more explications or specifications

🟪	No data on Eurostat file. However, the data exists in the ESPON one. The problem is that there is no specification about the source and/or the method of estimation used for this kind of situation .
🟥	Eurostat data ≠ ESPON data → no explication As compare to Eurostat official data, the ESPON one is overestimated (over 10 % of the Eurostat value)
🟧	Eurostat data ≠ ESPON data → no explication As compare to Eurostat official data, the ESPON one is overestimated (over 0-10 % of the Eurostat value)
🟦	Eurostat data ≠ ESPON data → no explication As compare to Eurostat official data, the ESPON one is overestimated (under 0-10 % of the Eurostat value)
🟦	Eurostat data ≠ ESPON data → no explication As compare to Eurostat official data, the ESPON one is overestimated (under 10 % of the Eurostat value)

Bellow (figure 5), the table summarizes the different scenarios encounter in this comparison. Most of the cases, the Eurostat data are the same that the latter used in ESPON project (dark green). If estimations are made, they are generally well referenced (green). We can expect that it will be the case of figure in most of the ESPON projects.

However the table shows some **incoherencies** (namely for the data of births). The main problem is that sources and/or estimations taken in ESPON data file are not referenced in metadata. As a consequence, it is impossible to come back to the data of origin and understand the choices made.

Figure 9 – Situation of this typology in the example used for the comparison



		pop_t_ 2000	pop_t_ 2001	pop_t_ 2002	pop_t_ 2003	pop_t_ 2004	pop_t_ 2005	BIRTHS 2001	BIRTHS 2002	BIRTHS 2003	BIRTHS 2 2004	BIRTHS 2005	DEATHS 2001	DEATHS 2002	DEATHS 2003	DEATHS 2004	DEATHS 2005
AT	at11	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	at12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.5	102.9	96.5	100.7	100.0	100.0	100.0	100.0	100.0
	at13	100.0	100.0	100.0	100.0	100.0	100.0	100.0	92.7	99.4	97.6	101.2	100.0	100.0	100.0	100.0	100.0
	at21	100.0	100.0	100.0	100.0	100.0	100.0	100.0	104.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	at22	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.2	101.0	99.0	100.0	100.0	100.0	100.0	100.0	100.0
	at31	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.4	103.0	97.1	102.2	100.0	100.0	100.0	100.0	100.0
	at32	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.3	103.8	98.1	100.0	100.0	100.0	100.0	100.0	100.0
	at33	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.1	102.9	97.1	101.4	100.0	100.0	100.0	100.0	100.0
	at34	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	105.3	95.0	100.0	100.0	100.0	100.0	100.0	100.0
BE	be10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	103.6	94.6	96.7	98.1	100.0	100.0	100.0	100.0	100.0
	be21	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.2	97.2	97.3	100.0	100.0	100.0	100.0	100.0
	be22	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.6	102.6	95.0	98.8	100.0	100.0	100.0	100.0	100.0
	be23	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.9	99.3	102.2	90.0	100.0	100.0	100.0	100.0	100.0
	be24	100.0	100.0	100.0	100.0	100.0	100.0	100.0	101.0	100.0	97.2	97.3	100.0	100.0	100.0	100.0	100.0
	be25	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.8	101.0	97.2	96.4	100.0	100.0	100.0	100.0	100.0
	be31	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.6	100.0	97.5	100.0	100.0	100.0	100.0	100.0	100.0
	be32	100.0	100.0	100.0	100.0	100.0	100.0	100.0	105.6	100.7	97.9	98.6	100.0	100.0	100.0	100.0	100.0
	be33	100.0	100.0	100.0	100.0	100.0	100.0	100.0	103.6	100.9	99.1	97.4	100.0	100.0	100.0	100.0	100.0
	be34	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	be35	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.0	98.1	98.1	100.0	100.0	100.0	100.0	100.0	100.0
BG	bg311	100.0	100.0	100.0	100.0	100.0	100.0	100.0	111.1	112.5	88.9	100.0	100.0	100.0	100.0	100.0	100.0
	bg312	100.0	100.0	100.0	100.0	100.0	100.0	100.0	93.3	115.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	bg313	100.0	100.0	100.0	100.0	100.0	100.0	100.0	105.6	100.0	100.0	105.9	100.0	100.0	100.0	100.0	100.0
	bg314	100.0	100.0	100.0	100.0	100.0	100.0	100.0	104.0	100.0	96.2	100.0	100.0	100.0	100.0	100.0	100.0
	bg315	100.0	100.0	100.0	100.0	100.0	100.0	100.0	107.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	bg321	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.7	100.0	100.0	100.0	100.0	100.0
	bg322	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	90.9	100.0	100.0	100.0	100.0	100.0
	bg323	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	105.3	90.5	105.0	100.0	100.0	100.0	100.0	100.0
	bg324	100.0	100.0	100.0	100.0	100.0	100.0	100.0	107.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	bg325	100.0	100.0	100.0	100.0	100.0	100.0	100.0	109.1	91.7	109.1	100.0	100.0	100.0	100.0	100.0	100.0
	bg331	100.0	100.0	100.0	100.0	100.0	100.0	100.0	104.7	93.5	100.0	92.0	100.0	100.0	100.0	100.0	100.0
	bg332	100.0	100.0	100.0	100.0	100.0	100.0	100.0	111.8	94.4	94.7	95.0	100.0	100.0	100.0	100.0	100.0
	bg333	100.0	100.0	100.0	100.0	100.0	100.0	100.0	105.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	bg334	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	92.9	107.7	100.0	100.0	100.0	100.0	100.0
	bg341	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.6	97.5	93.0	97.7	100.0	100.0	100.0	100.0	100.0
	bg342	100.0	100.0	100.0	100.0	100.0	100.0	100.0	108.3	92.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	bg343	100.0	100.0	100.0	100.0	100.0	100.0	100.0	92.3	108.3	92.3	100.0	100.0	100.0	100.0	100.0	100.0
	bg344	100.0	100.0	100.0	100.0	100.0	100.0	100.0	106.5	100.0	91.2	103.0	100.0	100.0	100.0	100.0	100.0
	bg411	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.1	96.3	97.3	90.9	100.0	100.0	100.0	100.0	100.0
	bg412	100.0	100.0	100.0	100.0	100.0	100.0	100.0	105.0	100.0	95.2	100.0	100.0	100.0	100.0	100.0	100.0
	bg413	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	93.8	100.0	100.0	100.0	100.0	100.0	100.0
	bg414	100.0	100.0	100.0	100.0	100.0	100.0	100.0	111.1	100.0	100.0	90.0	100.0	100.0	100.0	100.0	100.0
	bg415	100.0	100.0	100.0	100.0	100.0	100.0	100.0	109.1	100.0	91.7	100.0	100.0	100.0	100.0	100.0	100.0
	bg421	100.0	100.0	100.0	100.0	100.0	100.0	100.0	103.3	96.8	96.9	98.5	100.0	100.0	100.0	100.0	100.0
	bg422	100.0	100.0	100.0	100.0	100.0	100.0	100.0	109.5	100.0	91.3	100.0	100.0	100.0	100.0	100.0	100.0
	bg423	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.6	107.4	93.1	100.0	100.0	100.0	100.0	100.0	100.0
	bg424	100.0	100.0	100.0	100.0	100.0	100.0	100.0	90.0	111.1	90.0	100.0	100.0	100.0	100.0	100.0	100.0
	bg425	100.0	100.0	100.0	100.0	100.0	100.0	100.0	120.0	100.0	88.2	106.3	100.0	100.0	100.0	100.0	100.0
CH	ch01	ESTIM OK	100.0	100.0	100.0	100.0	100.0	100.0	102.1	104.3	96.5	97.3	100.0	100.0	100.0	100.0	100.0
	ch02	ESTIM OK	100.0	100.0	100.0	100.0	100.0	100.0	101.2	103.2	98.1	100.0	100.0	100.0	100.0	100.0	100.0
	ch03	ESTIM OK	100.0	100.0	100.0	100.0	100.0	100.0	102.2	96.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	ch04	ESTIM OK	100.0	100.0	100.0	100.0	100.0	100.0	102.4	96.9	97.0	99.2	100.0	100.0	100.0	100.0	100.0
	ch05	ESTIM OK	100.0	100.0	100.0	100.0	100.0	100.0	101.0	102.0	101.0	105.3	100.0	100.0	100.0	100.0	100.0
	ch06	ESTIM OK	100.0	100.0	100.0	100.0	100.0	100.0	101.4	98.6	98.6	102.9	100.0	100.0	100.0	100.0	100.0
	ch07	ESTIM OK	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.4	100.0	100.0	100.0	100.0	100.0	100.0
CY	cy000	100,0	100,0	100,0	100,0	100,0	100,0	100,0	103,8	97,5	97,6	101,2	100,0	100,0	100,0	100,0	100,0
CZ	cz010	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.0	91.0	93.3	100.0	100.0	100.0	100.0	100.0
	cz020	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.2	99.1	93.8	93.4	100.0	100.0	100.0	100.0	100.0
	cz031	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.2	100.0	96.6	96.7	100.0	100.0	100.0	100.0	100.0
	cz032	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.0	100.0	98.0	92.6	100.0	100.0	100.0	100.0	100.0
	cz041	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.7	100.0	100.0	100.0	100.0	100.0
	cz042	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.3	98.8	96.5	98.9	100.0	100.0	100.0	100.0	100.0
	cz051	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.6	102.5	93.0	100.0	100.0	100.0	100.0	100.0	100.0
	cz052	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.0	96.1	98.1	96.3	100.0	100.0	100.0	100.0	100.0
	cz053	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.7	102.2	95.8	98.0	100.0	100.0	100.0	100.0	100.0
	cz063	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.7	100.0	97.9	94.1	100.0	100.0	100.0	100.0	100.0
	cz064	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.0	100.0	94.4	95.5	100.0	100.0	100.0	100.0	100.0
	cz071	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.3	98.3	95.2	100.0	100.0	100.0	100.0	100.0
	cz072	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.0	98.1	91.2	100.0	100.0	100.0	100.0	100.0
	cz080	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.1	98.3	98.3	96.7	100.0	100.0	100.0	100.0	100.0
DE	de11	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.3	102.8	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de12	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	99.6	103.0	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de13	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.0	100.5	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de14	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	102.4	103.1	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de21	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	98.3	102.5	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de22	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.9	103.0	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de23	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.1	103.3	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de24	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	103.4	104.8	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de25	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.4	102.8	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de26	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.8	105.6	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de27	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.8	104.5	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de30	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	97.6	101.4	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de41	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.3	98.8	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de42	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	98.0	103.0	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de50	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	103.7	98.2	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de60	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	98.8	99.4	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de71	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	100.3	100.9	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de72	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	102.2	102.3	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de73	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	99.0	105.0	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de80	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	98.5	104.8	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de91	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	99.3	105.5	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de92	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	100.5	104.0	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de93	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.3	105.6	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	de94	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	100.0	105.4	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	dea1	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	99.5	102.8	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	dea2	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.8	102.1	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	dea3	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.3	105.4	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	dea4	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	102.6	102.1	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	dea5	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.9	103.6	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	deb1	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	101.6	105.0	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	deb2	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	102.3	102.4	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	deb3	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	102.4	100.6	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	dec0	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	98.7	102.7	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	ded1	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	98.2	101.8	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	ded2	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	97.0	102.3	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	ded3	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	95.2	100.0	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	dee0	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	97.7	100.6	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	def0	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	100.4	104.8	ESTIM OK	ESTIM OK	100.0	100.0	100.0
	deg0	100.0	100.0	100.0	100.0	100.0	100.0	CAS1	CAS1	100.0	97.7	103.6	ESTIM OK	ESTIM OK	100.0	100.0	100.0
DK	dk011	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
	dk012	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
	dk013	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
	dk014	100.0	100.0	100.0	100.0	100.0	CAS1	100.0	100.0	100.0	133.3	75.0	100.0	100.0	100.0	100.0	100.0
	dk021	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
	dk022	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
	dk031	100.0	100.0	100.0	100.0	100.0	CAS1	100.0	105.9	98.1	98.1	101.9	100.0	100.0	100.0	100.0	100.0
	dk032	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
	dk041	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
	dk042	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
	dk050	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA
EE	ee001	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.9	92.3	88.1	95.2	100.0	100.0	100.0	100.0	100.0
	ee004	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	ee006	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	107.7	92.9	100.0	100.0	100.0	100.0	100.0	100.0
	ee007	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.1	113.3	93.8	100.0	100.0	100.0	100.0	100.0	100.0
	ee008	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.1	100.0	100.0	97.2	100.0	100.0	100.0	100.0	100.0
ES	es111	100.0	100.0	100.0	100.0	100.0	100.0	100.0	101.3	94.0	98.8	96.6	100.0	100.0	100.0	100.0	100.0
	es112	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.0	100.0	95.2	100.0	100.0	100.0	100.0	100.0
	es113	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	90.0	105.3	100.0	100.0	100.0	100.0	100.0	100.0
	es114	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.3	97.6	98.8	100.0	100.0	100.0	100.0	100.0
	es120	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.4	100.0	96.0	100.0	100.0	100.0	100.0	100.0
	es130	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.6	93.8	94.1	96.2	100.0	100.0	100.0	100.0	100.0
	es211	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.0	92.6	96.4	100.0	100.0	100.0	100.0	100.0	100.0
	es212	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.4	92.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es213	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.7	96.0	99.0	99.0	100.0	100.0	100.0	100.0	100.0
	es220	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.3	93.5	98.4	101.6	100.0	100.0	100.0	100.0	100.0
	es230	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.0	89.3	96.6	96.7	100.0	100.0	100.0	100.0	100.0
	es241	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	88.9	100.0	100.0	100.0	100.0	100.0	100.0
	es242	100.0	100.0	100.0	100.0	100.0	100.0	100.0	111.1	90.0	90.9	100.0	100.0	100.0	100.0	100.0	100.0
	es243	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.5	92.9	100.0	97.7	100.0	100.0	100.0	100.0	100.0
	es300	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.5	93.9	97.5	99.4	100.0	100.0	100.0	100.0	100.0
	es411	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	91.7	92.3	100.0	100.0	100.0	100.0	100.0
	es412	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.4	103.7	90.0	96.8	100.0	100.0	100.0	100.0	100.0
	es413	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es414	100.0	100.0	100.0	100.0	100.0	100.0	100.0	109.1	91.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es415	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.2	96.3	100.0	103.8	100.0	100.0	100.0	100.0	100.0
	es416	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	92.3	100.0	92.9	100.0	100.0	100.0	100.0	100.0
	es417	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es418	100.0	100.0	100.0	100.0	100.0	100.0	100.0	92.9	95.5	100.0	95.7	100.0	100.0	100.0	100.0	100.0
	es419	100.0	100.0	100.0	100.0	100.0	100.0	100.0	91.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es421	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.3	97.4	102.7	97.4	100.0	100.0	100.0	100.0	100.0
	es422	100.0	100.0	100.0	100.0	100.0	100.0	100.0	102.3	97.8	95.7	95.9	100.0	100.0	100.0	100.0	100.0
	es423	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es424	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	89.5	86.4	95.7	100.0	100.0	100.0	100.0	100.0
	es425	100.0	100.0	100.0	100.0	100.0	100.0	100.0	101.9	92.9	88.9	98.4	100.0	100.0	100.0	100.0	100.0
	es431	100.0	100.0	100.0	100.0	100.0	100.0	100.0	103.2	96.9	101.6	98.5	100.0	100.0	100.0	100.0	100.0
	es432	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.1	100.0	100.0	100.0	100.0	100.0	100.0
	es511	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.5	94.4	95.7	97.3	100.0	100.0	100.0	100.0	100.0
	es512	100.0	100.0	100.0	100.0	100.0	100.0	100.0	101.7	86.8	90.7	97.4	100.0	100.0	100.0	100.0	100.0
	es513	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	89.5	95.0	93.0	100.0	100.0	100.0	100.0	100.0
	es514	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.5	94.3	93.3	91.5	100.0	100.0	100.0	100.0	100.0
	es521	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.8	92.4	96.0	97.8	100.0	100.0	100.0	100.0	100.0
	es522	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.3	94.6	98.2	95.0	100.0	100.0	100.0	100.0	100.0
	es523	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.5	92.7	96.5	96.6	100.0	100.0	100.0	100.0	100.0
	es531	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	91.7	92.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es532	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	95.2	98.8	97.7	100.0	100.0	100.0	100.0	100.0	100.0
	es533	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	88.9	100.0	100.0	90.0	100.0	100.0	100.0	100.0	100.0
	es611	100.0	100.0	100.0	100.0	100.0	100.0	100.0	101.6	98.5	86.7	96.2	100.0	100.0	100.0	100.0	100.0
	es612	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.2	96.3	96.4	96.6	100.0	100.0	100.0	100.0	100.0
	es613	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.3	101.2	97.6	100.0	100.0	100.0	100.0	100.0
	es614	100.0	100.0	100.0	100.0	100.0	100.0	100.0	103.5	93.5	96.8	97.9	100.0	100.0	100.0	100.0	100.0
	es615	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.0	96.2	100.0	94.5	100.0	100.0	100.0	100.0	100.0
	es616	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.5	97.0	98.5	103.0	100.0	100.0	100.0	100.0	100.0
	es617	100.0	100.0	100.0	100.0	100.0	100.0	100.0	97.3	91.9	96.4	96.5	100.0	100.0	100.0	100.0	100.0
	es618	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.0	95.7	97.2	95.9	100.0	100.0	100.0	100.0	100.0
	es620	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.8	94.5	98.2	96.5	100.0	100.0	100.0	100.0	100.0
	es630	100.0	100.0	100.0	100.0	100.0	100.0	100.0	90.9	91.7	109.1	100.0	100.0	100.0	100.0	100.0	100.0
	es640	100.0	100.0	100.0	100.0	100.0	100.0	100.0	91.7	92.3	130.0	100.0	100.0	100.0	100.0	100.0	100.0
	es703	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es704	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	90.0	90.9	157.1	70.0	100.0	100.0	100.0	100.0	100.0
	es705	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	101.2	98.8	97.7	106.1	100.0	100.0	100.0	100.0	100.0
	es706	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es707	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	es708	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	93.3	100.0	100.0	93.8	100.0	100.0	100.0	100.0	100.0
	es709	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	100.0	97.4	105.5	98.6	88.1	100.0	100.0	100.0	100.0	100.0
FI	fi131	100.0	100.0	100.0	100.0	100.0	100.0	100.0	107.1	100.0	107.7	100.0	100.0	100.0	100.0	100.0	100.0
	fi132	100.0	100.0	100.0	100.0	100.0	100.0	100.0	113.0	95.8	96.0	100.0	100.0	100.0	100.0	100.0	100.0
	fi133	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	106.7	100.0	93.8	100.0	100.0	100.0	100.0	100.0
	fi134	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	114.3	87.5	100.0	100.0	100.0	100.0	100.0
	fi181	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.8	97.6	99.4	100.0	100.0	100.0	100.0	100.0	100.0
	fi182	100.0	100.0	100.0	100.0	100.0	100.0	100.0	111.1	90.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	fi183	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.8	102.1	100.0	100.0	100.0	100.0	100.0
	fi184	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	fi185	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.0	100.0	100.0	100.0	100.0	100.0	100.0
	fi186	100.0	100.0	100.0	100.0	100.0	100.0	100.0	106.3	100.0	88.9	105.9	100.0	100.0	100.0	100.0	100.0
	fi187	100.0	100.0	100.0	100.0	100.0	100.0	100.0	108.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	fi193	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	96.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	fi194	100.0	100.0	100.0	100.0	100.0	100.0	100.0	105.0	100.0	95.2	105.0	100.0	100.0	100.0	100.0	100.0
	fi195	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	fi196	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.7	104.5	100.0	100.0	100.0	100.0	100.0
	fi197	100.0	100.0	100.0	100.0	100.0	100.0	100.0	104.3	94.0	94.3	100.0	100.0	100.0	100.0	100.0	100.0
	fi1a1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	88.9	112.5	88.9	100.0	100.0	100.0	100.0	100.0
	fi1a2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.0	98.1	98.1	94.6	100.0	100.0	100.0	100.0	100.0
	fi1a3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	fi200	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

1.2.3 Identification of the methodology used to transform raw data in indicator and apply the same methodology.

After validating raw data by different checks, it is important to understand the methodology used to transform raw data in indicator. It allows anyone to make sure that the construction of the indicator is **reproducible** and among others, to **recalculate** the indicator in other time span if necessary. Indeed, the method used to transform raw data into indicator has to be **clearly explained**.

In the case of the example chosen, the methods of calculation used to obtain the indicator are not described in the metadata but directly in the cells of the Excel sheet. As a consequence, it is not a problem to re-calculate the value of the indicators (annual growth rate of population, annual natural and migratory development) as bellow (figure 6).

Figure 10 – Methods used to transform raw data in core indicator

- **<u>Calculation of natural population development (n) and net migration development (m)</u>**

Theoretical formulation

$$n_{(2000\text{->}2005)} = 100 \text{ x } (\text{total naissances}_{(2001\text{->}2005)} - \text{total décès}_{(2001\text{->}2005)}) / \text{population totale}_{(2000)}$$

$$m_{(2000\text{->}2005)} = \left(100 \text{ x } (\text{population totale}_{(2005)} - \text{population totale}_{(2000)}) / \text{population totale}_{2000)}\right) - n_{(2000\text{->}2005)}$$

Excel formulation

=100*(T5-AB5)/C5

=(100*(H5-C5)/C5)-AE5

- **<u>Calculation of annual growth rate of the population (a)</u>**

Theoretical formulation

$$a_{(t\text{->}t+1)} = 100 \text{ x } \left((P_{t+1} - P_t)^{(1/t;t+1)} - 1)\right) \quad \text{<->} \quad a_{(2000\text{->}2005)} = 100 \text{ x } \left((P_{2005} - P_{2000})^{(1/5)} - 1)\right)$$

$$\text{<->}$$

$$a_{(t\text{->}t+1)} = 100 \text{ x } \left(e^{\ln((P_{t+1}/P_t) / (t;t+1))} - 1\right) \quad \text{<->} \quad a_{(2000\text{->}2005)} = 100 \text{ x } \left(e^{\ln((P_{2005}/P_{2000}) / 5)} - 1\right)$$

Excel formulation

=100*((H5/C5)^(1/5)-1)

=100*(EXP(LN(H5/C5)/5)-1)

- **Deduction of the annual natural population development (N) and the annual migratory population development (M) from (n), (m) and (a)**

Theoretical formulation

$$N_{(2000\text{->}2005)} = (n_{(2000\text{->}2005)} / (n_{(2000\text{->}2005)} + m_{(2000\text{->}2005)})) / a_{(2000\text{->}2005)}$$

$$M_{(2000\text{->}2005)} = (m_{(2000\text{->}2005)} / (n_{(2000\text{->}2005)} + m_{(2000\text{->}2005)})) / a_{(2000\text{->}2005)}$$

Excel formulation

=(AE5)/(AE5+AG5)*AK5

=(AG5)/(AG5+AE5)*AK5

Conclusion: How going further?

The objective of this demarche is not to apply the same methodology systematically to the data from ESPON projects. But in the case where there is a doubt on the value of a data; it would be interesting to make possible the comparison with data from official source and check the possible error in the ESPON data. Indeed, the creation of indicator is complex and needs different steps (downloading data, compilation of data creation of indicator). It is impossible to insure systematically the absence of error in a dataset. If a problem is detected in the treatment chain, it will necessarily affect the all data file.

This treatment has been done manually. Of course it takes a **lot of time** and is **not operational** in this form. On top of that the data file used for the example is characterised by an only source (Eurostat). How do we manage this kind of situation when the ESPON dataset compute different sources? This kind of figure would make the situation untreatable in this form.

That's why a priority for ESPON Database project is to transform this "manual error report" in "automatic error report".

Annexe 2 (challenge 2)

2.1 Regulation (EC) n° 1059/2003 of the European parliament and the council of 26 May 2003 on the establishment of a common classification of territorial units for statistics

REGULATION (EC) No 1059/2003 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Having regard to the Treaty establishing the European Community, and in particular Article 285 thereof,

Having regard to the proposal from the Commission ([1]),

Having regard to the opinion of the European Economic and Social Committee ([2]),

Having regard to the opinion of the Committee of the Regions ([3]),

Acting in accordance with the procedure laid down in Article 251 of the Treaty ([4]),

Whereas:

(1) Users of statistics express an increasing need for harmonisation in order to have comparable data across the European Union. In order to function, the internal market requires statistical standards applicable to the collection, transmission and publication of national and Community statistics so that all operators in the single market can be provided with comparable statistical data. In this context, classifications are an important tool for the collection, compilation and dissemination of comparable statistics.

(2) Regional statistics are a cornerstone of the European Statistical System. They are used for a wide range of purposes. For many years European regional statistics have been collected, compiled and disseminated on the basis of a common regional classification, called ‘Nomenclature of territorial units for statistics’ (hereinafter referred to as NUTS). It is now appropriate to fix this regional classification in a legal framework and to institute clear rules for future amendments of this classification. The NUTS classification should not preclude the existence of other subdivisions and classifications.

(3) Accordingly, all Member States' statistics transmitted to the Commission, which are broken down by territorial units, should use the NUTS classification, where applicable.

(4) In its analysis and dissemination, the Commission should use the NUTS classification for all statistics classified by territorial units, where applicable.

(5) Different levels are needed for regional statistics depending on the purpose of these statistics at national and European level. It is appropriate to have at least three hierarchical levels of detail in the European regional NUTS classification. Member States could have further levels of NUTS details, where they consider it necessary.

(6) Information on the current territorial composition of NUTS level 3 regions is necessary for the proper administration of the NUTS classification and should therefore be transmitted regularly to the Commission.

(7) Objective criteria for the definition of regions are necessary in order to ensure impartiality when regional statistics are compiled and used.

(8) Users of regional statistics need stability of the nomenclature over time. The NUTS classification should hence not be amended too frequently. This Regulation will ensure a greater stability of rules over time.

(9) Comparability of regional statistics requires that the regions be of a comparable size in terms of population. In order to achieve this goal, amendments of the NUTS classification should render the regional structure more homogeneous in terms of population size.

[1] OJ C 180 E, 26.6.2001, p. 108.
[2] OJ C 260, 17.9.2001, p. 57.
[3] OJ C 107, 3.5.2002, p. 54.
[4] Opinion of the European Parliament of 24 October 2001 (OJ C 112 E, 9.5.2002, p. 146), Council Common Position of 9 December 2002 (OJ C 32 E, 11.2.2003, p. 26) and Decision of the European Parliament of 8 April 2003 (not yet published in the Official Journal).

(10) The actual political, administrative and institutional situation must also be respected. Non-administrative units must reflect economic, social, historical, cultural, geographical or environmental circumstances.

(11) Reference should be made to the definition of the 'population' on which the classification is based.

(12) The NUTS classification is restricted to the economic territory of the Member States and does not provide complete coverage of the territory to which the Treaty establishing the European Community applies. Its use for Community purposes will therefore need to be assessed on a case-by-case basis. The economic territory of each country, as defined in Commission Decision 91/450/ EEC ([1]),also includes extraregio territory, made up of parts of the economic territory that cannot be attached to a certain region (air-space, territorial waters and the continental shelf, territorial enclaves, in particular embassies, consulates and military bases, and deposits of oil, natural gas, etc. in international waters, outside the continental shelf, worked by resident units). The NUTS classification must also provide the possibility of statistics for this extraregio territory.

(13) Amendments to the NUTS classification will require close consultations with the Member States.

(14) Since the objective of the proposed action, namely the harmonisation of regional statistics, cannot be sufficiently achieved by the Member States and can therefore be better achieved at Community level, the Community may adopt measures, in accordance with the principle of subsidiarity as set out in Article 5 of the Treaty. In accordance with the principle of proportionality, as set out in that Article, this Regulation does not go beyond what is necessary in order to achieve that objective.

(15) The NUTS classification laid down in this Regulation should replace the 'Nomenclature of territorial units for statistics (NUTS)' established to date by the Statistical Office of the European Communities in cooperation with the national statistical institutes. As a consequence, all references in Community acts to the 'Nomenclature of territorial units for statistics (NUTS)' should now be understood as referring to the NUTS classification laid down in this Regulation.

(16) Council Regulation (EC) No 322/97 of 17 February 1997 on Community Statistics ([2]) constitutes the reference framework for the provisions of this Regulation.

(17) The measures necessary for the implementation of this Regulation should be adopted in accordance with Council Decision 1999/468/EC of 28 June 1999 laying down the procedures for the exercise of implementing powers conferred on the Commission ([3]).

(18) The Statistical Programme Committee established by Council Decision 89/382/EEC, Euratom ([4]) has been consulted in accordance with Article 3 thereof,

HAVE ADOPTED THIS REGULATION:

Article 1

Subject matter

1. The purpose of this Regulation is to establish a common statistical classification of territorial units, hereinafter referred to as 'NUTS', in order to enable the collection, compilation and dissemination of harmonised regional statistics in the Community.

2. The NUTS classification laid down in Annex I shall replace the 'Nomenclature of territorial units for statistics (NUTS)' established by the Statistical Office of the European Communities in cooperation with the national statistical institutes of the Member States.

Article 2

Structure

1. The NUTS classification subdivides the economic territory of the Member States, as defined in Decision 91/450/EEC, into territorial units. It ascribes to each territorial unit a specific code and name.

2. The NUTS classification is hierarchical. It subdivides each Member State into NUTS level 1 territorial units, each of which is subdivided into NUTS level 2 territorial units, these in turn each being subdivided into NUTS level 3 territorial units.

3. However, a particular territorial unit may be classified at several NUTS levels.

4. At the same NUTS level, two different territorial units in the same Member State may not be identified by the same name. If two territorial units in different Member States have the same name, the country identifier is added to the territorial units' names.

5. In each Member State, there can be further hierarchical levels of detail, decided by the Member State, whereby NUTS level 3 is subdivided. Within two years from the entry into force of this Regulation, the Commission, after consulting the Member States, shall submit a communication to the European Parliament and the Council on the appropriateness of establishing rules on a Europe-wide basis for more detailed levels in the NUTS classification.

Article 3

Classification criteria

1. Existing administrative units within the Member States shall constitute the first criterion used for the definition of territorial units.

[1] OJ L 240, 29.8.1991, p. 36.
[2] OJ L 52, 22.2.1997, p. 1.
[3] OJ L 184, 17.7.1999, p. 23.
[4] OJ L 181, 28.6.1989, p. 47.

To this end, ‘administrative unit’ shall mean a geographical area with an administrative authority that has the power to take administrative or policy decisions for that area within the legal and institutional framework of the Member State.

2. In order to establish the relevant NUTS level in which a given class of administrative units in a Member State is to be classified, the average size of this class of administrative units in the Member State shall lie within the following population thresholds:

Level	Minimum	Maximum
NUTS 1	3 million	7 million
NUTS 2	800 000	3 million

If the population of a whole Member State is below the minimum threshold for a given NUTS level, the whole Member State shall be one NUTS territorial unit for this level.

3. For the purpose of this Regulation, the population of a territorial unit shall consist of those persons who have their usual place of residence in this area.

4. The existing administrative units that are used for the NUTS classification are laid down in Annex II. Amendments to Annex II shall be adopted in accordance with the regulatory procedure referred to in Article 7(2).

5. If for a given level of NUTS no administrative units of a suitable scale exist in a Member State, in accordance with the criteria referred to in paragraph 2, this NUTS level shall be constituted by aggregating an appropriate number of existing smaller contiguous administrative units. This aggregation shall take into consideration such relevant criteria as geographical, socio-economic, historical, cultural or environmental circumstances.

The resulting aggregated units shall hereinafter be referred to as ‘non-administrative units’. The size of the non-administrative units in a Member State for a given NUTS level shall lie within the population thresholds referred to in paragraph 2.

In accordance with the regulatory procedure referred to in Article 7(2), individual non-administrative units may however deviate from these thresholds because of particular geographical, socio-economic, historical, cultural or environmental circumstances, especially in the islands and the outermost regions.

Article 4

Components of NUTS

1. Within six months after the entry into force of this Regulation, the Commission shall publish the components of each NUTS level 3 territorial unit in terms of the smaller administrative units as laid down in Annex III, as transmitted to it by the Member States.

Amendments to Annex III shall be adopted in accordance with the regulatory procedure referred to in Article 7(2).

2. Within the first six months of each year, Member States shall transmit to the Commission all changes of the components for the previous year that may affect the NUTS level 3 boundaries and in so doing shall respect the electronic data format requested by the Commission.

Article 5

Amendments to NUTS

1. The Member States shall inform the Commission of:

(a) all changes that have occurred in administrative units, in so far as they may affect the NUTS classification, as laid down in Annex I, or the contents of Annexes II and III;

(b) all other changes at the national level that may affect the NUTS classification, in accordance with the classification criteria laid down in Article 3.

2. Changes to NUTS level 3 boundaries due to changes of smaller administrative units as laid down in Annex III:

(a) shall not be considered as amendments of NUTS if they involve a population transfer equal to or less than one percent of the NUTS 3 territorial units concerned;

(b) shall be considered as amendments of NUTS, in accordance with paragraph 3 of this Article, if they involve a population transfer of more than one percent of the NUTS 3 territorial units concerned.

3. Amendments to the NUTS for the non-administrative units in a Member State, as referred to in Article 3(5), may be made if, at the NUTS level in question, the amendment reduces the standard deviation of the size in terms of population of all EU territorial units.

4. Amendments to the NUTS classification shall be adopted in the second half of the calendar year in accordance with the regulatory procedure referred to in Article 7(2), not more frequently than every three years, on the basis of the criteria laid down in Article 3. Nevertheless, in the case of a substantial reorganisation of the relevant administrative structure of a Member State, the amendments to the NUTS classification may be adopted at intervals of less than three years.

The Commission implementing measures referred to in the first subparagraph shall enter into force, with regard to the transmission of the data to the Commission, on 1 January of the second year after their adoption.

5. When an amendment is made to the NUTS classification, the Member State concerned shall transmit to the Commission the time series for the new regional breakdown, to replace data already transmitted. The list of the time series and their length will be specified in accordance with the regulatory procedure referred to in Article 7(2) taking into account the feasibility of providing them. These time series are to be supplied within two years of the amendment to the NUTS classification.

Article 6

Management

The Commission shall take the necessary measures to ensure the consistent management of the NUTS classification. In particular, such measures may include:

(a) drafting and updating of explanatory notes on NUTS;

(b) examination of problems arising from the implementation of NUTS in the Member States' classifications of territorial units.

Article 7

Procedure

1. The Commission shall be assisted by the Statistical Programme Committee, established by Article 1 of Decision 89/382/EEC, Euratom (hereinafter referred to as the Committee).

2. Where reference is made to this paragraph, Articles 5 and 7 of Decision 1999/468/EC shall apply, having regard to the provisions of Article 8 thereof.

The period laid down in Article 5(6) of Decision 1999/468/EC shall be set at three months.

3. The Committee shall adopt its rules of procedure.

Article 8

Reporting

Three years after the entry into force of this Regulation, the Commission shall submit a report on its implementation to the European Parliament and the Council.

Article 9

Entry into force

This Regulation shall enter into force on the 20th day following that of its publication in the *Official Journal of the European Union*.

This Regulation shall be binding in its entirety and directly applicable in all Member States.

Done at Brussels, 26 May 2003.

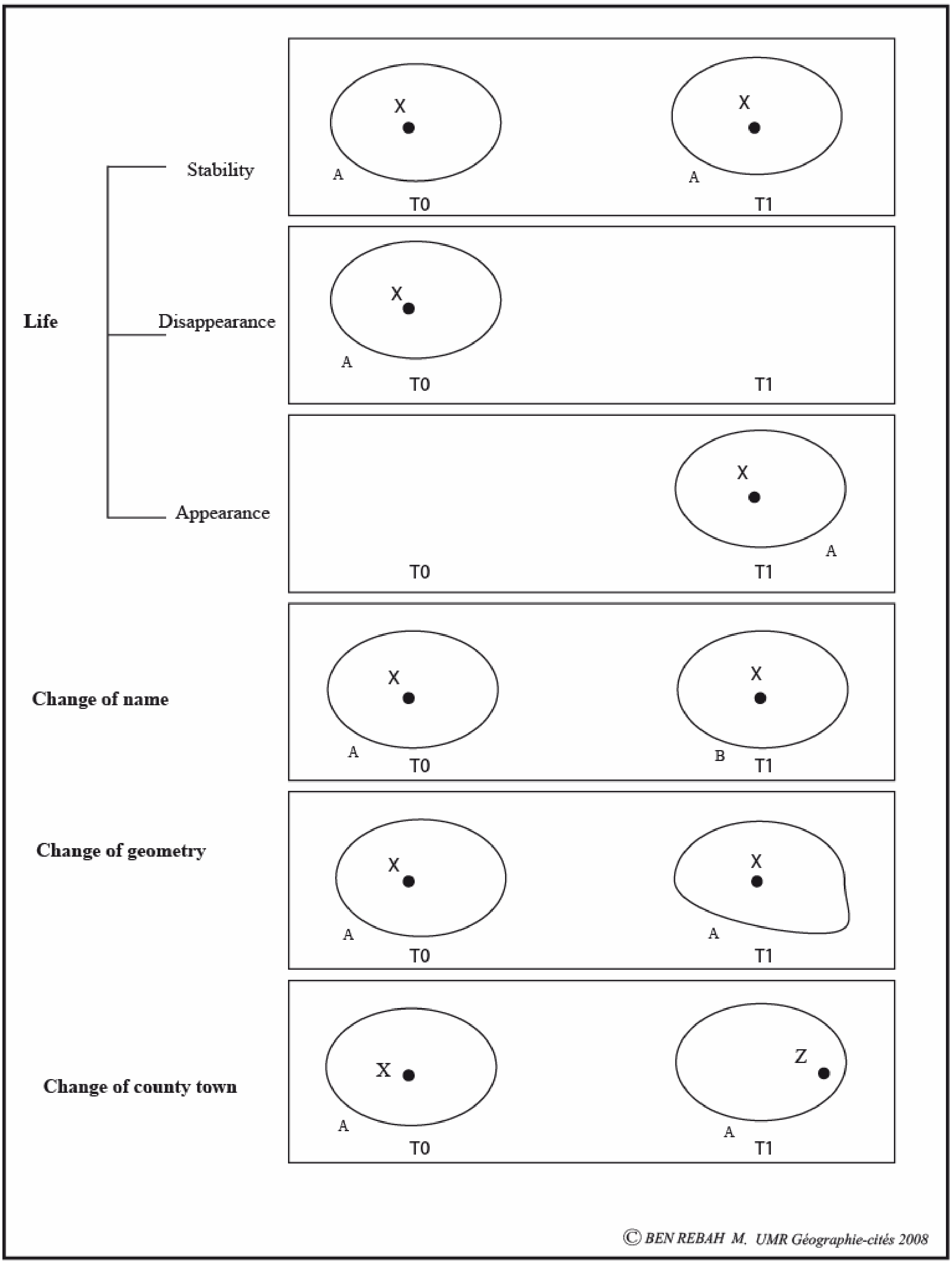
For the European Parliament

For the Council

2.2 Examples of incoherencies between versions of NUTS

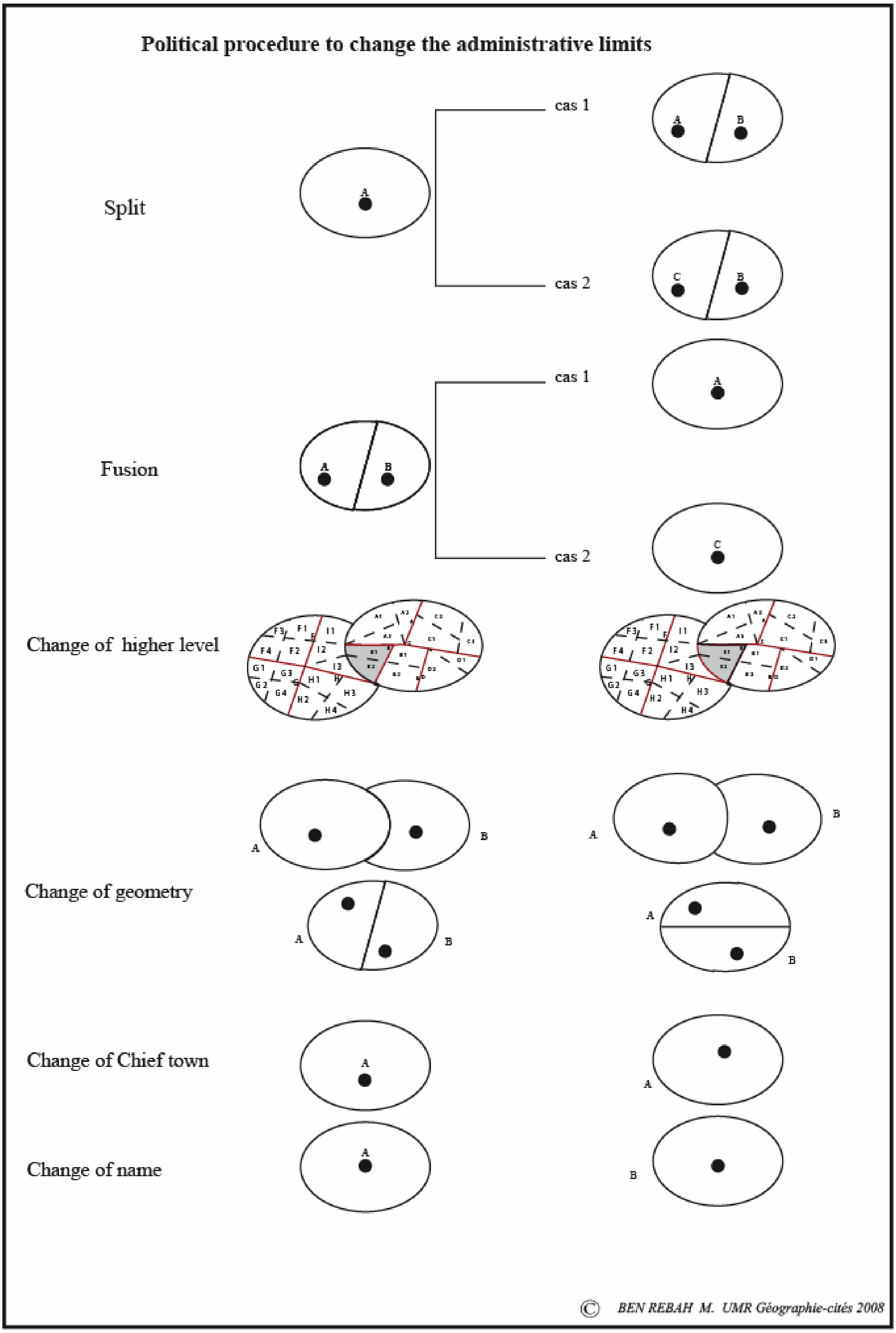
2.3 Classification of elementary changes of an administrative unit

Classification of elementary changes of an administrative unit



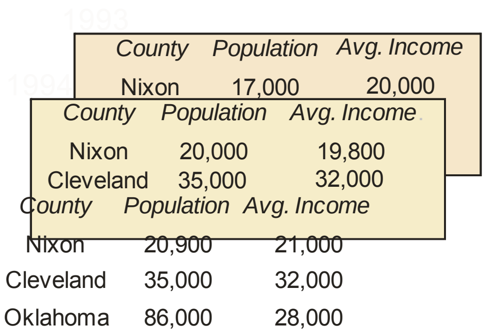
© BEN REBAH M. UMR Géographie-cités 2008

2.4 Examples of complex changes of administrative unit



© BEN REBAH M. UMR Géographie-cités 2008

2.5 Review of proposed temporal data models



1993

County	Population	Avg. Income
Nixon	17,000	20,000

1994

County	Population	Avg. Income
Nixon	20,000	19,800
Cleveland	35,000	32,000

County	Population	Avg. Income
Nixon	20,900	21,000
Cleveland	35,000	32,000
Oklahoma	86,000	28,000

Gadia and Vaishnav (1985):

Time-stamped values

Name	*Salary*	*Department*
[11,60] John	[11, 49] 15K [50, 54] 20K [55, 60] 25K	[11,44] Toys [45, 60] Shoes
[0,20] U [41,51] Tom	[0, 20] 20K [41, 51] 30K	[0, 20] Hardware [41, 51] Clothing
[0,44] U [50, Now] Mary	[0,44] U [50, Now] 25K	[0,44] U [50, Now] Credit

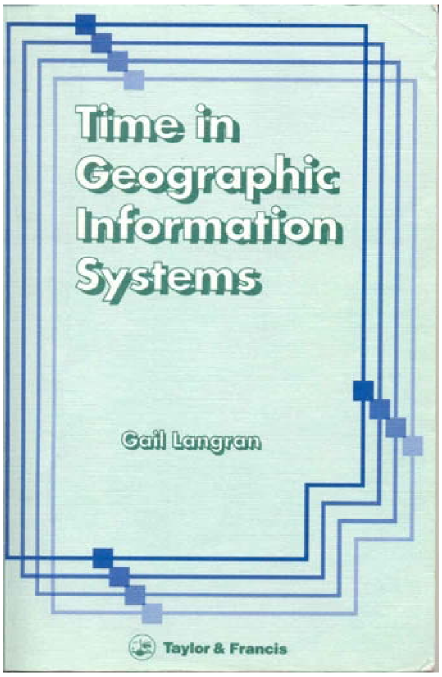
Gadia and Yeung (1988):

Time-stamped records

Stock	*Price*	*From*	*To*
IBM	16	10-7-91 10:07am	10-15-91 4:35pm
IBM	19	10-15-91 4:35pm	10-30-91 4:57pm
IBM	16	10-30-91 4:57pm	11-2-91 12:53pm
IBM	25	11-2-91 12:53pm	11-5-91 2:02pm

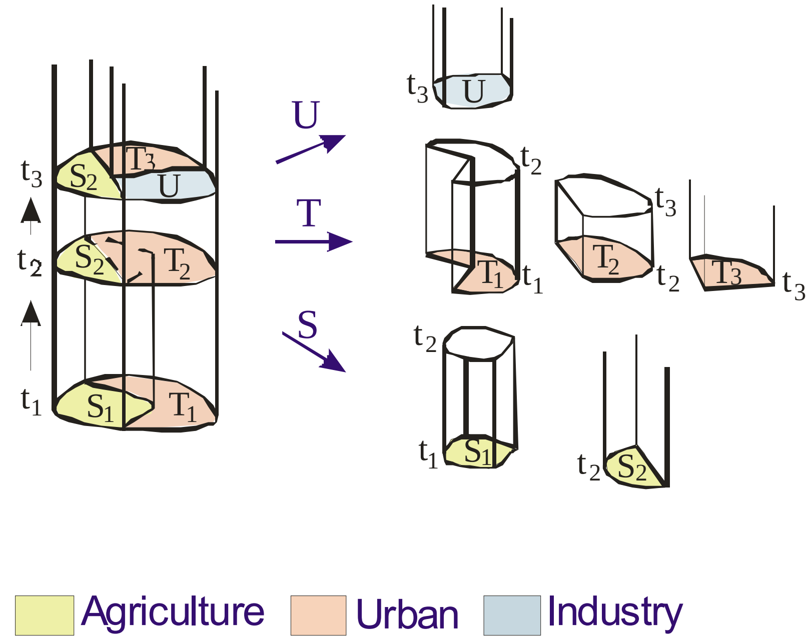
Snodgrass and Ahn (1985):

- Spatial change over time
  - History at location
  - Cadastral mapping



Poly id	*T_1*	*T_2*	*T_3*	*T_4*
1	Rural	Rural	Rural	Rural
2	Rural	Urban	Urban	Urban
3	Rural	Rural	Urban	Urban
4	Rural	Rural	Urban	Urban
5	Rural	Rural	Rural	Urban

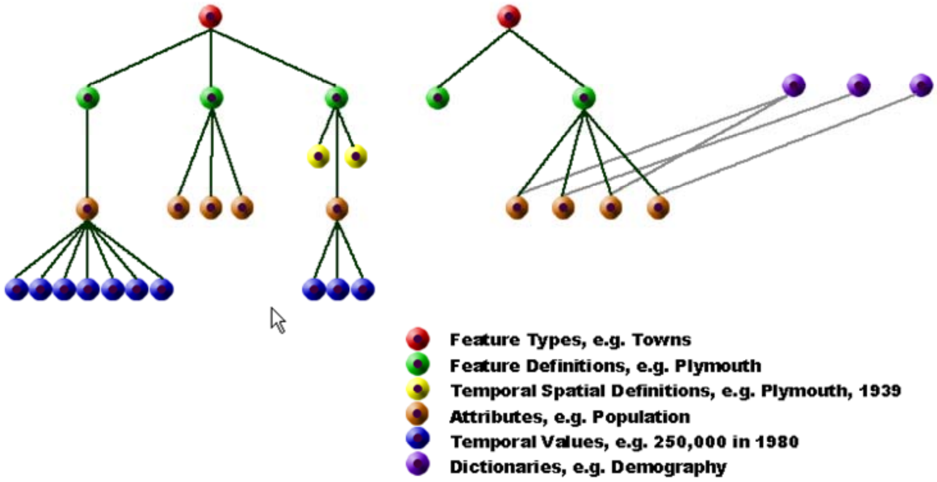
- Spatial objects with beginning time and ending time



Worboys (1992)

STEMgis (2003)

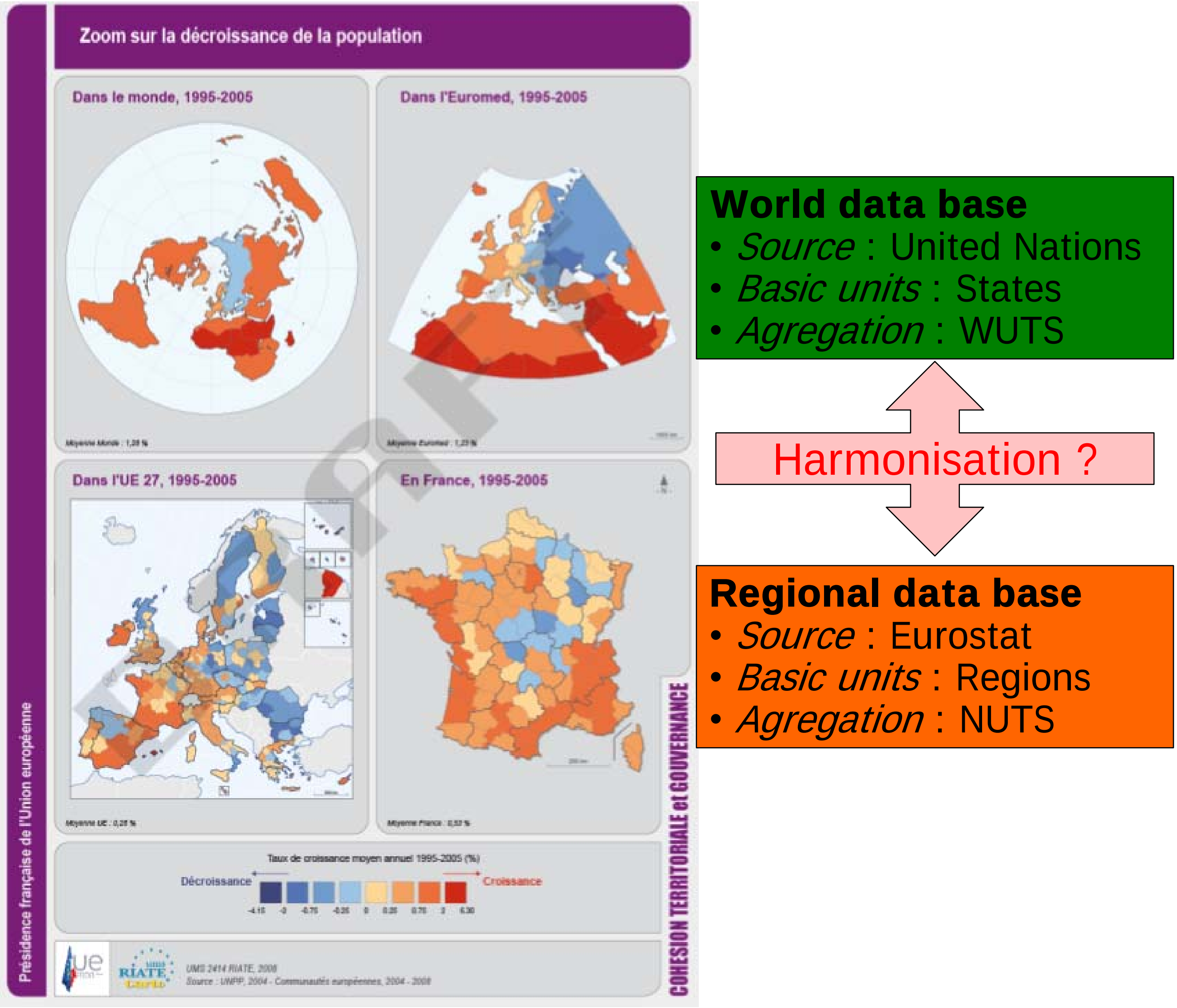
- Time-stamp spatial objects
- Hierarchical database


Feature Types, e.g. Towns
Feature Definitions, e.g. Plymouth
Temporal Spatial Definitions, e.g. Plymouth, 1939
Attributes, e.g. Population
Temporal Values, e.g. 250,000 in 1980
Dictionaries, e.g. Demography

Annexe 3 : Harmonistation of state level between European and World databases. (Challenge 3)

One crucial objective of the Challenge 3 is to be able to produce multiscalar analysis of vartiables of interest, combining World, pan-european, European and national analysis based on different levels of aggregation of data (World regions, States, NUTS2, NUTS3). As an example, we can illustrate the target by an analysis of the dercrease of population realized for the French presidency of EU in december 2008.

Figure 1 An example of multiscalar analysis


Zoom sur la décroissance de la population
Dans le monde, 1995-2005
Dans l'Euromed, 1995-2005
Moyenne Monde : 1,28 %
Moyenne Euromed : 1,23 %
Dans l'UE 27, 1995-2005
En France, 1995-2005
Moyenne UE : 0,28 %
Moyenne France : 0,57 %
Taux de croissance moyen annuel 1995-2005 (%)
Décroissance
Croissance
-4.16
-1
-0.75
-0.25
0
0.25
0.75
2
6.30
UMS 2414 RIATE, 2008
Source : UNPP, 2004 - Communautés européennes, 2004 - 2008
Présidence française de l'Union européenne
COHESION TERRITORIALE et GOUVERNANCE

World data base
- *Source* : United Nations
- *Basic units* : States
- *Agregation* : WUTS

Harmonisation ?

Regional data base
- *Source* : Eurostat
- *Basic units* : Regions
- *Agregation* : NUTS

The problem for this target is the fact that databases does not fit well, especially concerning the state level that is normally common to both typoe of datatsets. We can therefore expect contradictions in the results if we are not able to find procedure of harmonization between the two sources of information.

As an example of benchmarking, we have download on Eurostat Website the 4 Feb.2009 the mid-year population of each country of ESPON 31 for 2001, 2002, 2003, 2004 and 2005. The same variable has been download on the UNEP-GEO website. In both case, we have calculated the mean value of population for the 5 years of the period.

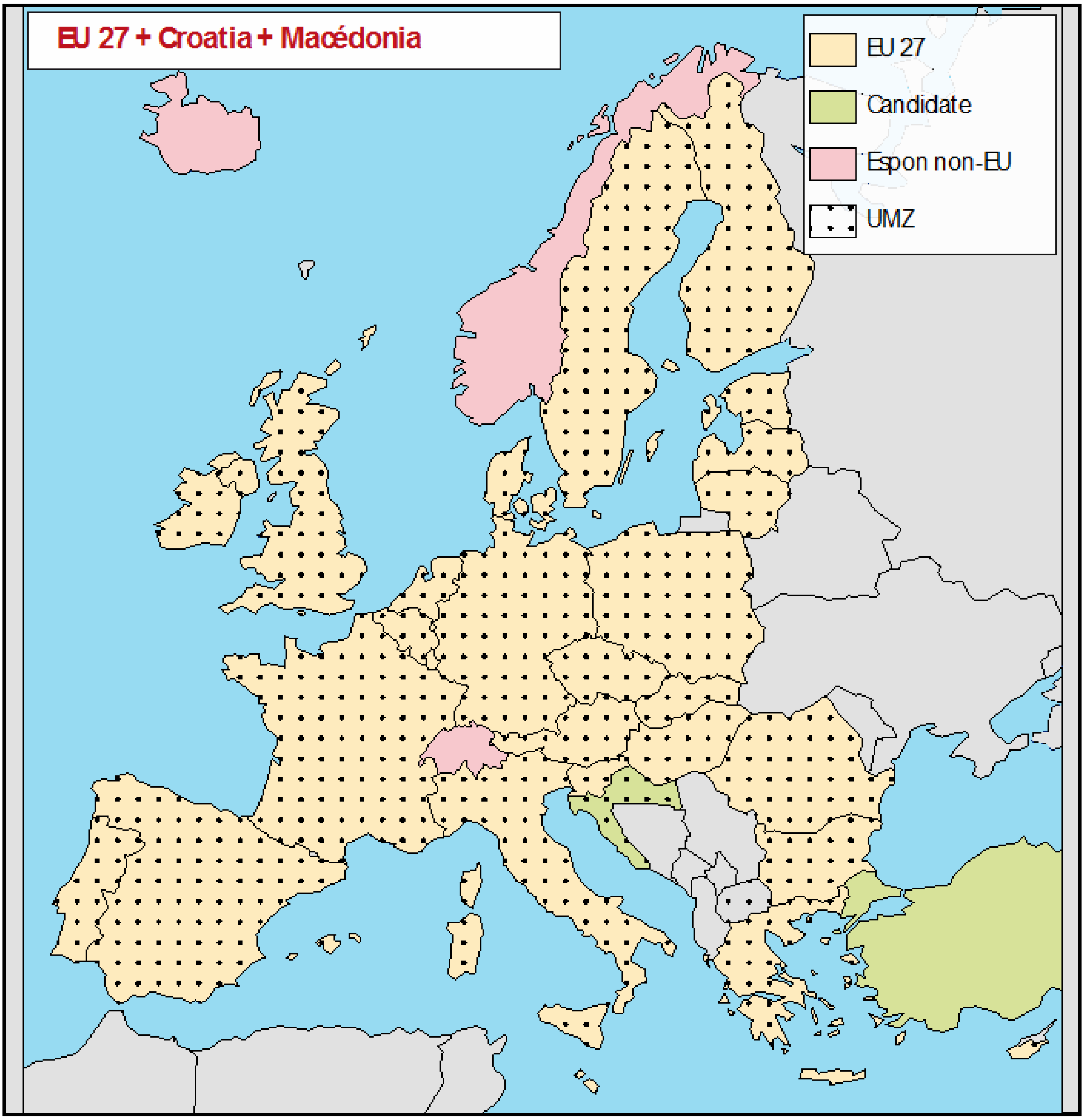
Table : Average population of ESPON31 Countries during the period 2001-2005 according to Eurostat and UN WPP 2006 (from UNEP-GEO Data portal)

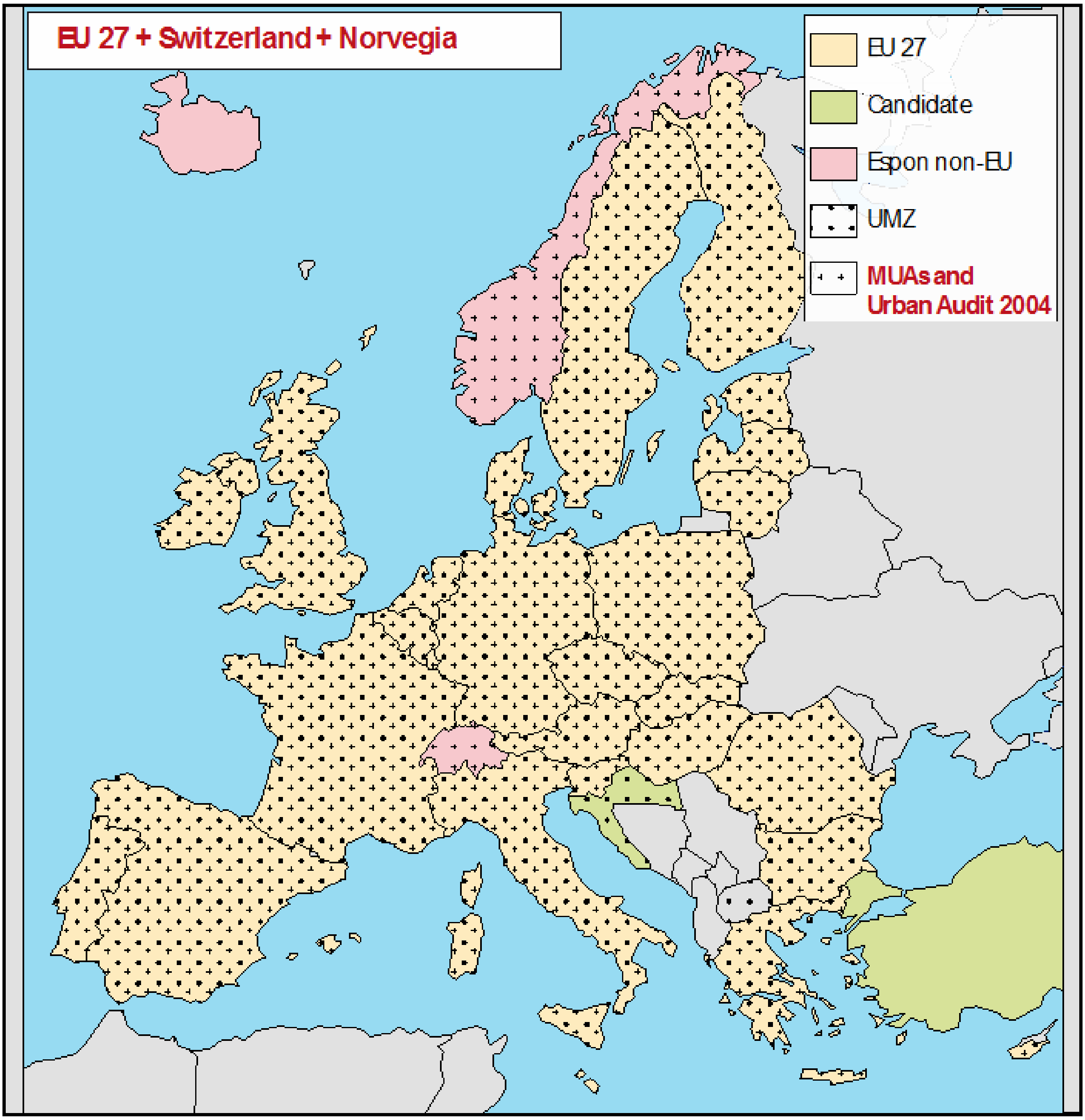
COUNTRY		AVERAGE POPULATION		DIFFERENCE	
Code	**Name**	**Eurostat**	**UNEP**	**Absolute**	**Relative**
cy	Cyprus	726,5	816,8	-90,4	-11,06%
at	Austria	8131,5	8213,3	-81,8	-1,00%
it	Italy	57705,0	58260,5	-555,5	-0,95%
ch	Switzerland	7326,2	7358,0	-31,8	-0,43%
uk	United Kingdom	59620,7	59825,9	-205,1	-0,34%
gr	Greece	11025,4	11057,6	-32,3	-0,29%
no	Norway	4566,4	4578,1	-11,6	-0,25%
bg	Bulgaria	7825,4	7844,6	-19,2	-0,24%
pl	Poland	38207,1	38291,7	-84,7	-0,22%
lv	Latvia	2326,4	2330,6	-4,2	-0,18%
es	Spain	42025,8	42085,3	-59,5	-0,14%
sk	Slovakia	5381,5	5387,2	-5,7	-0,11%
de	Germany	82468,3	82545,8	-77,4	-0,09%
lt	Lithuania	3450,9	3454,0	-3,1	-0,09%
ie	Ireland	3995,9	3998,6	-2,7	-0,07%
hu	Hungary	10134,0	10138,4	-4,4	-0,04%
fi	Finland	5215,2	5217,0	-1,9	-0,04%
se	Sweden	8960,5	8960,5	0,0	0,00%
fr	France+DOM-TOM	62020,4	62004,5	15,9	0,03%
dk	Denmark	5388,8	5386,8	2,0	0,04%
ee	Estonia	1354,3	1352,8	1,5	0,11%
li	Liechtenstein	34,0	33,9	0,0	0,12%
si	Slovenia	1996,2	1993,8	2,4	0,12%
ro	Romania	21855,1	21827,9	27,2	0,12%
is	Iceland	290,3	289,9	0,4	0,14%
nl	Netherlands	16204,4	16179,2	25,2	0,16%
cz	Czech Republic	10217,7	10198,9	18,8	0,18%
pt	Portugal	10430,8	10408,1	22,6	0,22%
mt	Malta	398,5	397,2	1,3	0,33%
be	Belgium	10379,0	10314,8	64,2	0,62%
lu	Luxembourg	452,5	448,8	3,7	0,83%
ESPON31		**500114,6**	**501200,6**	**-1086,0**	**-0,2%**

Comment : In the case of Cyprus, the hugest relative difference (-11%) is obviously related to a difference in the geographical delimitation of the territorial unit. UNEP take into account the population of the northern part of the island, which is not the case for Eurostat. Out of this specific case, the differences are more limited (plus or minus 1%) but sufficient to introduce substantial biases in the analysis

Annexe 4: (Challenge 6)

3.1 Data availability at local level for UMZ, MUA and Urban Audit database




EU 27 + Switzerland + Norvegia
EU 27
Candidate
Espon non-EU
UMZ
MUAs and Urban Audit 2004

Annexe 5: (Challenge 7)

EUROSTAT - ESPON Action Plan

Follow-up of meeting 24 November 2008

The meeting on 24 November 2008 was a follow-up of the meeting of April 2007. Both EUROSTAT and ESPON presented their activities followed by a discussion on the next steps of cooperation. Partnership between EUROSTAT and ESPON is of mutual benefit and cooperation will be handled flexible and focus on responses to demand. This may involve mutual information, requests for comments or clarifications, as well as topic related discussions and meetings. Both partners agreed to be active in involving one another and being responsive to requests.

A number of concrete action points for cooperation on specific topics have been identified. These points are listed in this Action Plan and have been divided into:

- Short term action point, i.e. points that can be dealt with at least before the next meeting,
- Medium term – strategic action points, i.e. points that have to be discussed and worked on and maybe can be the topic of a working meeting,
- Reoccurring action points, i.e. points that will reoccur during the next years.

For every task at least one person, but most of the times two persons (one from EUROSTAT and one from ESPON) will be responsible (indicated between square brackets). This should ensure an efficient implementation of the tasks. The last page of the action plan includes a contact list of people involved in the meetings and actions. (Note that in this first revised version of the Action Plan for some tasks the person(s) responsible [N.N.] has to be filled in and that contact details have to be completed.)

SHORT TERM

Task S1: **ESPON provision of information**

Timing: January 2008

Responsibility: ESPON [Marjan van Herwijnen]

Description: The following ESPON presentations and additional information will be sent to EUROSTAT:

- Presentation on ESPON 2013 Programme
- Presentation on ESPON 2013 Database project
- Presentation on ESPON FOCI project
- Generalized map layer [UMS RIATE]
- Environmental data for Urban Audit if available [FOCI]

Task S2: **EUROSTAT provision of information**

Timing: January 2008

Responsibility: EUROSTAT [César De Diego Diez, Teodora Brandmueller] & ESPON [Sandra Di Biaggio]

Description: For some points it has been mentioned that EUROSTAT can provide ESPON with latest information and/or data:

- Information on how ESPON can arrange a licence for GAUL (done)
- Agreement template to use EuroGeographics geometries (done)
- Naming conventions document [GISCO]
- Algorithms for data quality checks used for Urban Audit [Urban Audit Unit]
- Access and structure of the SIRE database [Urban Audit Unit]
- Access to the archiving of the EUROSTAT website (CD-rom with previous data versions archive) [Urban Audit Unit]
- Completion of the contact list (see last page)

Task S3: **Update of the ESPON-EUROSTAT re-dissemination agreement**

Timing: January 2007

Responsibility: EUROSTAT [Information and Dissemination Unit] & ESPON [Sandra Di Biaggio]

Description: The agreement between EUROSTAT and the Managing Authority of ESPON 2006 has to be updated according to the new ESPON 2013 programme's timings and needs. This agreement grants ESPON a licence to download, reproduce and publish EUROSTAT data and documents, including the GISCO database.

MEDIUM TERM – STRATEGIC

Task M1: Cooperation on grid data

Timing: 2009

Responsibility: EUROSTAT [N.N.] & ESPON [Marjan van Herwijnen] / WP B of ESPON Database project [N.N.]

Description: The use of grid data to overcome challenges related to NUTS has been approached by both EUROSTAT and ESPON during the past years. At the moment, the ESPON Database project is working on this topic. A discussion on how mutual benefit from the developments so far can be assured and closer cooperation for the future facilitated is needed. For this purpose EUROSTAT may provide ESPON with information on their work, in particularly with information on the population density grid developed in cooperation with JRC.

Task M2: Estimation of missing data

Timing: Autumn 2008, ongoing

Responsibility: ESPON [Marjan van Herwijnen]

Description: In general missing data is caused by the following reasons:

- EUROSTAT disseminates only data for the most recent NUTS version. Countries that changed their NUTS division are asked to make estimates to fit data from previous years to this most recent NUTS version. Often they do not deliver the data resulting in data gaps.
- Countries have not yet delivered data for the requested year due to various other reasons.

Various methods exist to fill the gaps and estimate the missing data. EUROSTAT has not the capacity to do this, but encourages ESPON to fill in the gaps and any data resulting from this is welcomed. The way this task can be carried out could be discussed in the meetings.

Task M3: INSPIRE

Timing: Autumn 2008, ongoing

Responsibility: ESPON [Marjan van Herwijnen] & EUROSTAT [N.N.]

Description: ESPON is both a potential user and input provider of INSPIRE. At the moment ESPON is registered as an INSPIRE Spatial Data Interest Group (SDIG). ESPON will check the possibility to participate as an expert in the working groups of INSPIRE. EUROSTAT and ESPON will keep each other mutually informed about the ongoing discussions so that ESPON can apply the necessary standards in its projects and comment on INSPIRE discussions from the perspective of applied territorial research at European level. The current situation is that:

- metadata regulations are finished and can be used,
- in May/June 2009 the metadata editor will be ready.

REOCURRING

Task R1: Routing meetings

Timing: Spring and autumn every year

Responsibility: EUROSTAT [Successor of Roger Cubitt] & ESPON [Peter Mehlbye]

Description: It was agreed to have two types of routing meetings throughout the year: 1. Management meetings and 2. Working meetings.

The management meetings will be held once or twice a year and discuss the more general topics. Possible points for the management meetings are: ownership of results, recent developments, main results, ...

Working meetings will be held about three times a year and discuss a specific topic more thoroughly with those people interested and involved (also from ESPON projects). Possible points for working meetings are: indicator development, estimation of missing data (see Task M2), urban boundaries, INSPIRE (see Task M3), grid-region data integration (see Task M1), statistics on labour market areas.

Task R2: Regular mutual information

Timing: Ongoing

Responsibility: ESPON [André Mueller] & EUROSTAT [N.N.]

Description: In order to facilitate future cooperation both partner will implement the necessary routines to assure that the respectively other party is regularly informed on new developments of interest for the cooperation.

Task R3: EUROSTAT involvement in ESPON activities

Timing: Ongoing

Responsibility: ESPON [Peter Mehlbye] & EUROSTAT [N.N.]

Description: In order to enhance the link between EUROSTAT and ESPON, EUROSTAT will be more involved in ESPON activities. Possibilities to do this are:

- Participation of EUROSTAT in ESPON Workshops, Seminars and other events
- Applying to the call for EoI for the Knowledge Support System of ESPON
- Inscription of Daniele Rizzi and Berthold Feldmann to the ESPON e-mail list for the newsletter, the stakeholders and other relevant lists (done)

Task R4: ESPON involvement in EUROSTAT activities

Timing: Ongoing

Responsibility: EUROSTAT [N.N.] & ESPON [Peter Mehlbye]

Description: In order to enhance the link between EUROSTAT and ESPON, ESPON will be more involved in EUROSTAT activities. Possibilities to do this are:

- Participation of ESPON in the Working Party meetings on "Geographical Information Systems for Statistics" (March each year)
- Participation of ESPON in the Working Party meetings on "Regional and Urban Statistics" (October each year)
- Participation of ESPON in relevant internal EU workshops and/or trainings

CONTACT LIST

Name	**E-mail**	**Telephone**	**Institute / Dep.**	**Topic**
ESPON				
Peter Mehlbye	peter.mehlbye@espon.eu	+352 54 55 80 710	ESPON CU	Director ESPON
André Mueller	andre.mueller@espon.eu	+352 54 55 80 708	ESPON CU	Cluster Coordinator ESPON
Sandra Di Biaggio	sandra.di.biaggio@espon.eu	+352 54 55 80 714	ESPON CU	ESPON 2013 Database Project
Marjan van Herwijnen	marjan.vanherwijnen@espon.eu	+352 54 55 80 698	ESPON CU	ESPON 2013 Database Project
René van der Lecq	rene.vanderlecq@espon.eu	+352 54 55 80 697	ESPON CU	ESPON FOCI Project
Claude Grasland	grasland@parisgeo.cnrs.fr	+33 1 44 27 99 83 +33 1 44 27 86 16 + 33 1 40 46 40 00	UMS RIATE	ESPON 2013 Database Project
Maher Ben Rebah	benreabah77@yahoo.fr	+331 40 46 40 00	UMS RIATE	ESPON 2013 Database Project
Nicolas Lambert	nicolas.lambert@ums-riate.fr		UMS RIATE	ESPON 2013 Database Project

Bogdan Moisuc	bogdan.moisuc@imag.fr	+33 4 76 82 72 11	Grenoble	ESPON 2013 Database Project
Geoffrey Caruso	geoffrey.caruso@uni.lu	+352 46 66 44 6625	UL	ESPON 2013 Database Project
Moritz Lennert	moritz.lennert@ulb.ac.be	+ 32 2 650 56 16	IGEAT	ESPON FOCI Project
EUROSTAT				
Roger Cubitt	roger.cubitt@ec.europa.eu	+352 43 01 33 088	D2	Head of Section D2
Daniele Rizzi	daniele.rizzi@ec.europa.eu	+352 43 01 38 201	GISCO	
Berthold Feldmann	berthold.feldmann@ec.europa.eu	+352 43 01 34 401	Regional and Urban Statistics	
César De Diego Diez	cesar.dediegodiez@ec.europa.eu	+352 43 01 34 992	GISCO	
Teodora Brandmueller	teodora.brandmueller@ec.europa.eu	+352 43 01 32 927		Urban Audit

.

Annex 6 : Example of data request addressed by ReRisk to ESPON DB project (FeB. 2009)

Task	Source	Year	Regional level	Comments / definition
2.1.1. Climate zone of the region and other specific regional factors that influence energy consumption				
Average annual minimum temperature	Nordregio/ Eurostat/ JRC	2007	**NUTS 2**	**Nordregio, "The NSPA in Europe"- data on harsh climate and JRC IPSC "Annual Report 2007" Grid information at EU level**
Average annual maximum temperature	Nordregio/ Eurostat/ JRC	2007	**NUTS 2**	**Nordregio, "The NSPA in Europe" - data on harsh climate JRC IPSC "Annual Report 2007" Grid information at EU level**
Average monthly humidity				
Average annual sun hours	Nordregio/ Eurostat/ JRC	2007	**NUTS 2**	**JRC IPSC "Annual Report 2007" Grid information at EU level**
Heating days per year	Eurostat		**NUTS 2**	Eurostat Energy Unit /JRC
2.1.2. Social indicators				
At risk of poverty rate	Espon	2006	**NUTS 0**	**Lisbon Indicator. Country level**
GDP per capita in pps	Eurostat	2005	**NUTS 2**	**NUTS 2. Regional gross domestic product (PPS per inhabitant)**
Unemployment rate	ESPON / Eurostat	2006 / 2007	**NUTS 3** **NUTS 2**	**ESPON core indicator. Available from Eurostat on NUTS 3 level in most countries (not Turkey, nor Switzerland)**
Indicators set of urban poverty	Eursotat	2003-2006	**NUTS 2/ NUTS 3**	**Urban audit**
2.1.3. Demographic indicators				
% elderly people	Eurostat	2006	**NUTS 2**	**NUTS 2. To be derived from "Average population by sex and age". Some data lacking for New Member States**
Number of one-person households				
Population density	ESPON		**NUTS 2**	**ESPON database ¿including Western Balkans?**
Urban sprawl				
2.1.4. Energy demand indicators				
Energy consumption by sources and sectors (industry, households	Eurostat		**NUTS 2**	**Will be facilitated by end 2008, no complete set of data**
Electricity consumption by sector (in gigawatt hours)	Eurostat – Datashop: New Cronos: Regio database: tran enr: energy en2 cons		**NUTS 2**	**NUTS 2 (NUTS 3 for Central European and Candidate Countries) according to ESPON Data Navigator**
Electricity consumption / GDP (kWh per 1000 Euro):			**NUTS 2**	**NUTS 2. Can be calculated combining information on electricity consumption and GDP**
Households' energy use (toe per capita) (including private transport)	DG Regio	2004	**NUTS**	**Estimates facilitated by DG Regio**
Gross Inland Consumption by fuel	Eurostat	2006	**NUTS 0**	**Country level**
2.1.5. Production capacity indicators*				
Electricity production capacity (in megawatt)	Eurostat - Datashop: New Cronos: Regio database: tran enr: energy en2 celec	2001	**NUTS 2**	**Energy sources: nuclear, hydroelectric, thermal, total** **NUTS 2 (NUTS 3 for Central European and Candidate Countries) according to ESPON Data Navigator**
Proportion of electricity generated by renewables (%)	Eurostat/ DG Tren		**NUTS 2**	**Data to be facilitated (if available) by Eurostat end of 2008)**
Proportion of electricity generated by liquid fossil fuels (%)	Cronos / Regio above	2001	**NUTS 2**	**Energy sources: thermal** **NUTS 2 (NUTS 3 for Central European and Candidate Countries)**
Proportion of electricity generated	Cronos / Regio	2001		

by solid fossil fuels (%)	above			**according to ESPON Data Navigator**
Proportion of electricity generated by natural gas (%):	Cronos / Regio above	2001		
Fossil fuel dependency (%)	Eurostat/ OECD		**NUTS 0**	**Country level**
Crude oil refined/fossil fuels primary consumption	Eurostat	2005	**NUTS 0**	**Country level**
Transmission capacity (bottlenecks)	UCTE - Union for the co-ordination of transmission of electricity	2007	**NUTS 0**	**"System Adequacy Retrospect 2007". Country level. Includes data on Western Balkan Countries**
Mapping of renewable resources	National / Regional energy agencies/ Eurostat/ JRC	2007	**NUTS 2**	**Available for solar energy at grid level. JRC IPSC "Annual Report 2007"**
Past speed of RES deployment (time frame?)	Eurostat	2001 - 2005	**NUTS 0**	**Country level (hydroelectric, wind, PV)**
2.1.6. Transport infrastructure: modal split of passenger and road transport				
Nº of daily trips by car	Eurostat	2003-2006	**NUTS 2/ NUTS 3**	**Urban audit**
Modal split of passenger transport				
Modal split of freight transport	Eurostat	2006	**NUTS 2**	**Railway infrastructure and transport flows by NUTS II region**
Total number of driven intra-regional trips (trucks / day)	Eurostat	2001	**NUTS 2**	**Available for NUTS 2, but not new Member States**
Total number of km produced by intra-regional trips (1000 Km / day)	Eurostat	2001	**NUTS 2**	**Available for NUTS 2, but not new Member States**
Nº of people working in the region vs nº of persons working in another region	Eurostat	2006	**NUTS 2**	**Available for NUTS 2 – must be related to size of region (area)**
Age of cars	National statistics		**NUTS 0/ NUTS 1/ NUTS 2**	**NUTS 2 for Eastern Countries. Available on Country level**
2.1.7. Regional Competitiveness & Elasticity Indicators				
Gross value added at basic prices	Eurostat	2005	**NUTS 2**	**Available for NUTs 2 and NACE 1 digit**
Gross energy consumption / GdP	Eurostat / EFTA		**NUTS 0**	**Lisbon indicator. Country level**
Regional energy costs	Eurostat (country data)/ National		**NUTS o / NUTS 2**	**Council Regulation (EC, Euratom) No 58/97 of 20 December 1996 concerning structural business statistics; Eurostat inquires on availability of regional data**
Regional gross value added in energy intensive industries / European gross value added in energy-intensive industries	Eurostat		**NUTS 2**	**NACE 24, 26, 27, 28**
Regional gross value added of transport-intensive sectors / European gross value added in transport-intensive industries	Eurostat		**NUTS 2**	**NACE 14, 17, 19, 20, 21, 26, 29, 31, 34, 45**
Employment in renewable energy sector	National / Regional sources		**NUTS 0/ NUTS 2**	
Household debt	Eurostat / OECD / ECB	2005	**NUTS 0**	**Country level**
Median disposable income	Eurostat	2004	**NUTS 2**	**Disposable income for NUTS 2, including New Member States**

Annex 7

Minutes of the meeting between the ESPON Database project's partners, the ESPON CU and DG REGIO GIS

16 January 2009

Main points of decision

1. Exchange of information and data
   a. Exchange or data and information between DG Regio-ESPON should be done in a informal way, i.e. without need for a formal agreement on this respect
   b. Exchange of information and data should transit by the ESPON Intranet. ESPON CU should be kept informed of such exchange.
   c. DG REGIO will upload possible updates on the ESPON Intranet. The ESPON CU will inform ESPON partners.
   d. Sources should always be thoroughly acknowledged.
   e. Updates or data from ESPON Database project should be made available as soon as they are considered trustworthy. The calendar of reports issuance should not constitute an obstacle.
   f. Contacts for the exchange of information and data are:
      - Sandra Di Biaggio and Marjan van Herwijnen for the ESPON CU
      - ESPON Database project functional mailbox (manager@espondb.eu) for the Lead partner
      - GIS functional mailbox (regio-gis@ec.europa.eu), Lewis Dijkstra and Philippe Monfort for DG REGIO

2. Legal aspects of data dissemination

In case some legal/copyright questions are tie to the dissemination of particular data, permission should be arranged with the parties concerned. This will be done on a case by case basis. As often as possible, permission to share data and/or results with ESPON network/DG Regio should be requested systematically.

3. ESPON Database

ESPON Database project will make sure that the data can be exported in a format compatible with the ones used by the EC (at the moment: ESRI © shapefiles and grids). Besides this, progress made within the ESPON database project on time series and changes on NUTS delimitations will be communicated to DG Regio

4. Data geographical scale
   Data should be made available at the most appropriate geographical scale. In particular, information should not be lost when reconstructing data into NUTS division. Therefore, raw data that served as basis to these processes, should as far as possible be made available.

5. List of items to be provided by DG REGIO
   5.1 Short-term actions
   - Population Grid (1Km), based on the Population Grid from JRC and including national population grids for Norway, Sweden and Finland. On the basis that the national population grids can be disseminated to the ESPON network (DG Regio needs to check this situation).
   - Contacts at JRC for climate change matters.
   - Data on the proximity to natural areas and data on territorial cohesion borrowed form the Green paper (underlying data from the maps included in the working document).
   - Data on metropolitan areas (NUTS-based typology, including field identificator from Urban Audit zones)
   - Accessibility to passenger flights and accessibility to medium-sized cities (also weighted by slope and congestion) and resulting urban-rual typology.
   - Key words for the GISCO naming convention adjusted by DG Regio (including additional domains).

   5.2 Medium-term actions
   - Information on the progress made on the metadata editor currently under development
   - List of indicators included in the DG Regio database
   - Making available specific datasets from future DG Regio publications and Reports