



HAL
open science

The Core Database Strategy – A new paradigm for data collection at regional level

Claude Grasland, Ronan Ysebaert

► To cite this version:

Claude Grasland, Ronan Ysebaert. The Core Database Strategy – A new paradigm for data collection at regional level. [Research Report] ESPON. 2011, 107 p. hal-03609358

HAL Id: hal-03609358

<https://hal.science/hal-03609358v1>

Submitted on 15 Mar 2022

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

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

Copyright

The Core Database Strategy – A new paradigm for data collection at regional level

SUMMARY

The aim of this report is twice:

Firstly, it proposes a **general strategy for data collection** inside M4D project and more generally at the level of ESPON program.

The Core Database Strategy (CDS) is an attempt to propose an innovative solution against the current situation where the ESPON database is adversely affected by the accumulation of heterogeneous data that are more and more difficult to manage. The report describes the general aims of this new strategy and the expected benefits, in particular when it comes to territorial monitoring.

Secondly, it **presents the preliminary tests and results** of this strategy in the case of regional data. We examine firstly the current **list of core indicators** likely to support the CDS. Then we propose **methods for the estimation of missing values** and building of long term time series of core indicators. Finally we demonstrate how such core indicators can be combined with accessibility measure in order to **produce innovative measure** of functional dynamics.



**ESPON M4D -
MULTI DIMENSIONAL DATABASE DESIGN & DEVELOPMENT**



LIST OF AUTHORS

Claude Grasland, UMS 2414 RIATE
Ronan Ysebaert, UMS 2414 RIATE

Contact

claude.grasland@parisgeo.cnrs.fr

ronan.ysebaert@ums-riate.fr

UMS 2414 RIATE

Tel. (+ 33) 1 57 27 65 35

TABLE OF CONTENT

1. The Core Database Strategy	3
1.1 Two possible scenarios for ESPON Database	3
1.2 Principles of the Core Database Strategy	5
1.3 Expected benefit of CDS for territorial monitoring	8
2 Preliminary tests and results	11
2.1 Collection of ESPON "Core" indicators through Eurostat website	12
2.2 Estimation of missing values using the ESTI framework and the ESPON metadata model 15	
2.3 Data quality control, example of population check by level	19
2.4 Transformation of indicators by functional smoothing.....	21
Conclusion	26
Bibliography.....	27
Annexe 1 – Data availability in Eurostat tables.....	28
Annexe 2 – Towards a dictionary of estimation methods for missing values? Working paper	70
Annexe 3 – R programmes for calculating potential of population in different car-time radius	104

1. The Core Database Strategy

During the second internal meeting of ESPON M4D project in October 2011, the coordinator of the thematic group has proposed a new method for the development of ESPON database called the "Core Database Strategy" (CDS). This strategy implies an important revision of data collection approach inside the ESPON M4D project but also in the whole ESPON program. That is the reason why it has been decided to launch during 6 months a test phase of the CDS strategy, in order to evaluate its potential to solve existing problem. The ESPON seminar in Denmark will be the milestone for the decision to adopt definitively, to modify or to give up this CDS strategy.

1.1 Two possible scenarios for ESPON Database

Based on the experience gained in the ESPON Database 2013 project (2008-2010) and the first months of work of the ESPON M4D Project (Feb-Sept 2011) it appears possible to propose two opposite strategies for the management, storage and diffusion of data produced in the ESPON program as a whole (figure 1).

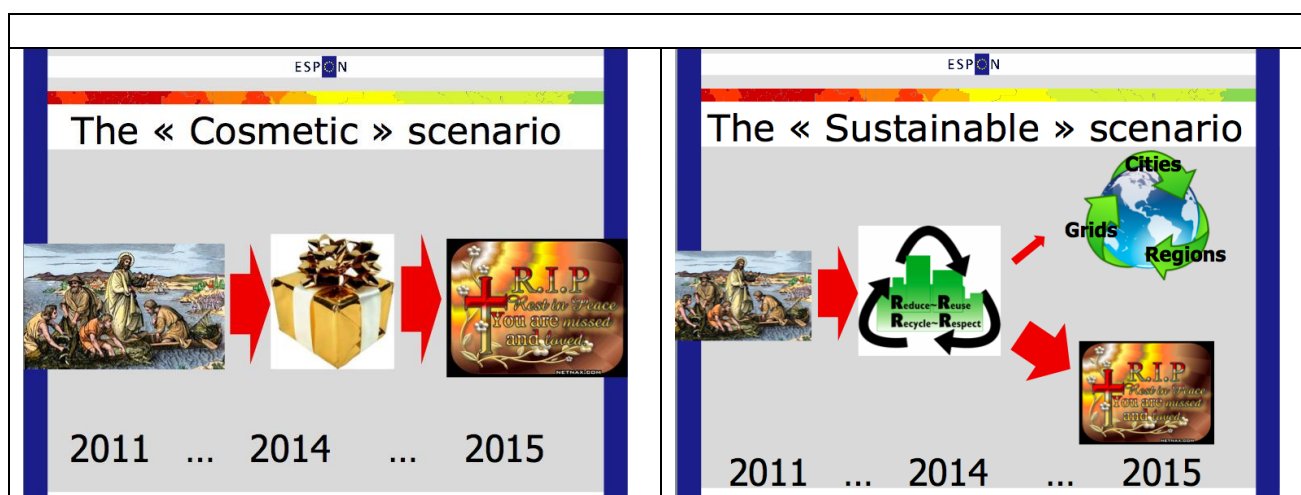


Figure 1 – Discussion of two scenarios for the achievement of ESPON Database (M4D meeting, October 2011)

- **The "Cosmetic" scenario** is driven by final evaluation of the ESPON program, in particular by the so-called quantitative indicators of success that are usually used in INTERREG programs. In the case of data, a classical indicator is the number of indicators or maps produced by each project. Following this strategy, it is important to produce as many indicators as possible and to publish regularly a list of new indicators produced by the ESPON program. It is also important to classify this indicators in relevant groups based on thematic topics (Agriculture, Economy, Transport, ...) but also, which is certainly the most important, in terms of policy relevance and linkage with EU policies (Lisbonne Strategy, Europe 2020, ...). Concerning the choice of geographical objects, the Cosmetic Scenario will obviously focus on official delineation for regions (NUTS2, NUTS3) and for cities (FUA) and will take care to make firstly use of official datasets produced by Eurostat and EEA or politically recognized document like OECD and DG Regio studies. Concerning estimation of missing values, the rule will be to stick to official data as long as they are available and to avoid any modification of official figures. Concerning innovation, the rule will

be to avoid any duplication of officially existing work like the OECD typology of urban-rural area. The keyword associated to this strategy is therefore “**cosmetic**” because what is at the top of the agenda is not the content of the database itself but the way it is promoted to external users or evaluators. When ESPON TIPTAP deliver more than 50 indicators related to variant of TIA scenarios, it increase formally the size of the database (it is good for quantitative evaluation) but it does not offer real added value in the long term as these 50 indicators cannot be updated regularly and are not useful for the majority of users. For the same reason, when ESPON database store gigabytes of regional data derived from Eurostat and add only values for non EU countries, it increases formally the size of the database (40 indicators by just storing the age structure by sex and 5-years age groups...) but it is of minor interest for the majority of external users who prefer to download directly data on Eurostat website or eventually through OECD explorer which present the interest to use more relevant territorial divisions based on mixture of NUTS2 and NUTS3. Last but not least, many ESPON projects has delivered so-called “Synthetic index” which are just elementary statistical transformation of a limited number of classical variable published regularly in the cohesion report. It is really of null interest to store indicators like “GDP per capita transformed into a qualitative index with value 1 to 5 ” but ESPON database is filled of such indexes that are useful in the context of a specific report but are certainly not of interest for any user and are just a pollution. Only the formula of such composite indicators should be stored, making possible to compute it and update in real time when new basic variables are delivered. With this strategy of storing the maximum of data, the ESPON Program appears as **a giant collection of data covering the maximum of subject of political interest and using as much as possible official data and definitions.**

- **The “Sustainable” Scenario** is driven by the assumption that ESPON program can contribute to major innovation in the statistical landscape of European Union and should provide not only answer to current political demand but also propose material for new innovative policies and strategies. The keywords are therefore the **sustainability and capacity of innovation** of the database i.e. the possibility to update and enrich the content of the database in past, present and future and to address new questions that are not necessarily currently present in the political agenda but are considered as likely to emerge in a near future. In this framework, the first question to ask for each data to be stored in the database is: *“Can it be estimated in the past and computed again in the future”*. If the answer is “no”, such a data is of limited interest because it cannot be used for territorial monitoring and is not likely to be of any help in the building of scenarios. The second question to be asked is: *“Can this indicator be estimated in different geometries like various regional delineations, various cities definitions, various grid types... or is it limited to a specific delineation of a specific type geographical object”*. If the answer is “no”, we can consider that the interest of the data are limited in time (because the delineation of regions and cities are changing through time) and limited in terms of potential users (e.g. stakeholder of cities and regions cannot share the same knowledge database). The third question to be asked is: *“Does this indicator open the door for new territorial policies at EU level and what is the added value as compared to existing work done out of ESPON ?”*. If the answer is “no”, the fact to store the data is of minor interest because it will introduce a confusion between what can be found elsewhere and what is really specific and original in ESPON production. It is even dangerous and risky for ESPON to store non original data without added value because the evaluators of the ESPON

program will certainly not evaluate only the quantitative dimension (“How many indicators?”) but also check the qualitative dimension (“What about consistency of the database?”) and finally the level of innovation (“What’s new as compared to Eurostat or OECD ?”). The difficulty with the sustainable approach is the fact that it has not been launched since the beginning of the ESPON 2013 program and that in the current database is stored a lot of data that does not fulfil the conditions of sustainability. Moreover, all the ESPON project are contractually obliged to deliver data that M4D is contractually obliged to check and store, whatever the interest to do it in a perspective of sustainability. In pragmatic terms, **the sustainable approach is possible only if we are ready to develop a separated part inside the ESPON database called ESPON Core Database.**

1.2 Principles of the Core Database Strategy

The Core Database Strategy is not something new in the ESPON Program. The principles behind this strategy have been elaborated and discussed during the ESPON 2006 Program, in particular in the final projects called *Modifiable Area Unit Problem (2006)* and *Data Navigator II (2007)*. The two versions of the *HyperAtlas (2004 and 2010)* have also contributed to the reflection on this topic, by demonstrating that a lot of indicators of major political interest could be derived from the same basic count variables. Finally, the ESPON Database 2013 program has elaborated a database structure which follows the central idea of the CDS which is to avoid the limitation of database to a single type of geographical object (NUTS2 and NUTS3 region). Out of ESPON, it is also worth to mention the *Atlas Interactif des Régions Européennes (AIRE)* realized by UMS RIATE with support of University Diderot’s Innovation Prize (2009) which proposed 30 cartographic visualizations of the same indicator, starting from an initial distribution at NUTS3 level. All this related work converges to a general model of data collection that can be briefly summarized by the figure elaborated during the M4D meeting of October 2011 and presented at the ESPON seminar in Cracow (Dec. 2011, figure 2).



Figure 2 – Methodology overview for sustainable solution

Without developing in details the CDS strategy, we can briefly explain in five bullet points the main ideas of the solution and the expected results.

1. **Start by eliminating all indicators except count variables.** In concrete terms, it means that a variable like GDP per capita is initially excluded from the core database as it is not a count variable but the ratio of two count variables which are total population and total amount of GDP (in \$ or pps). For the same reason, a variable like dependency ratio which is the ratio between "young" and "old" divided by "adult" will be eliminated. What will be stored is the pyramid of age at the most detailed level (population by 5-years age group), making possible in the future to test alternative delimitation of "young" (0-14 ? 0-19 ?, 0-25?) or old (60 and more ? 65 and more ? 70 and more ?).
2. **Store formula of indicators of interest derived from count variables.** All indicators that are derived from count variables are stored as formula using count variables located in the core database. They are therefore virtual data that are computed on request when a user requests it. This formula are not related to a particular geometry but can be computed for all type of objects (cities, regions, states, ...) as long as count variables are available for this particular geometry.
3. **Enlarge time series of count variables in past and future with estimation of missing values.** Count variables are in limited number which reduces the task of estimation of their time series. Moreover, they are more robust to statistical tests because it is not necessary to introduce weighting criteria as it is the case for ratio. They offer many opportunities of estimation in the ESTI framework and are likely to be checked in many ways (e.g. comparison of sums in lines and columns in the case of age structure). Once time series of count variable are completed, all the derived indicator can be immediately computed using their formula.
4. **Develop automatic procedure of exchange of count variables between geometries of various types.** Depending on the nature of information, it can be initially available at regional level (GDP), at grid or vector level (Land Cover), at dot level (Airport location), etc. The problem to be solved here is how to transfer count data from one geometry to another one by using the relevant method of aggregation/disaggregation, eventually coupled with the use of proxy and ancillary information. Considering experience gained in ESPON database 2013 project, we propose to focus this step on the use of OLAP cube technology which will be the "pivot" for exchange between geometries. For example, GDP data at NUTS 3 level will be firstly estimated at grid level (using ancillary variable like JRC grid of population) and then re-aggregated to new geometries like water basin, FUA, etc.
5. **Propose innovative procedures of multi-representation and multi-level analysis of indicators for territorial monitoring and political decision.** Whatever the geometry of interest (cities, regions, ...) an isolated indicator can be transformed into a family of variants that are targeted to answer to specific political questions. If we consider for example the unemployment rate of a specific FUA, it can be compared to the mean level of unemployment of EU, of the country where the city is located, of the cities located at a car distance of less than 2 hours, etc. The reader recognizes here the procedure of multilevel territorial analysis developed by HyperAtlas for regional units, but that can be also transposed to the case of other objects like dot or grids. Another example is the cartography of discontinuities between regional borders. As long as the target indicator is available in regional units (e.g. median age of population at NUTS2 level), we can link it with the matrix of contiguity and deliver a database on regional borders, where each limit is characterized by absolute or relative

differences of median age. One more time, it is not necessary to store this table of discontinuities that can be generated automatically for all indicators of the database.

The solution appears very exciting but is it really as simple and nice as it looks? In fact, some difficulties are to be expected and M4D project has to take into account some possible difficulties and objection. We discuss here some of them:

- **What will happen to the indicators currently contained in the database if CDS is adopted?** This is not a trivial question because many projects has delivered data that are not in line with CDS (lack of any new count variables, lack of clear formula of computation of indicators or too complex formula, ...). It is certainly necessary to maintain the storage of this data but this will implies to spend a lot of time that could be better used for CDS development.
- **What is the consequence for future data delivery and check by ESPON project?** The provisional solution elaborated by M4D with ESPON CU in order to limit the "data deluge" (after TIPTAP ...) was to limit the delivery of data by Priority 1 project to the ten best indicators. But it is not quite sure that this solution is relevant in the case of CDS. If we consider for example the ARTS project, our recent data check has revealed that this project has delivered absolutely no new count variable and has just proposed formula making use of existing data in order to propose sophisticated composite indexes. On the contrary, DEMIFER project has delivered many new count variables of high interest, in relation with different demographic scenarios. Limitation to ten best indicators is therefore not relevant in the case of DEMIFER (we would lost major contribution to core data toward the future) and, on the contrary, it would be stupid to store the composite indexes of ARTS which produce no added value in terms of count data : only formula should be stored in this case.
- **What will be the effect of CDS on the quantitative production of ESPON database?** The answer is dual. On one hand, there would be an apparent contraction of the database because count variables available in time series are of limited number, especially if we consider contingency table as one single variable (population by age and sex, active population by branch, land cover by type, ...). In this case, we can say that probably no more than 15 to 30 table of count variable will be finally stored. But this table of count variable are likely to produce several hundred of indicators derived from formula. And this hundreds of indicators should be multiplied by the number of geographical object where they will be available (cities, regions, states, grids) as well as they should be multiplied by the number of variants of this indicators that can be automatically generated (deviations to European, national or local context, evolution through time, discontinuities, ...). In reality, the CDS should provide ESPON with thousands of potential indicators, very innovative and nicely targeted to political needs at various scales.
- **What would be the added value of CDS as compare to OECD regional database?** In the case of regional database of OECD, the CDS is yet applied and the number of count variables proposed by this organization is very limited, and mostly to Eurostat data. The only added value of OECD regional database is the use of good territorial divisions NUTS2/3 (robust to the MAUP) and the nice visual application OECD Explorer. But it is precisely because the amount of data is limited in OECD database that the Explorer was easy to develop. If ESPON adopt the CDS strategy, it will be much easier to connect the database to existing application like HyperAtlas or to develop new visualization and

monitoring tools. In fact, ESPON database is potentially much more powerful than OECD regional database and involve genuine innovative indicators (like accessibility measure, telecom connections, demographic scenarios, ...) which are not simple extracts of Eurostat or National Statistical Institutes. The problem is that this innovative part of the ESPON database is currently hidden in an ocean of data of lower interest.

- **What about the question of “official” data if estimation is introduced by CDS?** The evaluation of count data for different geometries will produce a lot of estimated values that will be different from the official ones (when they exist). The lineage of data will be precisely described and the methods used for estimation will be precisely described, thanks to the ESTI framework described in the second part of this report. But it will necessarily happen that figures from the ESPON database will be different from the one published by Eurostat and National Statistical Institutes. This is a consequence of the harmonisation and check for internal consistency that will be used to remove holes and file time series. From this point of view, the status of the ESPON database should be equivalent to the one of the Maddisson database on GDP: the data of GDP and Population included in Maddisson database are not exactly the same than the one produced by IMF or UN, but they benefit from an internal coherency and are available in long term time series. In any case, the ESPON database will be able, thanks to its metadata model, to precise which figures are equivalent to the official and which data are estimated and by which method.
- **What is the agenda of the CDS experience? How will it be validated?** The teams involved in the M4D project has decided to launch a 6 months experience, in order to present results to the ESPON community at the meeting of Denmark in June 2012. During this period, we continue to maintain and develop the current structure of the ESPON database and we try to realize the expected improvement, in particular concerning the interface, the thesaurus, etc. We limit therefore the risk and will be able to stay on the “Cosmetic Strategy” if the CDS appears to be too difficult to launch. We are nevertheless convinced that it will not be the case and that the ESPON Program as a whole will strongly benefit from the new orientation we propose to enhance during the next months.

1.3 Expected benefit of CDS for territorial monitoring

We discuss here some expected benefits of the CDS for policymakers and stakeholder.

The ESPON project on *Territorial Monitoring* has been delayed a first time because of lack of applicant. At the same time, the linkage between *ESPON 2013 Database* project and *ESPON Interco* project has not been sufficient, making difficult the elaboration of a common set of territorial indicators of prior interest for policy makers. As a whole, it appears for the moment difficult to achieve the common project of priority 3 projects, which is to build an efficient tool for the monitoring of regions and cities of the ESPON area.

The CDS appears in our opinion as a potential candidate to the solution of this difficulties, because it focus precisely on the indicators that can be regularly updated in past and future and that are therefore the most useful for the understanding of current dynamics and the forecast of future evolutions. The fact that CDS stores only a limited number of count variables make it really easy to update each year. And as

long as the count variables are update, all the other derived indicators or complex indicators will be automatically recomputed as long as they are based on formula. Moreover, each time that an innovation is proposed and adopted by policy makers (like the maps of deviation, the analysis of discontinuities, the potential based on time distance, ...), it can be transposed to most recent data by application of formula to new count variables. Last but not least, when modification occurs in the geometry of spatial units (like the move of NUTS2006 to NUTS2010 delineation), the revision of data is easy because, one more time, the question is only to evaluate the count variables and to apply automatically the program of computation of derivated indicators.

A good illustration of this strategy is provided by the cartographic portal of UMS RIATE (figure 3):

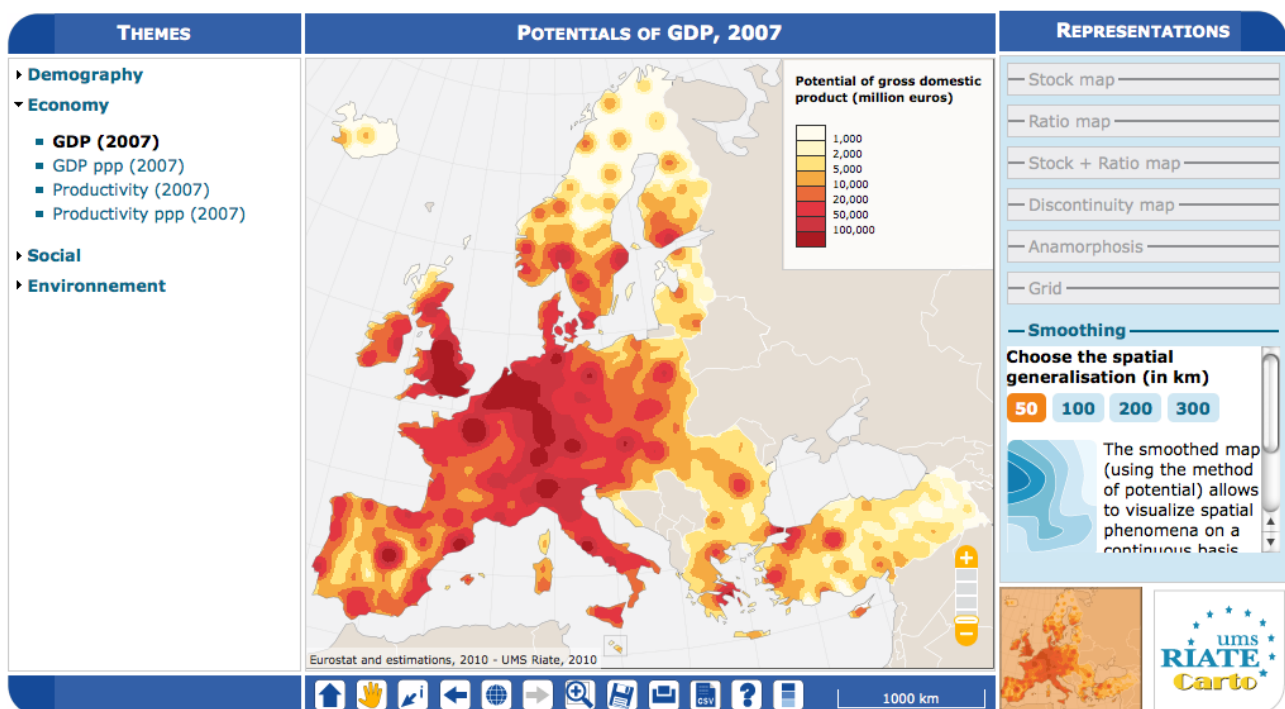


Figure 3 – AIRE, Interactive Atlas of European Regions
<http://aire.ums-riate.fr>

The aim of this application is to deliver simple maps of basic indicator characterising European Region to a great diversity of users (researchers, teachers, students, policy makers, local authorities). The originality of its approach is to propose a great diversity of map for each proposed indicator, making possible for the user to choose the map (or the maps) that are the most likely to fulfil their expectations. 7 types of maps are proposed and for each type of map, 4 to 5 different scales of aggregation, which means a choice between more than 30 solutions...

The difficulty is of course the fact that the users are generally interested by the last data published by Eurostat, which oblige to revise all the maps each year. It is easy to imagine that such a revision should be a nightmare if the authors are obliged to create as many maps as the number of indicator multiplied by 30 (numbers of variants ...). But it is not the case because the majority of maps (25 out of 30) are automatically generated through the connection with a database where only most recent count variables are stored. When new values are published by Eurostat, all the maps (except the maps of potential) are automatically regenerated and the cartographic expert has only to validate and introduce minor changes.

We have clearly here a model of what could be a tool for territorial monitoring with automatic update each year. Of course, AIRE is not specifically targeted to policy makers and the specification of the ESPON monitoring tool would be different. But there is certainly food for thought in the idea of periodic automatic update of results.

2 Preliminary tests and results

Having detailed the Core Database Strategy in the first part of the technical report, we want here to focus the discussion on the technical implementation of this strategy. During the last months, the ESPON M4D Project has achieved an experiment on data collection for total population, from 1990 to 2010, and NUTS0 to NUTS3 levels following the principles of the strategy described above. This experimentation has provided some documentation and methods, which we want to develop for the next steps of data collection if the CDS is approved by the ESPON Coordination Unit.

The work has firstly consisted by making an overview of available information on Eurostat website on a selected number of indicators (section 2.1). Then, estimation methods have been systematically listed and defined starting from the data navigator II framework and applied to the dataset in order to obtain a full dataset (Section 2.2). After, automatic checking methods will be implemented to test the quality of the resulting dataset (Section 2.3). Finally, methods of spatial analysis, using space-time matrixes have been implemented in a R programme to propose innovative indicators based on the resulting ESPON Core database (Section 2.4).

The presentation of the results obtained after two months of experimentation in this Technical Report gives, according to us, a good preview on what can be expected from the ESPON M4D Project in term of data collection and can answer to the most recurrent question asked to the project dealing with the ESPON database during each ESPON Seminar: "What kind of new indicator do you expect to deliver to the ESPON Community, and when?".

Nevertheless, it is important to define the main properties of the targeted indicators in order to comply with the CDS strategy. These properties constitute in fact the technical objectives of data collection at regional level.

- **Collecting count data:** We expect to provide the basic indicators for the ESPON Community. Instead of collecting an unemployment rate, we prefer to focus the work on "CORE INDICATORS", such as unemployed population and active population (and its derived, e.g. by age, sex etc.).
- **Targeting time-series data:** All count data is collected systematically over a defined time period. Generally, we will focus on the 2000-2010 years, but for some indicators (e.g. total population), we have extended this analysis to the period 1990-2010. It depends generally of the data availability on Eurostat website.
- **Data collection for the all ESPON Space + Candidate Countries.** The data collection will target the entire ESPON 31 countries and the Candidate Countries (Turkey, Croatia, FYROM, Montenegro). Moreover, we want to precise that the CDS is fully compatible with the strategy elaborated for the data collection in European neighbouring countries.
- **Full completeness of dataset:** In the datasets which will be provided by the ESPON M4D project, all missing values will be estimated, by combining information coming from various sources (Eurostat, National Statistical Institutes, OECD etc.) or by estimating missing values.
- **Systematic description of estimated data, following the ESPON metadata template and the work realized through the Data Navigator**

II framework. There is never a single solution for estimating missing value, but a couple of solutions combining the Space (E), the Source (S), the time (T) or the thematic dimensions (I). The idea of this work is to, for each estimated value, explain clearly what was the strategy followed in order to know exactly how the data has been obtained. The ESPON metadata template allows to integrate this information precisely.

- **Coherencies between the NUTS divisions:** If possible (when the data is not provided through a survey, such as the Labour Force Survey) The data is collected systematically in the NUTS3, NUTS2, NUTS1 and NUTS0 divisions. The values of NUTS3, NUTS2, NUTS1 and NUTS0 are also linked between them (e.g. the sum of NUTS3 vales are equals to the sum of NUTS2 values, to the NUTS1 values and so on).
- **Limit the statistical breakdowns in time-serie data:** New censuses occurring at country level are generally the moment where statistical breakdowns appear in time-serie data (observed both on Eurostat or National Statistical Institutes data)¹. In order to overcome these problems, temporal breakdown analyse and correction methods are necessarily to implement.
- **Proposing innovative indicators based on these count data.** Having collected count data of quality which have been meticulously checked, it is possible to implement systematic methods, derived from spatial analysis for instance, which allow to valorise the content of the core database.

2.1 Collection of ESPON “Core” indicators through Eurostat website

A selected number of indicators of interest have been extracted from Eurostat website (table 1). These indicators constitute the “data bank” and the main source for the building of the future “ESPON Core” database. All these indicators are available for different NUTS levels (at NUTS2 level at least), for a minimum time period (5 years at least) and are described by count values (km², inhabitants, euros, employees etc.)

Eurostat table	Indicator of interest contained in the dataset	Time-period covered by the table	NUTS levels	Last update	Possible ESPON Classification
demo_r_d2jan	Population by age-classes and sex	1990-2010	NUTS0 NUTS1 NUTS2	06/12/2011	0201 Population structure
demo_r_d3avg	Annual average population by sex	1990-2010	NUTS0 NUTS1 NUTS2 NUTS3	25/11/2011	0201 Population structure
demo_r_gind3	Total population Live births Deaths Net migration	2000-2010	NUTS2 NUTS3	07/12/2011	0202 Natural changes

¹ An impressive example related to this topic is given by the Nigerian census, where population has suddenly decrease from 114 million of inhabitants to 88 million in 1991. For more details, see “Population et Société, Le recensement du Nigéria, 1992, N°272”. The document (in French) is available at the following URL:
http://www.ined.fr/fr/publications/pop_soc/bdd/publication/147/

Eurostat table	Indicator of interest contained in the dataset	Time-period covered by the table	NUTS levels	Last update	Possible ESPON Classification
nama_r_e3empl95	Employment by NACE R1 activities (8 classes)	1995-2008	NUTS0 NUTS1 NUTS2 NUTS3	25/05/2011	0602 Labour market 0702 Employment
lfst_r_lfe2en1	Employment by NACE rev.1.1 activities (9 classes)	1999-2009	NUTS0 NUTS1 NUTS2	14/12/2010	0602 Labour market 0702 Employment
lfst_r_lfe2en2	Employment by NACE rev.2 activities (11 classes)	2008-2010	NUTS0 NUTS1 NUTS2	08/07/2011	0602 Labour market 0702 Employment
lfst_r_lfp3pop	Economically active population	1999-2010	NUTS0 NUTS1 NUTS2 NUTS3	08/07/2011	0602 Labour market
lfst_r_lfu3pers	Unemployment by sex and age	1999-2010	NUTS0 NUTS1 NUTS2 NUTS3	29/11/2011	0602 Labour market 0603 Living conditions 0702 Employment
lfst_r_lfu2ltu	Long-term unemployment	1999-2010	NUTS1 NUTS2	05/09/2011	0602 Labour market 0603 Living conditions 0702 Employment
edat_lfse_10	Persons aged 25-64 with upper secondary education attainment (in % only)	2008-2010	NUTS0 NUTS1 NUTS2	27/05/2011	0601 Education
educ_renrlrg1	Number of students by level of education by sex (ISCED0-6)	1998-2009	NUTS0 NUTS1 NUTS2	07/12/2011	0601 Education
nama_r_e3gdp	Gross Domestic Product in Euros and PPS	1995-2008	NUTS0 NUTS1 NUTS2 NUTS3	05/08/2011	0701 Aggregated accounts
nama_r_e3vabp95	Gross value added by economic branches, NACE rev.1.1 activities (9 classes)	1995-2008	NUTS0 NUTS1 NUTS2 NUTS3	30/06/2011	0703 Production and costs per sector
ef_r_nuts	Number of agricultural holdings (in area and economic weight)	2000-2003- 2005-2007	NUTS0 NUTS1 NUTS2 NUTS3	31/01/2011	0101 Farm structure
demo_r_d3area	Area of the region	1990-2010	NUTS0 NUTS1 NUTS2 NUTS3	07/10/2011	0501 Land use and land cover by types

Table 1: The basis of ESPON M4D data collection: Information extracted from Eurostat website – References and content of the tables

When all the data have been extracted from Eurostat, keeping their original format (figure 4), systematic measures of data availability for the ESPON space and Candidate Countries² are made in order to have an overview on what is feasible or not in term of data estimation: It is clear that it is currently impossible to propose the total population in 1970 at NUTS2/3 levels, considering the low level of harmonised information available. Differences introduced between the estimation and the reality might be too important.

Nevertheless, keeping the example of total population, it is possible, by combining different Eurostat tables, to make available a complete time-serie for the period 1990-2010. The amount of available information is sufficient, considering that:

- Data were complete at NUTS3 level for all ESPON 31 and Candidate Countries for 2 years (figure 5).
- At NUTS2 level, most of the data were available from 1990 to 2010 allowing the use of precise methods of data estimation (retropolations and hierarchy harmonization, cf part. 2).
- Total population is probably one of the most general statistical indicators, which is generally available when using National Statistical Institutes resources.

Bilan démographique et taux bruts - régions NUTS 2 et 3 [demo_r_gind3]											
Dernière mise à jour	18.02.11										
Date d'extraction	05.07.11										
Source des données	Eurostat										
INDIC_DE	Population au 1er janvier - totale										
GEO/TIME	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
BE1	959 318	964 405	978 384	992 041	999 899	1 006 749	1 018 804	1 031 215	1 048 491	1 068 532	1 089 538
BE2	5 940 251	5 952 552	5 972 781	5 995 553	6 016 024	6 043 161	6 078 600	6 117 440	6 161 600	6 208 877	6 251 983
BE3					3 380 498	3 395 942	3 413 978	3 435 879	3 456 775	3 475 671	3 498 384
BG3	4 364 632	4 328 514	4 192 490	4 151 356	4 110 906	4 075 053	4 038 920	4 008 299	3 979 885	3 953 367	
BG4	3 826 244	3 820 954	3 698 605	3 694 485	3 690 367	3 685 996	3 679 830	3 670 991	3 660 353	3 653 184	

Figure 4 – First step, extraction of Eurostat tables

COUNTRY	LEVEL	Total population										Live births										Deaths										Migratory balance																
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010			
ESPON 31	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
	NUTS1	34	43	50	61	62	83	83	100	100	88	11	88	100	33	33	100	100	38	33	85	5	0	88	100	33	33	100	100	38	33	85	5	0	23	40	50	61	62	83	83	83	85	85	5	0		
	NUTS2	35	41	51	61	61	83	83	100	100	87	13	85	98	37	37	38	38	39	100	83	7	0	85	98	37	37	38	38	39	100	83	7	0	21	38	51	61	61	83	83	89	100	83	7	0		
	NUTS3	27	32	43	50	50	31	31	100	100	30	12	83	93	38	38	83	83	83	90	87	8	0	83	93	38	38	83	83	83	90	87	8	0	17	30	43	50	42	83	83	90	87	8	0			
Candidates (HR, TR, MK)	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
	NUTS1	7	0	7	14	14	14	14	14	100	100	0	0	0	7	7	14	14	14	100	100	0	0	0	0	7	7	14	14	14	14	14	14	0	0	0	0	7	7	14	14	14	14	14	14	0	0	
	NUTS2	3	0	10	13	13	13	13	13	100	100	0	0	0	10	10	13	13	13	13	100	100	0	0	0	0	10	10	13	13	13	13	13	13	0	0	0	0	10	10	13	13	13	13	13	13	0	0
	NUTS3	7	0	13	26	26	26	26	26	100	100	0	0	0	13	13	26	26	26	26	100	100	0	0	0	0	13	13	26	26	26	26	26	26	0	0	0	0	13	13	26	26	26	26	26	26	0	0
BE	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
	NUTS1	67	67	67	67	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	67	67	67	67	100	100	100	100	100	100	0	0		
	NUTS2	31	31	31	31	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	31	31	31	31	100	100	100	100	100	100	0	0		
	NUTS3	35	35	35	35	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	35	35	35	35	100	100	100	100	100	100	0	0		
BG	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
	NUTS1	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0		
	NUTS2	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0		
	NUTS3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0		
CZ	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
	NUTS1	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		
	NUTS2	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		
	NUTS3	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		
DK	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
	NUTS1	0	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		
	NUTS2	0	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		
	NUTS3	0	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		
DE	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
	NUTS1	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		
	NUTS2	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		
	NUTS3	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	100	100	0	0		

Figure 5 – Second step, calculation of data availability in Eurostat tables

² Cf Annex 1 to see the data availability of selected tables.

2.2 Estimation of missing values using the ESTI framework and the ESPON metadata model

When analysing the degree of completeness of Eurostat's targeted indicators (cf Annex 1), in no case it is possible to have a complete dataset for a time-period longer than 2 two years for all the NUTS levels for the ESPON Area (even for the total population, probably the most complete dataset in Eurostat). In this context, Eurostat missing values are also not systematically provided by National Statistical Institutes, because of changing of territorial divisions (Poland, Denmark) or protecting historical data to free access (Romania, Slovakia).

Consequently, for building time-series data and estimating missing values, alternative solutions are required – out of Eurostat and NSI. By the past, ESPON has produced interesting materials for overcoming the problem raised by missing values. The ESTI framework in the Data Navigator 2 project is one of the milestones to consider in that topic (figure 6).

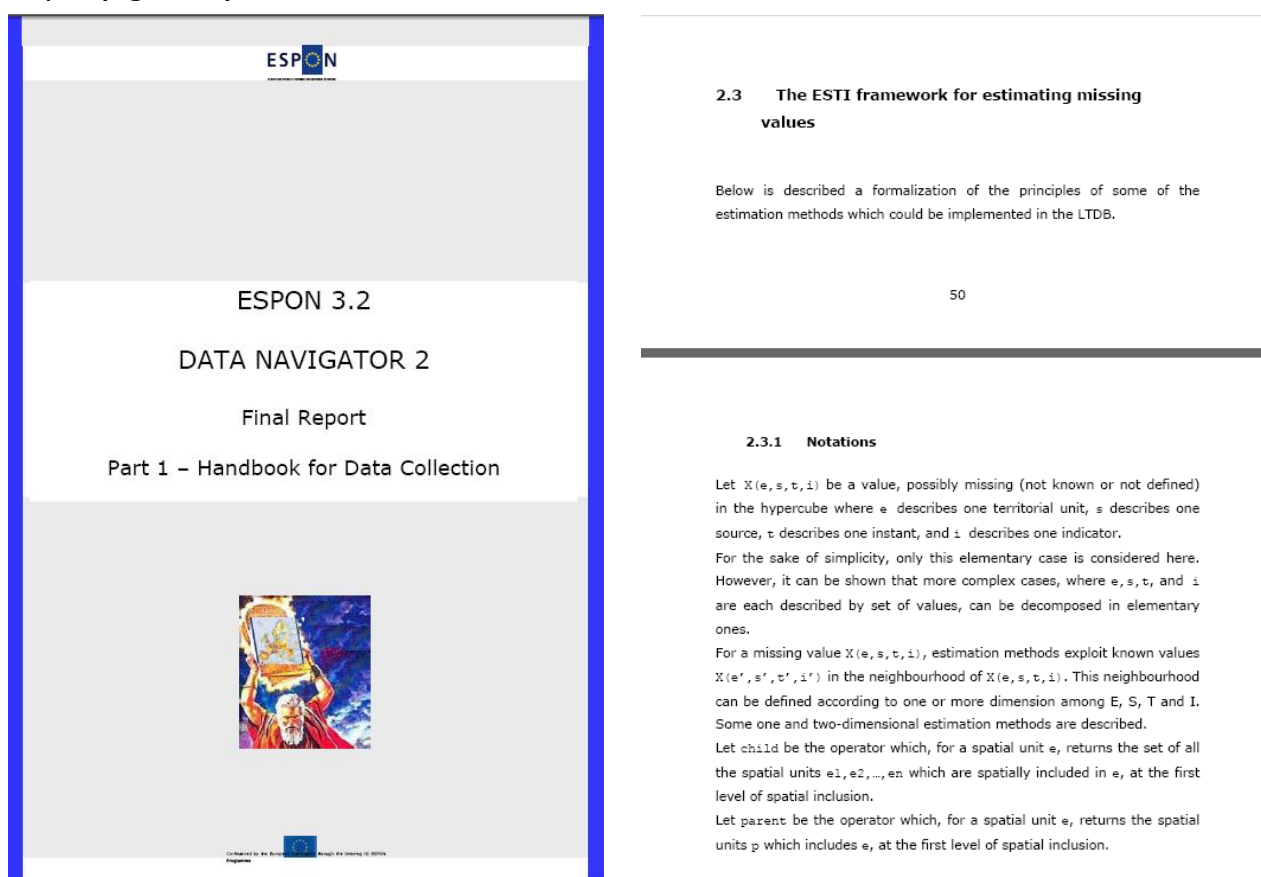


Figure 6 – The ESTI framework project, a milestone for estimating missing values

The ESTI framework proposed a couple of solutions, based on four dimensions:

- **The Spatial dimension (E):** Estimate the missing values thanks to values known at an upper, lower or at the same hierarchical level.
- **The Source dimension (S):** Alternative sources of information can be used when the main source does not provide the targeted information.
- **The Time dimension (T)** uses information provided in time before or after the value to estimate.

- **The Thematic dimension (I):** The idea is to replace the missing value by a known value coming from an alternative indicator, controlled by a correlation factor.

As regard to this useful background, we have tried to improve the ESTI framework on the basis of empirical situations, provided by the experience of the estimation of missing values on the total population 1990-2010 dataset coming from Eurostat.

The resulting document (entirely available in Annex 2), describes 12 one-dimensional and 8 multi-dimensional methods derived from the 4 dimensions of the ESTI framework allowing to estimate missing values in a couple of various situations. The aim of this document is twice. Firstly, to help users (who have not a specific background in statistics) to overcome problems raised by missing values. Secondly, to find an efficient system for describing the methodology used for the estimation in metadata.

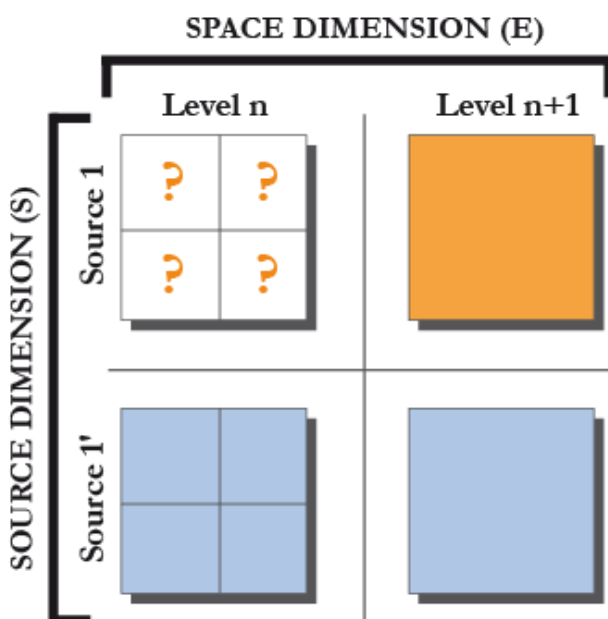
This documentation is scalable. It will be improved and expanded later regarding to user and experts feedbacks. Nevertheless, it is currently structured in different parts:

- **Conditions of use**

- ➔ No data at lower level of the hierarchy in $X(e,s,t,i)$.
- ➔ Information available at a upper level of the hierarchy $X((parent(e),s,t,i)$
- ➔ Data at lower level of the hierarchy available from another organism of statistics $X(e,s',t,i)$

This part describes some preliminary conditions for using a given method of data estimation. Indeed, it is impossible to estimate data without having a minimum information linked to the data to estimate.

- **Graphic illustration**



Considering that some estimation methods are sometimes difficult to explain (when using several dimensions in particular), we have tried to illustrate graphically each method simply to ease the understanding of them.

- **Textual explanation**

Estimation based on source and space dimension (S, E)
 The values have been collected from [*name of the organism of statistics*], then, adjusted by ESPON M4D Project in a way that the sum of the values of the children units (e1, e2...en) are equal to the value of the parent unit (e)
 In this case, [*name of the hierarchy, e.g. NUTS3*] values coming from [*name of the organism of statistics of X(e,s',t,i)*] is adjusted to the [*name of the hierarchy, e.g. NUTS0*] values delivered by [*Name of the organism of statistics of X(e,s,t,i), e.g. generally Eurostat*]

In this part, we aim to find a harmonized way for describing the methodology of the estimation method. For the moment, in the ESPON metadata, this description is made by a free-text. Consequently, there are too many different descriptions of data estimations, and very often there are too few elements of description for understanding correctly how the data has been estimated. With this information box, we propose a common definition of the estimation methodology. The only parameters to change are described in italics (e.g. the name of the organism of statistics, the level of data harmonization, the year used for the temporal extrapolation, etc.)

- **Mathematic formalization**

$$X(e, s, t, i) = aX(e, s', t, i) \text{ where } a = \frac{X(\text{parent}(e), s, t, i)}{X(\text{parent}(e), s', t, i)}$$

Following the mathematic syntax provided by the ESTI framework, we have provided the mathematic formalization of each estimation method³.

- **Example**

ID	Object type	Value
FR11	NUTS2	25
FR12	NUTS2	25
FR111	NUTS3	n/a
FR112	NUTS3	n/a
FR121	NUTS3	n/a
FR122	NUTS3	n/a

ID	Object type	Value
FR111	NUTS3	17
FR112	NUTS3	10
FR121	NUTS3	15
FR122	NUTS3	9

ID	Val.	Dist	Dev	Estim	Check
FR11	25	27(4)	0,93(5)		25 (7)
FR12	25	26(4)	0,96(5)		
FR111	17(3)			15,7 (6)	
FR112	10(3)			9,3 (6)	
FR121	15(3)			14,4 (6)	
FR122	9(3)			8,6 (6)	

HOW TO PROCEED ?

- (1) Download data at lower level from another organism of statistics (s').
- (2) Save the table downloaded from (s') in a dedicated space (s'')
- (3) Merge (s) and (s'') in the same table
- (4) Sum the values coming from (s'') of the childs (e1,e2...en) available in s for its parent (e)
 Ex: if the data is available at NUTS2 level for Eurostat data, aggregate the NUTS3 coming from the other organism of statistics in NUTS2.
- (5) Calculate the deviation (α) by dividing X(parent(e),s,t,i) and X(parent(e),s',t,i).
- (6) Use the deviation calculated above as correlation factor: X(e,s,t,i) = α X(e,s',t,i)
- (7) Check if the sum of the estimation at lower level is equal to the value delivered by Eurostat at upper level

In ESPON tables: Estimation of population in Scotland and Northern Ireland from 2001 to 2006, where NUTS3 values are missing and NUTS0 is available.

We propose here to formalize the procedures of data estimations as regard to regular concrete situations. We try systematically to explain step by step the methodology employed for estimating data by using the ESTI terminology. In that order, we expect that non-specialist of data estimation can appropriate themselves the methods explained in the documentation.

³ We precise that this formalization need to be improved with specialist of mathematic formalization.

• **“Intelligent” labels**

When a lot of estimation methods are used in a dataset, the risk is high to make confusion between existing labels, not making possible the differentiation between the information coming from official sources (Eurostat, National Statistical Institutes...) and estimated values. On top of that, we consider that it could be interesting to understand directly in the dataset what method of estimation has been used for filling missing values. This is why we propose a codification for labelling the values (figure 7).

		pop_tot		pop_tot	
		1990	source	1991	source
nit cod	Name				
HR01	Sjeverozapadna Hrvatska	1629932	TE1e	1646710	4a
HR02	Srednjsnja i Istocna (Pac)	1565290	TE1e	1557342	4a
HR03	Jadranska Hrvatska	1582778	TE1e	1580212	4a
HU10	Közép-Magyarország	2966523	1a	2970223	1a
HU21	Közép-Dunántúl	1117989	1a	1118848	1a
HU22	Nyugat-Dunántúl	1006781	1a	1006535	1a
HU23	Dél-Dunántúl	1017025	1a	1016707	1a
HU31	Észak-Magyarország	1323508	1a	1319422	1a
HU32	Észak-Alföld	1547520	1a	1546555	1a
HU33	Dél-Alföld	1395477	1a	1394850	1a
IE01	Border, Midland and V	949890	SE1a*	947629	SE1a*
IE02	Southern and Eastern	2557080	SE1a*	2573348	SE1a*

Value coming from a national statistical institute (the label begin by another number than "1")

Value estimated using the temporal (T) and space (E) dimensions

Value coming from Eurostat (the label begin by "1")

Value estimated using the source (S) and space (E) dimensions

Figure 7 – Extract of the dataset of population 1990-2010

For instance, when adopting the labelling for the table of total population 1990-2010 (figure 8), it is immediately possible to note that this file contains 4 labels related to Eurostat tables (labels beginning by 1a, 1b, etc), 35 labels are related to various National Statistical Institutes (beginning by 3a,4,13 etc), 9 labels related to one-dimensional estimation methods (beginning by E or S or T or I) and 43 multi-dimensional methods (beginning by a couple with at least two letters of the acronym ESTI). Having done this labelling, it is also easy to extract only values coming from Eurostat, NSI or estimations.

Source Reference		Source Reference	
Label	1	Label	E1a
Label	1a	Label	E1b
Label	1*a	Label	E1c
Label	1*b	Label	E2a
Label	2a	Label	E2b
Label	3a	Label	T1a
Label	4a	Label	T1b
Label	4b	Label	T4b
Label	4c	Label	T4c
Label	5a	Label	SE1a
Label	6a	Label	SE1a*
Label	8a	Label	SE1b
Label	9a	Label	SE1c
Label	10	Label	SE1d
Label	11	Label	SE1e
Label	12	Label	SE1f
Label	13	Label	SE1g
Label	14	Label	SE1h
Label	15	Label	SE1i
Label	16	Label	SE1j
Label	17	Label	SE1k
Label	18	Label	TE1a
Label	19	Label	TE1b
Label	20	Label	TE1c
Label	21		
Label	22		
Label	23		
Label	24		
Label	25		
Label	26		

Figure 8 – Extract of the labels contained in the metadata of the dataset of population 1990-2010

2.3 Data quality control, example of population check by level

As described in precise terms by Martin Charlton and Paul Harris in the ESPON Technical Report on outlier detection⁴, *“exceptional values may arise during the coding, transmission, manipulation or editing of data. They may not be noticed until late in the day or not at all, particularly if the data that is a candidate for input to the Database has been the outcome from a series of sequential manipulations. Exceptional values may also arise in the measurement of data; perhaps a sensor on some measuring equipment has been incorrectly calibrated or is faulty; a respondent can tick the wrong box in completing a survey form”*.

Anyway and even if the technical process of data creation is very precise, there is ever a risk of integrating involuntary inputs errors in the datasets created. This is the reason why the definition of automatic check of outliers and errors is too important.

The ESPON Database 1 has produced a battery of filters against which potential input data can be tested in R programme with experimental datasets. We propose in the second phase of the project to enlarge the process of data quality control in order to make possible systematic check for indicators that aim to integrate the ESPON Core Database.

To illustrate this operational phase of the quality control, the ESPON M4D project has implemented a **“hierarchical data check”** in a R programme (cf in Annexe 3). The idea of this simple check is to know whether the amount of data collected at NUTS3 level is equal to the data collected at other NUTS levels. This check provides two kind of tables:

The first one makes the sum of the count indicator at each NUTS level and compares in absolute (dif) and relative terms (pct) the gap between the different NUTS hierarchical levels 0-1-2-3 (figure 9). In this case, the table shows that a difference of 6835 inhabitants exists between the NUTS3 and the NUTS0 for the total population in 2001 and a gap of 3041992 inhabitants has been introduced between the NUTS0 and the NUTS1 nomenclature in 2003⁵.

⁴ CHARLTON M., HARRIS P., 2011, *Spatial analysis for quality control*, ESPON Technical Report, 81 p.

⁵ We precise that detected errors have been corrected thanks to this hierarchical check...

HIERARCHY	x0	x1	x2	x3	dif01	dif02	dif03	dif01pct	dif02pct	dif03pct
P1990	546126227	546126226	546126227	546126217	-1	0	-10	0	0	0
P1991	548905003	548905004	548905002	548904993	1	-1	-10	0	0	0
P1992	551014207	551014208	551014210	551013944	1	3	-263	0	0	0
P1993	554149996	554149995	554150001	554149173	-1	5	-823	0	0	0
P1994	556426840	556426841	556426845	556425828	1	5	-1012	0	0	0
P1995	558108454	558108454	558108457	558106964	0	3	-1490	0	0	0
P1996	560027827	560027829	560027829	560025783	2	2	-2044	0	0	0
P1997	562153481	562153482	562153484	562150644	1	3	-2837	0	0	-0,001
P1998	564016503	564016504	564016506	564012743	1	3	-3760	0	0	-0,001
P1999	566077808	566077808	566077809	566073413	0	1	-4395	0	0	-0,001
P2000	568004933	564930292	568004931	567999126	-3074641	-2	-5807	-0,541	0	-0,001
P2001	569767676	566701900	569767679	569760841	-3065776	3	-6835	-0,538	0	-0,001
P2002	571537613	568483541	571537610	571537610	-3054072	-3	-3	-0,534	0	0
P2003	573930976	570888984	573930976	573930971	-3041992	0	-5	-0,53	0	0
P2004	576500178	576500178	576500178	576500177	0	0	-1	0	0	0
P2005	579284958	579284958	579284959	579284962	0	1	4	0	0	0
P2006	581831623	581831624	581831625	581831628	1	2	5	0	0	0
P2007	584422763	584422764	584422763	584422766	1	0	3	0	0	0
P2008	587390215	587390215	587390215	587390215	0	0	0	0	0	0
P2009	590712839	590712839	590712839	590712839	0	0	0	0	0	0
P2010	592943266	592943263	592943264	592943251	-3	-2	-15	0	0	0

Figure 9 – Hierarchical data check 1
Comparison of the sum of the count indicator by hierarchical level and years

The second table (figure 10) is useful to target hierarchical incoherencies by country when a gap appears during the first test. In the total population example, errors have been introduced for Italy from 1990 to 2005 between the NUTS3 and NUTS0 levels for example. It is due to the fact that the NUTS0 population is delivered by Eurostat and the NUTS3 values by I-STAT without hierarchical harmonization. Thanks to this check, these logical input errors have been adjusted quickly.

CHECK	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
AT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	265167
BE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86825	0
BG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CZ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-304
HR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HU	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
IE	1070	-4723	-7700	-6900	-5900	-6600	-10500	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IT	66547	72493	73931	71591	68342	63803	57234	49668	39729	26110	14823	2836	-18267	-12427	-9156	-3223	0	0	0	0	1E+06
LI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LV	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	660	0	0	0	0
NL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PL	0	0	0	0	0	0	0	0	0	0	390256	0	0	0	0	0	0	0	0	0	0
PT	0	0	-2	-5	-5	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SE	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UK	8089	-1938	-19401	-26917	-32111	-32436	-38003	-48491	-51044	-58224	10869	-200	-118	-168	-130	-336	-138	-3666	-6	0	0

Figure 10 – Hierarchical data check 2
Comparison of the sum of the count indicator by hierarchical level, years and country (NUTS3-NUTS0 level in this case)

2.4 Transformation of indicators by functional smoothing

Since the beginning, one of the major strength of ESPON Program has always been the production of indicators based on multimodal accessibility of regions and cities by air, train or road. Thanks to the excellence of research teams involved in this area of research (S&W, RRG, MCRIT, INRETS, CESA, ...), ESPON has been able to produce regularly maps of potential of population or GDP under various assumption of friction of distance or opening of borders. In terms of comparative advantage, ESPON has clearly here a strong added value as compared to Eurostat, OECD and even DG Regio (Multimodal accessibility is one of the only ESPON map that has been reproduced in a Cohesion Report). This advantage is currently maintained by new projects in Priority 1 and Priority 2 dealing with local accessibility (TRACKS) or to simulation of the effect of various scenarios of accessibility in the future.

Considering this specificity of ESPON, we propose to push further the advantage by developing a family of new indicators based on functional smoothing of regional data available at NUTS2 or NUTS3 levels. This strategy is directly related to the Core Database Strategy because it proposes to transform all the count variables into functional potential and then to compute the related indicators derived from this count variables in order to achieve functional smoothing. Let us explain briefly the principle behind that proposal through a simple example, the analysis of population variation at NUTS3 level between 2000 and 2010, resulting of the data creation process described in part 2.1, 2.2 and 2.3.

- **The initial map of population variation at NUTS3 level is not correct because strongly dependant of the heterogeneity of territorial divisions (figure 11).** As explained by the ESPON 2006 project on Modifiable Area Unit Problem, such a map is not really useful for a functional approach of European territories because it is based on a mixture of regions of different size (big to small) and different nature (urban, mixed, rural). In many countries like Poland and Germany, the map displays demographic decreasing core-cities surrounded by demographic growing suburb areas. But in other countries like France or Spain this distinction is not visible because territorial units are mixing both type of regions in a single entity. As a result, the map is not useful at all for political decision because it is too heterogeneous in terms of territorial divisions.
- **The transformation into a map of population variation in a functional neighbourhood of 1 hour by road (figure 12)** eliminates the majority of problems identified previously. What is measured here is not only the internal situation of the regions but a combination between their internal dynamic and the dynamic of neighbouring regions (e.g. with a weight of 0.5 for regions located at 1 hour, 0.05 for regions located at 2 hours, ...). This map is easily obtained through a function of gaussian smoothing applied to time distance by car between NUTS3 regions (source: RRG). The time of computation is less than 2 minutes on a normal PC computer, which means that all count variables available at NUTS3 level could be easily functionally smoothed, producing related indicator of unemployment rate, GDP/capita in a functional neighbourhood of 1 hour by car. It is certainly of major political interest for regional authorities to measure not only their internal situation but also the one of the neighbouring area. It is indeed not the same to be a region with high unemployment rate surrounded by region with low unemployment rate (e.g. Seine-Saint-Denis in France) or a region with high unemployment rate surrounded by region with equivalent or highest levels of unemployment (e.g. majority of regions of eastern Germany).

- **Variations of time span and variations of opening of borders.** Following the same principle, we can derive alternative maps, for example by considering that the mobility of people is increase and that the influence of neighbouring regions is decreasing lower with distance (figure 13). We can also examine what happens when border are closed and when contributions to current situation is only national (figure 14) ...

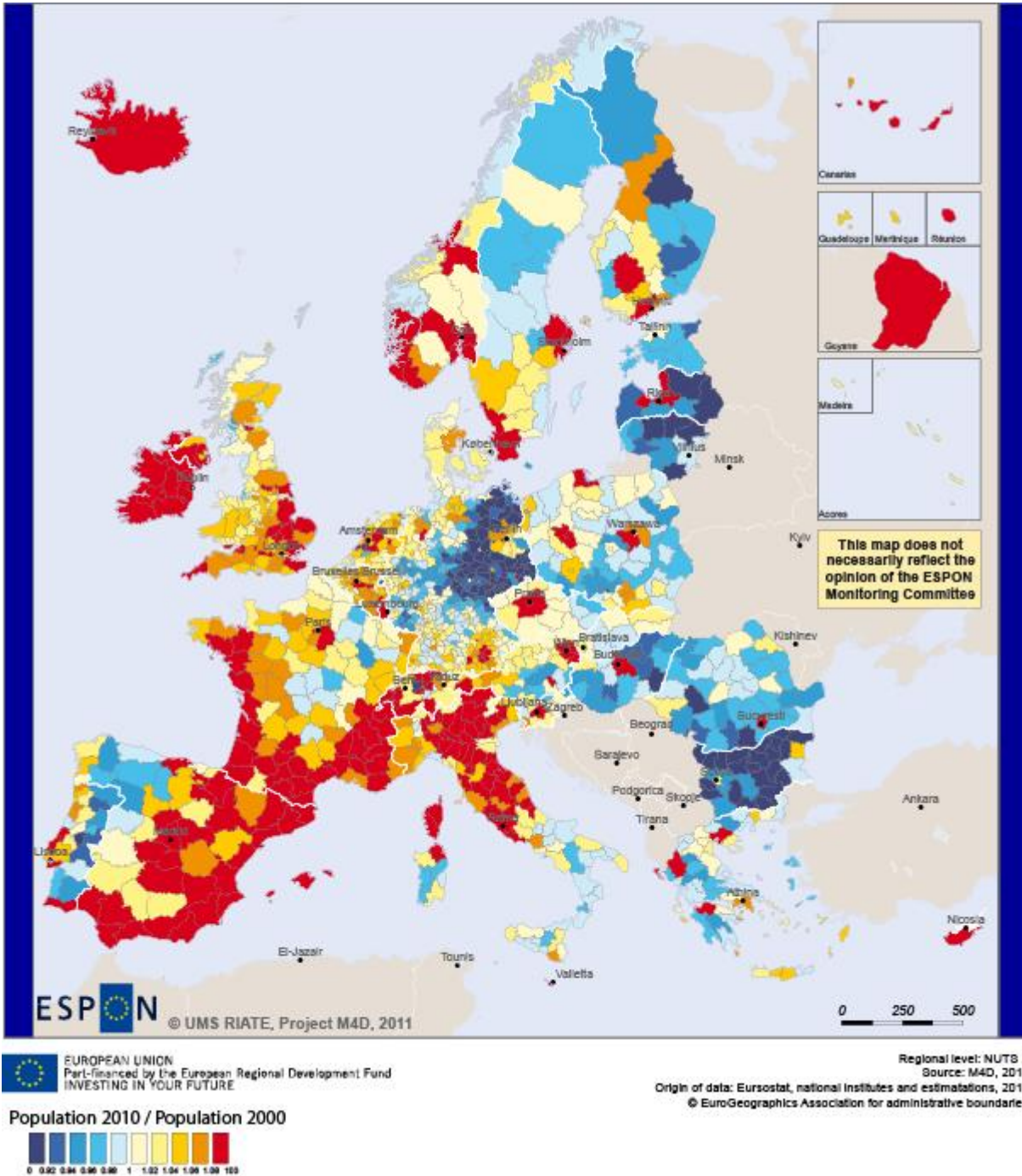
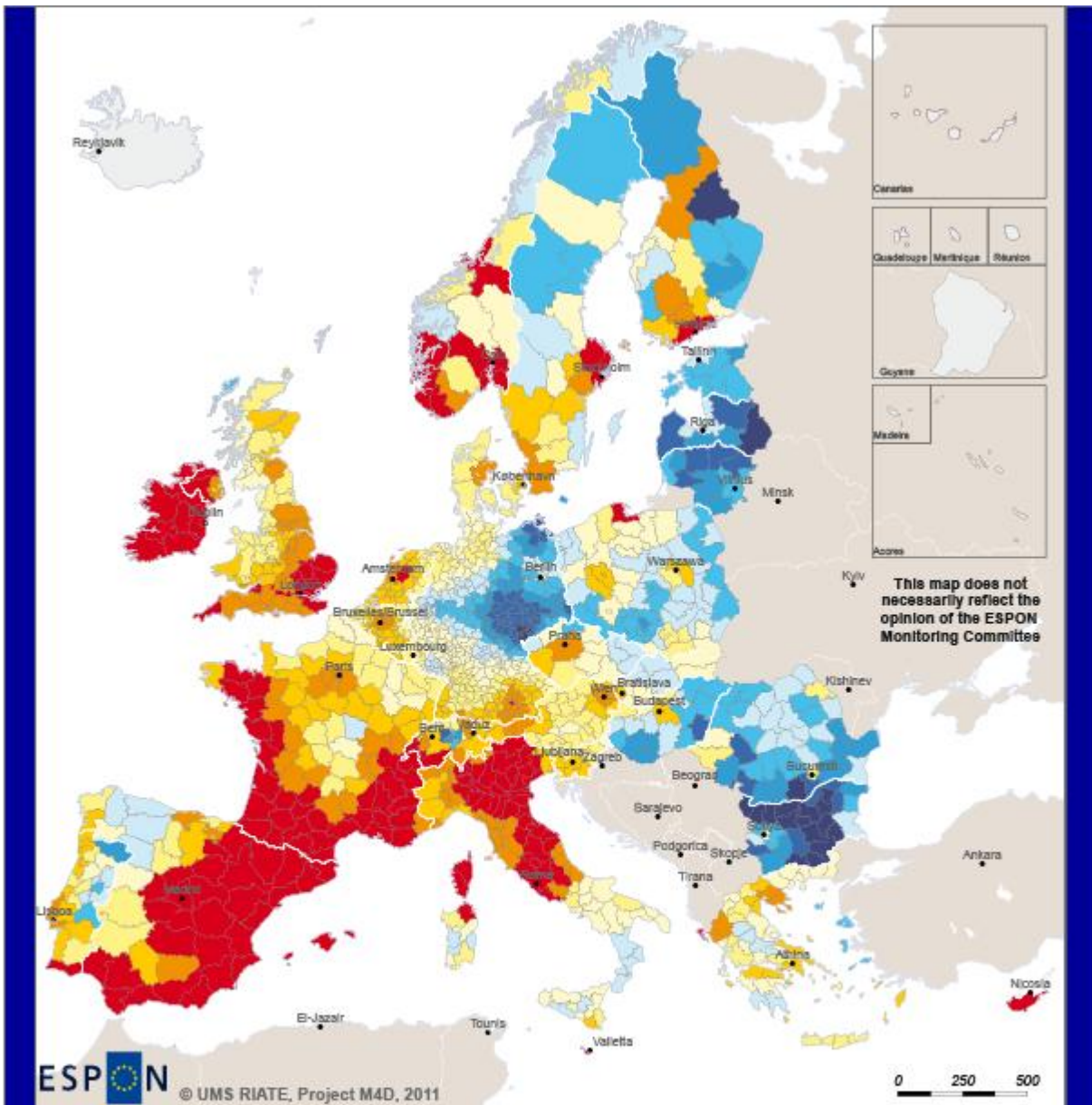


Figure 11 – Evolution of population between 2000 and 2010



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

Regional level: NUTS 3

Source: M4D, 2011

Origin of data: populations - Eurostat, national institutes and estimations, 2011

distance matrix - Schürmann et Spiekermann, ?

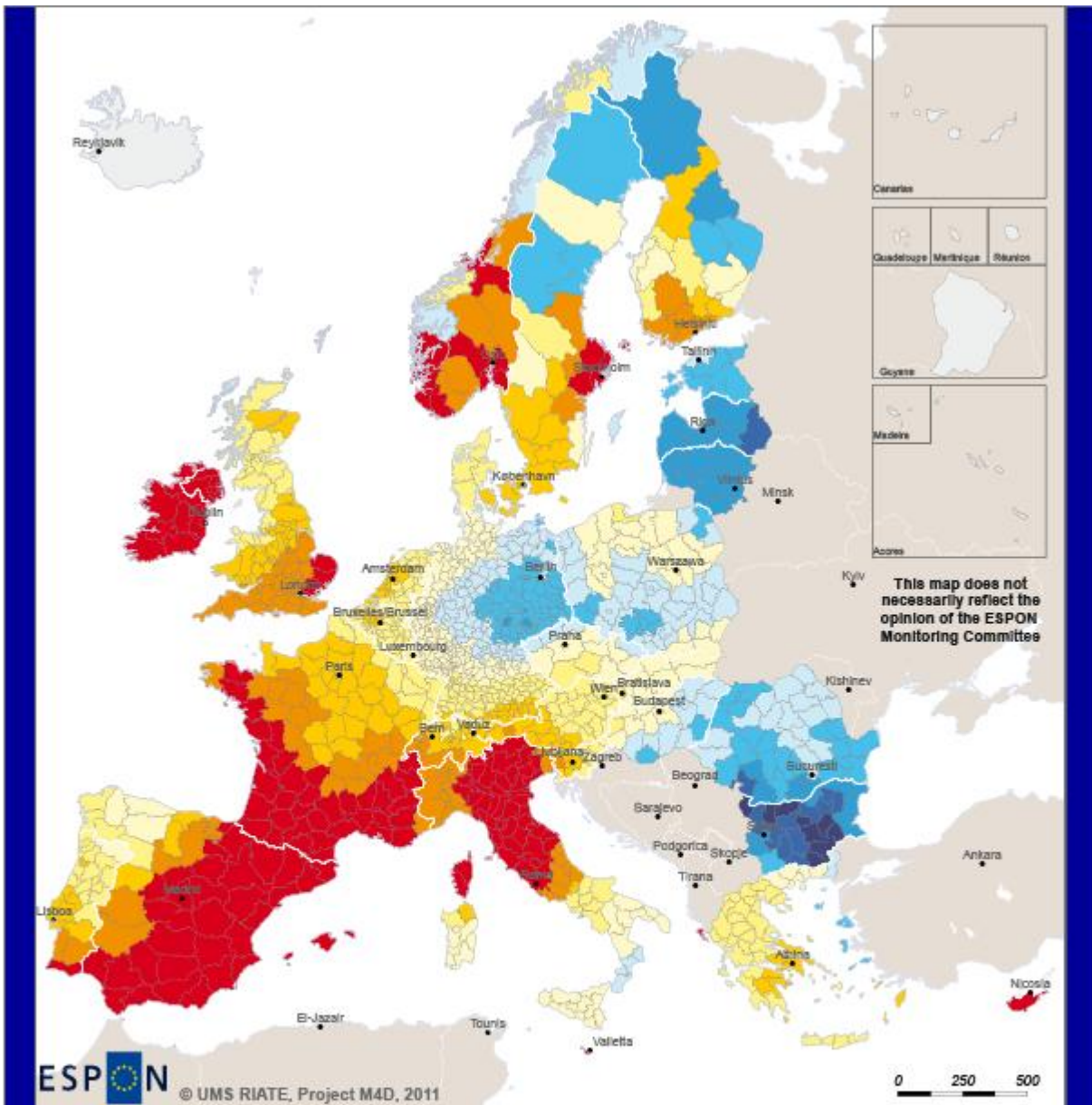
© EuroGeographics Association for administrative boundaries

Population 2010 / Population 2000



Populations are potential populations
in a gaussian smooth of a 60 min
radius by car with fully opened borders

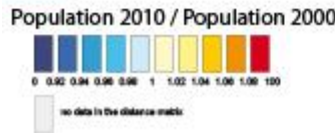
Figure 12 – Evolution of potential population between 2000 and 2010
in a 60 mn radius




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

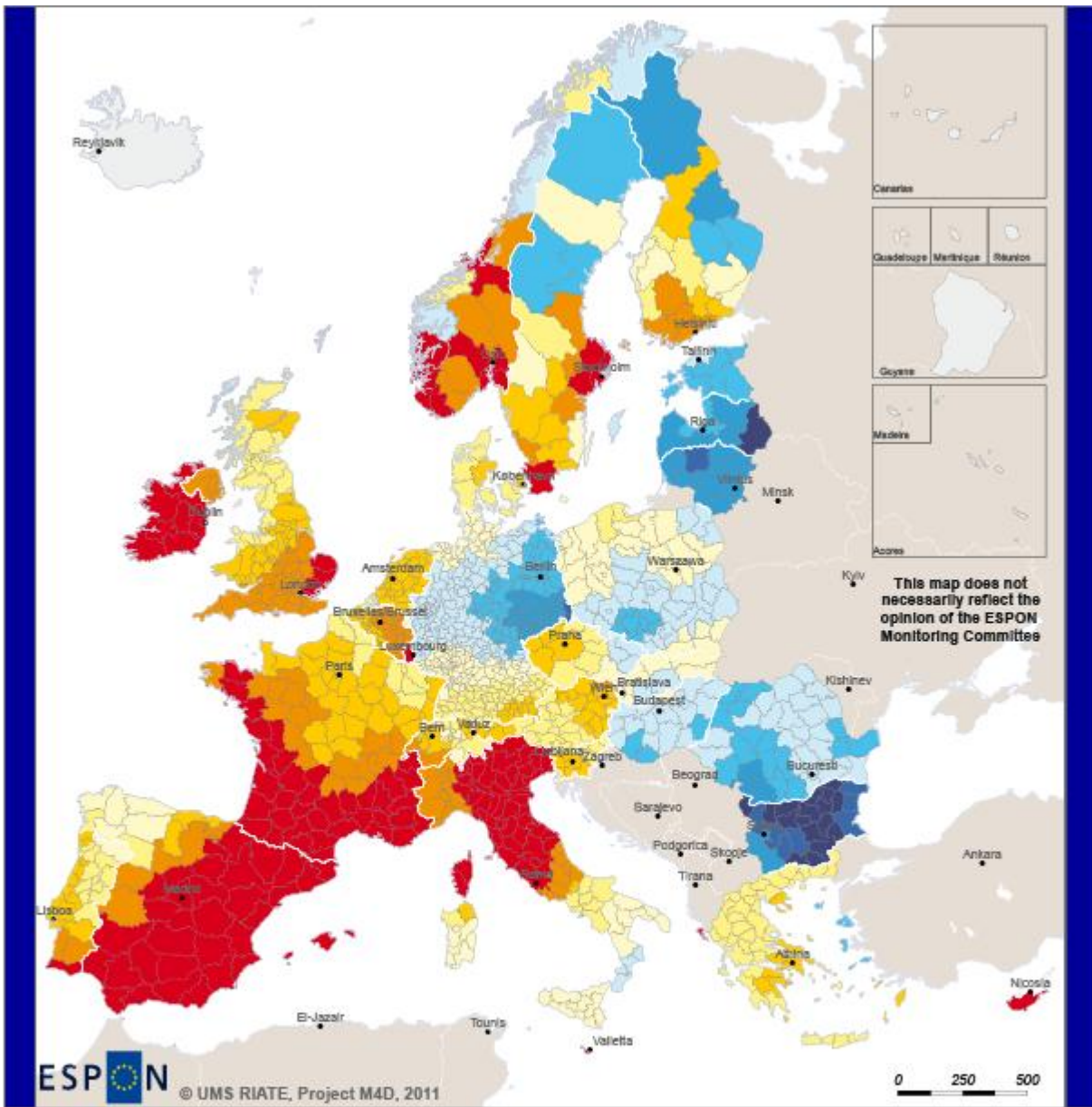
Regional level: NUTS 3
Source: M4D, 2011

Origin of data: populations - Eurostat, national institutes and estimations, 2011
distance matrix - Schürmann et Spiekermann, ?
© EuroGeographics Association for administrative boundaries



Populations are potential populations in a gaussian smooth of a 120 km radius by car with fully opened borders

Figure 13 – Evolution of potential population between 2000 and 2010 in a 120 km radius – opened borders



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

Regional level: NUTS 3

Source: M4D, 2011

Origin of data: populations - Eurostat, national institutes and estimations, 2011

distance matrix - Schürmann et Spiekermann, ?

© EuroGeographics Association for administrative boundaries

Population 2010 / Population 2000

Populations are potential populations
in a gaussian smooth of a 120 mn
radius by car with fully closed borders

no data in the distance matrix

Figure 14 – Evolution of potential population between 2000 and 2010
in a 120 mn radius – closed borders

Conclusion

*Yes, there are two paths you can go by, but in the long run
There's still time to change the road you're on
And it makes me wonder*

Led Zeppelin, Stairway to Heaven

It is certainly not obvious to choose between the two paths of development of the ESPON database that has been described at the beginning of this report in section 1.1 and there is no doubts that it makes us wonder ...

On the one hand, the « Cosmetic » approach offers a relative security because it follows a normative path where all the effort is put on registration of work produced by other ESPON TPG, with a minor role of M4D project which is just a kind of gate keeper and designer of results. It is certainly not a very exciting path for a TPG composed of researchers with advanced skills in cartography, spatial analysis, computer science... but with few skills in communication, public relation and accounting.

On the other hand the « Sustainable approach » proposes a more risky strategy as it suggests to follow an innovative path where, by definition, we cannot be sure of the results that will be finally obtained. We have tried to demonstrate by the work of the last months that the Core Database Strategy is feasible and likely to obtain results in a short delay. But time is not expansible and it will be difficult at the same time to push this strategy without reducing somewhere the effort on the pure compilation of data produced by TPGs.

What we propose to ESPON CU and ESPON MC is to let us develop the CDS experience during 6 months until de Denmark meeting. And to decide at this moment only if we are right or if we should have done better by sticking to a purely normative role.

Bibliography

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

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

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

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

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

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

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

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

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

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

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

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

PLUMEJEAUD C., VINCENT J.M., GRALAND C., BIMONTE S., MATHIAN H., GUELTON S., BOULIER J., GENSEL J., 2008, « *HyperSmooth: A System for Interactive Spatial Analysis Via Potential Maps* », Lecture Notes in Computer Science, Volume 5373, 4-16

Annexe 1 – Data availability in Eurostat tables

Situation in September 2011 (note: some tables from Eurostat have been updated since that time).

A.Table demo_r_gind3	29
B.Table demo_r_birth_deaths_N3.....	31
C.Table demo_r_d2jan	33
D.Table demo_r_d3avg	36
E.Table educ_renrlrg1	38
F.Table lfst_r_lfe2en1	41
G.Table lfst_r_lfe2en2	44
H.Table lfst_r_lfp3pop	46
I.Table lfst_r_lfu2ltu	50
J.Table lfst_r_lfu3pers	51
K.Table nama_r_e3empl95	56
L.Table nama_r_e3gdp.....	62
M.Table nama_r_e3vabp95	64

A. Table demo_r_gind3

COUNTRY	LEVEL	Total population									Live births									Deaths									Migratory balance																
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
ESPON 31	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	34	43	50	61	62	83	83	100	100	88	11	88	100	33	33	100	100	38	33	85	5	0	88	100	33	33	100	100	38	33	85	5	0	23	40	50	61	62	83	83	100	100	88	11
	NUTS2	35	41	51	61	61	83	83	100	100	87	13	85	98	97	97	38	38	93	100	83	7	0	85	98	97	97	38	38	93	100	83	7	0	21	38	51	61	61	83	83	100	100	83	7
	NUTS3	27	32	43	50	50	31	31	100	100	30	12	63	93	38	38	63	63	63	30	87	6	0	63	93	38	38	63	63	63	30	87	6	0	17	30	43	50	42	63	63	30	87	6	0
Candidates (HR, TR, MK)	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	7	0	7	14	14	14	14	14	14	100	100	0	0	7	7	14	14	14	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	14	14	14	14	14	100	100
	NUTS2	3	0	10	13	13	13	13	13	13	100	100	0	0	10	10	13	13	13	100	100	0	0	0	0	10	10	13	13	13	13	100	100	0	0	0	10	10	13	13	13	13	100	100	0
	NUTS3	7	0	13	26	26	26	26	26	26	100	100	0	0	13	13	26	26	26	100	100	0	0	0	0	13	13	26	26	26	26	100	100	0	0	0	13	13	26	26	26	26	100	100	0
BE	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	67	67	67	67	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	67	67	67	67	100	100	100	100	100	100	100	
	NUTS2	31	31	31	31	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	31	31	31	31	100	100	100	100	100	100	100	
	NUTS3	35	35	35	35	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	35	35	35	35	100	100	100	100	100	100	100	
BG	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	100	
	NUTS2	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	100	
	NUTS3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	100	100	
CZ	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	100	100	100	100	100	100	100	
	NUTS2	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	100	100	100	100	100	100	100	
	NUTS3	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	100	100	100	100	100	100	100	
DK	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	0	0	0	0	0	0	0	0	0	0	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0	0	0	100	100	100	
	NUTS2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	0	100	100	100
	NUTS3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	0	100	100	100
DE	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0	0	0	100	100	100	
	NUTS2	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0	0	0	100	100	100	
	NUTS3	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0	0	0	100	100	100	
EE	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0	0	0	100	100	100	
	NUTS2	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0	0	0	100	100	100	
	NUTS3	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0	0	0	100	100	100	
IE	NUTS0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	NUTS1	0	0	0	0	0	0	100	100	100	100	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	0	0	0	0	0	0	0	100	100	100
	NUTS2	0	0	0	0	0	0	0	100	100	100	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	0	0	0	0	0	0	0	100	100	100
	NUTS3	0	0	0	0	0	0	0	100	100	100	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	100	100	0	0	0	0	0	0	0	0	0	100	100	100

COUNTRY	LEVEL	Total population									Live births									Deaths									Migratory balance								
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2000	2001	2002	2003	200																				

C. Table demo_r_d2jan

COUNTRY		LEV	Age pyramid (5 years age class)																					
YEAR			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	
ESPON 31	NUTS0		81	90	97	97	97	97	97	97	97	100	100	100	100	100	100	100	100	100	100	100	94	6
	NUTS1		76	86	93	93	93	93	93	93	93	100	100	100	100	100	100	100	100	100	100	100	85	6
	NUTS2		66	75	92	92	92	94	95	96	96	98	98	98	98	98	98	98	98	98	100	100	83	8
Candidates (HR, TR, MK)	NUTS0		0	0	0	0	33	33	33	33	33	33	33	33	67	67	67	67	67	67	100	100	0	0
	NUTS1		0	0	0	0	7	7	7	7	7	7	7	7	14	14	14	14	14	14	100	100	0	0
	NUTS2		0	0	0	0	3	3	3	3	3	3	3	3	13	13	13	13	13	13	100	100	0	0
BE	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
BG	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
CZ	NUTS0		0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
DK	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	0	0
DE	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		87	87	87	87	87	100	100	100	100	100	35	100	100	100	100	100	100	100	100	100	0	0
EE	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
IE	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	NUTS2		0	0	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
GR	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
ES	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		83	83	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
FR	NUTS0		0	0	0	0	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		83	83	83	83	83	83	83	83	83	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		85	85	85	85	85	85	85	85	85	100	100	100	100	100	100	100	100	100	100	100	0	0

COUNTRY		LEV	Age pyramid (5 years age class)																				
YEAR			1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
SI	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
SK	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		0	0	0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
FI	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		100	40	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
SE	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
UK	NUTS0		0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS1		0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0	0
	NUTS2		0	0	83	95	95	95	95	95	95	95	100	100	100	100	100	100	100	100	100	0	0
IS	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
LI	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
NO	NUTS0		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
CH	NUTS0		0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
HR	NUTS0		0	0	0	0	0	0	0	0	0	0	0	0	100	100	100	100	100	100	100	0	
	NUTS1		0	0	0	0	0	0	0	0	0	0	0	0	100	100	100	100	100	100	100	0	
	NUTS2		0	0	0	0	0	0	0	0	0	0	0	0	100	100	100	100	100	100	100	0	
MK	NUTS0		0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS1		0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
	NUTS2		0	0	0	0	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	0
TR	NUTS0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	0
	NUTS1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	0
	NUTS2		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	0

D. Table demo_r_d3avg

COUNTRY	LEV	Total population										Total population, under 15 years										Total population, 15-64 years										Total population, above 65 years													
ESPN 31	NUTS3	27	32	43	50	50	31	31	100	100	90	12	15	17	28	39	42	43	75	37	39	30	12	15	17	28	33	42	43	75	37	39	30	12	15	17	28	39	42	43	75	37	39	30	12
Candidates (HR, TR, MK)	NUTS3	7	0	19	26	26	26	26	100	100	0	0	0	0	0	0	0	0	19	100	100	0	0	0	0	0	0	0	0	19	100	100	0	0	0	0	0	0	0	0	19	100	100	0	
BE	NUTS3	35	35	35	35	100	100	100	100	100	100	35	35	35	35	100	100	100	100	100	100	100	35	35	35	35	100	100	100	100	100	100	100	35	35	35	35	100	100	100	100	100	100	100	
BG	NUTS3	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	100	0	
CZ	NUTS3	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
DK	NUTS3	0	0	0	0	0	0	0	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	0					
DE	NUTS3	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
EE	NUTS3	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
IE	NUTS3	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
GR	NUTS3	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
ES	NUTS3	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
FR	NUTS3	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	0				
IT	NUTS3	0	0	33	33	33	33	33	100	100	100	0	0	33	33	33	33	100	100	100	100	0	0	33	33	33	33	100	100	100	100	0	0	33	33	33	33	100	100	100	100				
CY	NUTS3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				
LV	NUTS3	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0				
LT	NUTS3	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	0	100	100	100	100	100	100	100	100	100	0				
LU	NUTS3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				
HU	NUTS3	0	100	0	100	100	100	100	100	100	0	0	100	0	100	100	100	100	100	100	0	0	100	0	100	100	100	100	100	100	0	0	100	0	100	100	100	100	100	100	0				
MT	NUTS3	0	100	100	100	100	100	100	100	100	0	0	100	100	100	100	100	0	100	100	0	0	100	100	100	100	100	0	100	100	0	0	100	100	100	100	100	0	100	100	0				
NL	NUTS3	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0				
AT	NUTS3	0	0	100	100	100	100	100	100	100	0	0	0	100	100	100	100	100	100	100	0	0	0	100	100	100	100	100	100	100	0	0	0	100	100	100	100	100	100	100	0				
PL	NUTS3	0	100	100	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0				
PT	NUTS3	100	100	100	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0				
RO	NUTS3	0	0	100	100	100	100	100	100	100	0	0	0	100	100	100	100	100	100	100	0	0	0	100	100	100	100	100	100	100	0	0	0	100	100	100	100	100	100	100	0				
SI	NUTS3	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0				
SK	NUTS3	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0	0	0	0	100	100	100	100	100	100	0				
FI	NUTS3	0	0	0	0	0	0	0	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
SE	NUTS3	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100				
UK	NUTS3	79	79	79	79	79	79	79	100	100	0	0	0	0	0	0	0	83	100	0	0	0	0	0	0	0	0	0	83	100	0	0	0	0	0	0	0	83	100	0	0				
IS	NUTS3	0	0	0	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
LI	NUTS3	100	100	100	100	100	100	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
NO	NUTS3	0	0	0	0	0	100	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
CH	NUTS3	100	0	0	0	0	100	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
HR	NUTS3	0	0	100	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
MK	NUTS3	100	0	0	100	100	100	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	0	100	100	100	0	0	0	0	0	0	100	100	100	100	0				
TR	NUTS3	0	0	0	0	0	0	0	0	100	100	0	0	0	0	0	0	0	0	100	100	0	0	0	0	0	0	0	0	100	100	0	0	0	0	0	0	0	100	100	100	0			

Annexe 2 – Towards a dictionary of estimation methods for missing values? Working paper



NOTE: This documentation is provisional and will be improved thanks to feedbacks of persons trying to use this methodological paper.

Table of content

Forewords	72
Steps to follow – quick overview	73
The ESTI Framework and ESPON metadata	74
One dimensional estimations methods.....	76
Estimation based on the space dimension (E)	76
E1 - Deduction with available values at the same and an upper hierarchical level	77
E2 - Deduction with available values at a lower hierarchical level (average of children)	78
E3 - Approximation with available values at a lower, a same and an upper hierarchical level (min-max).....	79
E4 - Approximation with available values at a same hierarchical level	80
Specificities of the source dimension (S)	81
Estimation based on the time dimension (T)	84
T1 - Interpolation, 1-1 method (where $t-1 < t < t+1$)	85
T2 - Interpolation, n-n method (where $t-n < t-1 < t < t+1 < t+n$)	86
T3 - Retrospective method (where $t < t+1 < t+n$)	87
T4 - Prospective method (where $t-n < t-1 < t$)	88
Estimation based on the thematic dimension (I)	89
I1 - Deduction coming from indicators thematically coherent (sum, subtraction)	90
I2 - Proxy coming from indicators thematically coherent (sum, subtraction)	91
I3 - Estimation by using bivariate or multivariate regression	92
Multidimensional estimations methods	93
Estimation based on the source and spatial dimensions (SE)	93
SE1 – Estimation using other data sources and space harmonization	94
Estimation based on the space and time dimensions (TE)	95
TE1 - Retropolation and space harmonization	96
TE2 - Extrapolation and space harmonization	97
TE3 - Interpolation and space harmonization – evolution method	98
TE4 - Interpolation and space harmonization – distribution method	99
TE5 - Interpolation/retropolation and space harmonization based on distribution method ...	100
TE6 – Retropolation/extrapolation and space harmonization based on distribution method with available values at the same hierarchical level	101
TE7 - Interpolation/retropolation and space harmonization based on distribution method with available values at the same hierarchical level	103

FOREWORDS

The process of data collection / estimation is strongly linked to the development of the “core” database, which has the following properties:

❖ The core database properties

- Stock only count data / no ratios
- Collect systematically data for all the NUTS levels 0-1-2
- Data collected for NUTS3 if possible/available
- Coherence of hierarchies

$$\sum_{NUTS3} X(NUTS3, s, t, i) = \sum_{NUTS2} X(NUTS2, s, t, i) = \dots = \sum_{NUTS0} X(NUTS0, s, t, i)$$

- Eurostat is considered as a primary source for data collection.
- No changes on Eurostat data, unless identified error.
- No missing data.
- Targeted time-period: 1990-2010 (yearly) for primary stocks: GDP, population, area and 2000-2010 (yearly) for sub-components of primary stocks (cf below).
- Targeted space of analysis: ESPON Area + Candidate Countries

❖ The targeted indicators

Basic 1: Population (1990-2010)

Sub-components (2000-2010): death, birth, migration, by age and sex (5 years age-class), active population (by sex), unemployed population (by sex and age), employed population (by economic branch), long-term unemployment, education level of the population.

Basic 2: Money (GDP euros & PPS, 1990-2010)

Sub-components (2000-2010): GVA by economic branch, income

Basic 3: Area (1990-2010)

Sub-components (2000-2010): By type (Corine Land Cover)

❖ Main information sources

- Eurostat (primary source)
- NewCronos database (historical Eurostat database – NUTS 1999 version)
- National Statistical institutes
- United Nations databases (UNEP, UNPP)
- OECD database
- European DGs databases (DG AGRI...)
- Estimations (cf below)

STEPS TO FOLLOW – QUICK OVERVIEW...

a. Download data from Eurostat and transform it in the data model

The following tables have been downloaded from Eurostat (September 2011). They contains all the targeted count indicators.

List of tables downloaded from Eurostat

- agri_r_size.xls
- demo_r_births_by_age_mother.xls
- demo_r_biths_deaths_N3.xls
- demo_r_d2jan.xls
- demo_r_d3avg.xls
- demo_r_gind3.xls
- edat_lfse.xls
- educ_renrirg1.xls
- lfst_r_lfe2en1.xls
- lfst_r_lfe2en2.xls
- lfst_r_lfp3pop.xls
- lfst_r_lfu2tu.xls
- lfst_r_lfu3pers.xls
- nama_r_e3empl95.xls
- nama_r_e3gdp.xls
- nama_r_e3vabp95.xls

b. Transformation of the Eurostat tables into the ESPON data template. The table should contain both NUTS0, NUTS1, NUTS2 and NUTS3 data values (1927 territorial units in the NUTS 2006 division).

ESPON Data template

Unit code	Name	Object type	Version	pop_tot		pop_tot	
				1990	source	1991	source
BE	Belgium	NUTS0	2006	8947782	1a	8986375	1a
BG	Bulgaria	NUTS0	2006	8717308	1a	8652253	1a
CZ	Czech Republic	NUTS0	2006	10303000	UN		
DK	Denmark	NUTS0	2006	5185408	1a	5168463	1a
DE	Germany (including Form...	NUTS0	2006	73019322	1a	72973380	1a
EE	Estonia	NUTS0	2006	1170593	1a	1667749	1a
IE	Ireland	NUTS0	2006	3508370	1a	3520977	1a
GR	Greece	NUTS0	2006	10130582	1a	10162201	1a
ES	Spain	NUTS0	2006	38828237	1a	38874573	1a
FR	France	NUTS0	2006				
IT	Italy	NUTS0	2006	56694360	1a	5674419	1a
CY	Cyprus	NUTS0	2006	572955	1a	587941	1a
LV	Latvia	NUTS0	2006	2529140	1a	2529951	1a
LT	Lithuania	NUTS0	2006	3837708	1a	3701958	1a
LU	Luxembourg	NUTS0	2006	379300	1a	384400	1a
HU	Hungary	NUTS0	2006	10274623	1a	10273140	1a
MT	Malta	NUTS0	2006	352430	1a	355930	1a
NL	Netherlands	NUTS0	2006	14828274	1a	15010445	1a
AT	Austria	NUTS0	2006	7544618	1a	7730652	1a
PL	Poland	NUTS0	2006	37899300	8a	3893180	1a
PT	Portugal	NUTS0	2006	9570441	7a	9522640	1a
RO	Romania	NUTS0	2006	23212595	1a	23192274	1a
SI	Slovenia	NUTS0	2006	1986277	1a	1989945	1a
SK	Slovakia	NUTS0	2006	5207963	1a	5130711	1a
FI	Finland	NUTS0	2006	4974383	1a	4988478	1a
SE	Sweden	NUTS0	2006	8927059	1a	8950630	1a
UK	United Kingdom	NUTS0	2006				
IS	Iceland	NUTS0	2006	267395	1a	258866	1a
LI	Liechtenstein	NUTS0	2006	32452	1a	29932	1a
NO	Norway	NUTS0	2006	422316	1a	424830	1a
CH	Switzerland	NUTS0	2006			6750893	1a

c. Fulfillment of metadata as below (source sheet) for the information contained in the Eurostat table (cf part 1.2 for more information).

ESPON Source metadata template

Source Reference		
Label		1
Date	2011-07-26	
Copyright	© Eurostat	
Provider	Name	Eurostat
	URI	http://epp.eurostat.ec.europa.eu
Publication	Title	
	URI	
	Reference	
Methodology	Description	Population on January 1st, Table demo_r_gind3
	URI	
Access Rule	public	
Estimation	false	
Quality Level	high	

d. Completion of missing values thanks to the ESTI framework. A couple of solutions are proposed in this document.

e. Immediately after each estimation, fulfill the source sheet of the metadata template.

The ESTI Framework and ESPON metadata

All the development below is largely inspired from the Data Navigator 2 framework produced within the ESPON Project 3.2 (2007). The document proposes to formalize estimation methods by using the spatial (e), source (s), temporal (t), and thematic (i) dimensions:

Background - The ESTI framework

Data in the LTDB can therefore be characterized according to four dimensions E, S, T and I where:

- (E) represents the spatial dimension and refers to one or several territorial units;
- (S) represents the source dimension and refers to one or several organism of statistics;
- (T) represents the time dimension and refers to one or several instants or/and periods
- (I) represents the thematic dimension and refers to one or several indicators.

For instance, the value 83859 could be the result of a query concerning the dimension I and where:

E='Austria'

S='EUROSTAT'

T='1999'

I='Area in km²'.

Although the query is answered a lot of implicit assumptions are made here:

- Regarding the E dimension, the query is supposed to refer to the territorial unit whose shape corresponds to the "official" delimitation of Austria.
- Regarding the T dimension, the value returned corresponds to the value provided by the 'EUROSTAT' for the period ranging from 01/01/99 to 31/12/99.
- Regarding the I dimension, areas of lakes and rivers in Austria are supposed to be included...

In the previous example, a value is returned for the corresponding value of E, S, T and I, namely the tuple ('Austria', 'EUROSTAT', '1999', 'Area in km²'). It should be noted that values of E, S, T or I, or a combination of them might be missing when formulating the query. By default, such a missing value in one or more dimensions handled by the query, should be considered equivalent to a wildcard operator "*" meaning "all values". However, unavoidable incompleteness in the LTDB will lead to unanswered queries. This occurs when, no value is present in the LTDB, for a given tuple of values (included wildcard "*"). The objective is to overcome this case of missing value by proposing one or several estimation methods in order to compute the more probable (although not measured or sure) value and return it as an answer to the query. LTDB users should be warned when the returned value is a computed estimated value replacing a missing one.

Then, information in the LTDB (that is the future database implanting the data model presented in Figure 7) can be represented as a fourdimensional hypercube (the ESTI framework) with holes corresponding to missing values. Estimation methods help in filling-up these holes by considering information provided by the neighborhood of these holes. We present the main principles of the estimation methods which could be implemented as procedures together with the future LTDB. Then we describe why and how the model could moreover allow the elaboration of quality control procedures.

Background - The ESTI framework for estimating missing values

Let $X(e,s,t,i)$ be a value, possibly missing (not known or not defined) where *e* describes one territorial unit, *s* describes one source, *t* describes one instant, and *i* describes one indicator.

For the sake of simplicity, only this elementary case is considered here. However, it can be shown that more complex cases, where *e*, *s*, *t*, and *i* are each described by set of values, can be decomposed in elementary ones.

For a missing value $X(e,s,t,i)$, estimation methods exploit known values $X(e',s',t',i')$ in the neighbourhood of $X(e,s,t,i)$. This neighbourhood can be defined according to one or more dimension among E, S, T and I.

Some one and two-dimensional estimation methods are described. Let *child* be the operator which, for a spatial unit *e*, returns the set of all the spatial units e_1, e_2, \dots, e_n which are spatially included in *e*, at the first level of spatial inclusion.

Let *parent* be the operator which, for a spatial unit *e*, returns the spatial units *p* which includes *e*, at the first level of spatial inclusion.

ESPON 3.2, 2007, Data navigator 2, Final Report, Part1 – Handbook for data collection, p.51

These dimensions can be combined together, and provide indeed various methods for estimating missing values in a dataset.

As regard to this background, the aim of this document is double:

- Formalize procedures of data estimations as regard to regular concrete situations. We try systematically to explain step by step the methodology employed for estimating data by using the ESTI terminology.
- Provide information for fulfilling correctly the ESPON metadata in that order.

At the end, it is possible to find some concrete applications to the results of this document, linked to the improvements of the ESPON metadata model:

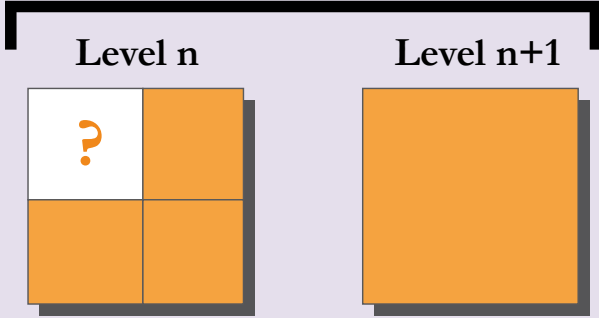
1. Harmonize the metadata of the source information: for the moment there are too many different descriptions of data estimations in the ESPON metadata.
2. Definition of “intelligent” labels for estimated data, which allow to understand directly in the dataset (without looking in the metadata) which estimation method has been used for estimating the data.
3. For the most common estimation methods (presented below), find an harmonized way to describe the methodology (which is currently developed in a free text mode).
4. After having inventoried the most current estimation methods, we can imagine in the next steps to implement it in a computer model to make possible the development of an automatic tool for estimating missing data.

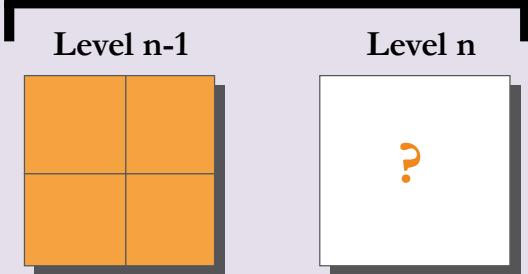
One dimensional estimations methods

Estimation based on the space dimension (E)

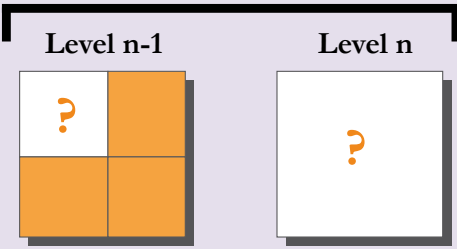


In this case, for a missing value $X(e,s,t,i)$, $s,t,$ and i being fixed, the idea is to use spatial units e' being at an upper, lower or same hierarchical spatial level as e , and for which $X(e',s,t,i)$ is known, in order to obtain information on the missing value.

Name	E1 - Deduction with available values at the same and an upper hierarchical level																																												
Conditions of use	If $X(\text{parent}(e),s,t,i)$ is known and if $X(e',s,t,i)$ is known for all e' so that $\text{parent}(e') = \text{parent}(e)$ and $e' \neq e$ (e' is at the same hierarchical level as e and has the same parent as e)																																												
Graphics illustration	<p style="text-align: center;">SPACE DIMENSION (E)</p> 																																												
Textual explanation	Estimation based on the space dimension (E) The value is estimated by subtracting the value available at an upper hierarchical level by the sum of units having the same hierarchical level and the same parent than the value estimated.																																												
Mathematic formalization	$X(e, s, t, i) = X(\text{parent}(e), s, t, i) - \sum_{e'} X(e', s, t, i)$																																												
Example	<table border="1" style="display: inline-table; margin-right: 20px;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">NATIVE TABLE</th> <th>ID</th> <th>Object type</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td></td> <td>UKM</td> <td>NUTS1</td> <td>100</td> </tr> <tr> <td></td> <td>UKM1</td> <td>NUTS2</td> <td>50</td> </tr> <tr> <td></td> <td>UKM2</td> <td>NUTS2</td> <td>n/a</td> </tr> <tr> <td></td> <td>UKM3</td> <td>NUTS2</td> <td>30</td> </tr> </tbody> </table> <table border="1" style="display: inline-table;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM TABLE</th> <th>ID</th> <th>Object type</th> <th>Estim</th> </tr> </thead> <tbody> <tr> <td></td> <td>UKM</td> <td>NUTS1</td> <td>100</td> </tr> <tr> <td></td> <td>UKM1</td> <td>NUTS2</td> <td>50</td> </tr> <tr> <td></td> <td>UKM2</td> <td>NUTS2</td> <td>20 (2)</td> </tr> <tr> <td></td> <td>UKM3</td> <td>NUTS2</td> <td>30</td> </tr> <tr> <td></td> <td>SUM</td> <td></td> <td>80</td> </tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (1) Sum the values of the children (e') (2) Subtract the value of the sum of children to the parent unit (e) 	NATIVE TABLE	ID	Object type	Value		UKM	NUTS1	100		UKM1	NUTS2	50		UKM2	NUTS2	n/a		UKM3	NUTS2	30	ESTIM TABLE	ID	Object type	Estim		UKM	NUTS1	100		UKM1	NUTS2	50		UKM2	NUTS2	20 (2)		UKM3	NUTS2	30		SUM		80
NATIVE TABLE	ID	Object type	Value																																										
	UKM	NUTS1	100																																										
	UKM1	NUTS2	50																																										
	UKM2	NUTS2	n/a																																										
	UKM3	NUTS2	30																																										
ESTIM TABLE	ID	Object type	Estim																																										
	UKM	NUTS1	100																																										
	UKM1	NUTS2	50																																										
	UKM2	NUTS2	20 (2)																																										
	UKM3	NUTS2	30																																										
	SUM		80																																										

Name	E2 - Deduction with available values at a lower hierarchical level (average of children)																																						
Conditions of use	If e'' is so that e = parent (e'') and X (e'', s, t, i) is defined → All the children of the value to estimate are known																																						
Graphics illustration	<p style="text-align: center;">SPACE DIMENSION (E)</p> <div style="text-align: center;">  </div>																																						
Textual explanation	Estimation based on the space dimension (E) Sum/average of the values available at a lower hierarchical level .																																						
Mathematic formalization	$X(e, s, t, i) = \frac{ child(e) }{ Y_{e'}\{e''\} } \sum_{e''} X(e'', s, t, i)$																																						
Example	<table border="1" style="display: inline-table; margin-right: 20px;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Object type</th> <th>Value</th> </tr> </thead> <tbody> <tr><td>FR4</td><td>NUTS1</td><td>n/a</td></tr> <tr><td>FR41</td><td>NUTS2</td><td>8</td></tr> <tr><td>FR42</td><td>NUTS2</td><td>5</td></tr> <tr><td>FR43</td><td>NUTS2</td><td>3</td></tr> <tr><td>FR44</td><td>NUTS2</td><td>5</td></tr> </tbody> </table> <table border="1" style="display: inline-table;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM. TABLE</th> <th>ID</th> <th>Object type</th> <th>Value</th> </tr> </thead> <tbody> <tr><td>FR4</td><td>NUTS1</td><td>21 (1)</td></tr> <tr><td>FR41</td><td>NUTS2</td><td>8</td></tr> <tr><td>FR42</td><td>NUTS2</td><td>5</td></tr> <tr><td>FR43</td><td>NUTS2</td><td>3</td></tr> <tr><td>FR44</td><td>NUTS2</td><td>5</td></tr> </tbody> </table> <p>HOW TO PROCEED ? (1) Sum the value of the children (be careful, only true for count data)</p>	ORIGINAL TABLE	ID	Object type	Value	FR4	NUTS1	n/a	FR41	NUTS2	8	FR42	NUTS2	5	FR43	NUTS2	3	FR44	NUTS2	5	ESTIM. TABLE	ID	Object type	Value	FR4	NUTS1	21 (1)	FR41	NUTS2	8	FR42	NUTS2	5	FR43	NUTS2	3	FR44	NUTS2	5
ORIGINAL TABLE	ID	Object type	Value																																				
FR4	NUTS1	n/a																																					
FR41	NUTS2	8																																					
FR42	NUTS2	5																																					
FR43	NUTS2	3																																					
FR44	NUTS2	5																																					
ESTIM. TABLE	ID	Object type	Value																																				
FR4	NUTS1	21 (1)																																					
FR41	NUTS2	8																																					
FR42	NUTS2	5																																					
FR43	NUTS2	3																																					
FR44	NUTS2	5																																					

Name	E3 - Approximation with available values at a lower, a same and an upper hierarchical level (min-max)																																																																																																																									
Conditions of use	If $X(\text{parent}(e),s,t,i)$ is known and if there exists at least one e' so that $\text{parent}(e') = \text{parent}(e)$ and $e' \neq e$ (e' is at the same hierarchical level as e and has the same parent as e), and $X(e',s,t,i)$ is not defined then three methods can be used.																																																																																																																									
Graphics illustration	<p style="text-align: center;">SPACE DIMENSION (E)</p>																																																																																																																									
Textual explanation	Estimation based on the space dimension (E) The value is estimated by putting in relation the maximum and the minimum values possible thanks to available information at the upper (n+1), the same (n), and lower (n-1) hierarchical levels.																																																																																																																									
Mathematic formalization	$X(s, s, t, i) = \left(\frac{(\sum_{e' \in \text{child}(e)} X(e', s, t, i)) + (X(\text{parent}(e), s, t, i) - \sum_{e'} X(e', s, t, i))}{2} \right)$																																																																																																																									
Example	<table border="1" style="display: inline-table; margin-right: 20px;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Object type</th> <th>Value</th> </tr> </thead> <tbody> <tr><td>FR</td><td>NUTS0</td><td></td><td>500</td></tr> <tr><td>FR1</td><td>NUTS1</td><td></td><td>100</td></tr> <tr><td>FR2</td><td>NUTS1</td><td></td><td>50</td></tr> <tr><td>FR3</td><td>NUTS1</td><td></td><td>n/a</td></tr> <tr><td>FR31</td><td>NUTS2</td><td></td><td>n/a</td></tr> <tr><td>FR32</td><td>NUTS2</td><td></td><td>90</td></tr> <tr><td>FR4</td><td>NUTS1</td><td></td><td>n/a</td></tr> <tr><td>FR41</td><td>NUTS2</td><td></td><td>n/a</td></tr> <tr><td>FR42</td><td>NUTS2</td><td></td><td>30</td></tr> <tr><td>FR43</td><td>NUTS2</td><td></td><td>10</td></tr> </tbody> </table> <table border="1" style="display: inline-table;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIMATION TABLE</th> <th>ID</th> <th>Object type</th> <th>Value</th> <th>Max</th> <th>Min</th> <th>Estim</th> </tr> </thead> <tbody> <tr><td>FR</td><td>NUTS0</td><td></td><td>500</td><td>/</td><td>/</td><td>500</td></tr> <tr><td>FR1</td><td>NUTS1</td><td></td><td>100</td><td>/</td><td>/</td><td>100</td></tr> <tr><td>FR2</td><td>NUTS1</td><td></td><td>50</td><td>/</td><td>/</td><td>50</td></tr> <tr><td>FR3</td><td>NUTS1</td><td></td><td>n/a</td><td>260 (1)</td><td>90 (2)</td><td>175 (3)</td></tr> <tr><td>FR31</td><td>NUTS2</td><td></td><td>n/a</td><td>/</td><td>/</td><td>n/a</td></tr> <tr><td>FR32</td><td>NUTS2</td><td></td><td>90</td><td>/</td><td>/</td><td>90</td></tr> <tr><td>FR4</td><td>NUTS1</td><td></td><td>n/a</td><td>310 (1)</td><td>40 (2)</td><td>175 (3)</td></tr> <tr><td>FR41</td><td>NUTS2</td><td></td><td>n/a</td><td>/</td><td>/</td><td>n/a</td></tr> <tr><td>FR42</td><td>NUTS2</td><td></td><td>30</td><td>/</td><td>/</td><td>30</td></tr> <tr><td>FR43</td><td>NUTS2</td><td></td><td>10</td><td>/</td><td>/</td><td>10</td></tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> Estimate the maximum value possible thanks to the parent unit value and the available values of the units of the same hierarchy (e') Estimate the minimum value possible thanks to the children (e'') $(\text{Min} + \text{Max}) / 2$ 	ORIGINAL TABLE	ID	Object type	Value	FR	NUTS0		500	FR1	NUTS1		100	FR2	NUTS1		50	FR3	NUTS1		n/a	FR31	NUTS2		n/a	FR32	NUTS2		90	FR4	NUTS1		n/a	FR41	NUTS2		n/a	FR42	NUTS2		30	FR43	NUTS2		10	ESTIMATION TABLE	ID	Object type	Value	Max	Min	Estim	FR	NUTS0		500	/	/	500	FR1	NUTS1		100	/	/	100	FR2	NUTS1		50	/	/	50	FR3	NUTS1		n/a	260 (1)	90 (2)	175 (3)	FR31	NUTS2		n/a	/	/	n/a	FR32	NUTS2		90	/	/	90	FR4	NUTS1		n/a	310 (1)	40 (2)	175 (3)	FR41	NUTS2		n/a	/	/	n/a	FR42	NUTS2		30	/	/	30	FR43	NUTS2		10	/	/	10
ORIGINAL TABLE	ID	Object type	Value																																																																																																																							
FR	NUTS0		500																																																																																																																							
FR1	NUTS1		100																																																																																																																							
FR2	NUTS1		50																																																																																																																							
FR3	NUTS1		n/a																																																																																																																							
FR31	NUTS2		n/a																																																																																																																							
FR32	NUTS2		90																																																																																																																							
FR4	NUTS1		n/a																																																																																																																							
FR41	NUTS2		n/a																																																																																																																							
FR42	NUTS2		30																																																																																																																							
FR43	NUTS2		10																																																																																																																							
ESTIMATION TABLE	ID	Object type	Value	Max	Min	Estim																																																																																																																				
FR	NUTS0		500	/	/	500																																																																																																																				
FR1	NUTS1		100	/	/	100																																																																																																																				
FR2	NUTS1		50	/	/	50																																																																																																																				
FR3	NUTS1		n/a	260 (1)	90 (2)	175 (3)																																																																																																																				
FR31	NUTS2		n/a	/	/	n/a																																																																																																																				
FR32	NUTS2		90	/	/	90																																																																																																																				
FR4	NUTS1		n/a	310 (1)	40 (2)	175 (3)																																																																																																																				
FR41	NUTS2		n/a	/	/	n/a																																																																																																																				
FR42	NUTS2		30	/	/	30																																																																																																																				
FR43	NUTS2		10	/	/	10																																																																																																																				

Name	E4 - Approximation with available values at a same hierarchical level																																						
Conditions of use	If e' so that $\text{parent}(e') = \text{parent}(e)$ and $e' \neq e$ and $X(e', s, t, i)$ is defined <ul style="list-style-type: none"> ➔ Have values for other units at a same hierarchical level ➔ Have a statistical distribution relatively homogeneous between the other units of a same hierarchical level 																																						
Graphics illustration	<p style="text-align: center;">SPACE DIMENSION (E)</p> <div style="text-align: center;">  </div>																																						
Textual explanation	Estimation based on the space dimension (E) Average of the values of spatial units of the same hierarchical level.																																						
Mathematic formalization	$X(e, s, t, i) = \frac{(X(\text{parent}(e), s, t, i) - \sum_{e'} X(e', s, t, i))}{ \text{child}(\text{parent}(e)) - Y_{e'}\{e'\} }$																																						
Example	<table border="1" style="display: inline-table; margin-right: 20px;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Object type</th> <th>Value</th> </tr> </thead> <tbody> <tr><td>FR</td><td>NUTS0</td><td>n/a</td></tr> <tr><td>FR4</td><td>NUTS1</td><td>n/a</td></tr> <tr><td>FR1</td><td>NUTS1</td><td>8</td></tr> <tr><td>FR2</td><td>NUTS1</td><td>5</td></tr> <tr><td>FR3</td><td>NUTS1</td><td>3</td></tr> </tbody> </table> <table border="1" style="display: inline-table;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM. TABLE</th> <th>ID</th> <th>Object type</th> <th>Value</th> </tr> </thead> <tbody> <tr><td>FR</td><td>NUTS0</td><td>n/a</td></tr> <tr><td>FR4</td><td>NUTS1</td><td>5,3 (1)</td></tr> <tr><td>FR1</td><td>NUTS1</td><td>8</td></tr> <tr><td>FR2</td><td>NUTS1</td><td>5</td></tr> <tr><td>FR3</td><td>NUTS1</td><td>3</td></tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <p>(1) Average of the values of spatial units of the same hierarchical level</p>	ORIGINAL TABLE	ID	Object type	Value	FR	NUTS0	n/a	FR4	NUTS1	n/a	FR1	NUTS1	8	FR2	NUTS1	5	FR3	NUTS1	3	ESTIM. TABLE	ID	Object type	Value	FR	NUTS0	n/a	FR4	NUTS1	5,3 (1)	FR1	NUTS1	8	FR2	NUTS1	5	FR3	NUTS1	3
ORIGINAL TABLE	ID	Object type	Value																																				
FR	NUTS0	n/a																																					
FR4	NUTS1	n/a																																					
FR1	NUTS1	8																																					
FR2	NUTS1	5																																					
FR3	NUTS1	3																																					
ESTIM. TABLE	ID	Object type	Value																																				
FR	NUTS0	n/a																																					
FR4	NUTS1	5,3 (1)																																					
FR1	NUTS1	8																																					
FR2	NUTS1	5																																					
FR3	NUTS1	3																																					

Specificities of the source dimension (S)



Data available from Eurostat or National Statistical Institutes are delivered after a long process of data creation (individual censuses, aggregation, harmonization at European using various methodologies). In that order, they cannot be considered as “primary sources” in a methodological point of view. However, it is difficult to obtain and synthesize all the steps of data creation in the methodological field of the source part of the ESPON metadata model. We consider indeed that this kind of information is not estimated (a).

This is the reason why the strategy of source completion has been defined as below:

- For European data at regional level, the most common source is Eurostat, who deliver official data for the European Community. Considering that this statistical organization makes available new tables and updates regularly, it is very important to mention systematically the date of the creation of the table (1) and the name of the table (2).
- If data comes from Eurostat, the label can for instance begin by “1” (which can be developed in 1a, 1b, 1c if the dataset contains different Eurostat tables or updates).

Label 1a, 1b, 1c...
Date 17/12/2010 (1)
Copyright © Eurostat
Provider Name: Eurostat
 URI:
<http://epp.eurostat.ec.europa.eu>
Publication Title: [Let this cell empty]
 URI: [Let this cell empty]
 Reference: [Let this cell empty]
Methodology Table demo_r_d3natmo (2)
Access rule Public
Estimation False (a)
Quality High
 ESPON metadata model (source part)

Births and deaths (1000) [demo_r_d3natm] (2)		
Last update	17.12.10	(1)
Extracted on	05.07.11	
Source of Data	Eurostat	
INDIC_DE		
Live births - total		
GEO/TIME	1990	1991
BE	123,6	125,4
BG	105,2	95,9
CZ	130,6	129,4
DK	63,4	64,4
DE	776,9	829,8

Eurostat table

The missing value in Eurostat tables could be completed by alternative sources, like information coming from National Statistical Institutes, United Nations, OCDE and so on. In this case, the information can be obtained through various formats: Statistical books, “my country in focus” format, information downloaded from the NSI website directly. When using this kind of information, the idea consists to fulfill as precise as possible the metadata template and save (if possible) the table or the document downloaded in a dedicated space by mentioning the name of the file. Some basic rules have to be followed:

- (1) The date corresponds to the date of publication of the report. If the data comes from the web page, mention the date of download
- (2) Mention systemically name, URI (home web page) and copyrights of the source used.
- (3) If the data comes from a publication, mention the title of the publication, its reference (web address or ISSN if not available) and if possible the page where the information has been downloaded.
- (4) If there is precision on the data (e.g. resident population estimated the 31th March 2000, mention it in the methodology field).
- (5) The data quality depends on the nature of the data (exhaustive census? Surveys?) but basically will be medium or high.

Label 7a (1)
Date 01/12/2010 (2)
Copyright © 2006 Republic of Croatia – Croatian Bureau of Statistics (3)
Provider Name: Croatian Bureau of Statistics (3)
URI: http://www.dzs.hr/default_e.htm (3)
Publication Title: Statistical Yearbook 2010 of the Republic of the Republic of Croatia (4)
URI: *ISSN 1333-3305(4)*
Reference: page 108
Methodology Mid-year estimate
Access rule False (a)
Estimation true
Quality High
ESPON metadata model with National Statistical Institute source

Estimation based on the time dimension (T)

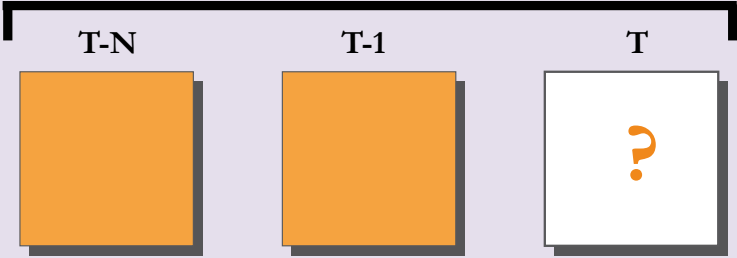


Various time interpolation methods using linear or non-linear assumption, prospective or retrospective computations of tendency can be used. Three of them are described here. The idea here is to estimate the missing value $X(e,s,t,i)$ by using two known values $X(e,s,t_1,i)$ and $X(e,s,t_2,i)$, who t_1 and t_2 are placed in time after or before t .

Name	T1 - Interpolation, 1-1 method (where $t-1 < t < t+1$)																																																															
Conditions of use	If $X(e,s,t-1,i)$ and $X(e,s,t+1,i)$ are known → Have other values at the same hierarchical level at $t-1$ and $t+1$ → Don't have too much statistical breakdown in the temporal serie (adapted for territorial units who have more or less a linear evolution). → Do not have any information at a upper level in the hierarchy (parent(X))																																																															
Graphics illustration	<p style="text-align: center;">TIME DIMENSION (T)</p> <div style="text-align: center;"> </div>																																																															
Textual explanation	Estimation based on the time dimension (T) – [linear/power] interpolation. This method uses the two closest neighbours placed in time before ([mention the year of reference]) and after ([mention the year of reference]) the value estimated.																																																															
Mathematic formalization	<ul style="list-style-type: none"> • Linear model $X(e,s,t,i) = X(e,s,t-1,i) + \frac{X(e,s,t+1,i) - X(e,s,t-1,i)}{(t+1) - (t-1)} (t - (t-1))$ <ul style="list-style-type: none"> • Power model $X(e,s,t,i) = X(e,s,t-1,i) \left(\frac{X(e,s,t+1,i)}{X(e,s,t-1,i)} \right)^{\frac{(t-(t-1))}{((t+1)-(t-1))}}$																																																															
Example	<table border="1" style="margin-bottom: 10px;"> <thead> <tr> <th rowspan="5" style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Value 2000</th> <th>Value 2001</th> <th>Value 2002</th> </tr> </thead> <tbody> <tr> <td>FR1</td> <td>8</td> <td>n/a</td> <td>12</td> </tr> <tr> <td>FR2</td> <td>5</td> <td>n/a</td> <td>7</td> </tr> <tr> <td>FR3</td> <td>1</td> <td>n/a</td> <td>4</td> </tr> <tr> <td>FR4</td> <td>4</td> <td>n/a</td> <td>8</td> </tr> </tbody> </table> <p>HOW TO PROCEED ? For the linear model (1), put the following formula in the table cell: =value(2000)+((value (2002)-value(2000))/(2002-2000))*(2001-2000)</p> <p>For the power model (2) put the following formula in the table cell: value(2000)*((value(2002)/value(2000))^(2001-2000)/(2002-2000))</p> <table border="1" style="margin-top: 10px;"> <caption>Estimation – linear model (1)</caption> <thead> <tr> <th rowspan="5" style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM TABLE</th> <th>ID</th> <th>Value 2000</th> <th>Value 2001</th> <th>Value 2002</th> </tr> </thead> <tbody> <tr> <td>FR1</td> <td>8</td> <td>10</td> <td>12</td> </tr> <tr> <td>FR2</td> <td>5</td> <td>6</td> <td>7</td> </tr> <tr> <td>FR3</td> <td>1</td> <td>2,5</td> <td>4</td> </tr> <tr> <td>FR4</td> <td>4</td> <td>6</td> <td>8</td> </tr> </tbody> </table> <table border="1" style="margin-top: 10px;"> <caption>Estimation – power model (2)</caption> <thead> <tr> <th rowspan="5" style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM TABLE</th> <th>ID</th> <th>Value 2000</th> <th>Value 2001</th> <th>Value 2002</th> </tr> </thead> <tbody> <tr> <td>FR1</td> <td>8</td> <td>9,8</td> <td>12</td> </tr> <tr> <td>FR2</td> <td>5</td> <td>5,9</td> <td>7</td> </tr> <tr> <td>FR3</td> <td>1</td> <td>2,0</td> <td>4</td> </tr> <tr> <td>FR4</td> <td>4</td> <td>5,7</td> <td>8</td> </tr> </tbody> </table>	ORIGINAL TABLE	ID	Value 2000	Value 2001	Value 2002	FR1	8	n/a	12	FR2	5	n/a	7	FR3	1	n/a	4	FR4	4	n/a	8	ESTIM TABLE	ID	Value 2000	Value 2001	Value 2002	FR1	8	10	12	FR2	5	6	7	FR3	1	2,5	4	FR4	4	6	8	ESTIM TABLE	ID	Value 2000	Value 2001	Value 2002	FR1	8	9,8	12	FR2	5	5,9	7	FR3	1	2,0	4	FR4	4	5,7	8
ORIGINAL TABLE	ID		Value 2000	Value 2001	Value 2002																																																											
	FR1		8	n/a	12																																																											
	FR2		5	n/a	7																																																											
	FR3		1	n/a	4																																																											
	FR4	4	n/a	8																																																												
ESTIM TABLE	ID	Value 2000	Value 2001	Value 2002																																																												
	FR1	8	10	12																																																												
	FR2	5	6	7																																																												
	FR3	1	2,5	4																																																												
	FR4	4	6	8																																																												
ESTIM TABLE	ID	Value 2000	Value 2001	Value 2002																																																												
	FR1	8	9,8	12																																																												
	FR2	5	5,9	7																																																												
	FR3	1	2,0	4																																																												
	FR4	4	5,7	8																																																												

Name	T2 - Interpolation, n-n method (where t-n < t-1 < t < t+1 < t+n)																																									
Conditions of use	If $X(e,s,t-1,i)$, $X(e,s,t-n,i)$, $X(e,s,t+1,i)$ and $X(e,s,t+n,i)$ are known. → Have other values at the same hierarchical level at $t-1$, $t-n$, $t+1$ and $t+1$ → Method more adapted to temporal series characterized by statistical breakdowns. → Do not have any information at a upper level in the hierarchy (parent(X))																																									
Graphics illustration	<p style="text-align: center;">TIME DIMENSION (T)</p> <div style="text-align: center;"> </div>																																									
Textual explanation	Estimation based on the time dimension (T) – [linear/power] interpolation. This method uses the closest neighbours placed in time before ([mention the years of reference]) and after ([mention the years of reference]) the value estimated (average value on the time period).																																									
Mathematic formalization	<ul style="list-style-type: none"> Linear model $X(e,s,t,i) = X(e,s,t\alpha,i) + \frac{X(e,s,t\beta,i) - X(e,s,t\alpha,i)}{t\beta - t\alpha} (t - t\alpha)$ Power model $X(e,s,t,i) = X(e,s,t\alpha,i) \left(\frac{X(e,s,t\beta,i)}{X(e,s,t\alpha,i)} \right)^{\frac{(t-t\alpha)}{(t\beta-t\alpha)}}$ <p>Where</p> $X(e,s,t\alpha,i) = \frac{\sum_n X(e,s,t-1,i) + \dots + X(e,s,t-n,i)}{n}$ $X(e,s,t\beta,i) = \frac{\sum_n X(e,s,t+1,i) + \dots + X(e,s,t+n,i)}{n}$ <p>And $t\alpha = \frac{\sum_n t-1 + \dots + t-n}{n}$; $t\beta = \frac{\sum_n t+1 + \dots + t+n}{n}$</p>																																									
Example	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="9" style="background-color: #002060; color: white;">ORIGINAL TABLE</th> </tr> <tr> <th style="background-color: #002060; color: white;">ID</th> <th style="background-color: #002060; color: white;">Value 2000</th> <th style="background-color: #002060; color: white;">Value 2001</th> <th style="background-color: #002060; color: white;">Value 2002</th> <th style="background-color: #002060; color: white;">Value 2003</th> <th style="background-color: #002060; color: white;">Value 2004</th> <th style="background-color: #002060; color: white;">Value 2005</th> <th style="background-color: #002060; color: white;">Value 2006</th> <th style="background-color: #002060; color: white;">Value 2007</th> </tr> </thead> <tbody> <tr> <td>RS00</td> <td>10</td> <td>20</td> <td>30</td> <td>20</td> <td>n/a</td> <td>50</td> <td>40</td> <td>30</td> </tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> Estimate the average time period for t-1 ...t-n Estimate the average time period for t+1 ...t+n Calculation of the average for the $X(e,s,t-1,i) \dots X(e,s,t-n,i)$ values Calculation of the average for the $X(e,s,t+1,i) \dots X(e,s,t+n,i)$ values Estimation thanks a linear model, following the same logical than T1 method Estimation thanks a power model, following the same logical than T1 method <table border="1" style="width: 50%; margin-left: auto; margin-right: 0; border-collapse: collapse; text-align: center;"> <thead> <tr> <th colspan="2" style="background-color: #e67e22; color: white;">ESTIMATION TABLE</th> </tr> </thead> <tbody> <tr> <td style="background-color: #e67e22; color: white;">$t\alpha$</td> <td>2001,5 (1)</td> </tr> <tr> <td style="background-color: #e67e22; color: white;">$t\beta$</td> <td>2006 (2)</td> </tr> <tr> <td style="background-color: #e67e22; color: white;">$X(e,s,t\alpha,i)$</td> <td>22,5 (3)</td> </tr> <tr> <td style="background-color: #e67e22; color: white;">$X(e,s,t\beta,i)$</td> <td>40 (4)</td> </tr> <tr> <td style="background-color: #e67e22; color: white;">Linear model</td> <td>28,3 (5)</td> </tr> <tr> <td style="background-color: #e67e22; color: white;">Exp. model</td> <td>27,3 (6)</td> </tr> </tbody> </table>	ORIGINAL TABLE									ID	Value 2000	Value 2001	Value 2002	Value 2003	Value 2004	Value 2005	Value 2006	Value 2007	RS00	10	20	30	20	n/a	50	40	30	ESTIMATION TABLE		$t\alpha$	2001,5 (1)	$t\beta$	2006 (2)	$X(e,s,t\alpha,i)$	22,5 (3)	$X(e,s,t\beta,i)$	40 (4)	Linear model	28,3 (5)	Exp. model	27,3 (6)
ORIGINAL TABLE																																										
ID	Value 2000	Value 2001	Value 2002	Value 2003	Value 2004	Value 2005	Value 2006	Value 2007																																		
RS00	10	20	30	20	n/a	50	40	30																																		
ESTIMATION TABLE																																										
$t\alpha$	2001,5 (1)																																									
$t\beta$	2006 (2)																																									
$X(e,s,t\alpha,i)$	22,5 (3)																																									
$X(e,s,t\beta,i)$	40 (4)																																									
Linear model	28,3 (5)																																									
Exp. model	27,3 (6)																																									

Name	T3 - Retrospective method (where $t < t+1 < t+n$)																																																																																																																							
Conditions of use	If $X(e,s,t+1,i)$ and $X(e,s,t+n,i)$ are known <ul style="list-style-type: none"> ➔ Have other values at the same hierarchical level at $t+1$ and $t+n$ ➔ Don't have too much statistical breakdown in the temporal serie. ➔ Do not have any information at a upper level in the hierarchy (parent(X)) ➔ With the linear model, it is possible to obtain negative value after estimation. For most of indicators (e.g. population), it is better to use an exponential model, which exclude automatically negative values in the results of the estimation ➔ Do not have any information at a upper level in the hierarchy (parent(X)) 																																																																																																																							
Graphics illustration	<p style="text-align: center;">TIME DIMENSION (T)</p> <div style="text-align: center;"> </div>																																																																																																																							
Textual explanation	Estimation based on the time dimension (T) – [linear/power] retropolation. This method uses the two closest neighbours placed in time after ([mention the two years of reference]) the value estimated.																																																																																																																							
Mathematic formalization	<ul style="list-style-type: none"> • Linear model $X(e,s,t,i) = X(e,s,t+1,i) - \frac{X(e,s,t+n,i) - X(e,s,t+1,i)}{(t+n) - (t+1)} ((t+1) - t)$ <ul style="list-style-type: none"> • Power model $X(e,s,t,i) = X(e,s,t+1,i) \left(\frac{X(e,s,t+n,i)}{X(e,s,t+1,i)} \right)^{\frac{(t-(t+1))}{((t+n)-(t+1))}}$																																																																																																																							
Example	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Value 1990</th> <th>Value 1995</th> <th>Value 2000</th> <th>Value 2005</th> <th>Value 2010</th> </tr> </thead> <tbody> <tr><td>FR1</td><td>n/a</td><td>n/a</td><td>n/a</td><td>n/a</td><td>18</td><td>24</td></tr> <tr><td>FR2</td><td>n/a</td><td>n/a</td><td>n/a</td><td>n/a</td><td>29</td><td>35</td></tr> <tr><td>FR3</td><td>n/a</td><td>n/a</td><td>n/a</td><td>n/a</td><td>32</td><td>44</td></tr> <tr><td>FR4</td><td>n/a</td><td>n/a</td><td>n/a</td><td>n/a</td><td>10</td><td>15</td></tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <p>For the linear model (1) and in this case, put the following formula in the table cell for estimating the 1990 value: $=value(2005) - ((value(2010) - value(2005))/((2010-2005)*(2005-1990)))$</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th colspan="7">Estimation – linear model (1)</th> </tr> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM TABLE</th> <th>ID</th> <th>Estim 1990</th> <th>Estim 1995</th> <th>Estim 2000</th> <th>Value 2005</th> <th>Value 2010</th> </tr> </thead> <tbody> <tr><td>FR1</td><td></td><td>0</td><td>6</td><td>12</td><td>18</td><td>24</td></tr> <tr><td>FR2</td><td></td><td>11</td><td>17</td><td>23</td><td>29</td><td>35</td></tr> <tr><td>FR3</td><td></td><td>-4</td><td>8</td><td>20</td><td>32</td><td>44</td></tr> <tr><td>FR4</td><td></td><td>-5</td><td>0</td><td>5</td><td>10</td><td>15</td></tr> </tbody> </table> <p>For the power model (2) and in this case, put the following formula in the table cell for estimating the 1990 value: $= value(2005)*((value(2010)/value2005)^((1990-2005)/(2010-2005)))$</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th colspan="7">Estimation – power model (2)</th> </tr> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM TABLE</th> <th>ID</th> <th>Estim 1990</th> <th>Estim 1995</th> <th>Estim 2000</th> <th>Value 2005</th> <th>Value 2010</th> </tr> </thead> <tbody> <tr><td>FR1</td><td></td><td>7,6</td><td>10,1</td><td>13,5</td><td>18</td><td>24</td></tr> <tr><td>FR2</td><td></td><td>16,5</td><td>19,9</td><td>24,0</td><td>29</td><td>35</td></tr> <tr><td>FR3</td><td></td><td>12,3</td><td>16,9</td><td>23,3</td><td>32</td><td>44</td></tr> <tr><td>FR4</td><td></td><td>3,0</td><td>4,4</td><td>6,7</td><td>10</td><td>15</td></tr> </tbody> </table>	ORIGINAL TABLE	ID	Value 1990	Value 1995	Value 2000	Value 2005	Value 2010	FR1	n/a	n/a	n/a	n/a	18	24	FR2	n/a	n/a	n/a	n/a	29	35	FR3	n/a	n/a	n/a	n/a	32	44	FR4	n/a	n/a	n/a	n/a	10	15	Estimation – linear model (1)							ESTIM TABLE	ID	Estim 1990	Estim 1995	Estim 2000	Value 2005	Value 2010	FR1		0	6	12	18	24	FR2		11	17	23	29	35	FR3		-4	8	20	32	44	FR4		-5	0	5	10	15	Estimation – power model (2)							ESTIM TABLE	ID	Estim 1990	Estim 1995	Estim 2000	Value 2005	Value 2010	FR1		7,6	10,1	13,5	18	24	FR2		16,5	19,9	24,0	29	35	FR3		12,3	16,9	23,3	32	44	FR4		3,0	4,4	6,7	10	15
ORIGINAL TABLE	ID	Value 1990	Value 1995	Value 2000	Value 2005	Value 2010																																																																																																																		
FR1	n/a	n/a	n/a	n/a	18	24																																																																																																																		
FR2	n/a	n/a	n/a	n/a	29	35																																																																																																																		
FR3	n/a	n/a	n/a	n/a	32	44																																																																																																																		
FR4	n/a	n/a	n/a	n/a	10	15																																																																																																																		
Estimation – linear model (1)																																																																																																																								
ESTIM TABLE	ID	Estim 1990	Estim 1995	Estim 2000	Value 2005	Value 2010																																																																																																																		
FR1		0	6	12	18	24																																																																																																																		
FR2		11	17	23	29	35																																																																																																																		
FR3		-4	8	20	32	44																																																																																																																		
FR4		-5	0	5	10	15																																																																																																																		
Estimation – power model (2)																																																																																																																								
ESTIM TABLE	ID	Estim 1990	Estim 1995	Estim 2000	Value 2005	Value 2010																																																																																																																		
FR1		7,6	10,1	13,5	18	24																																																																																																																		
FR2		16,5	19,9	24,0	29	35																																																																																																																		
FR3		12,3	16,9	23,3	32	44																																																																																																																		
FR4		3,0	4,4	6,7	10	15																																																																																																																		

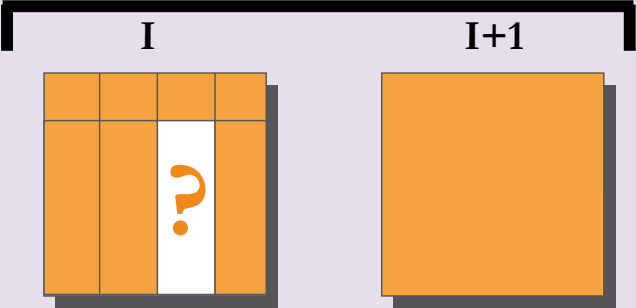
Name	T4 - Prospective method (where $t-n < t-1 < t$)																																																																																													
Conditions of use	If $X(e,s,t-1,i)$ and $X(e,s,t-n,i)$ are known <ul style="list-style-type: none"> ➔ Have other values at the same hierarchical level at least at $t-1$ and $t-n$ ➔ Don't have too much statistical breakdown in the temporal serie. ➔ With the exponential model, it is possible to obtain very high values when the estimation is made for a long time-period (e.g. projection of population). ➔ Do not have any information at a upper level in the hierarchy (parent(X)) 																																																																																													
Graphics illustration	<p style="text-align: center;">TIME DIMENSION (T)</p> 																																																																																													
Textual explanation	Estimation based on the time dimension (T) – [linear/power] extrapolation. This method uses the two closest neighbours placed in time before (<i>mention the two years of reference</i>) the value estimated.																																																																																													
Mathematic formalization	<ul style="list-style-type: none"> • Linear model $X(e,s,t,i) = X(e,s,t-1,i) - \frac{X(e,s,t-1,i) - X(e,s,t-n,i)}{(t-1) - (t-n)} (t - (t-1))$ • Power model $X(e,s,t,i) = X(e,s,t-1,i) \left(\frac{X(e,s,t-1,i)}{X(e,s,t-n,i)} \right)^{\frac{(t-(t-1))}{((t-1)-(t-n))}}$ 																																																																																													
Example	<table border="1" data-bbox="502 1126 960 1254"> <thead> <tr> <th rowspan="2">ORIGINAL TABLE</th> <th>ID</th> <th>Value 2005</th> <th>Value 2010</th> <th>Value 2020</th> <th>Value 2030</th> <th>Value 2040</th> </tr> </thead> <tbody> <tr> <td>FR1</td> <td>18</td> <td>24</td> <td>n/a</td> <td>n/a</td> <td>n/a</td> </tr> <tr> <td>FR2</td> <td>29</td> <td>35</td> <td>n/a</td> <td>n/a</td> <td>n/a</td> </tr> <tr> <td>FR3</td> <td>32</td> <td>44</td> <td>n/a</td> <td>n/a</td> <td>n/a</td> </tr> <tr> <td>FR4</td> <td>10</td> <td>15</td> <td>n/a</td> <td>n/a</td> <td>n/a</td> </tr> </tbody> </table> <p>HOW TO PROCEED ? For the linear model (1) and in this case, put the following formula in the table cell for estimating the 1990 value: =value(2010) + ((value(2010) – value(2005))/(2010-2005))*(2040-2010)</p> <table border="1" data-bbox="986 1299 1423 1444"> <caption>Estimation – linear model (1)</caption> <thead> <tr> <th rowspan="2">ESTIM TABLE</th> <th>ID</th> <th>Value 2005</th> <th>Value 2010</th> <th>Estim 2020</th> <th>Estim 2030</th> <th>Estim 2040</th> </tr> </thead> <tbody> <tr> <td>FR1</td> <td>18</td> <td>24</td> <td>36</td> <td>60</td> <td>108</td> </tr> <tr> <td>FR2</td> <td>29</td> <td>35</td> <td>47</td> <td>71</td> <td>119</td> </tr> <tr> <td>FR3</td> <td>32</td> <td>44</td> <td>68</td> <td>116</td> <td>212</td> </tr> <tr> <td>FR4</td> <td>10</td> <td>15</td> <td>25</td> <td>45</td> <td>85</td> </tr> </tbody> </table> <p>For the power model (2) and in this case, put the following formula in the table cell for estimating the 2040 value: =value(2010)*(value(2010)/value(2005))^((2040-2010)/(2010-2005))</p> <table border="1" data-bbox="986 1478 1423 1624"> <caption>Estimation – power model (2)</caption> <thead> <tr> <th rowspan="2">ESTIM TABLE</th> <th>ID</th> <th>Value 2005</th> <th>Value 2010</th> <th>Estim 2020</th> <th>Estim 2030</th> <th>Estim 2040</th> </tr> </thead> <tbody> <tr> <td>FR1</td> <td>18</td> <td>24</td> <td>42,7</td> <td>75,9</td> <td>134,8</td> </tr> <tr> <td>FR2</td> <td>29</td> <td>35</td> <td>51,0</td> <td>74,3</td> <td>108,2</td> </tr> <tr> <td>FR3</td> <td>32</td> <td>44</td> <td>83,2</td> <td>157,3</td> <td>297,4</td> </tr> <tr> <td>FR4</td> <td>10</td> <td>15</td> <td>33,8</td> <td>75,9</td> <td>170,9</td> </tr> </tbody> </table>	ORIGINAL TABLE	ID	Value 2005	Value 2010	Value 2020	Value 2030	Value 2040	FR1	18	24	n/a	n/a	n/a	FR2	29	35	n/a	n/a	n/a	FR3	32	44	n/a	n/a	n/a	FR4	10	15	n/a	n/a	n/a	ESTIM TABLE	ID	Value 2005	Value 2010	Estim 2020	Estim 2030	Estim 2040	FR1	18	24	36	60	108	FR2	29	35	47	71	119	FR3	32	44	68	116	212	FR4	10	15	25	45	85	ESTIM TABLE	ID	Value 2005	Value 2010	Estim 2020	Estim 2030	Estim 2040	FR1	18	24	42,7	75,9	134,8	FR2	29	35	51,0	74,3	108,2	FR3	32	44	83,2	157,3	297,4	FR4	10	15	33,8	75,9	170,9
ORIGINAL TABLE	ID		Value 2005	Value 2010	Value 2020	Value 2030	Value 2040																																																																																							
	FR1	18	24	n/a	n/a	n/a																																																																																								
FR2	29	35	n/a	n/a	n/a																																																																																									
FR3	32	44	n/a	n/a	n/a																																																																																									
FR4	10	15	n/a	n/a	n/a																																																																																									
ESTIM TABLE	ID	Value 2005	Value 2010	Estim 2020	Estim 2030	Estim 2040																																																																																								
	FR1	18	24	36	60	108																																																																																								
FR2	29	35	47	71	119																																																																																									
FR3	32	44	68	116	212																																																																																									
FR4	10	15	25	45	85																																																																																									
ESTIM TABLE	ID	Value 2005	Value 2010	Estim 2020	Estim 2030	Estim 2040																																																																																								
	FR1	18	24	42,7	75,9	134,8																																																																																								
FR2	29	35	51,0	74,3	108,2																																																																																									
FR3	32	44	83,2	157,3	297,4																																																																																									
FR4	10	15	33,8	75,9	170,9																																																																																									

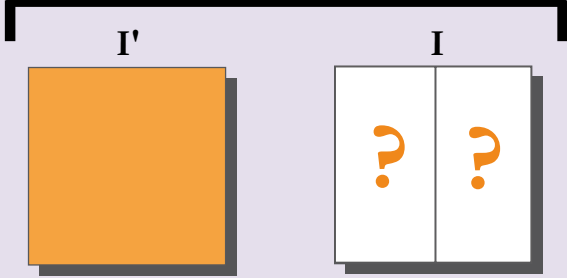
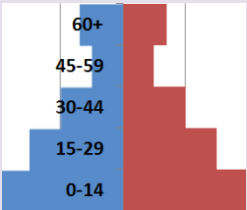
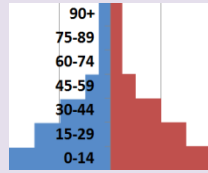
Estimation based on the thematic dimension (I)

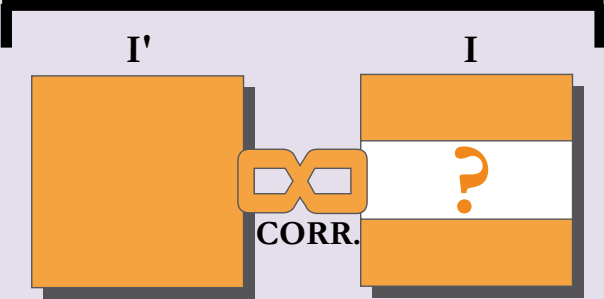
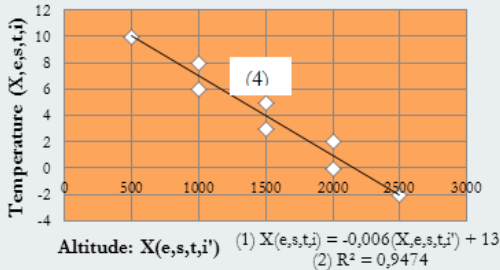


The idea is to replace the missing value $X(e,s,t,i)$ by a known value $X(e,s,t,i')$ where i' is another indicator.

$X(e,s,t,i) = \alpha (e,s,t,i')$ where α is a correlation factor empirically fixed.

Name	I1 - Deduction coming from indicators thematically coherent (sum, subtraction)																								
Conditions of use	If $X(e,s,t,i)$ and $X(e,s,t,i')$ are components a contingency table and $X(e,s,t,i+1)$ where $i+1$ is the sum of a indicators of the contingency table and i' are known → No data for one dimension of the contingency table (i). → Data available for the other dimension of the contingency table (i') → Sum of the contingency table available ($i+1$)																								
Graphics illustration	<p style="text-align: center;">THEMATIC DIMENSION (I)</p> 																								
Textual explanation	Estimation based on thematic dimension (I) The values have been deducted by subtracting the total of a contingency table by the sum of the different indicators available in this contingency table.																								
Mathematic formalization	$X(e, s, t, i) = X(e, s, t, i + 1) - \sum_{i'} X(e, s, t, i')$																								
Example	<table border="1" data-bbox="507 1070 1369 1115"> <thead> <tr> <th>ORIGINAL TABLE</th> <th>ID</th> <th>POPTOT (i+1)</th> <th>0-14 (i')</th> <th>15-64 (i')</th> <th>65-79 (i')</th> <th>80+ (i)</th> </tr> </thead> <tbody> <tr> <td></td> <td>X</td> <td>70</td> <td>20</td> <td>30</td> <td>10</td> <td>n/a</td> </tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> Sum the data available for the other dimensions of the contingency table (i') Subtract the total of the contingency table (I) by the sum of the other dimensions of the table (i') <table border="1" data-bbox="507 1285 1193 1330"> <thead> <tr> <th>ESTIMATION TABLE</th> <th>ID</th> <th>i+1</th> <th>Sum of i'</th> <th>i</th> </tr> </thead> <tbody> <tr> <td></td> <td>X</td> <td>70</td> <td>60 (1)</td> <td>10 (2)</td> </tr> </tbody> </table>	ORIGINAL TABLE	ID	POPTOT (i+1)	0-14 (i')	15-64 (i')	65-79 (i')	80+ (i)		X	70	20	30	10	n/a	ESTIMATION TABLE	ID	i+1	Sum of i'	i		X	70	60 (1)	10 (2)
ORIGINAL TABLE	ID	POPTOT (i+1)	0-14 (i')	15-64 (i')	65-79 (i')	80+ (i)																			
	X	70	20	30	10	n/a																			
ESTIMATION TABLE	ID	i+1	Sum of i'	i																					
	X	70	60 (1)	10 (2)																					

Name	I2 - Proxy coming from indicators thematically coherent (sum, subtraction)																																																			
Conditions of use	If $X(e,s,t,i')$ is a known component of a contingency table and the aim of the estimation is to split $X(e,s,t,i')$ into several ones $n \cdot X(e,s,t,i)$ <ul style="list-style-type: none"> ➔ $X(e,s,t,i')$ is known ➔ The user know in how many indicators he want to split $X(e,s,t,i')$ 																																																			
Graphics illustration	<p style="text-align: center;">THEMATIC DIMENSION (I)</p> 																																																			
Textual explanation	Estimation based on thematic dimension (I) The indicator [<i>name of the indicator</i>] has been split into [<i>n</i>] indicators [<i>name of the indicators</i>] by a division process.																																																			
Mathematic formalization	$X(e,s,t,i) = \frac{X(e,s,t,i')}{n}$ where n is the number of indicators required.																																																			
Example	<table border="1" data-bbox="555 974 774 1182"> <thead> <tr> <th colspan="3">ORIGINAL TABLE</th> </tr> <tr> <th>Age</th> <th>Males</th> <th>Females</th> </tr> </thead> <tbody> <tr><td>0-14</td><td>20</td><td>20</td></tr> <tr><td>15-29</td><td>15</td><td>15</td></tr> <tr><td>30-44</td><td>10</td><td>10</td></tr> <tr><td>45-59</td><td>5</td><td>5</td></tr> <tr><td>60+</td><td>7</td><td>7</td></tr> <tr><td>60-74</td><td>n/a</td><td>n/a</td></tr> <tr><td>75-89</td><td>n/a</td><td>n/a</td></tr> <tr><td>90+</td><td>n/a</td><td>n/a</td></tr> </tbody> </table>  <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (1) Define the targeted indicators for the split operation $X(e,s,t,i')$ (2) Define in how many indicators the operation should be made (3) Divide the stock of $X(e,s,t,i')$ by the number of classes required <table border="1" data-bbox="542 1339 774 1503"> <thead> <tr> <th colspan="3">ESTIMATION TABLE</th> </tr> <tr> <th>Age</th> <th>Males</th> <th>Females</th> </tr> </thead> <tbody> <tr><td>60+</td><td>7 (1)</td><td>7 (1)</td></tr> <tr><td>N</td><td>3 (2)</td><td>3 (2)</td></tr> <tr><td>60-74</td><td>2,33 (3)</td><td>2,33 (3)</td></tr> <tr><td>75-89</td><td>2,33 (3)</td><td>2,33 (3)</td></tr> <tr><td>90+</td><td>2,33 (3)</td><td>2,33 (3)</td></tr> </tbody> </table> 	ORIGINAL TABLE			Age	Males	Females	0-14	20	20	15-29	15	15	30-44	10	10	45-59	5	5	60+	7	7	60-74	n/a	n/a	75-89	n/a	n/a	90+	n/a	n/a	ESTIMATION TABLE			Age	Males	Females	60+	7 (1)	7 (1)	N	3 (2)	3 (2)	60-74	2,33 (3)	2,33 (3)	75-89	2,33 (3)	2,33 (3)	90+	2,33 (3)	2,33 (3)
ORIGINAL TABLE																																																				
Age	Males	Females																																																		
0-14	20	20																																																		
15-29	15	15																																																		
30-44	10	10																																																		
45-59	5	5																																																		
60+	7	7																																																		
60-74	n/a	n/a																																																		
75-89	n/a	n/a																																																		
90+	n/a	n/a																																																		
ESTIMATION TABLE																																																				
Age	Males	Females																																																		
60+	7 (1)	7 (1)																																																		
N	3 (2)	3 (2)																																																		
60-74	2,33 (3)	2,33 (3)																																																		
75-89	2,33 (3)	2,33 (3)																																																		
90+	2,33 (3)	2,33 (3)																																																		

Name	I3 - Estimation by using bivariate or multivariate regression																																																														
Conditions of use	<ul style="list-style-type: none"> → If $X(e, s, t, i)$, is statistically correlated with $X(e, s, t, i')$, $X(e, s, t, i'')$ and $X(e, s, t, i''')$: significance and correlation → $X(e, s, t, i')$, $X(e, s, t, i'')$ and $X(e, s, t, i''')$ are known → Understand the significance of the relation between i and i' 																																																														
Graphics illustration	<p style="text-align: center;">THEMATIC DIMENSION (I)</p> 																																																														
Textual explanation	<p>Estimation based on thematic dimension (I) The value has been deducted from the parameters of a regression model putting in relation [mention the name of the indicator $X(e, s, t, i)$] and [mention the names of the other indicators of the regression]. The significance of the model has been carefully checked.</p>																																																														
Mathematic formalization	<p>Bivariate regression</p> $X(e, s, t, i) = a \times X(e, s, t, i') + b$ <p>Multivariate regression</p> $X(e, s, t, i) = a_1 X(e, s, t, i') + a_2 X(e, s, t, i'') + \dots + a_n X(e, s, t, i''') + b$																																																														
Example	<table border="1" style="display: inline-table; margin-right: 20px;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Temp. (i)</th> <th>Alt. (i')</th> </tr> </thead> <tbody> <tr><td>X1</td><td>0</td><td>2000</td></tr> <tr><td>X2</td><td>3</td><td>1500</td></tr> <tr><td>X3</td><td>6</td><td>1000</td></tr> <tr><td>X4</td><td>10</td><td>500</td></tr> <tr><td>X5</td><td>8</td><td>1000</td></tr> <tr><td>X6</td><td>5</td><td>1500</td></tr> <tr><td>X7</td><td>2</td><td>2000</td></tr> <tr><td>X8</td><td>-2</td><td>2500</td></tr> <tr><td>X9</td><td>n/a</td><td>800</td></tr> </tbody> </table> <table border="1" style="display: inline-table;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIMATION TABLE</th> <th>ID</th> <th>Temp. (i)</th> <th>Alt. (i')</th> </tr> </thead> <tbody> <tr><td>X1</td><td>0</td><td>2000</td></tr> <tr><td>X2</td><td>3</td><td>1500</td></tr> <tr><td>X3</td><td>6</td><td>1000</td></tr> <tr><td>X4</td><td>10</td><td>500</td></tr> <tr><td>X5</td><td>8</td><td>1000</td></tr> <tr><td>X6</td><td>5</td><td>1500</td></tr> <tr><td>X7</td><td>2</td><td>2000</td></tr> <tr><td>X8</td><td>-2</td><td>2500</td></tr> <tr><td>X9</td><td>8,2 (5)</td><td>800</td></tr> </tbody> </table> <div style="text-align: center; margin-top: 20px;">  <p>Altitude: $X(e, s, t, i')$ (1) $X(e, s, t, i) = -0,006(X(e, s, t, i')) + 13$ (2) $R^2 = 0,9474$</p> </div> <p>HOW TO PROCEED?</p> <ol style="list-style-type: none"> (1) Estimate parameters by using known observations and evaluate their precision. (2) Evaluate the R-Square and the significance of the model. (3) In case of multivariate regression, evaluate the influence of the indicators in the model (globally and individually). (4) Detect observations who can influence the result (residuals). (5) If (1), (2), (3) and (4) proves that the statistical relation exists, estimate $X(e, s, t, i)$ thanks to the parameters calculated by the regression. 	ORIGINAL TABLE	ID	Temp. (i)	Alt. (i')	X1	0	2000	X2	3	1500	X3	6	1000	X4	10	500	X5	8	1000	X6	5	1500	X7	2	2000	X8	-2	2500	X9	n/a	800	ESTIMATION TABLE	ID	Temp. (i)	Alt. (i')	X1	0	2000	X2	3	1500	X3	6	1000	X4	10	500	X5	8	1000	X6	5	1500	X7	2	2000	X8	-2	2500	X9	8,2 (5)	800
ORIGINAL TABLE	ID	Temp. (i)	Alt. (i')																																																												
X1	0	2000																																																													
X2	3	1500																																																													
X3	6	1000																																																													
X4	10	500																																																													
X5	8	1000																																																													
X6	5	1500																																																													
X7	2	2000																																																													
X8	-2	2500																																																													
X9	n/a	800																																																													
ESTIMATION TABLE	ID	Temp. (i)	Alt. (i')																																																												
X1	0	2000																																																													
X2	3	1500																																																													
X3	6	1000																																																													
X4	10	500																																																													
X5	8	1000																																																													
X6	5	1500																																																													
X7	2	2000																																																													
X8	-2	2500																																																													
X9	8,2 (5)	800																																																													

Multidimensional estimations methods

Multi-dimensional estimation methods are results from the combination of two or more one-dimensional methods. Generally they are more accurate and capitalize on more information.



Estimation based on the source and spatial dimensions (SE)

Alternative sources of information can be used when the main source does not provide the targeted information. The idea here is to replace the missing value $X(e,s,t,i)$ by a known value $X(e,s',t,i)$ where s' is another organism of statistics. However, it is very important that the values coming from another statistics provider fits correctly with the values and the hierarchy of the primary source. This is the reason why it is important to operate a space harmonization in a second step. In that order, $X(\text{parent}(e),s,t,i) = X(\text{parent}(e),s',t,i)$ by defining a correlation factor α depending of the deviation between $X(\text{parent}(e),s,t,i)$ and $X(\text{parent}(e),s',t,i)$.

Name	SE1 – Estimation using other data sources and space harmonization																																																																																				
Conditions of use	<ul style="list-style-type: none"> ➔ No data at lower level of the hierarchy in X(e,s,t,i). ➔ Information available at a upper level of the hierarchy X((parent(e),s,t,i) ➔ Data at lower level of the hierarchy available from another organism of statistics X(e,s',t,i) 																																																																																				
Graphics illustration																																																																																					
Textual explanation	<p>Estimation based on source and space dimension (S, E) The values have been collected from [name of the organism of statistics] , then, adjusted by ESPON M4D Project in a way that the sum of the values of the children units (e1, e2...en) are equal to the value of the parent unit (e) In this case, [name of the hierarchy, e.g. NUTS3] values coming from [name of the organism of statistics of X(e,s',t,i)] is adjusted to the [name of the hierarchy, e.g. NUTS0] values delivered by [Name of the organism of statistics of X(e,s,t,i), e.g. generally Eurostat]</p>																																																																																				
Mathematic formalization	$X(e, s, t, i) = \alpha X(e, s', t, i) \text{ where } \alpha = \frac{X(\text{parent}(e), s, t, i)}{X(\text{parent}(e), s', t, i)}$																																																																																				
Example	<table border="1" data-bbox="499 1361 874 1525"> <thead> <tr> <th></th> <th>ID</th> <th>Object type</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td rowspan="6" style="writing-mode: vertical-rl; transform: rotate(180deg);">s TABLE</td> <td>FR11</td> <td>NUTS2</td> <td>25</td> </tr> <tr> <td>FR12</td> <td>NUTS2</td> <td>25</td> </tr> <tr> <td>FR111</td> <td>NUTS3</td> <td>n/a</td> </tr> <tr> <td>FR112</td> <td>NUTS3</td> <td>n/a</td> </tr> <tr> <td>FR121</td> <td>NUTS3</td> <td>n/a</td> </tr> <tr> <td>FR122</td> <td>NUTS3</td> <td>n/a</td> </tr> </tbody> </table> <table border="1" data-bbox="890 1312 1241 1424"> <thead> <tr> <th></th> <th>ID</th> <th>Object type</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td rowspan="4" style="writing-mode: vertical-rl; transform: rotate(180deg);">s' TABLE</td> <td>FR111</td> <td>NUTS3</td> <td>17</td> </tr> <tr> <td>FR112</td> <td>NUTS3</td> <td>10</td> </tr> <tr> <td>FR121</td> <td>NUTS3</td> <td>15</td> </tr> <tr> <td>FR122</td> <td>NUTS3</td> <td>9</td> </tr> </tbody> </table> <table border="1" data-bbox="890 1447 1358 1585"> <thead> <tr> <th></th> <th>ID</th> <th>Val.</th> <th>Dist</th> <th>Dev</th> <th>Estim</th> <th>Check</th> </tr> </thead> <tbody> <tr> <td rowspan="7" style="writing-mode: vertical-rl; transform: rotate(180deg);">EST. TABLE</td> <td>FR11</td> <td>25</td> <td>27(4)</td> <td>0,93(5)</td> <td></td> <td>25 (7)</td> </tr> <tr> <td>FR12</td> <td>25</td> <td>26(4)</td> <td>0,96(5)</td> <td></td> <td>25 (7)</td> </tr> <tr> <td>FR111</td> <td>17(3)</td> <td></td> <td></td> <td>15,7 (6)</td> <td></td> </tr> <tr> <td>FR112</td> <td>10(3)</td> <td></td> <td></td> <td>9,3 (6)</td> <td></td> </tr> <tr> <td>FR121</td> <td>15(3)</td> <td></td> <td></td> <td>14,4 (6)</td> <td></td> </tr> <tr> <td>FR122</td> <td>9(3)</td> <td></td> <td></td> <td>8,6 (6)</td> <td></td> </tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (1) Download data at lower level from another organism of statistics (s'). (2) Save the table downloaded from (s') in a dedicated space (s') (3) Merge (s) and (s') in the same table (4) Sum the values coming from (s') of the childs (e1,e2,..en) available in s for its parent (e) Ex: if the data is available at NUTS2 level for Eurostat data, aggregate the NUTS3 coming from the other organism of statistics in NUTS2. (5) Calculate the deviation (α) by dividing X(parent(e),s,t,i) and X(parent(e),s',t,i). (6) Use the deviation calculated above as correlation factor: X(e,s,t,i) = α X(e,s',t,i) (7) Check if the sum of the estimation at lower level is equal to the to the value delivered by Eurostat at upper level <p><i>In ESPON tables: Estimation of population in Scotland and Northern Ireland from 2001 to 2006, where NUTS3 values are missing and NUTS0 is available.</i></p>		ID	Object type	Value	s TABLE	FR11	NUTS2	25	FR12	NUTS2	25	FR111	NUTS3	n/a	FR112	NUTS3	n/a	FR121	NUTS3	n/a	FR122	NUTS3	n/a		ID	Object type	Value	s' TABLE	FR111	NUTS3	17	FR112	NUTS3	10	FR121	NUTS3	15	FR122	NUTS3	9		ID	Val.	Dist	Dev	Estim	Check	EST. TABLE	FR11	25	27(4)	0,93(5)		25 (7)	FR12	25	26(4)	0,96(5)		25 (7)	FR111	17(3)			15,7 (6)		FR112	10(3)			9,3 (6)		FR121	15(3)			14,4 (6)		FR122	9(3)			8,6 (6)	
	ID	Object type	Value																																																																																		
s TABLE	FR11	NUTS2	25																																																																																		
	FR12	NUTS2	25																																																																																		
	FR111	NUTS3	n/a																																																																																		
	FR112	NUTS3	n/a																																																																																		
	FR121	NUTS3	n/a																																																																																		
	FR122	NUTS3	n/a																																																																																		
	ID	Object type	Value																																																																																		
s' TABLE	FR111	NUTS3	17																																																																																		
	FR112	NUTS3	10																																																																																		
	FR121	NUTS3	15																																																																																		
	FR122	NUTS3	9																																																																																		
	ID	Val.	Dist	Dev	Estim	Check																																																																															
EST. TABLE	FR11	25	27(4)	0,93(5)		25 (7)																																																																															
	FR12	25	26(4)	0,96(5)		25 (7)																																																																															
	FR111	17(3)			15,7 (6)																																																																																
	FR112	10(3)			9,3 (6)																																																																																
	FR121	15(3)			14,4 (6)																																																																																
	FR122	9(3)			8,6 (6)																																																																																

Estimation based on the space and time dimensions (TE)



In Eurostat tables, the following situation occurs very often: it is possible to have information at t_1 or t_2 for units at the lower hierarchical level $X(e,s,t_1,i)$, but this information is available at t only for parent units $X(\text{parent}(e),s,t,i)$ and/or for some units of the same hierarchical level ($X(e',s,t,i)$). There is potentially an infinity of situations and solutions but we can formalize some concrete examples of configurations with dedicated solutions correctly documented.

Name	TE1 - Retropotation and space harmonization																																																																																																																																																			
Conditions of use	<ul style="list-style-type: none"> ➔ No data at lower level of the hierarchy at T. ➔ Information available at a upper level of the hierarchy at T (parent(e)) ➔ Data at lower level available at $T+1$ and $T+N$ 																																																																																																																																																			
Graphics illustration																																																																																																																																																				
Textual explanation	<p>Estimation based on the time and space dimensions (ET)</p> <p>1/ Time dimension - [linear/power] retropotation. This method uses the two closest neighbours placed in time after ([mention the two years of reference]) the value estimated.</p> <p>2/ Space dimension – The estimated values have been adjusted in a way that the sum of values of children units (e1,e2...en) are equal to the value of the parent unit (parent(e1,e2..en)).</p> <p>In this case, the [mention the lower level, e.g.] NUTS3 values coming from the estimation are adjusted to [mention the upper level, e.g. NUTS0 value].</p>																																																																																																																																																			
Mathematic formalization	<ul style="list-style-type: none"> • Linear model $X(e,s,t,i) = a \left(X(e,s,t+1,i) - \frac{X(e,s,t+n,i) - X(e,s,t+1,i)}{(t+n) - (t+1)} ((t+1) - t) \right)$ <ul style="list-style-type: none"> • Power model $X(e,s,t,i) = a \left(X(e,s,t+1,i) \left(\frac{X(e,s,t+n,i)}{X(e,s,t+1,i)} \right)^{\frac{(t-(t+1))}{((t+n)-(t+1))}} \right)$ <p>where $a = \frac{X(\text{parent}(e),s,t,i)}{X(\text{parent}(e),s,t,i)}$ and $X(\text{parent}(e),s,t',i) = \sum_{e'} X(e,s,t',i)$</p>																																																																																																																																																			
Example	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Object type</th> <th>Value 1995</th> <th>Value 2000</th> <th>Value 2005</th> <th>Value 2010</th> </tr> </thead> <tbody> <tr><td></td><td>FR</td><td>NUTS0</td><td>160</td><td>180</td><td>200</td><td>220</td></tr> <tr><td></td><td>FR111</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>10</td><td>12</td></tr> <tr><td></td><td>FR112</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>20</td><td>23</td></tr> <tr><td></td><td>FR121</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>10</td><td>14</td></tr> <tr><td></td><td>FR122</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>60</td><td>62</td></tr> <tr><td></td><td>FR211</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>10</td><td>8</td></tr> <tr><td></td><td>FR212</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>10</td><td>13</td></tr> <tr><td></td><td>FR221</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>30</td><td>32</td></tr> <tr><td></td><td>FR222</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>50</td><td>56</td></tr> </tbody> </table> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM TABLE</th> <th>ID</th> <th>Object type</th> <th>Est. T 1995</th> <th>Est.T 2000</th> <th>Adj (E) 1995</th> <th>Adj (E) 2000</th> </tr> </thead> <tbody> <tr><td></td><td>FR</td><td>NUTS0</td><td>171,1(2)</td><td>183,9(2)</td><td>0,93(3)</td><td>0,98(3)</td></tr> <tr><td></td><td>FR111</td><td>NUTS3</td><td>6,9(1)</td><td>8,3(1)</td><td>6,5(4)</td><td>8,2(4)</td></tr> <tr><td></td><td>FR112</td><td>NUTS3</td><td>15,1(1)</td><td>17,4(1)</td><td>14,1(4)</td><td>17,0(4)</td></tr> <tr><td></td><td>FR121</td><td>NUTS3</td><td>5,1(1)</td><td>7,1(1)</td><td>4,8(4)</td><td>7,0(4)</td></tr> <tr><td></td><td>FR122</td><td>NUTS3</td><td>56,2(1)</td><td>58,1(1)</td><td>52,5(4)</td><td>56,8(4)</td></tr> <tr><td></td><td>FR211</td><td>NUTS3</td><td>15,6(1)</td><td>12,5(1)</td><td>14,6(4)</td><td>12,2(4)</td></tr> <tr><td></td><td>FR212</td><td>NUTS3</td><td>5,9(1)</td><td>7,7(1)</td><td>5,5(4)</td><td>7,5(4)</td></tr> <tr><td></td><td>FR221</td><td>NUTS3</td><td>26,4(1)</td><td>28,1(1)</td><td>24,7(4)</td><td>27,5(4)</td></tr> <tr><td></td><td>FR222</td><td>NUTS3</td><td>39,9(1)</td><td>44,6(1)</td><td>37,3(4)</td><td>43,7(4)</td></tr> <tr><td></td><td></td><td></td><td></td><td>SUM</td><td>160(5)</td><td>180</td></tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (1) Estimate the evolution at T thanks to available information at lower level at $T+1$ and $T+N$. (cf estimation method T3, p.16) (2) Calculate the sum of (1) for the units of the same hierarchy (e') in order to have value the of $X(\text{parent}(e),s,t',i)$ (3) Estimate the deviation between $X(\text{parent}(e),s,t,i)$ and $X \text{parent}(e),s,t',i)$ (4) Multiply the value of the deviation (3) by the temporal estimation (1) (5) Check if the sum of values estimated is equal to the value delivered at upper level. 	ORIGINAL TABLE	ID	Object type	Value 1995	Value 2000	Value 2005	Value 2010		FR	NUTS0	160	180	200	220		FR111	NUTS3	n/a	n/a	10	12		FR112	NUTS3	n/a	n/a	20	23		FR121	NUTS3	n/a	n/a	10	14		FR122	NUTS3	n/a	n/a	60	62		FR211	NUTS3	n/a	n/a	10	8		FR212	NUTS3	n/a	n/a	10	13		FR221	NUTS3	n/a	n/a	30	32		FR222	NUTS3	n/a	n/a	50	56	ESTIM TABLE	ID	Object type	Est. T 1995	Est.T 2000	Adj (E) 1995	Adj (E) 2000		FR	NUTS0	171,1(2)	183,9(2)	0,93(3)	0,98(3)		FR111	NUTS3	6,9(1)	8,3(1)	6,5(4)	8,2(4)		FR112	NUTS3	15,1(1)	17,4(1)	14,1(4)	17,0(4)		FR121	NUTS3	5,1(1)	7,1(1)	4,8(4)	7,0(4)		FR122	NUTS3	56,2(1)	58,1(1)	52,5(4)	56,8(4)		FR211	NUTS3	15,6(1)	12,5(1)	14,6(4)	12,2(4)		FR212	NUTS3	5,9(1)	7,7(1)	5,5(4)	7,5(4)		FR221	NUTS3	26,4(1)	28,1(1)	24,7(4)	27,5(4)		FR222	NUTS3	39,9(1)	44,6(1)	37,3(4)	43,7(4)					SUM	160(5)	180
ORIGINAL TABLE	ID	Object type	Value 1995	Value 2000	Value 2005	Value 2010																																																																																																																																														
	FR	NUTS0	160	180	200	220																																																																																																																																														
	FR111	NUTS3	n/a	n/a	10	12																																																																																																																																														
	FR112	NUTS3	n/a	n/a	20	23																																																																																																																																														
	FR121	NUTS3	n/a	n/a	10	14																																																																																																																																														
	FR122	NUTS3	n/a	n/a	60	62																																																																																																																																														
	FR211	NUTS3	n/a	n/a	10	8																																																																																																																																														
	FR212	NUTS3	n/a	n/a	10	13																																																																																																																																														
	FR221	NUTS3	n/a	n/a	30	32																																																																																																																																														
	FR222	NUTS3	n/a	n/a	50	56																																																																																																																																														
ESTIM TABLE	ID	Object type	Est. T 1995	Est.T 2000	Adj (E) 1995	Adj (E) 2000																																																																																																																																														
	FR	NUTS0	171,1(2)	183,9(2)	0,93(3)	0,98(3)																																																																																																																																														
	FR111	NUTS3	6,9(1)	8,3(1)	6,5(4)	8,2(4)																																																																																																																																														
	FR112	NUTS3	15,1(1)	17,4(1)	14,1(4)	17,0(4)																																																																																																																																														
	FR121	NUTS3	5,1(1)	7,1(1)	4,8(4)	7,0(4)																																																																																																																																														
	FR122	NUTS3	56,2(1)	58,1(1)	52,5(4)	56,8(4)																																																																																																																																														
	FR211	NUTS3	15,6(1)	12,5(1)	14,6(4)	12,2(4)																																																																																																																																														
	FR212	NUTS3	5,9(1)	7,7(1)	5,5(4)	7,5(4)																																																																																																																																														
	FR221	NUTS3	26,4(1)	28,1(1)	24,7(4)	27,5(4)																																																																																																																																														
	FR222	NUTS3	39,9(1)	44,6(1)	37,3(4)	43,7(4)																																																																																																																																														
				SUM	160(5)	180																																																																																																																																														

Name	TE2 - Extrapolation and space harmonization																																																																																																																																																			
Conditions of use	<ul style="list-style-type: none"> ➔ No data at lower level of the hierarchy at T. ➔ Information available at a upper level of the hierarchy at T (parent(e)) ➔ Data at lower level available at $T-1$ and $T-N$ 																																																																																																																																																			
Graphics illustration																																																																																																																																																				
Textual explanation	<p>Estimation based on the time and space dimensions (ET)</p> <p>1/ Time dimension - [linear/power] extrapolation. This method uses the two closest neighbours placed in time before (<i>mention the two years of reference</i>) the value estimated.</p> <p>2/ Space dimension – The estimated values have been adjusted in a way that the sum of values of children units (e1,e2...en) are equal to the value of the parent unit (parent(e1,e2..en)).</p> <p>In this case, the <i>mention the lower level, e.g.] NUTS3 values coming from the estimation are adjusted to [mention the upper level, e.g. NUTS0 value].</i></p>																																																																																																																																																			
Mathematic formalization	<ul style="list-style-type: none"> • Linear model $X(e, s, t, i) = a \left(X(e, s, t-1, i) - \frac{X(e, s, t-1, i) - X(e, s, t-n, i)}{(t-1) - (t-n)} (t - (t-1)) \right)$ <ul style="list-style-type: none"> • Power model $X(e, s, t, i) = a \left(X(e, s, t-1, i) \left(\frac{X(e, s, t-1, i)}{X(e, s, t-n, i)} \right)^{\frac{t-(t-1)}{(t-1)-(t-n)}} \right)$ <p>where $a = \frac{X(\text{parent}(e), s, t, i)}{X(e, s, t, i)}$ and $X(\text{parent}(e), s, t', i) = \sum_{e'} X(e, s, t', i)$</p>																																																																																																																																																			
Example	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Object type</th> <th>Value 1995</th> <th>Value 2000</th> <th>Value 2005</th> <th>Value 2010</th> </tr> </thead> <tbody> <tr><td></td><td>FR</td><td>NUTS0</td><td>200</td><td>220</td><td>270</td><td>290</td></tr> <tr><td></td><td>FR111</td><td>NUTS3</td><td>10</td><td>12</td><td>n/a</td><td>n/a</td></tr> <tr><td></td><td>FR112</td><td>NUTS3</td><td>20</td><td>23</td><td>n/a</td><td>n/a</td></tr> <tr><td></td><td>FR121</td><td>NUTS3</td><td>10</td><td>14</td><td>n/a</td><td>n/a</td></tr> <tr><td></td><td>FR122</td><td>NUTS3</td><td>60</td><td>62</td><td>n/a</td><td>n/a</td></tr> <tr><td></td><td>FR211</td><td>NUTS3</td><td>10</td><td>8</td><td>n/a</td><td>n/a</td></tr> <tr><td></td><td>FR212</td><td>NUTS3</td><td>10</td><td>13</td><td>n/a</td><td>n/a</td></tr> <tr><td></td><td>FR221</td><td>NUTS3</td><td>30</td><td>32</td><td>n/a</td><td>n/a</td></tr> <tr><td></td><td>FR222</td><td>NUTS3</td><td>50</td><td>56</td><td>n/a</td><td>n/a</td></tr> </tbody> </table> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIM TABLE</th> <th>ID</th> <th>Object type</th> <th>Est. T 1995</th> <th>Est. T 2000</th> <th>Adj (E) 1995</th> <th>Adj (E) 2000</th> </tr> </thead> <tbody> <tr><td></td><td>FR</td><td>NUTS0</td><td>244,7 (2)</td><td>275,1 (2)</td><td>1,10 (3)</td><td>1,05 (3)</td></tr> <tr><td></td><td>FR111</td><td>NUTS3</td><td>14,4 (1)</td><td>17,3 (1)</td><td>15,9 (4)</td><td>18,2 (4)</td></tr> <tr><td></td><td>FR112</td><td>NUTS3</td><td>26,5 (1)</td><td>30,4 (1)</td><td>29,2 (4)</td><td>32,1 (4)</td></tr> <tr><td></td><td>FR121</td><td>NUTS3</td><td>19,6 (1)</td><td>27,4 (1)</td><td>21,6 (4)</td><td>28,9 (4)</td></tr> <tr><td></td><td>FR122</td><td>NUTS3</td><td>64,1 (1)</td><td>66,2 (1)</td><td>70,7 (4)</td><td>69,8 (4)</td></tr> <tr><td></td><td>FR211</td><td>NUTS3</td><td>6,4 (1)</td><td>5,1 (1)</td><td>7,1 (4)</td><td>5,4 (4)</td></tr> <tr><td></td><td>FR212</td><td>NUTS3</td><td>16,9 (1)</td><td>22,0 (1)</td><td>18,6 (4)</td><td>23,2 (4)</td></tr> <tr><td></td><td>FR221</td><td>NUTS3</td><td>34,1 (1)</td><td>36,4 (1)</td><td>37,7 (4)</td><td>38,4 (4)</td></tr> <tr><td></td><td>FR222</td><td>NUTS3</td><td>62,7 (1)</td><td>70,2 (1)</td><td>69,2 (4)</td><td>74,1 (4)</td></tr> <tr><td></td><td></td><td></td><td></td><td><i>SUM</i></td><td>270 (5)</td><td>290 (5)</td></tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (1) Estimate the evolution at T thanks to available information at lower level at $T+1$ and $T+N$. (cf estimation method T4, p.17, here the power model has been used) (2) Calculate the sum of (1) for the units of the same hierarchy (e') in order to have value the of $X(\text{parent}(e), s, t', i)$ (3) Estimate the deviation between $X(\text{parent}(e), s, t, i)$ and $X(\text{parent}(e), s, t', i)$ (4) Multiply the value of the deviation (3) by the temporal estimation (1) (5) Check if the sum of values estimated is equal to the value delivered at upper level. 	ORIGINAL TABLE	ID	Object type	Value 1995	Value 2000	Value 2005	Value 2010		FR	NUTS0	200	220	270	290		FR111	NUTS3	10	12	n/a	n/a		FR112	NUTS3	20	23	n/a	n/a		FR121	NUTS3	10	14	n/a	n/a		FR122	NUTS3	60	62	n/a	n/a		FR211	NUTS3	10	8	n/a	n/a		FR212	NUTS3	10	13	n/a	n/a		FR221	NUTS3	30	32	n/a	n/a		FR222	NUTS3	50	56	n/a	n/a	ESTIM TABLE	ID	Object type	Est. T 1995	Est. T 2000	Adj (E) 1995	Adj (E) 2000		FR	NUTS0	244,7 (2)	275,1 (2)	1,10 (3)	1,05 (3)		FR111	NUTS3	14,4 (1)	17,3 (1)	15,9 (4)	18,2 (4)		FR112	NUTS3	26,5 (1)	30,4 (1)	29,2 (4)	32,1 (4)		FR121	NUTS3	19,6 (1)	27,4 (1)	21,6 (4)	28,9 (4)		FR122	NUTS3	64,1 (1)	66,2 (1)	70,7 (4)	69,8 (4)		FR211	NUTS3	6,4 (1)	5,1 (1)	7,1 (4)	5,4 (4)		FR212	NUTS3	16,9 (1)	22,0 (1)	18,6 (4)	23,2 (4)		FR221	NUTS3	34,1 (1)	36,4 (1)	37,7 (4)	38,4 (4)		FR222	NUTS3	62,7 (1)	70,2 (1)	69,2 (4)	74,1 (4)					<i>SUM</i>	270 (5)	290 (5)
ORIGINAL TABLE	ID	Object type	Value 1995	Value 2000	Value 2005	Value 2010																																																																																																																																														
	FR	NUTS0	200	220	270	290																																																																																																																																														
	FR111	NUTS3	10	12	n/a	n/a																																																																																																																																														
	FR112	NUTS3	20	23	n/a	n/a																																																																																																																																														
	FR121	NUTS3	10	14	n/a	n/a																																																																																																																																														
	FR122	NUTS3	60	62	n/a	n/a																																																																																																																																														
	FR211	NUTS3	10	8	n/a	n/a																																																																																																																																														
	FR212	NUTS3	10	13	n/a	n/a																																																																																																																																														
	FR221	NUTS3	30	32	n/a	n/a																																																																																																																																														
	FR222	NUTS3	50	56	n/a	n/a																																																																																																																																														
ESTIM TABLE	ID	Object type	Est. T 1995	Est. T 2000	Adj (E) 1995	Adj (E) 2000																																																																																																																																														
	FR	NUTS0	244,7 (2)	275,1 (2)	1,10 (3)	1,05 (3)																																																																																																																																														
	FR111	NUTS3	14,4 (1)	17,3 (1)	15,9 (4)	18,2 (4)																																																																																																																																														
	FR112	NUTS3	26,5 (1)	30,4 (1)	29,2 (4)	32,1 (4)																																																																																																																																														
	FR121	NUTS3	19,6 (1)	27,4 (1)	21,6 (4)	28,9 (4)																																																																																																																																														
	FR122	NUTS3	64,1 (1)	66,2 (1)	70,7 (4)	69,8 (4)																																																																																																																																														
	FR211	NUTS3	6,4 (1)	5,1 (1)	7,1 (4)	5,4 (4)																																																																																																																																														
	FR212	NUTS3	16,9 (1)	22,0 (1)	18,6 (4)	23,2 (4)																																																																																																																																														
	FR221	NUTS3	34,1 (1)	36,4 (1)	37,7 (4)	38,4 (4)																																																																																																																																														
	FR222	NUTS3	62,7 (1)	70,2 (1)	69,2 (4)	74,1 (4)																																																																																																																																														
				<i>SUM</i>	270 (5)	290 (5)																																																																																																																																														

Name	TE3 - Interpolation and space harmonization – evolution method																																																																																																
Conditions of use	<ul style="list-style-type: none"> ➔ No data at lower level of the hierarchy at <i>T</i>. ➔ Information available at a upper level of the hierarchy at <i>T</i> (parent(e)). ➔ Data at lower level available at <i>T-1</i> and <i>T+1</i>. 																																																																																																
Graphics illustration																																																																																																	
Textual explanation	<p>Estimation based on the time and space dimensions (ET)</p> <p>1/ Time dimension - [linear/power] interpolation. This method uses the two closest neighbours placed in time before and after ([mention the two years of reference]) the value estimated.</p> <p>2/ Space dimension – The estimated values have been adjusted in a way that the sum of values of children units (e1,e2...en) are equal to the value of the parent unit (parent(e1,e2..en)).</p> <p>In this case, the [mention the lower level, e.g.] NUTS3 values coming from the estimation are adjusted to [mention the upper level, e.g. NUTS0 value].</p>																																																																																																
Mathematic formalization	<ul style="list-style-type: none"> • Linear model $X(e, s, t, i) = a \left(X(e, s, t - 1, i) + \frac{X(e, s, t + 1, i) - X(e, s, t - 1, i)}{(t + 1) - (t - 1)} (t - (t - 1)) \right)$ <ul style="list-style-type: none"> • Power model $X(e, s, t, i) = a \left(X(e, s, t - 1, i) \left(\frac{X(e, s, t + 1, i)}{X(e, s, t - 1, i)} \right)^{\frac{(t - (t - 1))}{((t + 1) - (t - 1))}} \right)$ <p>where $a = \frac{X(\text{parent}(e), s, t, i)}{X(\text{parent}(e), s, t, i)}$ and $X(\text{parent}(e), s, t', i) = \sum_e X(e, s, t', i)$</p>																																																																																																
Example	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th rowspan="10" style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Object type</th> <th>Value 2000</th> <th>Value 2001</th> <th>Value 2002</th> </tr> </thead> <tbody> <tr><td>FR</td><td>NUTS0</td><td>160</td><td>180</td><td>200</td></tr> <tr><td>FR111</td><td>NUTS3</td><td>6</td><td>n/a</td><td>10</td></tr> <tr><td>FR112</td><td>NUTS3</td><td>18</td><td>n/a</td><td>20</td></tr> <tr><td>FR121</td><td>NUTS3</td><td>13</td><td>n/a</td><td>10</td></tr> <tr><td>FR122</td><td>NUTS3</td><td>43</td><td>n/a</td><td>60</td></tr> <tr><td>FR211</td><td>NUTS3</td><td>5</td><td>n/a</td><td>10</td></tr> <tr><td>FR212</td><td>NUTS3</td><td>13</td><td>n/a</td><td>10</td></tr> <tr><td>FR221</td><td>NUTS3</td><td>22</td><td>n/a</td><td>30</td></tr> <tr><td>FR222</td><td>NUTS3</td><td>40</td><td>n/a</td><td>50</td></tr> </tbody> </table> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center; margin-top: 5px;"> <thead> <tr> <th rowspan="10" style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIMATION TABLE</th> <th>ID</th> <th>Object type</th> <th>Estim 2001</th> <th>Adjust 2001</th> </tr> </thead> <tbody> <tr><td>FR</td><td>NUTS0</td><td>177,8 (2)</td><td>1,01 (3)</td></tr> <tr><td>FR111</td><td>NUTS3</td><td>7,7 (1)</td><td>7,8 (4)</td></tr> <tr><td>FR112</td><td>NUTS3</td><td>19,0 (1)</td><td>19,2 (4)</td></tr> <tr><td>FR121</td><td>NUTS3</td><td>11,4 (1)</td><td>11,5 (4)</td></tr> <tr><td>FR122</td><td>NUTS3</td><td>50,8 (1)</td><td>51,4 (4)</td></tr> <tr><td>FR211</td><td>NUTS3</td><td>7,1 (1)</td><td>7,2 (4)</td></tr> <tr><td>FR212</td><td>NUTS3</td><td>11,4 (1)</td><td>11,5 (4)</td></tr> <tr><td>FR221</td><td>NUTS3</td><td>25,7 (1)</td><td>26,0 (4)</td></tr> <tr><td>FR222</td><td>NUTS3</td><td>44,7 (1)</td><td>45,3 (4)</td></tr> <tr><td>SUM</td><td></td><td></td><td>180 (5)</td></tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (1) Estimate the evolution at <i>T</i> thanks to available information at lower level at <i>T-1</i> and <i>T+1</i>. (cf estimation method T1, p.14, here the power model has been used). (2) Calculate the sum of (1) for the units of the same hierarchy (e') in order to have value the of X(parent(e),s,t',i). (3) Estimate the deviation between X(parent(e),s,t,i) and X parent(e),s,t',i). (4) Multiply the value of the deviation (3) by the temporal estimation (1). (5) Check if the sum of values estimated is equal to the value delivered at upper level. 	ORIGINAL TABLE	ID	Object type	Value 2000	Value 2001	Value 2002	FR	NUTS0	160	180	200	FR111	NUTS3	6	n/a	10	FR112	NUTS3	18	n/a	20	FR121	NUTS3	13	n/a	10	FR122	NUTS3	43	n/a	60	FR211	NUTS3	5	n/a	10	FR212	NUTS3	13	n/a	10	FR221	NUTS3	22	n/a	30	FR222	NUTS3	40	n/a	50	ESTIMATION TABLE	ID	Object type	Estim 2001	Adjust 2001	FR	NUTS0	177,8 (2)	1,01 (3)	FR111	NUTS3	7,7 (1)	7,8 (4)	FR112	NUTS3	19,0 (1)	19,2 (4)	FR121	NUTS3	11,4 (1)	11,5 (4)	FR122	NUTS3	50,8 (1)	51,4 (4)	FR211	NUTS3	7,1 (1)	7,2 (4)	FR212	NUTS3	11,4 (1)	11,5 (4)	FR221	NUTS3	25,7 (1)	26,0 (4)	FR222	NUTS3	44,7 (1)	45,3 (4)	SUM			180 (5)
ORIGINAL TABLE	ID		Object type	Value 2000	Value 2001	Value 2002																																																																																											
	FR		NUTS0	160	180	200																																																																																											
	FR111		NUTS3	6	n/a	10																																																																																											
	FR112		NUTS3	18	n/a	20																																																																																											
	FR121		NUTS3	13	n/a	10																																																																																											
	FR122		NUTS3	43	n/a	60																																																																																											
	FR211		NUTS3	5	n/a	10																																																																																											
	FR212		NUTS3	13	n/a	10																																																																																											
	FR221		NUTS3	22	n/a	30																																																																																											
	FR222	NUTS3	40	n/a	50																																																																																												
ESTIMATION TABLE	ID	Object type	Estim 2001	Adjust 2001																																																																																													
	FR	NUTS0	177,8 (2)	1,01 (3)																																																																																													
	FR111	NUTS3	7,7 (1)	7,8 (4)																																																																																													
	FR112	NUTS3	19,0 (1)	19,2 (4)																																																																																													
	FR121	NUTS3	11,4 (1)	11,5 (4)																																																																																													
	FR122	NUTS3	50,8 (1)	51,4 (4)																																																																																													
	FR211	NUTS3	7,1 (1)	7,2 (4)																																																																																													
	FR212	NUTS3	11,4 (1)	11,5 (4)																																																																																													
	FR221	NUTS3	25,7 (1)	26,0 (4)																																																																																													
	FR222	NUTS3	44,7 (1)	45,3 (4)																																																																																													
SUM			180 (5)																																																																																														

Name	TE4 - Interpolation and space harmonization – distribution method																																																																																																																												
Conditions of use	<ul style="list-style-type: none"> ➔ No data at lower level of the hierarchy at T. ➔ Information available at a upper level of the hierarchy at T (parent(e)). ➔ Data at lower level available at $T-1$ and $T+1$. 																																																																																																																												
Graphics illustration																																																																																																																													
Textual explanation	<p>Estimation based on the time and space dimensions (ET)</p> <p>1/ Space/time dimension: Calculation of the share of the unit as regard to a given upper level of belonging [mention the hierarchical level, e.g. NUTS0] placed in time before and after the estimated value [mention the years].</p> <p>2/ Average of the share of units (share of the unit before and the share of the unit after the estimated value)</p> <p>2/ Time/space dimension – The ratio calculated in 2/ is multiplied with a value available at upper level [mention the hierarchical level, e.g. NUTS0] at the targeted time period.</p>																																																																																																																												
Mathematic formalization	$X(e, s, t, i) = a \left(\frac{X(e, s, t-1)}{\sum_{e'} X(e', s, t-1)} + \frac{X(e, s, t+1)}{\sum_{e'} X(e', s, t+1)} \right) \text{ where } a = X(\text{parent}(e), s, t, i)$																																																																																																																												
Example	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th rowspan="2" style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Object type</th> <th>Value 2000</th> <th>Value 2001</th> <th>Value 2002</th> <th rowspan="2" style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIMATION TABLE</th> <th>ID</th> <th>Object type</th> <th>Distrib 2000</th> <th>Distrib 2002</th> <th>Average distrib</th> <th>Estim 2002</th> </tr> </thead> <tbody> <tr> <td>FR</td> <td>NUTS0</td> <td>160</td> <td>180</td> <td>200</td> <td>FR</td> <td>NUTS0</td> <td>1 (2)</td> <td>1 (2)</td> <td>1 (4)</td> <td>180</td> </tr> <tr> <td>FR111</td> <td>NUTS3</td> <td>6</td> <td>n/a</td> <td>10</td> <td>FR111</td> <td>NUTS3</td> <td>0,04 (1)</td> <td>0,05 (1)</td> <td>0,04 (3)</td> <td>7,2 (5)</td> </tr> <tr> <td>FR112</td> <td>NUTS3</td> <td>18</td> <td>n/a</td> <td>20</td> <td>FR112</td> <td>NUTS3</td> <td>0,11 (1)</td> <td>0,1 (1)</td> <td>0,11 (3)</td> <td>19,8 (5)</td> </tr> <tr> <td>FR121</td> <td>NUTS3</td> <td>13</td> <td>n/a</td> <td>10</td> <td>FR121</td> <td>NUTS3</td> <td>0,08 (1)</td> <td>0,05 (1)</td> <td>0,07 (3)</td> <td>12,6 (5)</td> </tr> <tr> <td>FR122</td> <td>NUTS3</td> <td>43</td> <td>n/a</td> <td>60</td> <td>FR122</td> <td>NUTS3</td> <td>0,27 (1)</td> <td>0,3 (1)</td> <td>0,28 (3)</td> <td>50,4 (5)</td> </tr> <tr> <td>FR211</td> <td>NUTS3</td> <td>5</td> <td>n/a</td> <td>10</td> <td>FR211</td> <td>NUTS3</td> <td>0,03 (1)</td> <td>0,05 (1)</td> <td>0,04 (3)</td> <td>7,2 (5)</td> </tr> <tr> <td>FR212</td> <td>NUTS3</td> <td>13</td> <td>n/a</td> <td>10</td> <td>FR212</td> <td>NUTS3</td> <td>0,08 (1)</td> <td>0,05 (1)</td> <td>0,07 (3)</td> <td>12,6 (5)</td> </tr> <tr> <td>FR221</td> <td>NUTS3</td> <td>22</td> <td>n/a</td> <td>30</td> <td>FR221</td> <td>NUTS3</td> <td>0,14 (1)</td> <td>0,15 (1)</td> <td>0,14 (3)</td> <td>25,2 (5)</td> </tr> <tr> <td>FR222</td> <td>NUTS3</td> <td>40</td> <td>n/a</td> <td>50</td> <td>FR222</td> <td>NUTS3</td> <td>0,25 (1)</td> <td>0,25 (1)</td> <td>0,25 (3)</td> <td>45 (5)</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>SUM</td> <td></td> <td></td> <td></td> <td></td> <td>180 (6)</td> </tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (1) Estimate the distribution of the stock at $t-1$ and $t+1$ as regard to an upper unit in the hierarchy (Take the lower level of the hierarchy available at t). (2) Check if the sum of the distribution at lower level is equal to 1. (3) Estimate the average distribution between $t-1$ and $t+1$. (4) Check if the sum of the distribution at lower level is equal to 1. (5) At t, multiply the values calculated in (3) by the stock available to the unit available at t at a upper level (the same than the one chosen for calculating (1)). (6) Check if the sum of values estimated at lower level at t is equal to value delivered by at upper level. 	ORIGINAL TABLE	ID	Object type	Value 2000	Value 2001	Value 2002	ESTIMATION TABLE	ID	Object type	Distrib 2000	Distrib 2002	Average distrib	Estim 2002	FR	NUTS0	160	180	200	FR	NUTS0	1 (2)	1 (2)	1 (4)	180	FR111	NUTS3	6	n/a	10	FR111	NUTS3	0,04 (1)	0,05 (1)	0,04 (3)	7,2 (5)	FR112	NUTS3	18	n/a	20	FR112	NUTS3	0,11 (1)	0,1 (1)	0,11 (3)	19,8 (5)	FR121	NUTS3	13	n/a	10	FR121	NUTS3	0,08 (1)	0,05 (1)	0,07 (3)	12,6 (5)	FR122	NUTS3	43	n/a	60	FR122	NUTS3	0,27 (1)	0,3 (1)	0,28 (3)	50,4 (5)	FR211	NUTS3	5	n/a	10	FR211	NUTS3	0,03 (1)	0,05 (1)	0,04 (3)	7,2 (5)	FR212	NUTS3	13	n/a	10	FR212	NUTS3	0,08 (1)	0,05 (1)	0,07 (3)	12,6 (5)	FR221	NUTS3	22	n/a	30	FR221	NUTS3	0,14 (1)	0,15 (1)	0,14 (3)	25,2 (5)	FR222	NUTS3	40	n/a	50	FR222	NUTS3	0,25 (1)	0,25 (1)	0,25 (3)	45 (5)							SUM					180 (6)
ORIGINAL TABLE	ID		Object type	Value 2000	Value 2001	Value 2002	ESTIMATION TABLE		ID	Object type	Distrib 2000	Distrib 2002	Average distrib	Estim 2002																																																																																																															
	FR	NUTS0	160	180	200	FR		NUTS0	1 (2)	1 (2)	1 (4)	180																																																																																																																	
FR111	NUTS3	6	n/a	10	FR111	NUTS3	0,04 (1)	0,05 (1)	0,04 (3)	7,2 (5)																																																																																																																			
FR112	NUTS3	18	n/a	20	FR112	NUTS3	0,11 (1)	0,1 (1)	0,11 (3)	19,8 (5)																																																																																																																			
FR121	NUTS3	13	n/a	10	FR121	NUTS3	0,08 (1)	0,05 (1)	0,07 (3)	12,6 (5)																																																																																																																			
FR122	NUTS3	43	n/a	60	FR122	NUTS3	0,27 (1)	0,3 (1)	0,28 (3)	50,4 (5)																																																																																																																			
FR211	NUTS3	5	n/a	10	FR211	NUTS3	0,03 (1)	0,05 (1)	0,04 (3)	7,2 (5)																																																																																																																			
FR212	NUTS3	13	n/a	10	FR212	NUTS3	0,08 (1)	0,05 (1)	0,07 (3)	12,6 (5)																																																																																																																			
FR221	NUTS3	22	n/a	30	FR221	NUTS3	0,14 (1)	0,15 (1)	0,14 (3)	25,2 (5)																																																																																																																			
FR222	NUTS3	40	n/a	50	FR222	NUTS3	0,25 (1)	0,25 (1)	0,25 (3)	45 (5)																																																																																																																			
						SUM					180 (6)																																																																																																																		

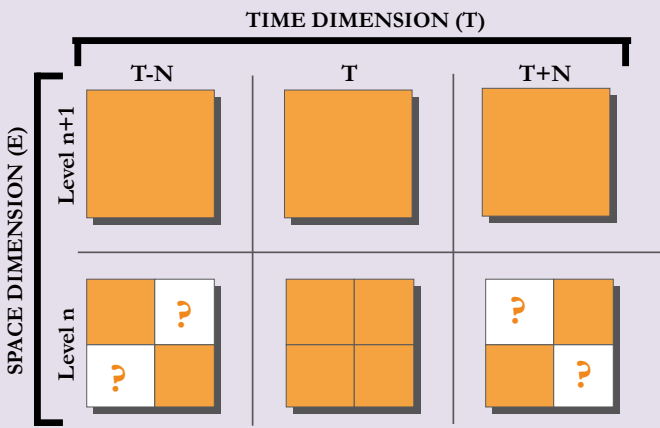
Name	TE5 - Interpolation/retropolation and space harmonization based on distribution method																																																																																																																								
Conditions of use	<ul style="list-style-type: none"> ➔ At lower level of the hierarchy, information is available for a single year (t) ➔ At upper level (parent(e)), the information is available at (t-n) and/or (t+n). 																																																																																																																								
Graphics illustration	<div style="text-align: center;"> <p>TIME DIMENSION (T)</p> </div>																																																																																																																								
Textual explanation	<p>Estimation based on the space and time dimensions (TE)</p> <p>1/ Space/time dimension: Calculation of the share of the unit as regard to its immediate upper level of belonging available [mention the hierarchical level of the nomenclature, e.g. NUTS0] placed in time after or before the estimated value [mention the year]</p> <p>2/ Time/space dimension – The ratio calculated in 1/ is multiplied with its immediate upper level of belonging available available [mention the hierarchical level of the nomenclature, e.g. NUTS0] at the targeted time period.</p>																																																																																																																								
Mathematic formalization	$X(e, s, t - N, i) = a \left(\frac{X(e, s, t, i)}{\sum_{e'} X(e', s, t, i)} \right)$ <p>where $a = X(\text{parent}(e), s, t - N, i)$</p>																																																																																																																								
Example	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</th> <th>ID</th> <th>Object type</th> <th>Value t-2</th> <th>Value t-1</th> <th>Value t</th> </tr> </thead> <tbody> <tr><td></td><td>FR</td><td>NUTS0</td><td>160</td><td>180</td><td>200</td></tr> <tr><td></td><td>FR111</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>10</td></tr> <tr><td></td><td>FR112</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>20</td></tr> <tr><td></td><td>FR121</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>10</td></tr> <tr><td></td><td>FR122</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>60</td></tr> <tr><td></td><td>FR211</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>10</td></tr> <tr><td></td><td>FR212</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>10</td></tr> <tr><td></td><td>FR221</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>30</td></tr> <tr><td></td><td>FR222</td><td>NUTS3</td><td>n/a</td><td>n/a</td><td>50</td></tr> </tbody> </table> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <thead> <tr> <th style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIMATION TABLE</th> <th>ID</th> <th>Object type</th> <th>Distrib t</th> <th>Estim t-2</th> <th>Estim t-1</th> </tr> </thead> <tbody> <tr><td></td><td>FR</td><td>NUTS0</td><td>1 (2)</td><td>160 (4)</td><td>180 (4)</td></tr> <tr><td></td><td>FR111</td><td>NUTS3</td><td>0,05 (1)</td><td>8 (3)</td><td>9 (3)</td></tr> <tr><td></td><td>FR112</td><td>NUTS3</td><td>0,1 (1)</td><td>16 (3)</td><td>18 (3)</td></tr> <tr><td></td><td>FR121</td><td>NUTS3</td><td>0,05 (1)</td><td>8 (3)</td><td>9 (3)</td></tr> <tr><td></td><td>FR122</td><td>NUTS3</td><td>0,3 (1)</td><td>48 (3)</td><td>54 (3)</td></tr> <tr><td></td><td>FR211</td><td>NUTS3</td><td>0,05 (1)</td><td>8 (3)</td><td>9 (3)</td></tr> <tr><td></td><td>FR212</td><td>NUTS3</td><td>0,05 (1)</td><td>8 (3)</td><td>9 (3)</td></tr> <tr><td></td><td>FR221</td><td>NUTS3</td><td>0,15 (1)</td><td>24 (3)</td><td>27 (3)</td></tr> <tr><td></td><td>FR222</td><td>NUTS3</td><td>0,25 (1)</td><td>40 (3)</td><td>45 (3)</td></tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (1) Estimate the distribution at lower level at t as regard to its immediate (if possible) upper level of belonging. (2) Check if the sum of the distribution at lower level is equal to 1. (3) At t-1 and t-2, multiply the values calculated in (1) by the stock available with its upper level of belonging (take the same hierarchical level than in (1)). For instance: $poptot(T, UKM65) = \text{Distribution}(T+1, UKM65/UKM) * poptot(T, UKM)$ (4) Check if the sum of values estimated at lower level at t and t-1 is equal to value delivered by Eurostat at upper level. 	ORIGINAL TABLE	ID	Object type	Value t-2	Value t-1	Value t		FR	NUTS0	160	180	200		FR111	NUTS3	n/a	n/a	10		FR112	NUTS3	n/a	n/a	20		FR121	NUTS3	n/a	n/a	10		FR122	NUTS3	n/a	n/a	60		FR211	NUTS3	n/a	n/a	10		FR212	NUTS3	n/a	n/a	10		FR221	NUTS3	n/a	n/a	30		FR222	NUTS3	n/a	n/a	50	ESTIMATION TABLE	ID	Object type	Distrib t	Estim t-2	Estim t-1		FR	NUTS0	1 (2)	160 (4)	180 (4)		FR111	NUTS3	0,05 (1)	8 (3)	9 (3)		FR112	NUTS3	0,1 (1)	16 (3)	18 (3)		FR121	NUTS3	0,05 (1)	8 (3)	9 (3)		FR122	NUTS3	0,3 (1)	48 (3)	54 (3)		FR211	NUTS3	0,05 (1)	8 (3)	9 (3)		FR212	NUTS3	0,05 (1)	8 (3)	9 (3)		FR221	NUTS3	0,15 (1)	24 (3)	27 (3)		FR222	NUTS3	0,25 (1)	40 (3)	45 (3)
ORIGINAL TABLE	ID	Object type	Value t-2	Value t-1	Value t																																																																																																																				
	FR	NUTS0	160	180	200																																																																																																																				
	FR111	NUTS3	n/a	n/a	10																																																																																																																				
	FR112	NUTS3	n/a	n/a	20																																																																																																																				
	FR121	NUTS3	n/a	n/a	10																																																																																																																				
	FR122	NUTS3	n/a	n/a	60																																																																																																																				
	FR211	NUTS3	n/a	n/a	10																																																																																																																				
	FR212	NUTS3	n/a	n/a	10																																																																																																																				
	FR221	NUTS3	n/a	n/a	30																																																																																																																				
	FR222	NUTS3	n/a	n/a	50																																																																																																																				
ESTIMATION TABLE	ID	Object type	Distrib t	Estim t-2	Estim t-1																																																																																																																				
	FR	NUTS0	1 (2)	160 (4)	180 (4)																																																																																																																				
	FR111	NUTS3	0,05 (1)	8 (3)	9 (3)																																																																																																																				
	FR112	NUTS3	0,1 (1)	16 (3)	18 (3)																																																																																																																				
	FR121	NUTS3	0,05 (1)	8 (3)	9 (3)																																																																																																																				
	FR122	NUTS3	0,3 (1)	48 (3)	54 (3)																																																																																																																				
	FR211	NUTS3	0,05 (1)	8 (3)	9 (3)																																																																																																																				
	FR212	NUTS3	0,05 (1)	8 (3)	9 (3)																																																																																																																				
	FR221	NUTS3	0,15 (1)	24 (3)	27 (3)																																																																																																																				
	FR222	NUTS3	0,25 (1)	40 (3)	45 (3)																																																																																																																				

Name	TE6 – Retropolation/extrapolation and space harmonization based on distribution method with available values at the same hierarchical level																																																																																																																																															
Conditions of use	<ul style="list-style-type: none"> ➔ Having at least one data available for a unit having the same parent and the same position in the hierarchy than the data to estimate at t-n and/or t+n. ➔ At lower level of the hierarchy, all information is available for two years (t and t+n or t and t-n) ➔ Data available for the parent unit both at t-n and/or t+n. 																																																																																																																																															
Graphics illustration																																																																																																																																																
Textual explanation	<p>Estimation based on the time and space dimensions (TE)</p> <p>1/ Time dimension - [linear/power] extrapolation. This method uses the two closest neighbours placed in time [before/after] ([mention the two years of reference]) the value estimated.</p> <p>2/ Space dimension: Estimate the stock to estimate for missing values at t-n and/or t+n, by subtracting the value available for the unit of belonging (parent(e)) by the sum of available values for values of the same hierarchy and having the same parents (X(e',s,t,i))</p> <p>3/ Space harmonization: In that order, the sum of the data of units of the same hierarchy is equal to the value of their parent unit (which is already available).</p>																																																																																																																																															
Mathematic formalization	$X(s, s, t, i) = a \times \frac{\sum a}{b}$ <p>Where a =</p> <ul style="list-style-type: none"> • Linear model $X(s, s, t, i) = \left(X(s, s, t-1, i) - \frac{X(s, s, t-1, i) - X(s, s, t-n, i)}{(t-1) - (t-n)} (t - (t-1)) \right)$ <ul style="list-style-type: none"> • Power model $X(s, s, t, i) = a \left(X(s, s, t-1, i) \left(\frac{X(s, s, t-1, i)}{X(s, s, t-n, i)} \right)^{\frac{(t-(t-1))}{((t-1)-(t-n))}} \right)$ <p>And where $b = X(\text{parent}(s), s, t-N, i) - \sum_{e'} X(e', s, t-N, i)$</p>																																																																																																																																															
Example	<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>ID</th> <th>Object type</th> <th>Value 2000</th> <th>Value 2001</th> <th>Value 2002</th> <th>Estim (2) 2000</th> <th>Adjust (6) 2000</th> <th>Vérif (7)</th> </tr> </thead> <tbody> <tr> <td>DEE0</td> <td>NUTS2</td> <td>2 649</td> <td>2 615</td> <td>2 581</td> <td></td> <td></td> <td></td> </tr> <tr> <td>DEE01</td> <td>NUTS3</td> <td>n/a</td> <td>100</td> <td>98</td> <td>102,5</td> <td>Stock (4) to estimate</td> <td>1915</td> </tr> <tr> <td>DEE02</td> <td>NUTS3</td> <td>254 (1)</td> <td>248</td> <td>243</td> <td></td> <td>1915</td> <td></td> </tr> <tr> <td>DEE03</td> <td>NUTS3</td> <td>235 (1)</td> <td>233</td> <td>230</td> <td></td> <td></td> <td></td> </tr> <tr> <td>DEE04</td> <td>NUTS3</td> <td>101 (1)</td> <td>101</td> <td>100</td> <td></td> <td></td> <td></td> </tr> <tr> <td>DEE05</td> <td>NUTS3</td> <td>n/a</td> <td>207</td> <td>204</td> <td>210,3</td> <td></td> <td>210,3</td> </tr> <tr> <td>DEE06</td> <td>NUTS3</td> <td>n/a</td> <td>108</td> <td>107</td> <td>109,3</td> <td>Stock (3) estimated</td> <td>109,1</td> </tr> <tr> <td>DEE07</td> <td>NUTS3</td> <td>n/a</td> <td>198</td> <td>197</td> <td>198,8</td> <td></td> <td>198,5</td> </tr> <tr> <td>DEE08</td> <td>NUTS3</td> <td>n/a</td> <td>222</td> <td>219</td> <td>225,6</td> <td>1918</td> <td>225,3</td> </tr> <tr> <td>DEE09</td> <td>NUTS3</td> <td>n/a</td> <td>261</td> <td>258</td> <td>263,1</td> <td></td> <td>262,7</td> </tr> <tr> <td>DEE0A</td> <td>NUTS3</td> <td>n/a</td> <td>176</td> <td>174</td> <td>179,1</td> <td></td> <td>178,8</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>Deviation (5)</td> <td></td> </tr> <tr> <td>DEE0B</td> <td>NUTS3</td> <td>n/a</td> <td>217</td> <td>216</td> <td>219,1</td> <td></td> <td>218,7</td> </tr> <tr> <td>DEE0C</td> <td>NUTS3</td> <td>n/a</td> <td>245</td> <td>241</td> <td>248,4</td> <td>0,998</td> <td>248,0</td> </tr> <tr> <td>DEE0D</td> <td>NUTS3</td> <td>142 (1)</td> <td>141</td> <td>139</td> <td></td> <td></td> <td></td> </tr> <tr> <td>DEE0E</td> <td>NUTS3</td> <td>n/a</td> <td>158</td> <td>156</td> <td>161,2</td> <td></td> <td>161,0</td> </tr> </tbody> </table>								ID	Object type	Value 2000	Value 2001	Value 2002	Estim (2) 2000	Adjust (6) 2000	Vérif (7)	DEE0	NUTS2	2 649	2 615	2 581				DEE01	NUTS3	n/a	100	98	102,5	Stock (4) to estimate	1915	DEE02	NUTS3	254 (1)	248	243		1915		DEE03	NUTS3	235 (1)	233	230				DEE04	NUTS3	101 (1)	101	100				DEE05	NUTS3	n/a	207	204	210,3		210,3	DEE06	NUTS3	n/a	108	107	109,3	Stock (3) estimated	109,1	DEE07	NUTS3	n/a	198	197	198,8		198,5	DEE08	NUTS3	n/a	222	219	225,6	1918	225,3	DEE09	NUTS3	n/a	261	258	263,1		262,7	DEE0A	NUTS3	n/a	176	174	179,1		178,8							Deviation (5)		DEE0B	NUTS3	n/a	217	216	219,1		218,7	DEE0C	NUTS3	n/a	245	241	248,4	0,998	248,0	DEE0D	NUTS3	142 (1)	141	139				DEE0E	NUTS3	n/a	158	156	161,2		161,0
ID	Object type	Value 2000	Value 2001	Value 2002	Estim (2) 2000	Adjust (6) 2000	Vérif (7)																																																																																																																																									
DEE0	NUTS2	2 649	2 615	2 581																																																																																																																																												
DEE01	NUTS3	n/a	100	98	102,5	Stock (4) to estimate	1915																																																																																																																																									
DEE02	NUTS3	254 (1)	248	243		1915																																																																																																																																										
DEE03	NUTS3	235 (1)	233	230																																																																																																																																												
DEE04	NUTS3	101 (1)	101	100																																																																																																																																												
DEE05	NUTS3	n/a	207	204	210,3		210,3																																																																																																																																									
DEE06	NUTS3	n/a	108	107	109,3	Stock (3) estimated	109,1																																																																																																																																									
DEE07	NUTS3	n/a	198	197	198,8		198,5																																																																																																																																									
DEE08	NUTS3	n/a	222	219	225,6	1918	225,3																																																																																																																																									
DEE09	NUTS3	n/a	261	258	263,1		262,7																																																																																																																																									
DEE0A	NUTS3	n/a	176	174	179,1		178,8																																																																																																																																									
						Deviation (5)																																																																																																																																										
DEE0B	NUTS3	n/a	217	216	219,1		218,7																																																																																																																																									
DEE0C	NUTS3	n/a	245	241	248,4	0,998	248,0																																																																																																																																									
DEE0D	NUTS3	142 (1)	141	139																																																																																																																																												
DEE0E	NUTS3	n/a	158	156	161,2		161,0																																																																																																																																									

HOW TO PROCEED ?

This example has been developed with available information at $t+1$ and $t+n$. But the methodology would be the same for data available at $t-1$ and $t-n$

- (1) Identify the units of the same hierarchy and having the same parent for which data is not available at $t-1$.
- (2) **For missing values only**, calculation of the 2000's value thanks to available information at lower level in 2001 and 2002 ($t+1$ and $t+2$) Cf estimation method T1, p.14, here the power model has been used).
- (3) Sum of (2)
- (4) Estimate the quantity to estimate (subtraction of the immediate upper level of belonging by available values of the same hierarchical level and having the same parents at t). *Then*
- (5) Divide (4) by (3)
- (6) Multiply (2) by (5).
- (7) Check if the sum of values available and estimated at lower level at $t-1$ are the same than the value available at the immediate upper level of belonging.

Name	TE7 - Interpolation/retopulation and space harmonization based on distribution method with available values at the same hierarchical level																																																																									
Conditions of use	<ul style="list-style-type: none"> ➔ Having at least one data available for a unit having the same parent and the same position in the hierarchy than the data to estimate at t-n and/or t+n. ➔ At lower level of the hierarchy, information is available for a single year (t) ➔ Data available for the parent unit both at t-n and/or t+n. 																																																																									
Graphics illustration																																																																										
Textual explanation	<p>Estimation based on the space and time dimensions (TE)</p> <p>1/ Space/time dimension: At t [<i>mention the year of reference</i>], calculation of the share that the count data of X(e,s,t,i) represents as regard to units of the same hierarchy and having the same parents (X(e',s,t,i)), who have also missing data at t-n and/or t+n.</p> <p>2/ Space dimension: Estimate the stock to estimate for missing values at t-n and/or t+n, by subtracting the value available for the unit of belonging (parent(e)) by the sum of available values for values of the same hierarchy and having the same parents (X(e',s,t,i))</p> <p>3/ Space harmonization: Multiplication of 1/ and 2/. In that order, the sum of the data of units of the same hierarchy is equal to the value of their parent unit.</p>																																																																									
Mathematic formalization	$X(e, s, t - N, i) = a \left(\frac{X(e, s, t, i)}{\sum_{e'} X(e', s, t, i)} \right)$ <p>where $a = X(\text{parent}(e), s, t - N, i) - \sum_{e'} X(e', s, t - N, i)$</p>																																																																									
Example	<table border="1" style="display: inline-table; margin-right: 20px;"> <thead> <tr> <th></th> <th>ID</th> <th>Object type</th> <th>Value T-1</th> <th>Value T</th> </tr> </thead> <tbody> <tr> <td rowspan="6" style="writing-mode: vertical-rl; transform: rotate(180deg);">ORIGINAL TABLE</td> <td>PL1</td> <td>NUTS1</td> <td>80</td> <td>100</td> </tr> <tr> <td>PL11</td> <td>NUTS2</td> <td>n/a (1)</td> <td>20 (2)</td> </tr> <tr> <td>PL12</td> <td>NUTS2</td> <td>n/a (1)</td> <td>15 (2)</td> </tr> <tr> <td>PL13</td> <td>NUTS2</td> <td>n/a (1)</td> <td>10 (2)</td> </tr> <tr> <td>PL14</td> <td>NUTS2</td> <td>20</td> <td>30</td> </tr> <tr> <td>PL15</td> <td>NUTS2</td> <td>20</td> <td>25</td> </tr> </tbody> </table> <table border="1" style="display: inline-table;"> <thead> <tr> <th></th> <th>ID</th> <th>Object type</th> <th>Distrib T (missing)</th> <th>Stock to estimate (T-1)</th> <th>Estim T-1</th> </tr> </thead> <tbody> <tr> <td rowspan="6" style="writing-mode: vertical-rl; transform: rotate(180deg);">ESTIMATION TABLE</td> <td>PL1</td> <td>NUTS1</td> <td>/</td> <td>80 (4)</td> <td>80</td> </tr> <tr> <td>PL11</td> <td>NUTS2</td> <td>0,45 (2)</td> <td>/</td> <td>18 (5)</td> </tr> <tr> <td>PL12</td> <td>NUTS2</td> <td>0,33 (2)</td> <td>/</td> <td>13,2 (5)</td> </tr> <tr> <td>PL13</td> <td>NUTS2</td> <td>0,22 (2)</td> <td>/</td> <td>8,8 (5)</td> </tr> <tr> <td>PL14</td> <td>NUTS2</td> <td>/</td> <td>20 (4)</td> <td>20</td> </tr> <tr> <td>PL15</td> <td>NUTS2</td> <td>/</td> <td>20 (4)</td> <td>20</td> </tr> <tr> <td></td> <td>SUM</td> <td></td> <td>1 (3)</td> <td>40 (4)</td> <td>80 (6)</td> </tr> </tbody> </table> <p>HOW TO PROCEED ?</p> <ol style="list-style-type: none"> (8) Identify the units of the same hierarchy and having the same parent for which data is not available at t-1. (9) For these units, estimate the share they represent individually at t as regard to the other missing value of the same hierarchy and having the same parents. (10) Check if the sum of the distribution at lower level is equal to 1. (11) Estimate the quantity to estimate (subtraction of the immediate upper level of belonging by available values of the same hierarchical level and having the same parents at t). (12) Multiply (2) and (4). (13) Check if the sum of values available and estimated at lower level at t-1 are the same than the value available at the immediate upper level of belonging. 		ID	Object type	Value T-1	Value T	ORIGINAL TABLE	PL1	NUTS1	80	100	PL11	NUTS2	n/a (1)	20 (2)	PL12	NUTS2	n/a (1)	15 (2)	PL13	NUTS2	n/a (1)	10 (2)	PL14	NUTS2	20	30	PL15	NUTS2	20	25		ID	Object type	Distrib T (missing)	Stock to estimate (T-1)	Estim T-1	ESTIMATION TABLE	PL1	NUTS1	/	80 (4)	80	PL11	NUTS2	0,45 (2)	/	18 (5)	PL12	NUTS2	0,33 (2)	/	13,2 (5)	PL13	NUTS2	0,22 (2)	/	8,8 (5)	PL14	NUTS2	/	20 (4)	20	PL15	NUTS2	/	20 (4)	20		SUM		1 (3)	40 (4)	80 (6)
	ID	Object type	Value T-1	Value T																																																																						
ORIGINAL TABLE	PL1	NUTS1	80	100																																																																						
	PL11	NUTS2	n/a (1)	20 (2)																																																																						
	PL12	NUTS2	n/a (1)	15 (2)																																																																						
	PL13	NUTS2	n/a (1)	10 (2)																																																																						
	PL14	NUTS2	20	30																																																																						
	PL15	NUTS2	20	25																																																																						
	ID	Object type	Distrib T (missing)	Stock to estimate (T-1)	Estim T-1																																																																					
ESTIMATION TABLE	PL1	NUTS1	/	80 (4)	80																																																																					
	PL11	NUTS2	0,45 (2)	/	18 (5)																																																																					
	PL12	NUTS2	0,33 (2)	/	13,2 (5)																																																																					
	PL13	NUTS2	0,22 (2)	/	8,8 (5)																																																																					
	PL14	NUTS2	/	20 (4)	20																																																																					
	PL15	NUTS2	/	20 (4)	20																																																																					
	SUM		1 (3)	40 (4)	80 (6)																																																																					

Annexe 3 – R programmes for calculating potential of population in different car-time radius

```
# =====
# Program M4D_TIMEDIST_001.R
#
# Extraction of part of the distance
# matrix and agregation at NUTS2 level
#=====
#-----
# (A) DATA PREPARATION
#-----

setwd("X:/M4D/CORE DB/DIST")
list.files()

# (A.1) Read file
library(foreign)
tab<-read.dbf("nuts3matrix.dbf", as.is = FALSE)

# (A.2) Create NUTS2 code

tab$N2CODEA<-substr(tab$N3CODEA,1,4)
tab$N2CODEB<-substr(tab$N3CODEB,1,4)
dim(tab)
head(tab)

#-----
# (B) AGGREGATE CAR_TIME BY MIN
#-----

# (B.1) Select Road time < 10h (600 mn)
tabN3N3<-
data.frame(tab$N3CODEA,tab$N3CODEB,tab$N2CODEA,tab$N2CODEB,tab$CAR_TIME)
names(tabN3N3)<-c("NUTS3_i","NUTS3_j","NUTS2_i","NUTS2_j","TIME_ij")
tabN3N3<-tabN3N3[tabN3N3$TIME_ij<=600,]

# (B.2) NUTS3_NUTS3 Distance < 10h
head(tabN3N3)
dim(tabN3N3)
write.table(tabN3N3,"CAR_TIME_N3N3_inf600.txt",quote=FALSE,row.names=FALSE,sep="\t",
dec=",")

# (B.3) NUTS3_NUTS2 min(Distance) < 10h

tabN3N2<-aggregate(tabN3N3$TIME_ij
,by=list(as.character(tabN3N3$NUTS3_i),as.character(tabN3N3$NUTS2_j)), FUN=min)
names(tabN3N2)<-c("NUTS3_i","NUTS2_j","min_TIME_ij")
head(tabN3N2)
```

```

dim(tabN3N2)
write.table(tabN3N2,"CAR_TIME_N3N2_min_inf600.txt",quote=FALSE,row.names=FALSE,sep
="\t", dec=",")

# (B.4) NUTS2_NUTS2 min(Distance) < 10h

tabN2N2<-aggregate(tabN3N3$TIME_ij
,by=list(as.character(tabN3N3$NUTS2_i),as.character(tabN3N3$NUTS2_j)), FUN=min)
names(tabN2N2)<-c("NUTS3_i","NUTS2_j","min_TIME_ij")
head(tabN2N2)
dim(tabN2N2)
write.table(tabN2N2,"CAR_TIME_N2N2_min_inf600.txt",quote=FALSE,row.names=FALSE,sep
="\t", dec=",")

#-----
# (C) AGGREGATE LORRY_TIME BY MIN
#-----

# (C.1) Select Lorry time < 10h (600 mn)
tabN3N3<-
data.frame(tab$N3CODEA,tab$N3CODEB,tab$N2CODEA,tab$N2CODEB,tab$LORRY_TIME)
names(tabN3N3)<-c("NUTS3_i","NUTS3_j","NUTS2_i","NUTS2_j","TIME_ij")
tabN3N3<-tabN3N3[tabN3N3$TIME_ij<=600,]

# (C.2) NUTS3_NUTS3 Distance < 10h
head(tabN3N3)
dim(tabN3N3)
write.table(tabN3N3,"LORRY_TIME_N3N3_inf600.txt",quote=FALSE,row.names=FALSE,sep="
\t", dec=",")

# (C.3) NUTS3_NUTS2 min(Distance) < 10h

tabN3N2<-aggregate(tabN3N3$TIME_ij
,by=list(as.character(tabN3N3$NUTS3_i),as.character(tabN3N3$NUTS2_j)), FUN=min)
names(tabN3N2)<-c("NUTS3_i","NUTS2_j","min_TIME_ij")
head(tabN3N2)
dim(tabN3N2)
write.table(tabN3N2,"LORRY_TIME_N3N2_min_inf600.txt",quote=FALSE,row.names=FALSE,s
ep="\t", dec=",")

# (C.4) NUTS2_NUTS2 min(Distance) < 10h

tabN2N2<-aggregate(tabN3N3$TIME_ij
,by=list(as.character(tabN3N3$NUTS2_i),as.character(tabN3N3$NUTS2_j)), FUN=min)
names(tabN2N2)<-c("NUTS3_i","NUTS2_j","min_TIME_ij")
head(tabN2N2)
dim(tabN2N2)
write.table(tabN2N2,"LORRY_TIME_N2N2_min_inf600.txt",quote=FALSE,row.names=FALSE,s
ep="\t", dec=",")

```

```

# =====
# Program M4D_POTENTIAL_RATIO_001.R
#
# Compute potential between NUTS units
# on the basis of time distance
# for a ratio  $Z=V/P$ 
# =====

#-----
# (A) DATA PREPARATION
#-----

# (A.1) Define workspace

setwd("X:/M4D/CORE DB/TESTPOT")
list.files()

# (A.2)Read Distance file

time<-read.table("CAR_TIME_N3N3_inf600.txt", header=TRUE, sep="\t", dec=",")
dist<-time[,c(1,2,5)]
names(dist)<-c("codei", "codej", "Dij")
# head(dist)

# (A.3)Read Stock file

pop<-read.table("tabpop3.txt", header=TRUE, sep="\t", dec=",")
# head(pop)
stock<-pop[,c(1,22,12)]
names(stock)<-c("code", "V", "P")
stock$Z<-stock$V/stock$P
# head(stock)

# (A.4) Merge Distance and Stock

tab<-merge(dist,stock, by.x="codei",by.y="code", all.x=FALSE, all.y=FALSE)
# head(tab)

#-----
# (B) POTENTIAL COMPUTATION
#-----

# (B.1) Choose Interaction function (Exponential)

# Exponent of distance (2=Gaussian)
beta<-2

# Threshold of interaction for  $f(D)=0.5$ 

```

```

scale<-120
alpha<--log(1/beta)/(scale**beta)

# Definition of territorial borders
# (example : national border)
tab$Aij<-as.numeric(substr(tab$codei,1,2)!=substr(tab$codej,1,2))

# Definition of Barrier effect
# Open border = 1.0 <=====> Closed border = 0.0
gamma<- 1

# (B.2) Compute probability of interaction

tab$int<-round(exp(-alpha*(tab$Dij**beta))*(gamma**tab$Aij),8)
# head(tab)

# (B.3) Compute potential contributions

matstock<-as.matrix(tab[,4:5])
dim(matstock)
matint<-tab[,8]
dim(matint)
mat<-matint*matstock
dim(mat)
row.names(mat)<-tab[,2]
tab2<-as.data.frame(mat)
# head(tab2)

# (B.4) Agregate potential by units

tabpot<-aggregate(tab2,list(tab[,2]),sum)
names(tabpot)<-c("code", "Pot_V", "Pot_P")
tabpot$Pot_Z<-tabpot$Pot_V/tabpot$Pot_P
# head(tabpot,10)

#-----
# (C) EXPORT RESULTS
#-----
dim(stock)
# head(stock)
dim(tabpot)
# head(tabpot)
tabfin<-merge(stock,tabpot, by="code", all.x=TRUE)
dim(tabfin)

```