TECHNICAL
REPORT

ESP●N

Disaggregation of socioeconomic data into a regular grid and combination with other types of data

CONTENT

- **Introduction and methodology description.** This section describes the background details and methodological solution adopted with regard to the disaggregation of data.

- **Integration of socio-economic and environmental information**. This part presents the OLAP Cube solution implemented in order to combine data from several data sources and formats, mainly socioeconomic statistics together with continuous data, such as land cover.

- **Future steps**. Taking into account the results from the challenge along the ESPON DB projects this section drafts the following steps that could be developed as a continuation.

OLAP CUBE

ESPON 2013 DATABASE
MARCH 2011

37 PAGES (incl. annexes)

LIST OF AUTHORS

Roger Milego, Universitat Autònoma de Barcelona, UAB 08193 Bellaterra

Maria José Ramos, Universitat Autònoma de Barcelona , UAB 08193 Bellaterra

Contact

roger.milego@uab.cat
mariajose.ramos@uab.cat

tel. + 34 93 581 35 46

TABLE OF CONTENT

1   Introduction

The ESPON 2013 DB project has been structured in several challenges in order to fulfil its objectives. The challenge to which this technical report refers to is the challenge number 5: "Combining socio-economic data measured for administrative zoning (NUTS level) and environmental data defined on a regular grid (like Corine Land cover)". The UAB (Universitat Autònoma de Barcelona) is the responsible partner with regard to this challenge.

Most of the socioeconomic variables or indicators are typically given by administrative unit, i.e. NUTS regions, whereas the environmental data is usually not following those boundaries, but given by natural units or regular grid cells.

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

The ESPON 2006 program developed some indicators in which the environmental data was transposed to NUTS division by means of GIS tools, in order to make them comparable to socioeconomic data. The results from this integration strategy, not always convincing, make clear the necessity of implementing a new integration process based on grid methods as it is said in the tender of the Espon 2013 Database project and in the Modifiable Areal Unit Problem study.

ESPON 2013 Database challenges

According to the E.S.T.I (space, Source, Time, Indicator) framework presented in the "Handbook for data collection" (ESPON 3.2, Final Report, Annex) four main objectives were identified in the tender of the ESPON 2013 DB Programme.

This challenge is included in the second key-question**: "Combination of heterogeneous sources- balancing Eurostat data"**, that emphasizes the need of the integration between different types and sources of data. The harmonization of the database in a fixed spatial division (NUTS3) solution that was chosen by many ESPON 2006 projects presented some doubts and not always convincing results. It is in this scenario that **a new integration methodology based on the reverse operation** is mentioned, **a projection of socio-economic information into units elaborated for the monitoring of natural resources.**

Modifiable Areal Unit Problem (ESPON 3.4.3)

The MAUP study, in its chapter 4 "*Exploration of gridding methods",* highlights the integration of heterogeneous databases as one of the most promising application of gridding methods for ESPON.

Two potential fields of applications are distinguished for gridding methods:

- **Time harmonisation of changing territorial units.**

"The use of grid help to build an harmonised territorial framework where all changing territorial divisions are harmonised and can further be used for the analysis of time variation" MAUP study (ESPON 3.4.3)

- **Thematic harmonisation and combination of heterogeneous spatial sources.**

The ESPON 2006 integration strategy, called "Eurostat oriented" by MAUP study, based on transferring all the information that it is not delivered on the basis of administrative units (NUTS 2 or NUTS 3) toward administrative units, is questioned, and the use of a new strategy is proposed.

"Information of good quality (as CLC) is therefore transformed into information of bad quality when projected in spatial units which are not adapted" MAUP study (ESPON 3.4.3).

" "Eurostat oriented" strategy could be replaced by another strategy that could be called the "EEA oriented" where all data would be transformed into grid and integrated on this basis" MAUP study (ESPON 3.4.3).

Methodology Proposal

In the First Interim Report of the project (Feb 2009), a methodology proposal was made on the basis of existing applications made by other institutions, such as:

- *"A Downscaled Population Density Map of the EU from Commune Data and Land Cover Information"* by Javier Gallego, JRC-ISPRA**.**

   A combination of commune population data with Corine Land Cover to produce an EU-wide grid with 1 ha resolution of downscaled population density[1].

- **G-Econ Research project of the University of Yale to develop a geophysically based data set on economic activity.**

   Estimation of gross output at a 1-degree longitude by 1-degree latitude resolution at a global scale for virtually all terrestrial grid cells based on spatial rescaling settled on proportional allocation**[2]**.

- FARO-EU (Foresight Analysis of Rural areas Of Europe)

   The project is aimed to analyse Rural Development in Europe by analysing patterns and trends of a selection of territorial indicators specific for rural areas within a Spatial Regional Reference Framework[3]**.**

---

[1]

http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S14P3_JAVIER_GALLEGO_DOWNSCALED_POPULATION_DENSITY.pdf

[2] *"New Metrics for Environmental Economics: Gridded Economic Dats"* by William D. Nordhaus.
http://www.oecd.org/dataoecd/44/7/37117455.pdf

[3] www.faro-eu.org

- **"*Transforming Population Data for Interdisciplinary Usages: From census to grid*"** by Deborah Balk & Greg Yetman from Columbia University.

   Creation of the Gridded Population of the World (GPW) data base implementing **a proportional allocation** of population from administrative units to grid cells [4].

The objectives established in the tender of the ESPON 2013 DB, the MAUP study results and recommendations, the bibliography research on existing methodologies and our experience at the UAB, as European Topic Centre on Land Use and Spatial Information, supporting the EEA in monitoring the land use/land cover change in Europe and analyzing the environmental consequences; lead us to the conclusion that the best way to downscale socioeconomic data and make them comparable with other kind of data, is **using a regular grid structure**, in which each cell takes a figure of the indicator or variable.

It was also concluded that depending on the nature of each variable, a different integration method should be applied. In other words, the way of calculating the actual figure for each grid cell might differ between different types of data, according to their definition.

The European Reference Grid

The EEA recommends the use of EEA reference grids for projection ETRS89-LAEA 52N 10E. The recommendation is based on proposal at the 1st European Workshop on Reference Grids[5].

The 1st Workshop on European Reference Grids was organized by the Joint Research Centre of the European Commission following a request of the EEA and the request of the INSPIRE Implementing Strategies Working Group that recommended the adaptation of a Europe-wide reference grid to facilitate the management and analyses of spatial information. The interest of the creation of a common coordinate reference system and a common equal-area grid to represent EU and Pan-Europe was also expressed by the National Statistical Institutes.

Taking into account this recommendations and our UAB/ETC-LUSI experience under some EEA projects such as LEAC (Land and Ecosystem Accounting), we proposed to disaggregate socioeconomic data into the **1 km European Reference Grid[6]**, as it is the way in which valuable data for the ESPON projects, such as the Corine Land Cover changes, are stored as well.

Testing process

After making all these decisions, we have carried out several tests with different data using different integration methods, trying to achieve useful results, on the one hand,

---

[4] http://sedac.ciesin.columbia.edu/gpw-v2/GPWdocumentation.pdf
[5] http://eusoils.jrc.ec.europa.eu/projects/alpsis/Docs/ref_grid_sh_proc_draft.pdf
[6] http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=760

but aiming at preparing the basis for the automation of the processes and integration with environmental data in an OLAP (On-Line Analytical Processing) cube.

Whenever an ESPON project would like to make the comparison of any kind of socioeconomic data together with land cover or environmental data not measured by administrative unit, the ESPON 2013 DB project, through the challenge 5 outcomes, will be ready to facilitate this task and provide them with the expected results or methodological tools to be applied.

The goal of the results presented in this report is to highlight the high potential of the three methodologies proposed. At this point the layers distributed by 1km grid resulting from applying the different methods are not available; their future distribution will be based on the building of OLAP's cubes using the most updated data.

Objectives

This technical report is aimed at explaining the details of the methodology and tests undertaken regarding the challenge 5 of the ESPON 2013 DB and it has, in particular, the following objectives:

- Review and summarise the background of the challenge 5.
- Give a detailed description of the methodology to be applied in order to downscale socioeconomic data.
- Describe the different disaggregation methods.
- Explain the tests undertaken and the results that have been achieved so far.
- Make some conclusions about all the processes undertaken so far.
- Define the next steps to be carried out.

2 Methodology description

This section describes the details of the methodology proposed to face the challenge of combining data measured using different reporting units.

According to what it has been explained in the introduction, and after reviewing several studies and taking into account our experience at the UAB (ETC-LUSI) and the EEA, we propose to integrate socioeconomic data in the 1 km European Reference Grid, because, besides some other reasons, this unit is used to summarise land cover data and other types of environmental data processed at the EEA.
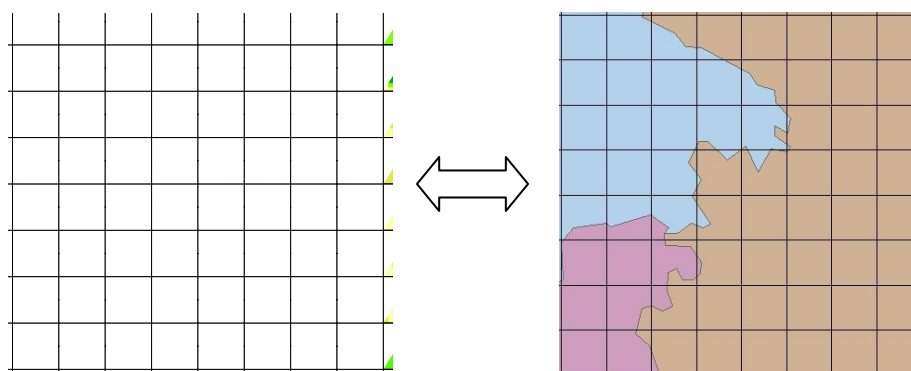
Figure 1. *The 1 km European Reference Grid will hold both environmental and socioeconomic information.*

Therefore, the first step to be carried out should be the **intersection** between the 1 km European Reference Grid and the administrative units by which the indicator is given. This is done by a physical overlay of both layers in vector format, by means of the ArcGIS tool Intersect. This tool creates a new layer holding both the geometries and the attributes of the source layers.
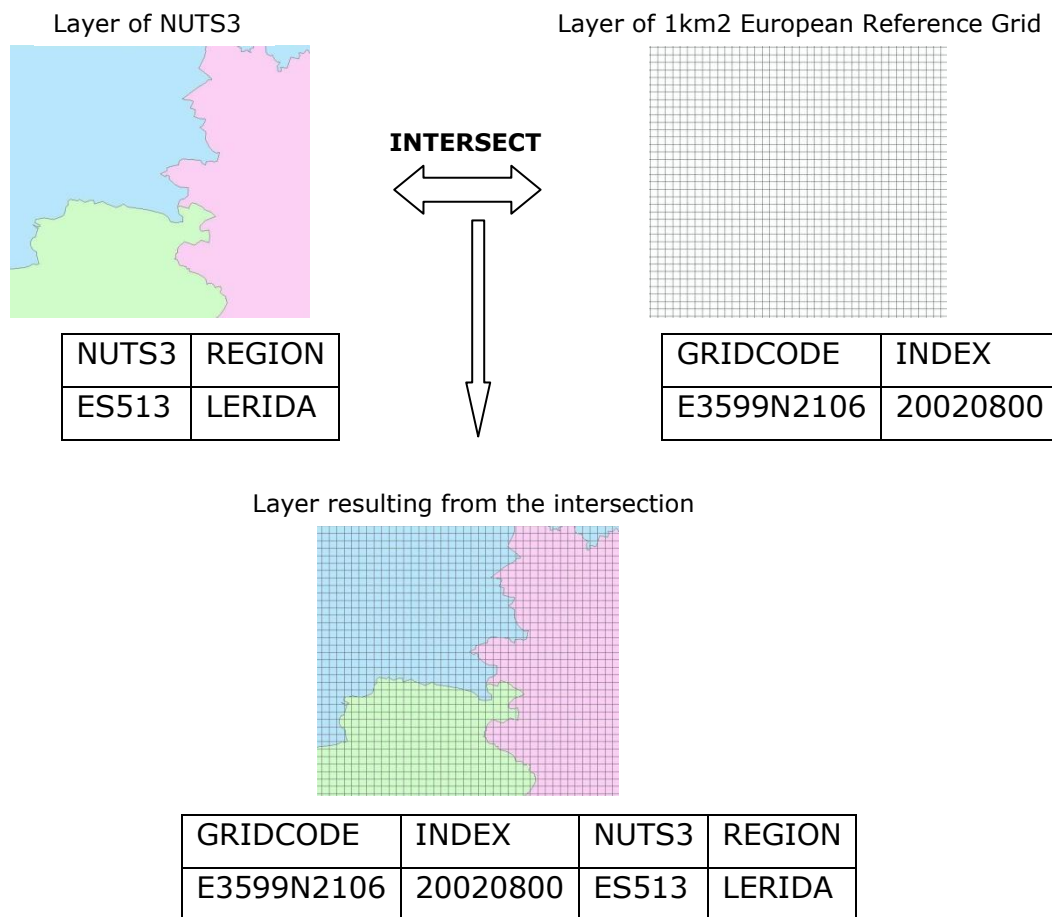
Figure 2. *Example of the intersection tool.*

Once the intersection has been computed, a way of computing a single figure by each reference grid cell should be defined. It has been stated that depending on the nature of each indicator or variable, a different kind of integration procedure should be defined. In this regard, we have defined and tested with different data the following three integration methods:

2.1 Integration methods

Maximum area criteria: the cell takes the value of the unit which covers most of the cell area. It should be a good option for uncountable variables.
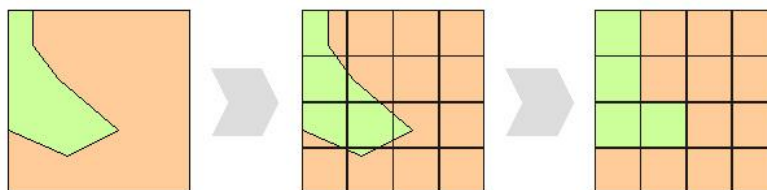
Proportional calculation: the cell takes a calculated value depending on the values of the units falling inside and their share within the cell. This method seems very appropriate for countable variables.
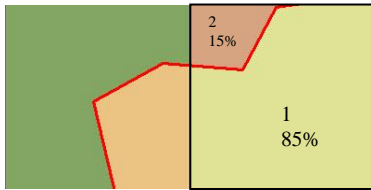
Cell value = Σ (V_i * $Share_i$)

Where: V_i = Value of unit i

$Share_i$ = Share of unit i within the cell

In the example: V_1 * 0.85 + V_2 * 0.15

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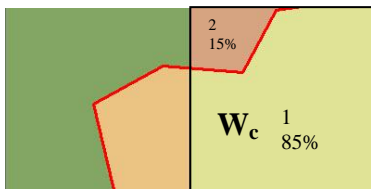
Cell value = W_c Σ (V_i * $Share_i$)

Where: V_i = Value of unit i

$Share_i$ = Share of unit i within the cell

W_c = weight assigned to cell c

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

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

Whenever it is possible, the third method has been used and should be used to disaggregate the data, as it gives an added value to the source data, providing more interesting results when different data are put together on a cell-by-cell basis.

Annex 1 shows the test results for the three integration methods.

3 Integration of socio-economic and environmental information

Once the variable has been distributed by 1 km cell, it can be compared to other data on a cell-by-cell basis, and it can be integrated into an **OLAP (Online Analytical Processing) cube**.

The OLAP technology[7] use a multidimensional data model, allowing complex analytical and ad-hoc queries with a rapid execution time. This solution facilitates the integrated analysis of several types of geographic and statistical data to users with or without GIS knowledge.

The next schema presents the general process to make possible the integration within an OLAP cube

Dimensions		Measures	
Nuts3 Code	Corine Class Level 1	Ha	Unemployment Total
AT111	Artificial surfaces	3795	697
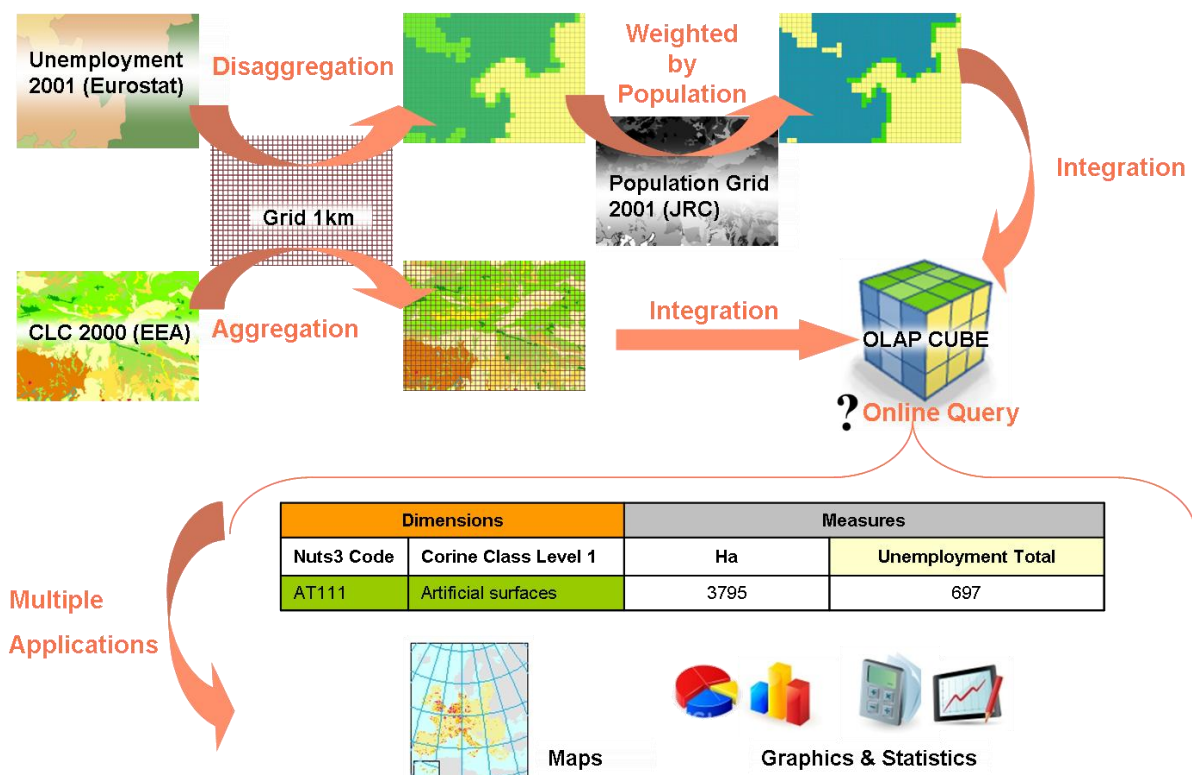
Figure 13. *Simplified schema of data integration in an OLAP Cube*

In the case of ESPON OLAP cube, it consists on the **ESPON socio-economic** variables as numerical attributes or measures that will be aggregated using a set of dimensions. The **dimensions** or topics of interest for the user are generally represented by different types: **spatial dimensions,** usually represented by

---

[7]Some OLAP information resources:

1- http://en.wikipedia.org/wiki/Online_analytical_processing

2- http://www.cs.sfu.ca/CC/459/han/papers/chaudhuri97.pdf

3- http://es.wikipedia.org/wiki/Cubo_OLAP

administrative units for Europe (NUTS), a number of **thematic dimensions**: land use data or dominant land cover types; and a third type, which is the **temporal dimension** that shows the difference between two years time.

In this way, we are able to ask to the ESPON-OLAP cube some questions taking into account the socioeconomic variables or indicators. For example, having the GDP and CLC changes in the ESPON OLAP cube, we could analyse which land cover flows occur by different GDP ranges, and, in the end, get the results on a NUTS3, NUTS2 or country (NUTS0) basis.

The OLAP Cube delivered together with the ESPON 2013 DB Final Report is the version 3.0, and holds the following dimensions and measures:

- Socioeconomic variables
  - GDP 2003
  - GDP 2006
  - Active population 2003
  - Active population 2006
  - Unemployment 2003
  - Unemployment 2006
- Land cover
  - Corine Land Cover 1990
  - Corine Land Cover 2000
  - Corine Land Cover 2006
- Land cover changes
  - Land Cover Flows 1990-2000
  - Land Cover Flows 1990-2006
  - Land Cover Flows 2000-2006
- Measures
  - Population density 2001 (inhab/km^2)
  - Area (ha)
- Geographical dimensions
  - Elevation Breakdown
  - Biogeographic Regions
  - Large Urban Zones and City Names
  - Massifs
  - NUTS 2006
  - NUTS 2003
  - River Basin Districts UE

The ESPON OLAP Cube can be accessed either online or offline. Online connection has not been implemented yet in the framework of the ESPON DB. So far, a CUB file has been provided in order to use the ESPON OLAP Cube offline. **Annex 2** explains how to use this file and make queries to the Cube by means of MS-Excel.

4   Future steps

There are several improvements and future implementations we have in mind to make this methodological approach and derived tools grow up in the future. We list below some of them:

a) Methodological improvements: Some aspects that have to be deeply analysed are:

   - Treatment of administrative units with no data values.
   - Differences between geographical extents, for example between Nuts 3 2006 layer and Corine Land Cover.
   - Improvement of the disaggregation performance in terms of time and manageability of the final layer.
   - Calculation of deviations with respect to the source data.

b) Extent increase: It is foreseen to increase the number of neighbours, countries and areas to process. The possible candidates will comprise Maghreb and part of Saharan Africa together with a portion of the CIS[8].

c) Follow-up of the EFGS: It might be very useful to keep following up the outcomes of the European Forum for Geostatistics[9] in order to contrast the proposed method, and, eventually, improve it.

d) Questionnaire: As the ESPON OLAP Cube is a standalone database we would like to explain to the ESPON projects the practical possibilities of the tool and ask them feedback about the data they would like to have combined with other data sources. For example, many projects might like to have some of their datasets combined with Corine Land Cover. It is important for us to identify such datasets in order to think how aggregate/disaggregate them into the grid and, eventually, integrate them within the ESPON OLAP Cube.

e) Integration with the ESPON 2013 Database: It should be deeply studied the most suitable solution to integrate grid data in the ESPON 2013 Database and to make the ESPON OLAP Cube available to users through the database interface. The main idea is to facilitate the comparison and analysis of socio-economic data with environmental data that usually not follows an administrative distribution. At this stage, the ESPON OLAP Cube is provided as a standalone file which is accessed from the Database by means of a link. In order to use it, a connection from MS Excel should be established as it is explained in the ESPON OLAP Cube User Manual (Annex 2 of this Technical Report).

f) Visualisation tool: At this stage, the ESPON OLAP Cube queries are made through MS Excel Pivot Tables application, which allows to get the results in table format. A suitable web visualisation tool could be adapted and integrated into the ESPON database framework allowing the visualisation of grid and socio-

---

[8] Commonwealth of Independent States
[9] http://www.efgs.ssb.no/

economic ESPON OLAP data and, eventually, the other ESPON database data. Two specific environments could be defined depending on the specific needs of the user: one more simple, designed for a basic user just for visualisation purposes, and a more advanced one, being able to show comprehensive results like interactive maps, graphs or tables.

g) A set of ESPON OLAP Cubes: The ESPON OLAP Cube, as a standalone database, is built by means of the integration of specific datasets within the 1 km2 European Reference Grid, which should be previously selected according to the analytical purposes. Therefore, if several purposes are identified, there should be several thematic Cubes to fit them. Moreover, some ESPON OLAP Cubes could also be built on the basis of a better resolution grid only for some specific region, such as urban areas to better analyse urban processes.

h) in order to integrate the gridded data and make them ready to be queried and analysed together with other types of data, like land cover, natural units, protected areas, and so on, for different scales (i.e. Urban Cube).

5   Conclusions

After having defined a methodology in order to be able to put together and analyse socioeconomic and environmental data, and having made several tests using different datasets, we can make now some conclusions about the main outcomes of the work done and the things we have learned so far:

- Disaggregating socioeconomic data by a regular grid is the best solution in order to downscale such information reported by administrative areas.
- The 1 km European Reference Grid is a good option to undertake the disaggregation because:
    - o It has an European coverage
    - o It follows Inspire specifications
    - o It is used for several institutions as the reference grid
    - o Its resolution is optimal in order not to lose data precision
- For uncountable data (non-numeric values), the best aggregation method is the "maximum area criterion".
- For countable data, the best method is the proportional one, which calculates the final value according to the area share of each of the values.
- Whenever it is possible, it is better to weight the final figures when using a proportional method, e.g. by population.
- The "proportional and weighted" aggregation method is the one that gives better results, plus some added value to the downscaling.
- The different methods are independent from the source data format and can be applied to vector and raster format.
- In order to achieve good results following this methodology it is important to use data sources which follow the same spatial and temporal specifications (extent, spatial resolution, temporal resolution…).
- This methodology allows the integration of socio-economic in an OLAP cube, which facilitates the comparison and analysis of such data together with land cover data, for example.
- The benefits of having the ESPON OLAP Cube built up has been proven, although the pre-processing and data selection steps take quite a long time and resources. Furthermore, once the Cube is ready, it is not possible for the users to add data dynamically.

To sum up, it can be added that any kind of socioeconomic data can be processed using the methodology proposed and tested, in order to have them downscaled and stored by 1 km grid, facilitating their comparison with many other data not reported by administrative units.

Our next steps will be aimed at improving the performance of the methodologies proposed but also to analyse the introduction of some changes based on our results and new projects at European context.

6 References

- ***Literature***

Arévalo J., *Land and Ecosystem Accounting. Technical Procedure, Internal Report v.2,* 2009, ETC-LUSI, European Environmental Agency.

Chaudhuri S., Dayal U., *An overview of Data Warehousing and OLAP Technology,* Simon Fraser University Canada (SFU.CA).

Deichmann U., Balk D., Yetman G., 2001, *Transforming Population Data for Interdisciplinary Usages: From census to grid,* NASA Socioeconomic Data and Applications Center (SEDAC).

Gallego J., *A Downscaled Population Density Map of the EU from Commune Data and Land Cover Information,* JRC-Ispra.

Gallego J., *Downscaling population density in the European Union with a land cover map and a point survey,* JRC-Ispra.

Gallego J., *Population density grid of EU-27+, version 4. Summary of the downscaling method,* JRC-Ispra.

Malinowski E., Zimányi E., 2009, *Advanced Data Warehouse Design- From Conventional to Spatial and Temporal Applications*, Springer.

Short Proceedings of the 1ˢᵗ European Workshop on Reference Grids, Ispra, 27-19 October 2003, JRC- Institute for Environmental and Sustainability, Ispra

William D. Nordhaus, 2006, *New Metrics for Environmental Economics: Gridded Economic Data,* Yale University

- ***Websites***

Espon programme: http://www.espon.eu/

- **Future Orientation for Cities (FOCI). ESPON programme**
http://www.espon.eu/mmp/online/website/content/programme/1455/2233/2236/2239/index_EN.html

- **The modifiable areas unit problem (MAUP). ESPON Scientific Support Project 3.4.3**
http://www.espon.eu/mmp/online/website/content/projects/261/431/index_EN.html

European Environment Agency (EEA). European Commission:
http://www.eea.europa.eu/

- **Population density disaggregated with Corine land cover 2000.**
http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-1

- **EEA reference grids**
http://www.eea.europa.eu/data-and-maps/data/eea-reference-grids

European forum for Geostatistics 2010. http://www.efgs.ssb.no/

Eurostat, European Commission:
http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/

Foresight Analysis of Rural areas of Europe (Faro-eu.org) http://www.faro-eu.org

Geographically based Economic data (G-Econ) project, University of Yale
http://gecon.yale.edu/

Joint Research Centre (JRC). European Comission:
http://ec.europa.eu/dgs/jrc/index.cfm

ANNEX 1. Testing the methodology

In order to test the methodology proposed and the different methods of data integration into the reference grid, we have chosen some socioeconomic variables or other data not being measured by grid cell, but stored using administrative units or other kind of irregular delimitation. Thus, we have chosen variables such as GDP, unemployment or urban dominance. The next sections show the steps undertaken and results achieved for any kind of test done.

Testing the maximum area criterion: Urban dominance (2000) based on the Urban Morphological Zones.

The maximum area criterion is suitable for uncountable data or data given by ranges. In particular, this approach can be applied to better differentiate the urban/non urban character of the land as has been highlighted by the team of **ESPON FOCI**[10] project. The advantage of taking percentage of Urban Morphological Zones over pure land cover classes is that UMZ indicates if certain area of urban fabric is part of a greater entity (equivalent to a city). Then, the urban dominance could be used to characterise certain areas and to relate them to other descriptors.

   a) Source data:

In order to make this test we have used the Urban Morphological Zones 2000 (Map1. Urban Morphological Zones 2000), an EEA's dataset which is a delimitation of urban areas according to a functional definition. The UMZ2000 come from a reclassification of different land cover classes of CLC2000 following different criteria and, therefore, they are not following administrative boundaries but an artificial boundary created by addition of land cover polygons.

For further details about UMZ, please visit:

http://dataservice.eea.europa.eu/download.asp?id=17335&filetype=.pdf

---

[10] http://www.espon.eu/mmp/online/website/content/programme/1455/2233/2236/2239/index_EN.html

Map 1 . Urban Morphological Zones 2000
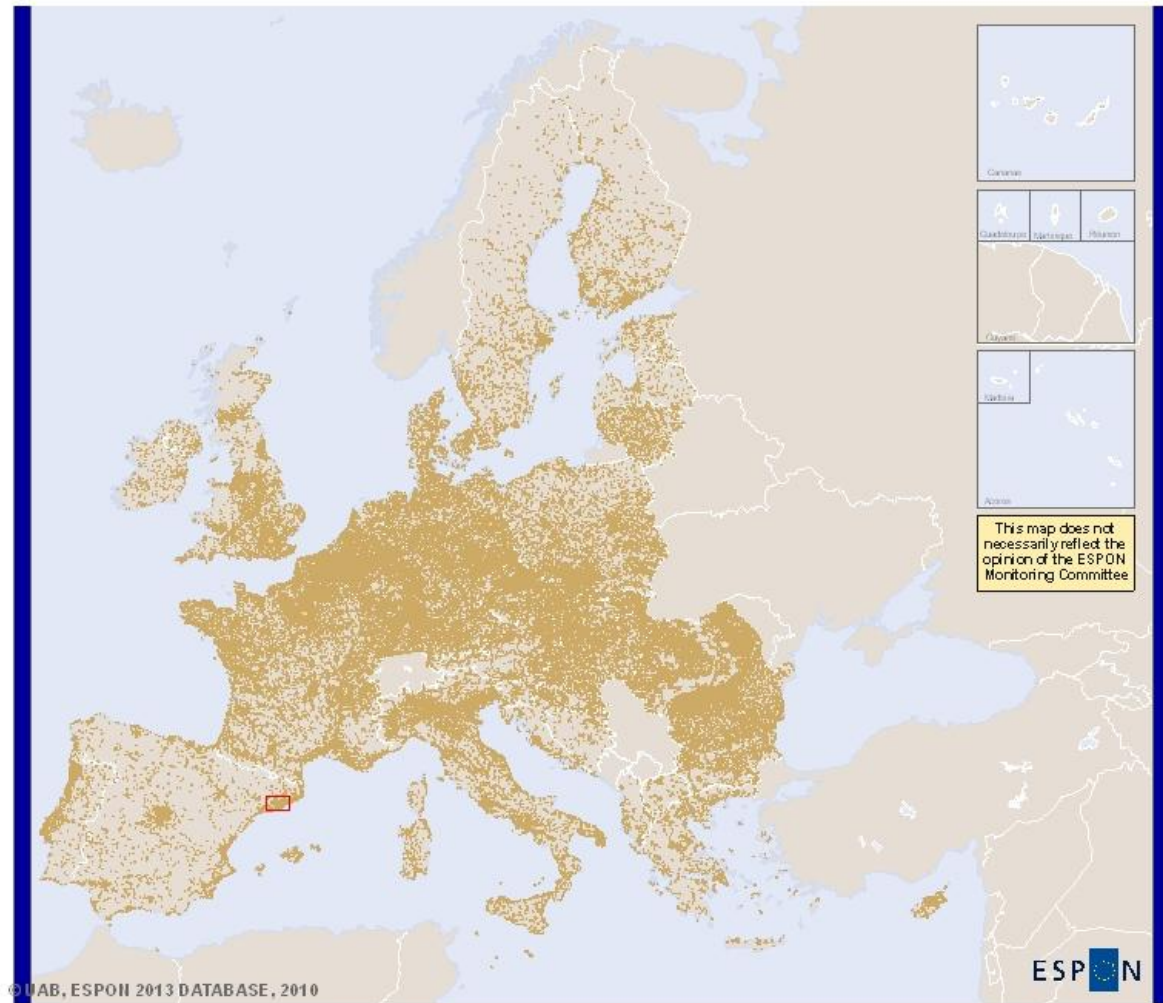
© UAB, ESPON 2013 DATABASE, 2010

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

ESP❋N

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

Regional level: UMZ
Source: Corine Land Cover, 2000
Origin of data: Corine Land Cover, 2000
© EuroGeographics Association for administrative boundaries

Urban Morphological Zones 2000

Urban Morphological Zones 2000

b) Process steps:

As it has been stated in the methodology section, the first step has been the intersection between the Reference Grid and the UMZ2000. In this way, we are able to calculate which share of grid cell is occupied by an UMZ.
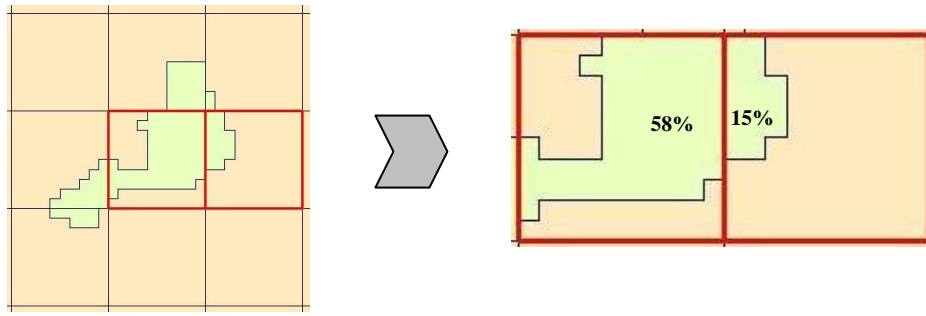
Figure 3. *Intersection of UMZ2000 and the Reference Grid*.

If the cell has more than its half (50%) covered by an UMZ polygon, we define it as an urban-dominant cell, whereas if it has less than 50% of UMZ inside, it is defined as a non-urban cell.
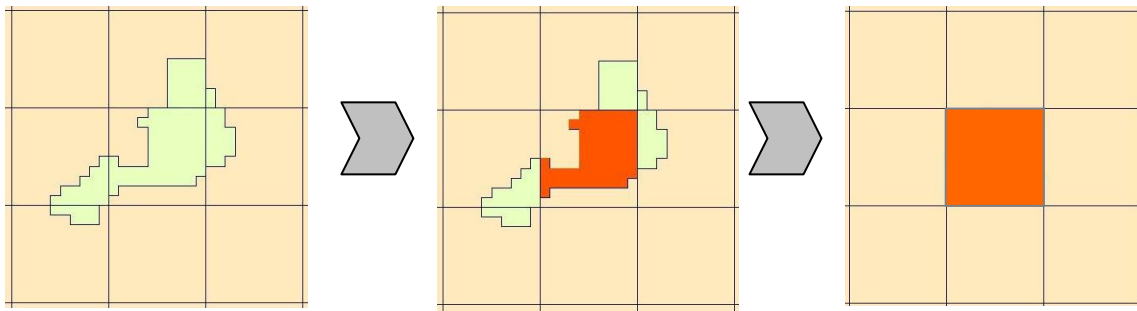
Figure 4. *Urban dominance definition process*.

Finally, this kind of map can be elaborated, where we can see the urban dominance in Europe by 1 km grid cell being able to identify quicker the main points of urban surfaces in Europe:

Map 2 . Urban Dominance layout
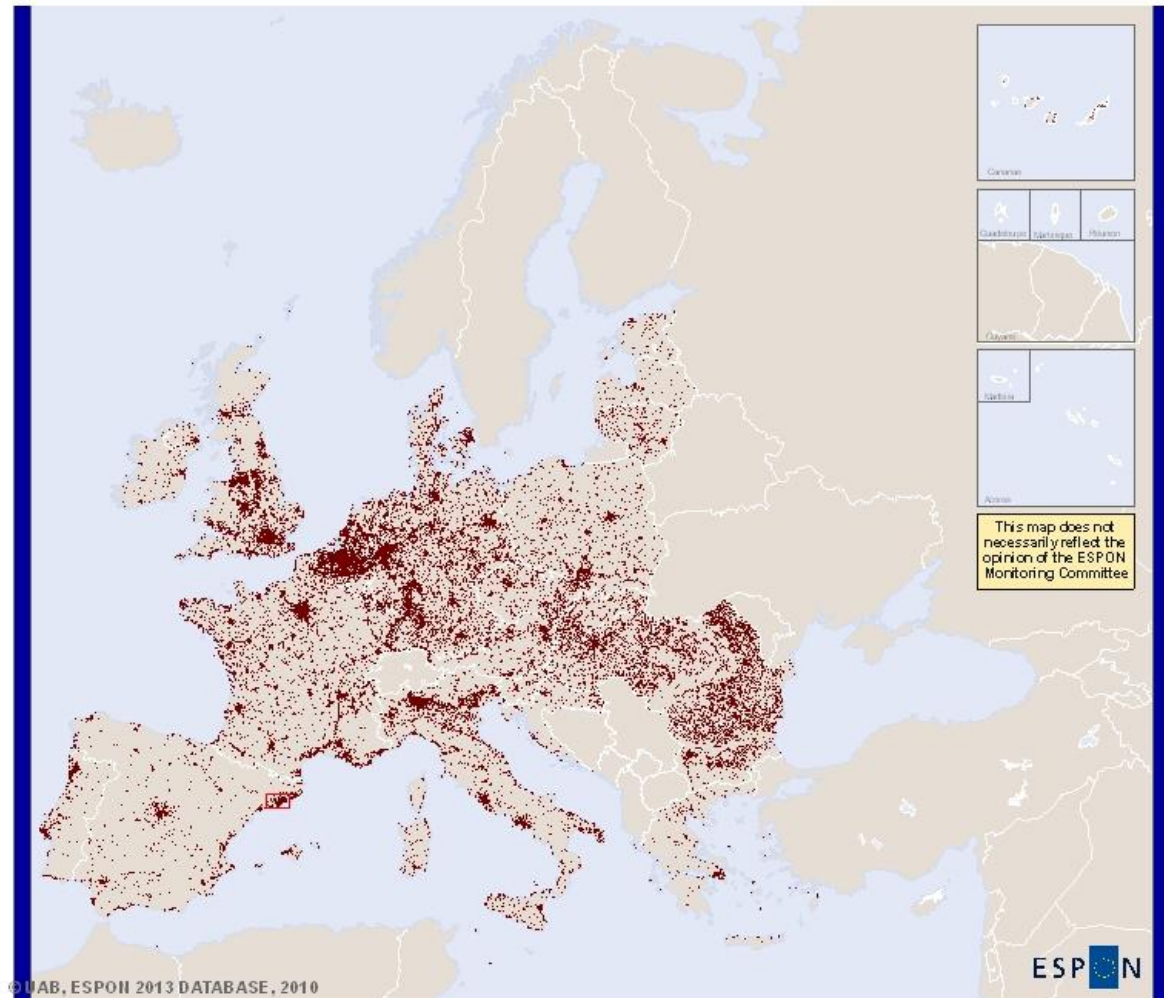
This map does not necessarily reflect the opinion of the ESPON Monitoring Committee

© UAB, ESPON 2013 DATABASE, 2010

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

Regional level: Grid
Source: UMZ, 2000
Origin of data: Corine Land Cover, 2000
© EuroGeographics Association for administrative boundaries

Urban Morphological Zones 2000

■ Urban Dominance
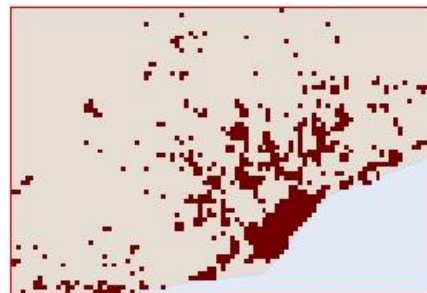
What is interesting is that now we are able to analyse land cover changes and flows by urban or non-urban cells, or any other indicator. Furthermore, we can create an urban classification by percentage ranges, e.g. 0-20%, 20-60%, 60-100% (low urban dominance, mid urban dominance, high urban dominance):

Map 3 . Urban Classificatiion by density ranges

© UAB, ESPON 2013 DATABASE, 2010

Regional level: Grid
Source: UMZ, 2000
Origin of data: Corine Land Cover, 2000
© EuroGeographics Association for administrative boundaries

Urban Classification

- 0- 20%
- 20-60%
- 60-100%
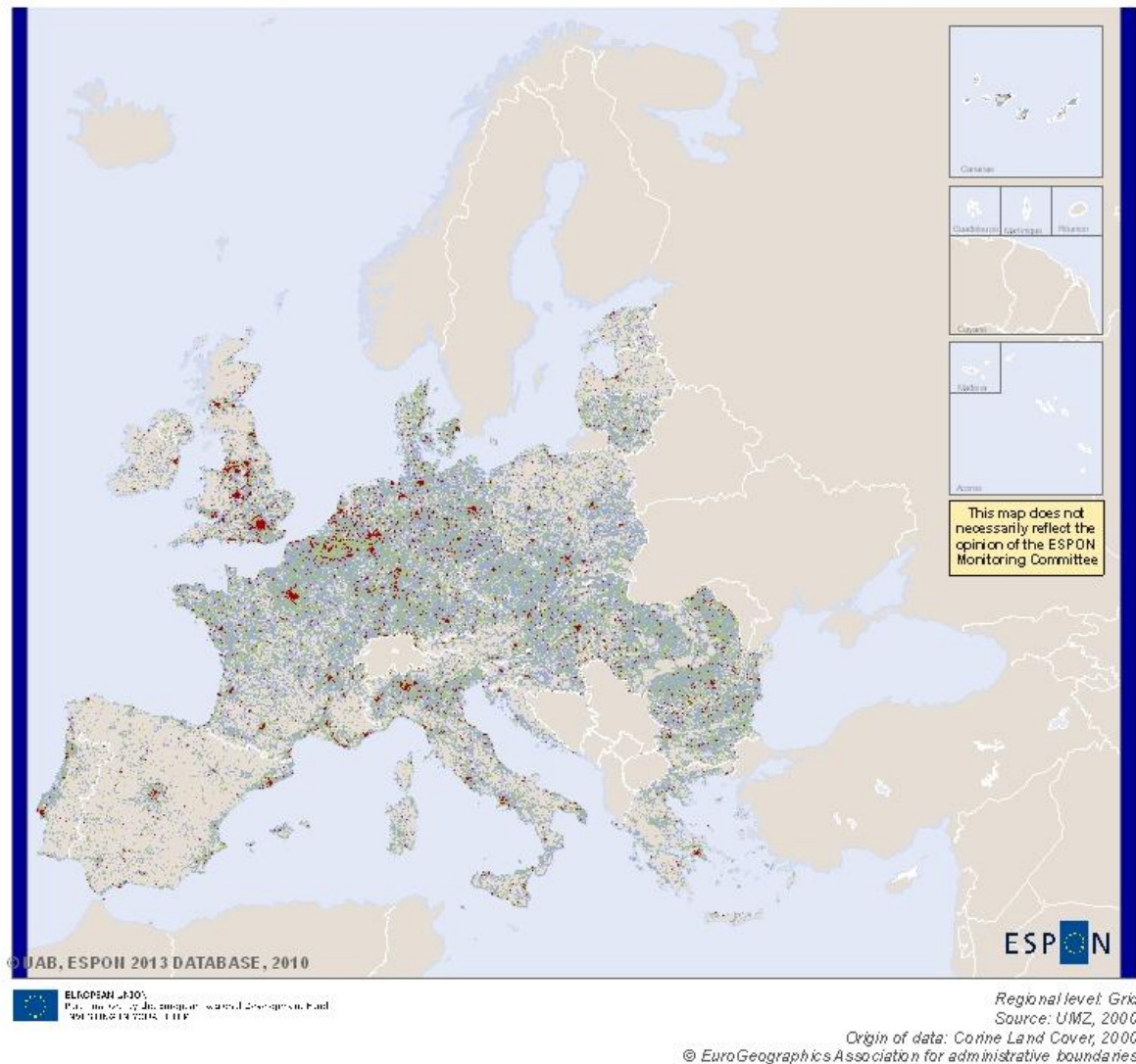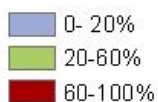
Conclusion: the maximum area criteria is useful for non-numeric values or in case we are interested in a discrete classification of grid cells, either by a thematic attribute or a value range. Moreover, the **ESPON FOCI** project found interesting this kind of approach applied to UMZ because it can be used as a criteria to define urban dominance and integrate it in the urban-hinterland analysis.

Testing the "Proportional calculation" method:

Unemployment rate total (2001)

a) Source data:

As for the proportional calculation method, we have chosen an indicator of unemployment: the unemployment rate total (2001) from Eurostat. It represents unemployed people as a percentage of the economically active population, and it is measured by NUTS3 region.

Although the most suitable methodology to be applied on this indicator is the third one, proportional and weighted calculation, we have selected the unemployment rate total as an example of the possible results that can be obtained by applying the second methodology.

	A	B	C	D	E	F	G	H	I	J
1	Employment and Labour Market									
2	NUTS level 3 (version 1999)									
3	Subtheme:		Unemployment	Unemployment	Unemployment	Unemployment	Unemployment	Unemployment	Unemployment	Unemployment
4	Indicator:		Unemployment rate total	Unemployment rate female	Unemployment rate male	Unemployment rate under 25 years	Development of unemployment rate 1998-2001	Development of unemployment rate, female, 1998-2001	Development of unemployment rate, male, 1998-2001	Development of unemployment rate, <25 years, 1998-2001
5	Description:		in %	in %	in %	in %	in percentage points	in percentage points	in percentage points	in percentage points
6	Time:		2001	2001	2001	2001	1998-2001	1998-2001	1998-2001	1998-2001
7	Source:		Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices
8	Comment:									
9	NUTS_3_99	Region	UNRT01N3	UNRF01N3	UNRM01N3	UNRU2501N3	UNRT98N3	UNRF98N3	UNRM98N3	UNRU2598N3
10	AT111	Mittelburgenland	3,1	4,8	2,2	3,4	-1,0	-1,9	-0,5	-0,1
11	AT112	Nordburgenland	2,3	2,8	2,0	4,0	-0,8	-1,2	-0,5	-0,6
12	AT113	Südburgenland	4,3	5,2	3,5	7,6	-0,9	-1,4	-0,5	0,0
13	AT121	Mostviertel-Eisenwurzen	2,1	2,7	1,5	3,5	-0,6	-0,8	-0,6	-0,5
14	AT122	Niederösterreich-Süd	3,4	3,8	3,0	5,2	-0,9	-1,2	-0,8	-0,1
15	AT123	Sankt Pölten	3,3	3,7	2,9	6,0	-0,8	-1,5	-0,4	0,7
16	AT124	Waldviertel	3,0	3,7	2,4	4,7	-0,9	-1,8	-0,3	-0,1
17	AT125	Weinviertel	2,9	3,5	2,4	4,1	-0,6	-1,1	-0,3	-0,5
18	AT126	Wiener Umland/Nordteil	2,3	2,7	2,1	3,6	-0,7	-1,1	-0,4	0,0
19	AT127	Wiener Umland/Südteil	2,9	3,3	2,8	4,3	-1,0	-1,2	-0,8	-0,2
20	AT13	Wien	4,9	4,9	4,9	7,2	-1,7	-1,8	-1,7	-1,9
21	AT211	Klagenfurt-Villach	4,1	5,2	3,2	8,1	-1,1	-1,3	-1,0	-1,2
22	AT212	Oberkärnten	5,6	8,5	3,6	9,4	-1,4	-2,0	-0,9	-2,4
23	AT213	Unterkärnten	3,7	5,1	2,6	5,7	-1,2	-1,7	-0,9	-2,0

Figure 5. *Unemployment source data.*

b) Process steps:

We start joining the figures of the indicator on unemployment with the layer of NUTS3 using the unique identifier of the NUTS3 regions, this operation allows to have a geometry representation of the information.

In order to have a single dataset holding the NUTS3 and the Reference Grid geometries, including the attribute information, we carried out an overlay process.
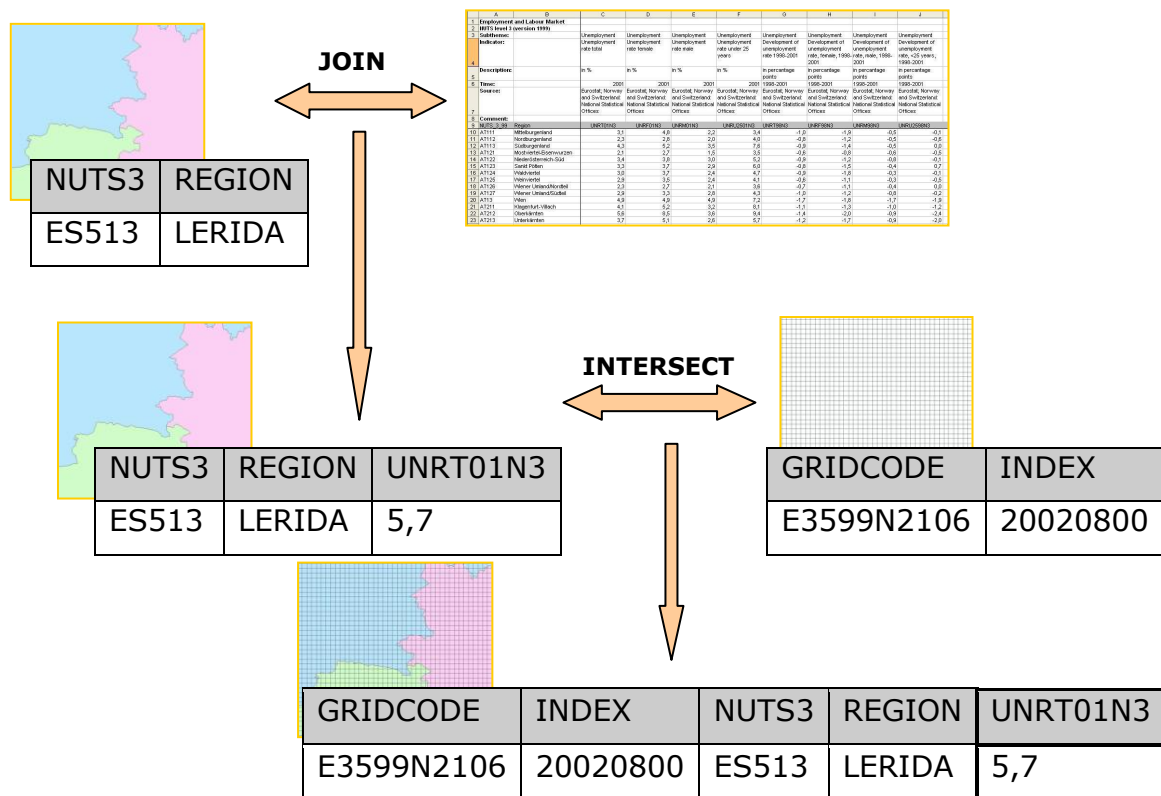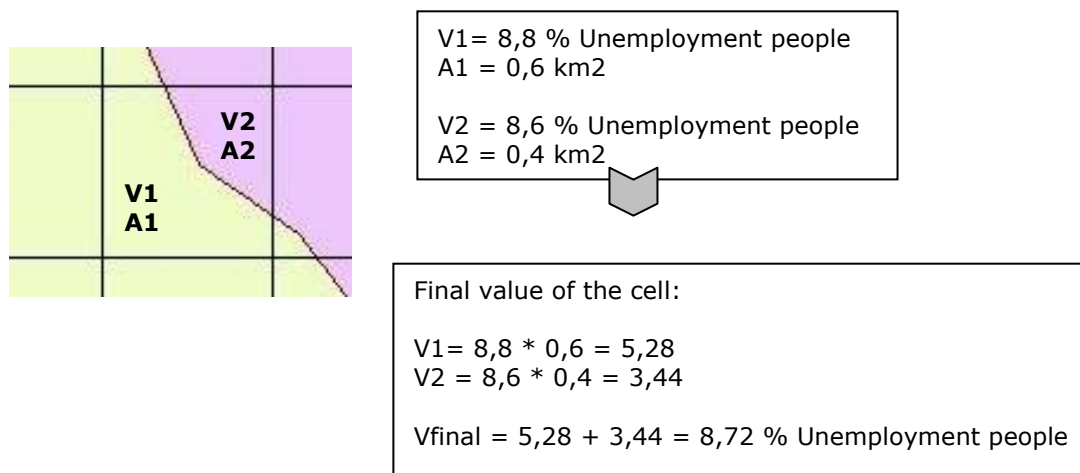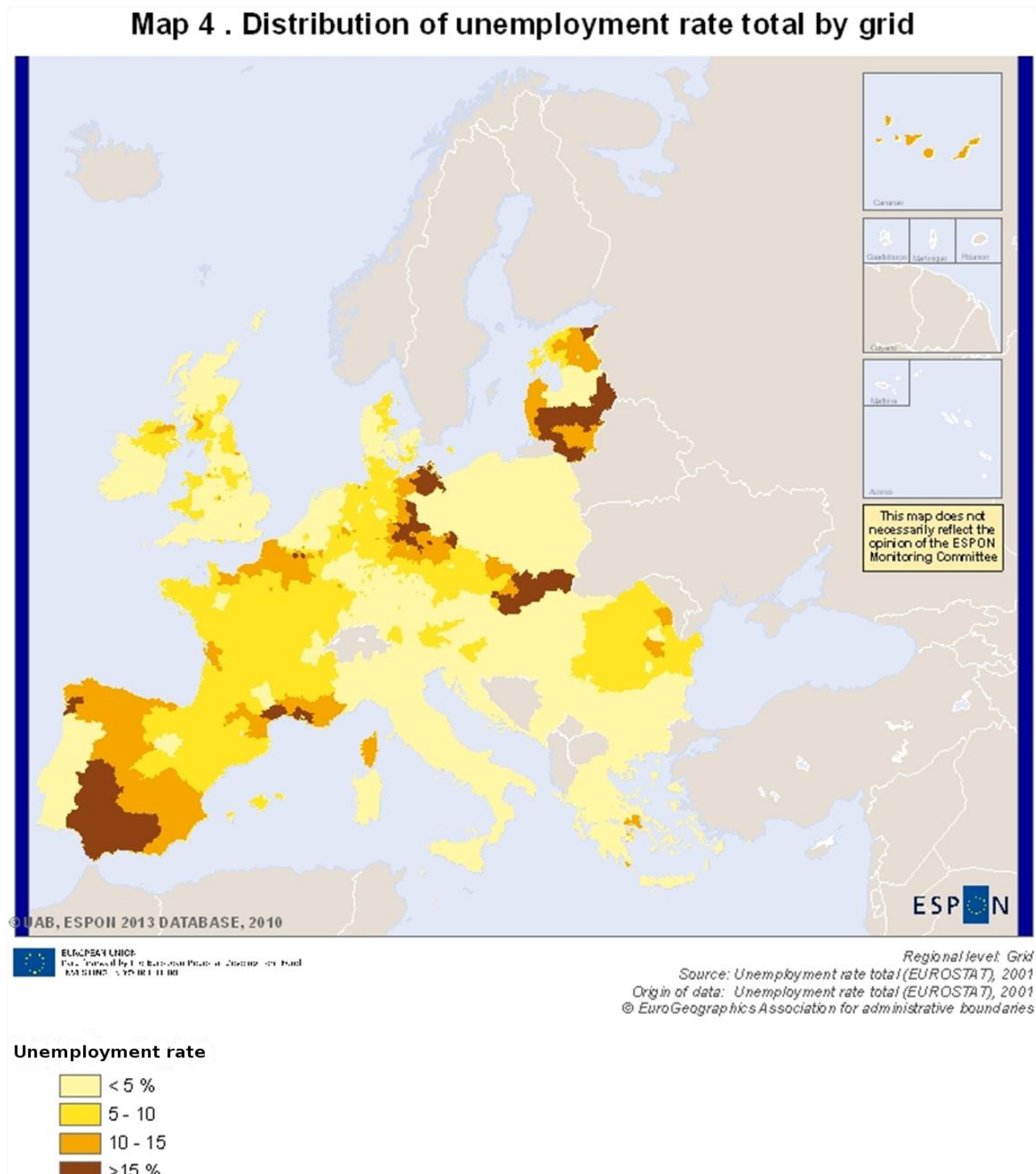
Figure 6. *Creation process schema for unemployment downscaling.*

After that, we should recalculate a single value for each grid cell, based on the NUTS3 value or values that fall inside. If a cell contains different NUTS3 values, the final value will be calculated depending on the share of surface of each NUTS3 region within the cell. For example, if a cell has 0.2 out of 1 covered by one NUTS3 region and 0.8 covered by a second NUTS3, the figures should be calculated accordingly (multiplying the first figure by 0.2 and the second one by 0.8).

V1= 8,8 % Unemployment people
A1 = 0,6 km2

V2 = 8,6 % Unemployment people
A2 = 0,4 km2

Final value of the cell:

V1= 8,8 * 0,6 = 5,28
V2 = 8,6 * 0,4 = 3,44

Vfinal = 5,28 + 3,44 = 8,72 % Unemployment people

c) Results:

The results put on a map have a look like this:

Map 4 . Distribution of unemployment rate total by grid

Unemployment rate

	< 5 %
	5 - 10
	10 - 15
	>15 %

Although it seems a map where NUTS3 have been coloured as such, the data are stored by 1 km grid cell and, therefore, they can be compared with other data stored under the same grid coding, such as land cover data.

Testing the "Proportional and weighted calculation" method: GDP – Wealth and Production (2002)

a) Source data: In order to test the third aggregation method, we have chosen an economic variable, the GDP in euro per inhabitant 2002 (Eurostat) (Map 5. GDP €/inhab. 2002), and decided to weight its values by the population living in each 1 km grid cell. In this way, the GDP value is downscaled in a more realistic manner. As for population, we have used the JRC's population density grid dataset[11] for the year 2001. In this grid, population data for communes is remapped based on Corine Land Cover classes and a quite complex algorithm[12].

Map 5 . GDP €/inhab. 2002 distributed by Nuts3 2003

Regional level: Nuts3 2003
Source: GDP in euros per inhabitant (EUROSTAT), 2002
Origin of data: GDP in euros per inhabitant (EUROSTAT), 2002
© EuroGeographics Association for administrative boundaries

Wealth and Production

GDP €/inhab. 2002

- < 10000
- 10000 - 20000
- 20000 - 35000
- > 35000

---
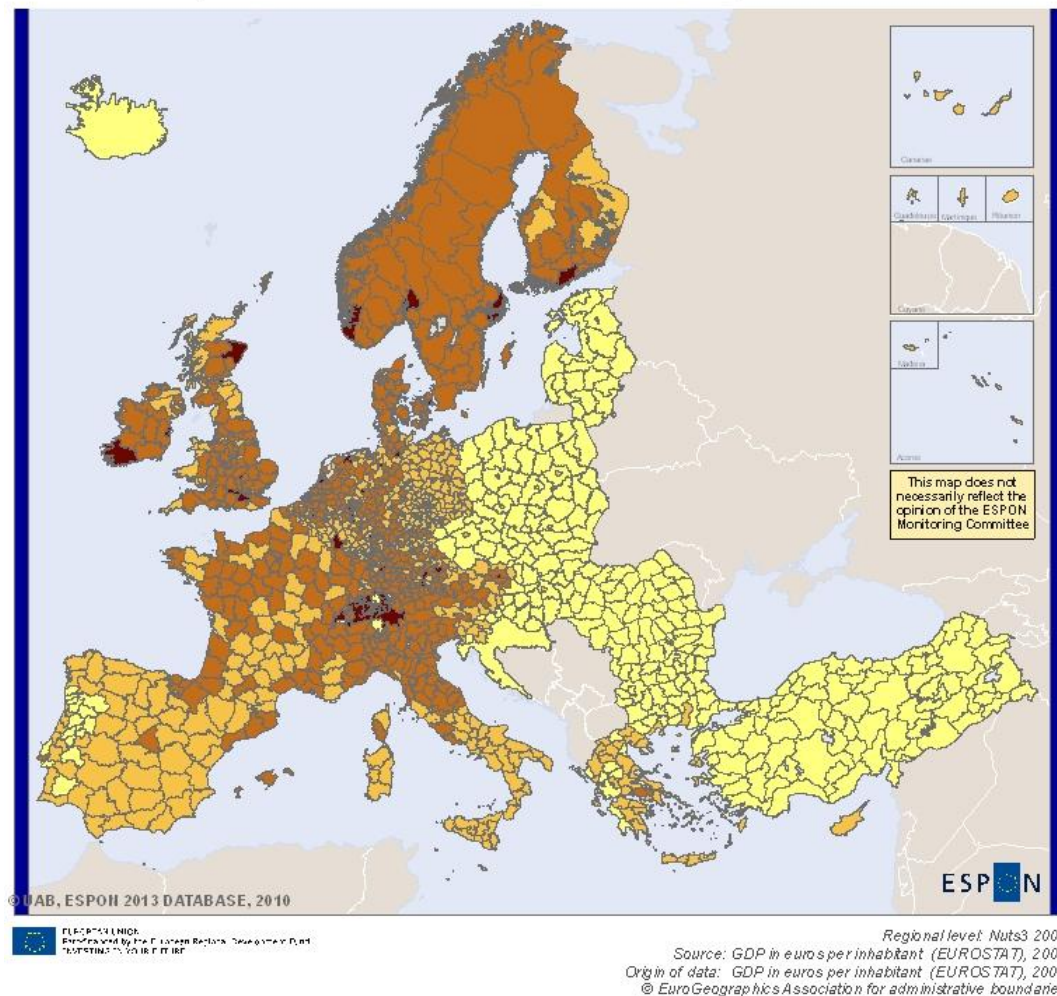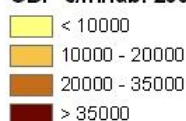
[11] http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP_RESEARCH/PGE_RESEARCH_NTTS/S14P3%20-%20JAVIER%20GALLEGO%20-%20%20DOWNSCALED%20POPULATION%20DENSITY.PDF
12 http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-1

b) Process steps: In this case, the GDP is measured by NUTS3 regions. Therefore, the first step, as in the previous two cases, is overlaying the layer in which the data is given with the 1 km Reference Grid. After that, a single figure should be calculated for each 1 km grid cell, depending on the values coming from the NUTS3 regions overlaying it. If more than one value is shared by a grid cell, the final figure is calculated proportionally with regard to the area that each value occupies within the cell.
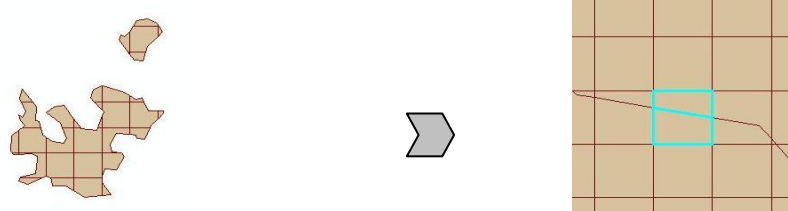
GRID CODE	AREA	GDP	GDP * AREA
GRIDCODE1	A1	GDP1	A1 * GDP1
GRIDCODE1	A2	GDP2	A2 * GDP2

GRID CODE	GDP
GRIDCODE1	A1 * GDP1 + A2*GDP2

Finally, the value of GDP per capita that has been calculated by each grid cell is multiplied by the population figure in that cell, giving a final GDP value which reflects not only the richness of the region but also the distribution of that richness amongst the inhabitants. The next figure presents the general schema followed to calculate the indicator.
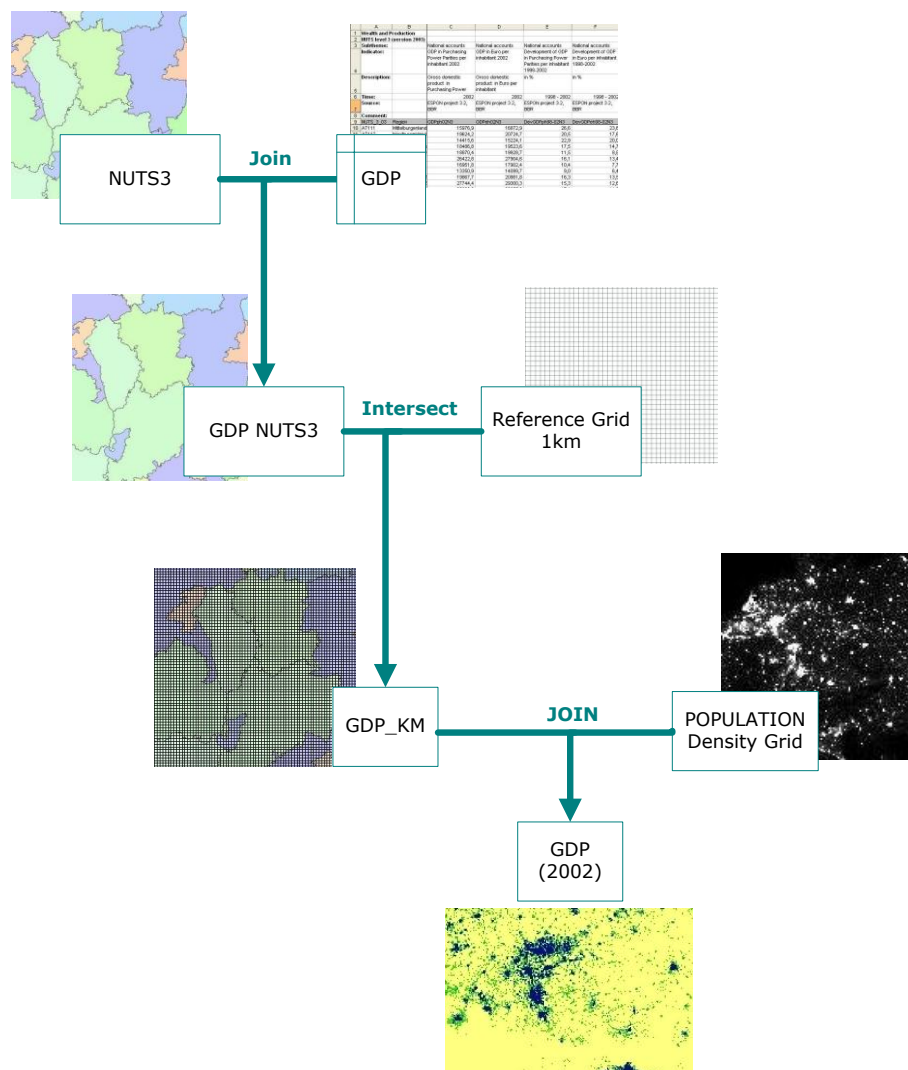
Figure 7. *Creation process schema for GDP downscaling.*

Exceptions: To obtain the final GDP (2002) weighted by the population it has been necessary to use the population density grid but also the population 2003 distributed by NUTS3 (Eurostat). The reason is that the population density grid doesn't cover all the extension of the layer NUTS3 , this mainly happened in islands zones like Canary islands. In these regions, to be able to calculate the GDP total, the weighted process has been made with the information of the population from Eurostat.

	A	B	C	D	E	F	G
1	Population						
2	NUTS level 3 (version 2003)						
3	Subtheme:		Population structure	Population structure	Population structure	Population structure	Population structure
4	Indicator:		Average Population 2003	Average male Population, share in %, 2003	Average female Population, share in %, 2003	Population density 2002	Development average population 1995-2003 in %
5	Description:						
6	Time:		2003	2003	2003	2002	2003
7	Source:		Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices
8	Comment:		UKM + UKN = 2002	UKM + UKN = 2002	UKM + UKN = 2002		UKM + UKN = 2002
9	NUTS_3_03	Region	AvgPopN303	AvgmPopN303	AvgfPopN303	PopdensN302	DavgPop9503N3
25	AT222	Liezen	81.800	48,4	51,5	24,9	1,1
26	AT223	Östliche Obersteiermark	173.700	48,4	51,6	57,4	-10,6
27	AT224	Oststeiermark	268.400	49,3	50,7	77,3	4,8
28	AT225	West- und Südsteiermark	190.700	49,1	50,9	84,5	3,0
29	AT226	Westliche Obersteiermark	108.200	49,0	50,9	36,2	-4,0
30	AT311	Innviertel	273.200	49,2	50,9	96,8	1,9
31	AT312	Linz-Wels	531.300	48,5	51,5	302,4	0,6
32	AT313	Mühlviertel	202.700	50,0	50,0	75,3	4,9
33	AT314	Steyr-Kirchdorf	152.600	49,0	51,0	69,3	1,1
34	AT315	Traunviertel	227.100	48,7	51,3	89,7	2,9
35	AT321	Lungau	21.300	49,3	50,7	21,5	-1,8

Figure 8. *Eurostat's 2003 population source data.*

The next figure highlights in blue and red the main zones not covered by the population density grid and where it has been necessary to use the population information provided by Eurostat.
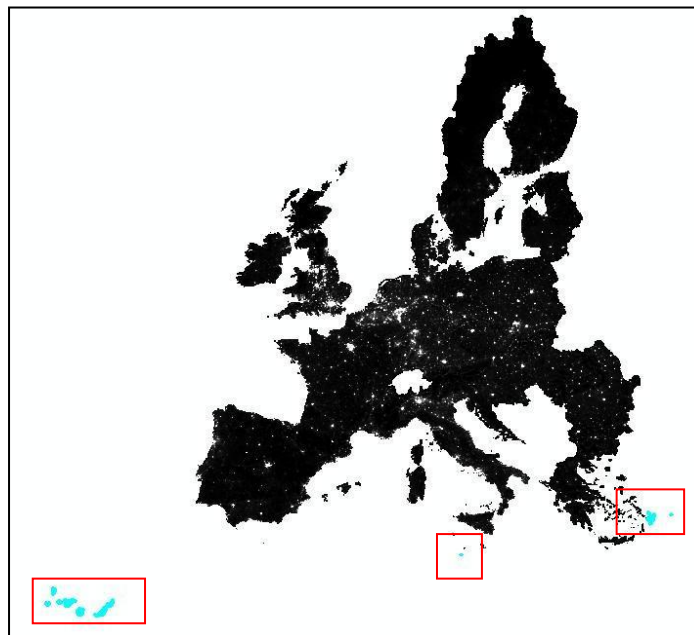
Figure 9. *Location of areas out of population grid's scope.*

Results: when we put the results on a map, we have the following layouts:

Map 6 . Distribution of GDP in Euro 2002 by grid
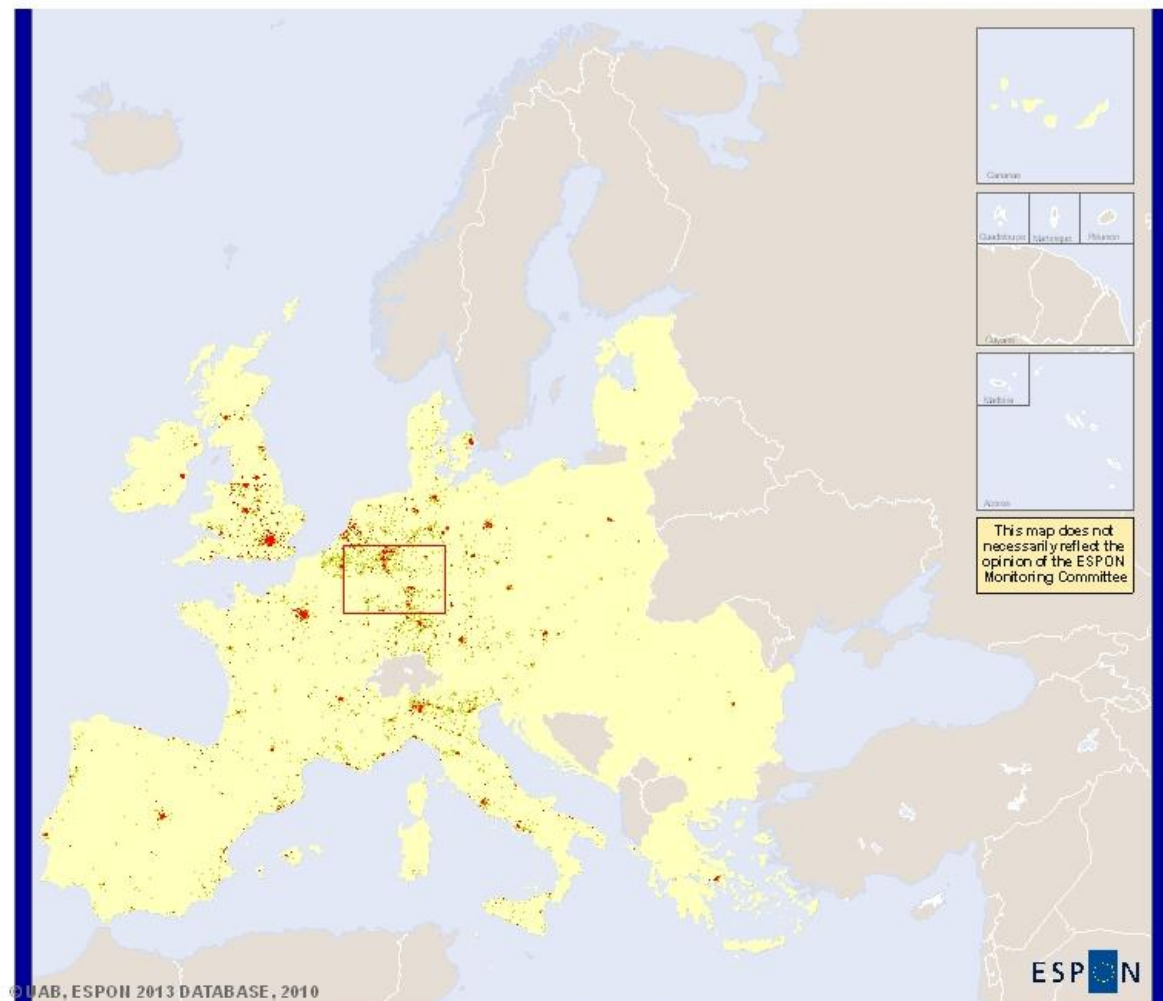
This map does not necessarily reflect the opinion of the ESPON Monitoring Committee

ESP⬤N

© UAB, ESPON 2013 DATABASE, 2010

Wealth and Production

GDP in Euro 2002

☐	< 10 million euros
◻	10 - 30
◻	30- 50
◻	> 50 million euros
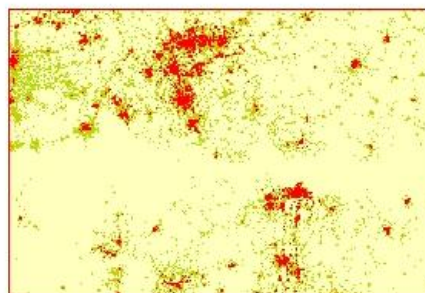
As it is obvious according to the method, the GDP is concentrated in the biggest urban areas, where most of the people are living and somehow higher in the grid cells belonging to the richest regions in Europe. Consequently, this method of redistributing and weighting data by grid cells is useful to be somehow independent of the administrative (arbitrary) divisions. This case is highlighted for example in the south-west of Ireland and in the north of Italy.

a) South-west of Ireland:
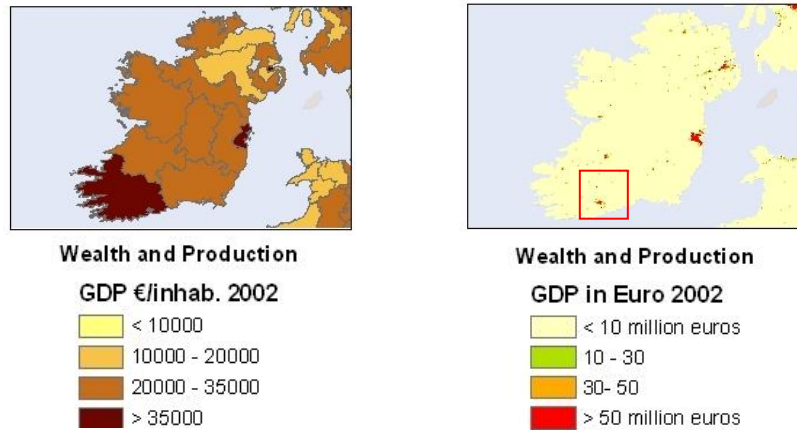
Figure 10. *Original GDP €/inhab distributed by Nuts3 2003 vs. GDP in Euro distributed by grid at the South-west of Ireland.*

In this case the Nuts3 region (IE025) is very big, but the richness is concentrated mainly around the Cork city (a small dot at the mapped scale, highlighted with a red square).
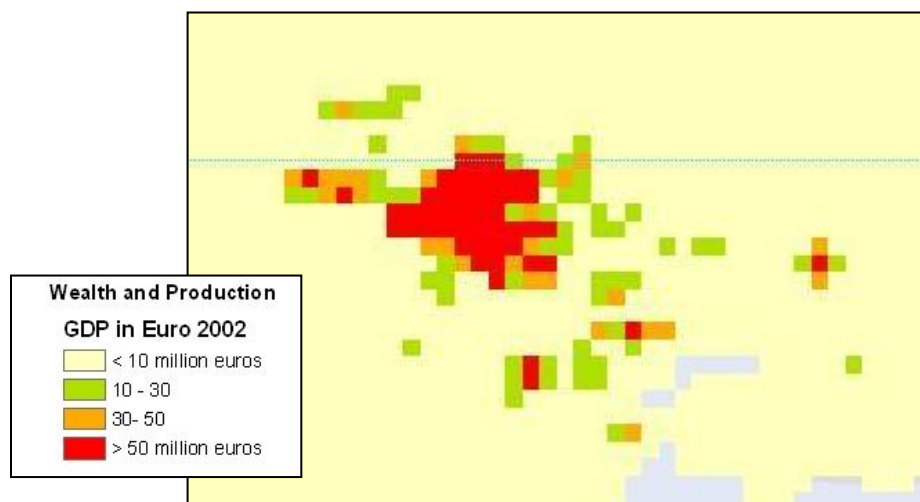
Figure 11. *Zoom in on the Cork City at the south-west of Ireland*

b) North of Italy:

Wealth and Production
GDP €/inhab. 2002
- < 10000
- 10000 - 20000
- 20000 - 35000
- > 35000

Wealth and Production
GDP in Euro 2002
- < 10 million euros
- 10 - 30
- 30- 50
- > 50 million euros
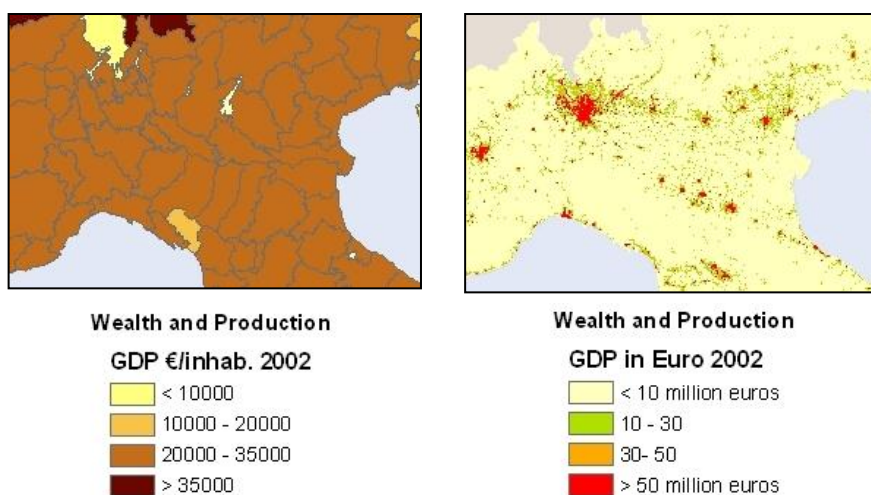
Figure 12. *Original GDP €/inhab distributed by Nuts3 2003 vs. GDP in Euro distributed by grid at the North of Italy.*

In the case of the north of Italy, the regions are much smaller and have quite a high GDP/inhab. When theses values are weighted by population and distributed by grid cells they better show the concentration of richness in the big cities, like Milano in this case.

ANNEX 2. ESPON OLAP Cube User Manual

This Annex is aimed at explaining how to query the ESPON OLAP Cube file by means of the default PivotTable application included in MS Excel.

An OLAP Cube can be queried **online** and **offline**. So far, the online connection has not been implemented. In order to use the cube, a single file **.CUB**, which works offline, has been provided.

The .CUB file can be connected to and queried from Microsoft Excel with a few steps detailed in the following pages:

- Select "Pivot Table…" in the Data Menu
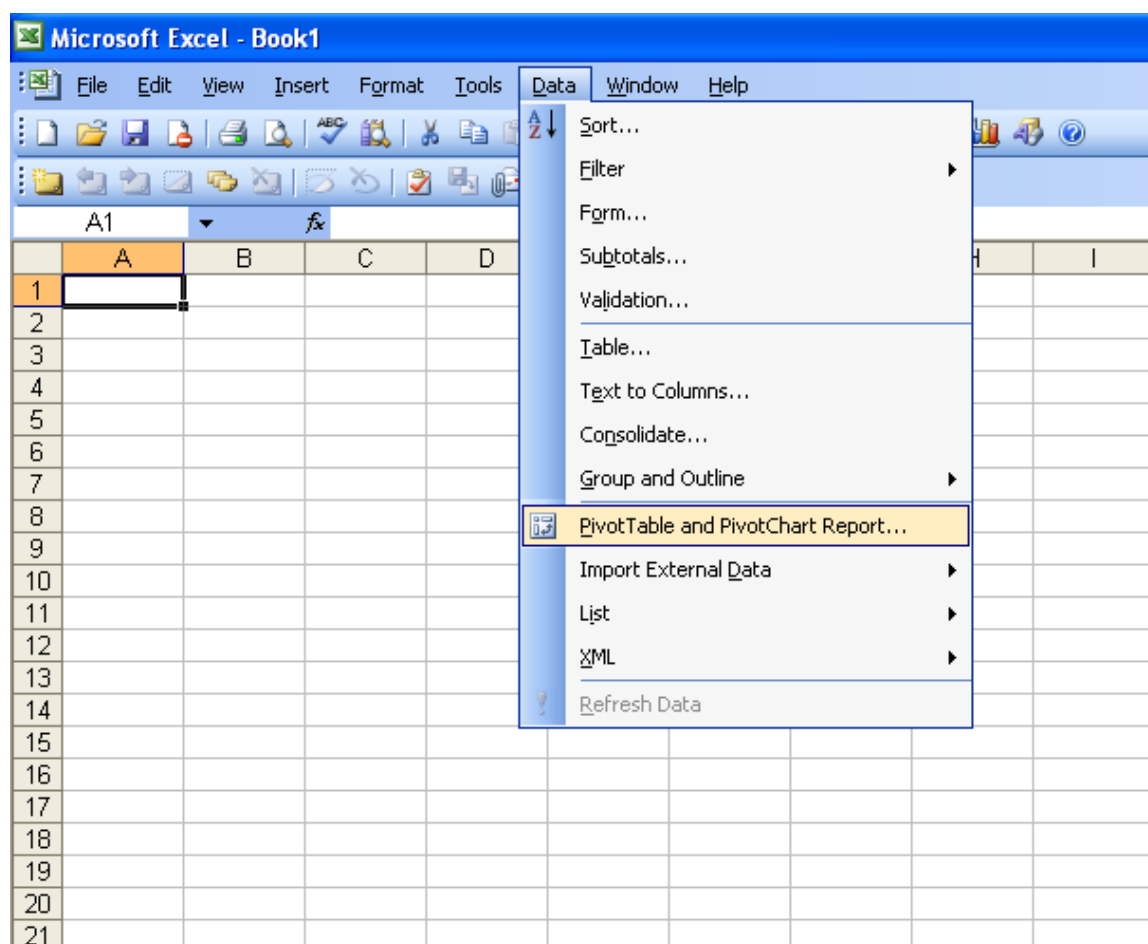
- Select "External data source"[13] and Pivot Table as report type
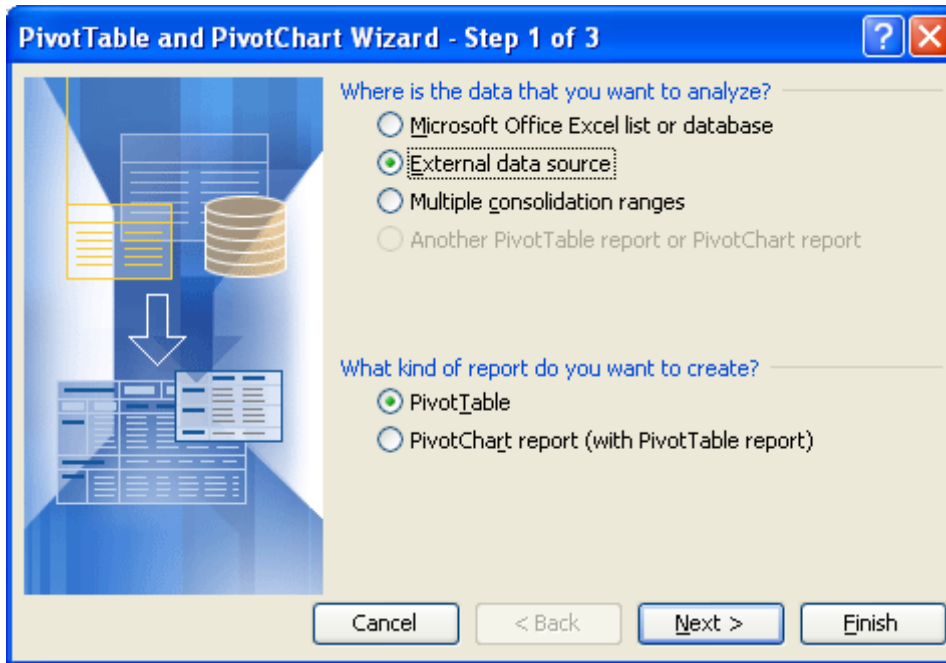
- Click on "Get Data…"
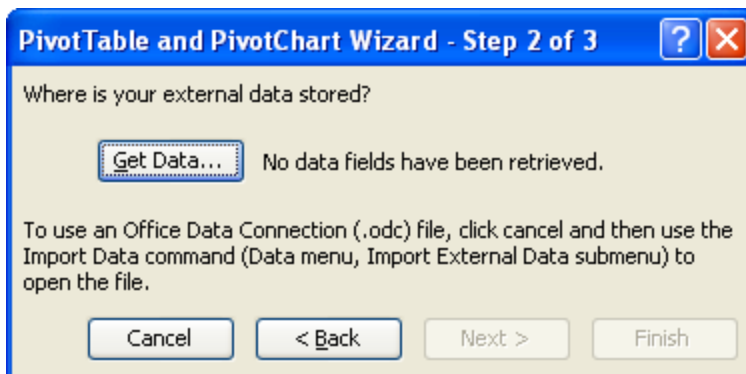
- Go to "OLAP Cubes" Tab. Choose <New Data Source>. Click OK

---
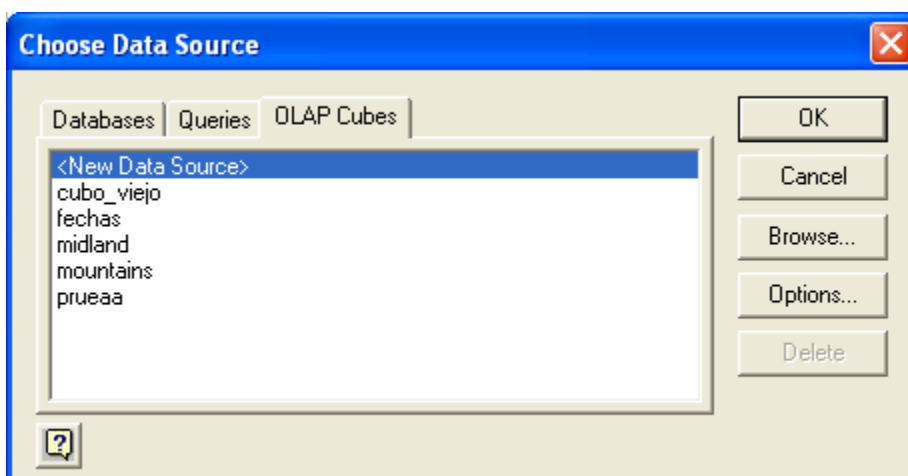
[13] In this case, "external data source" means that the data we are going to build the pivot table from is not coming from an Excel sheet, but from an "external" data source. In our case, this external data source is the .CUB file, i.e. the ESPON OLAP Cube.

- Write down a name for your connection. Choose "MS OLE DB Provider for OLAP Services 8.0 (note: this component should be installed in order to connect to an OLAP Cube). Click Connect… button.
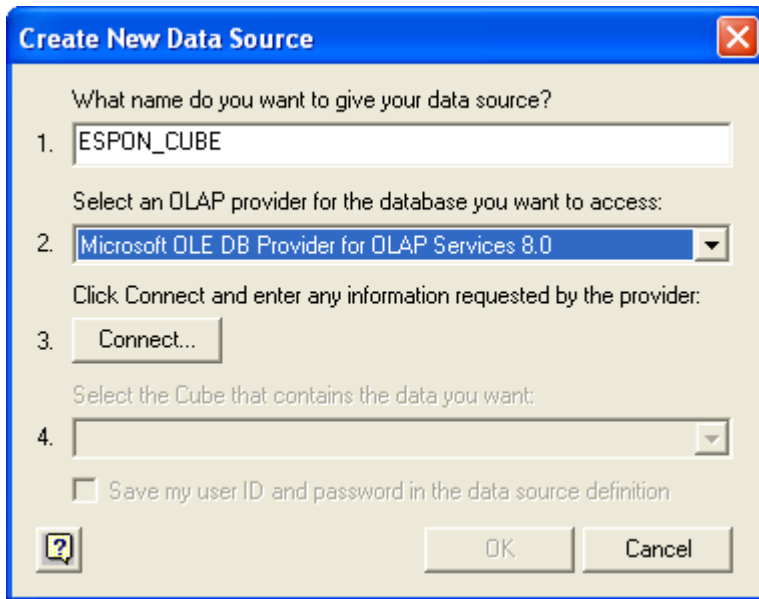
- Choose Cube file, and browse and choose the .cub file in your computer. Click Finish.
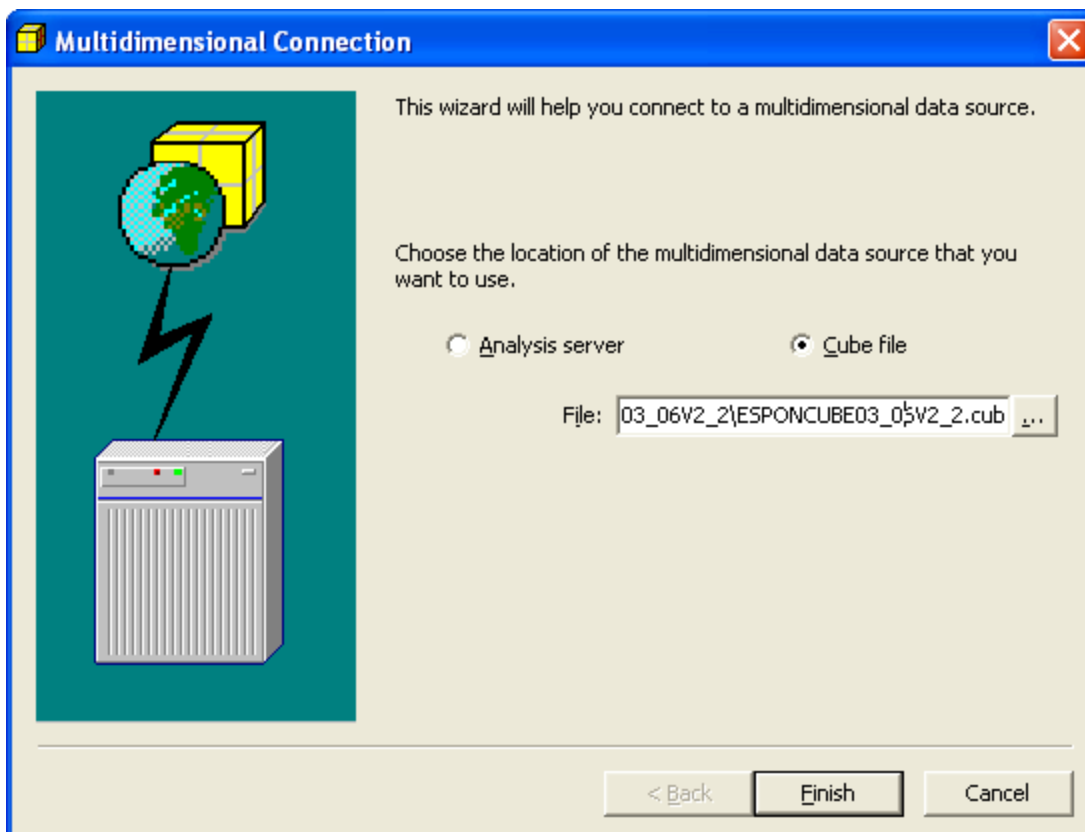
- Click OK
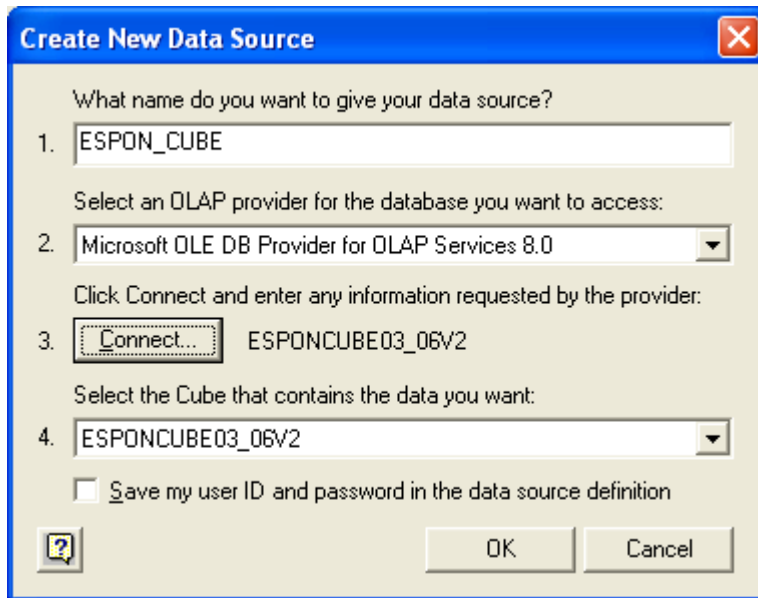
- Choose the connection just created. Click OK
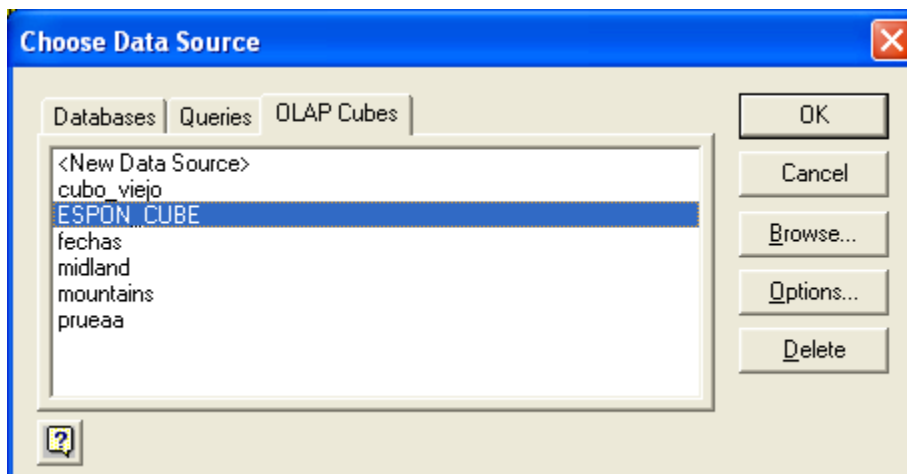
- Click "Next >"
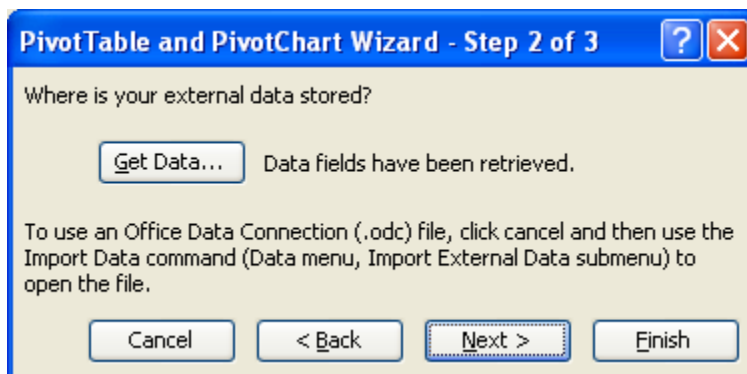
- Choose either a new or existing worksheet. Click Finish.
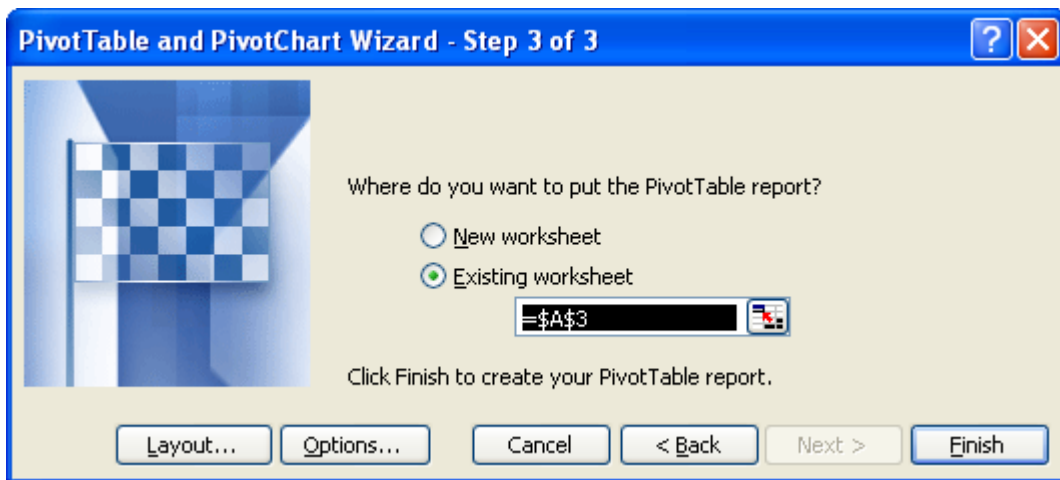
- The Cube is ready!
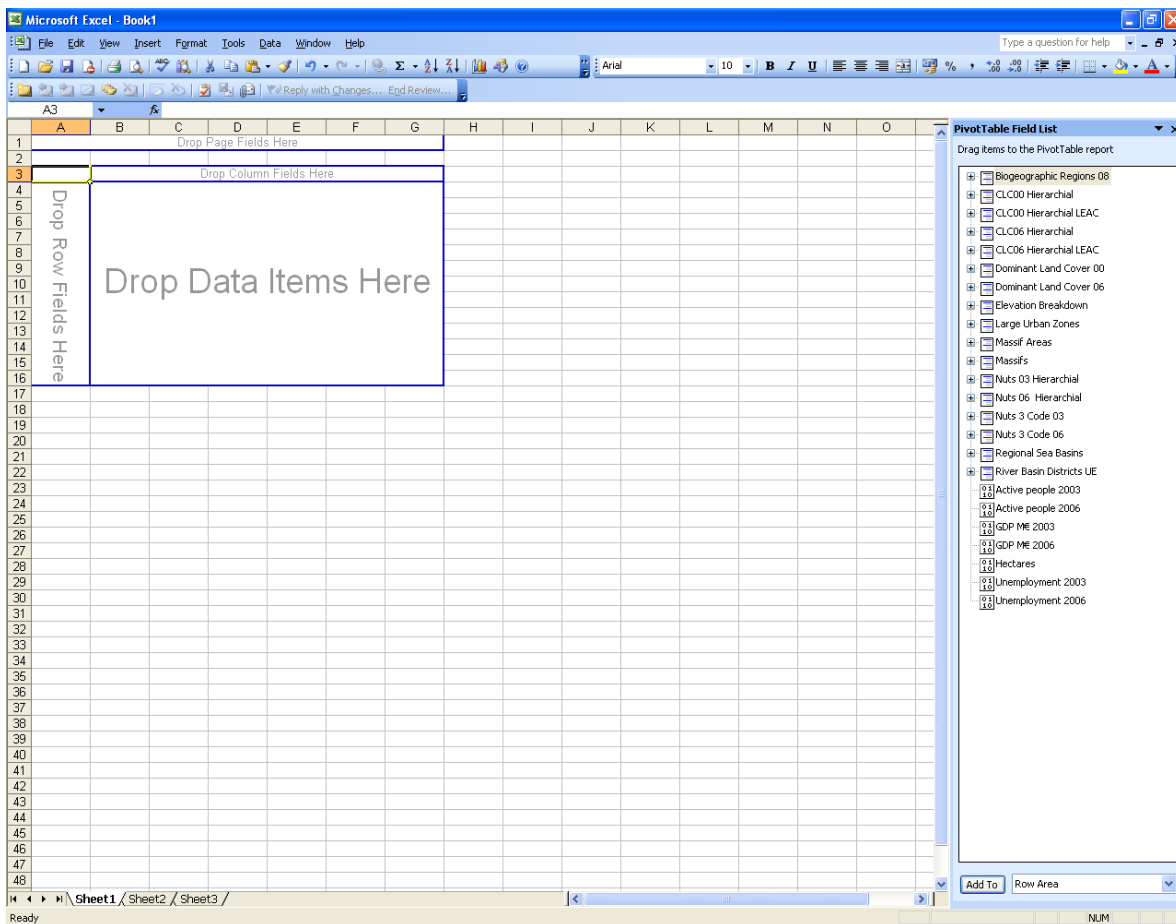
How to query the Cube

- Drag & Drop dimensions (e.g. Nuts 3 Code 06) in the Row Area
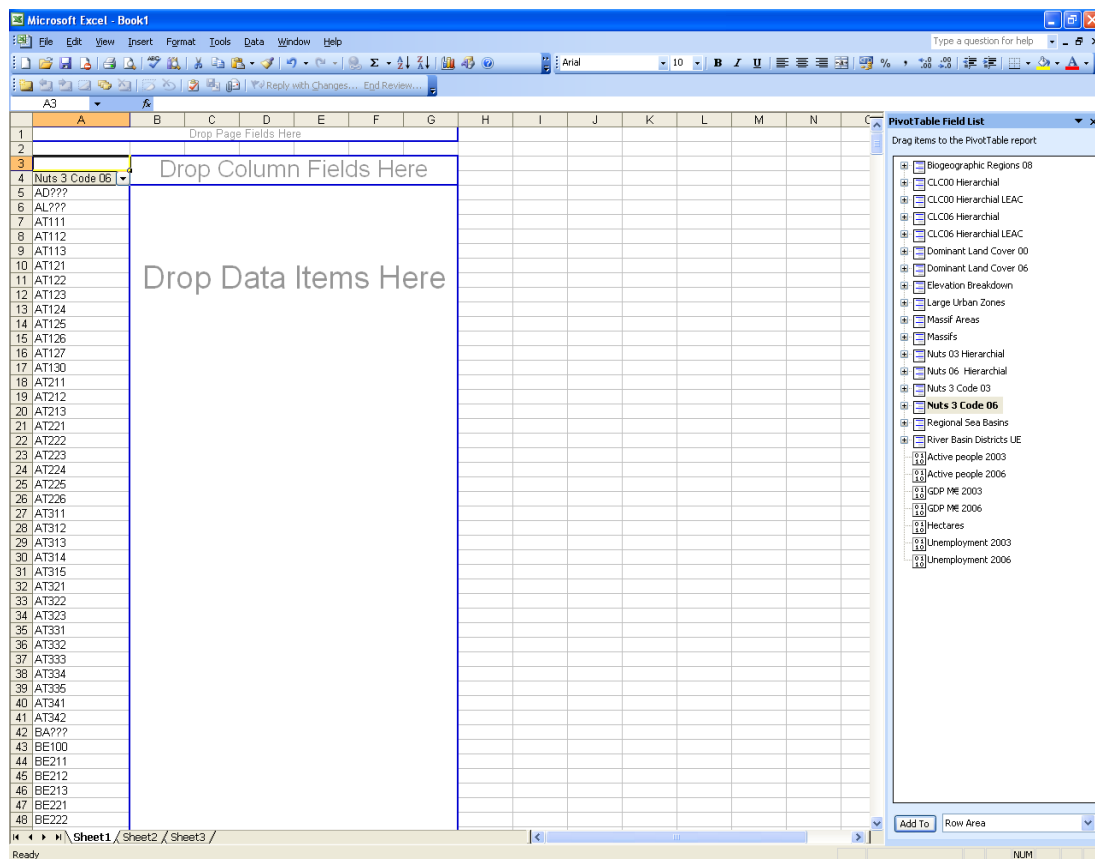
- Drag & Drop measures (e.g. GDP M€ 06) in the Data Area
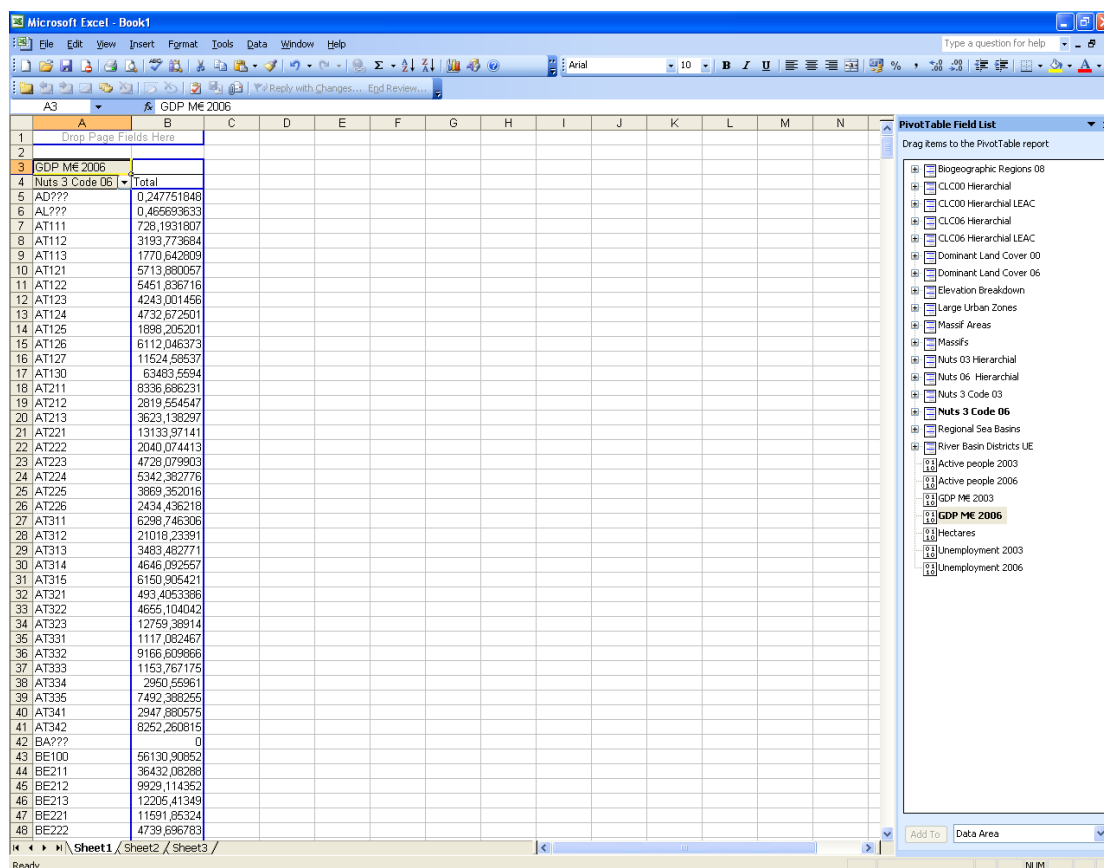
- Drag &Drop other measures (e.g. CLC06 hierarchical Level 1) in the Column Area

- Play around with the ESPON OLAP Cube to build new queries.
- You can also build Pivot charts.
- If you are using MS Excel 2007, please check:

http://office.microsoft.com/en-gb/excel/CH062528071033.aspx

- Documentation about OLAP and OLAP Cubes:

http://en.wikipedia.org/wiki/Online_analytical_processing

http://en.wikipedia.org/wiki/OLAP_cube