



HAL
open science

OLAP technologies applied to the integration of geographic, thematic and socioeconomic data in the context of ESPON

Roger Milego, Maria-José Ramos, Francisco Domingues, César Martinez

► **To cite this version:**

Roger Milego, Maria-José Ramos, Francisco Domingues, César Martinez. OLAP technologies applied to the integration of geographic, thematic and socioeconomic data in the context of ESPON. [Research Report] ESPON. 2014. hal-03609708

HAL Id: hal-03609708

<https://hal.science/hal-03609708v1>

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



OLAP technologies applied to the integration of geographic, thematic and socioeconomic data in the context of ESPON

CONTENT

This technical report details the methodology that has been developed within ESPON M4D in order to combine and integrate in a single database data from heterogeneous sources, so that they can be queried to obtain valuable results.

Data are integrated through a Reference Grid structure, at 1 km² resolution or 100 m² in the case of the urban data.

This method has resulted in the production of different ESPON OLAP Cubes, at European, Urban and Neighbourhood levels.

An ESPON OLAP webtool has been produced to facilitate querying the OLAP Cubes and get maps and charts in a swift way.

50 pages



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



LIST OF AUTHORS

Roger Milego, Universitat Autònoma de Barcelona (UAB)

María José Ramos, Universitat Autònoma de Barcelona (UAB)

Francisco Domingues, Universitat Autònoma de Barcelona (UAB)

César Martínez, Universitat Autònoma de Barcelona (UAB)

Contact

roger.milego@uab.cat

mariajose.ramos@uab.cat

francisco.domingues@uab.cat

cesar.martinez@uab.cat

Tel. (+ 34) 93 581 35 20

TABLE OF CONTENT

1	Setting the scene	4
2	Background and introductory concepts	6
	Modifiable Areal Unit Problem (ESPON 3.4.3)	6
	Existing experiences	7
	The 1 km ² European Reference Grid	8
	GEOSTAT Population Grid	8
3	Methodology description	10
3.1	Integration methods	11
3.2	OLAP Cubes creation.....	14
3.2.1	Specificities of the Urban OLAP Cube.....	16
3.2.2	Specificities of the Neighbourhood OLAP Cube.....	19
4	Results	23
4.1	ESPON OLAP Cubes	23
4.1	ESPON OLAP Cubes user-oriented scenarios	26
5	ESPON OLAP Webtool	34
6	Conclusion and further developments	39
	Annex 1 – ESPON OLAP Cube User Manual	41
	Bibliography	52

LIST OF FIGURES

Figure 1. GEOSTAT Population Grid 1 km ² (2006).....	9
Figure 2. General schema of data processing for OLAP integration.....	10
Figure 3. The 1 km ² European Reference Grid will hold both environmental and socioeconomic information.....	10
Figure 4. Example of the intersection process.....	11
Figure 5. Example of unique combination of units within a grid cell.....	12
Figure 6. Example of the selected star schema.....	13
Figure 7. Schema of OLAP fact table built-up.....	14
Figure 8. Sample result of an OLAP Cube query within MS Excel.....	15
Figure 9. Attribute table of a vector grid generated with the command ‘genvecgrid’ of GME tool.....	20
Figure 10. Union of GlobCover (vector) projected into ETRS89 with the RefGrid (1km).....	21
Figure 11. Simplified method schema, example with GlobCover (GlobCover Raster: GCR and GlobCover Vector: GCV).....	21
Figure 12. Window to select the OLAP Cube within the ESPON OLAP Cube Viewer.....	34
Figure 13. Screenshot of the OLAP web tool homepage.....	35
Figure 14. Subselection of spatial dimensions in the webtool.....	36
Figure 15. Map, table and charts displayed after a query.....	36
Figure 16. Screenshots of the OLAP web tool results after a query.....	37
Figure 17. Example of output from the webtool without map.....	37
Figure 18. Example of output from the webtool without table.....	38

1 Setting the scene

This technical report has been prepared by the UAB in the framework of the ESPON M4D project. It is an update of a former technical report entitled "Disaggregation of socioeconomic data into a regular grid: Results of the methodology testing phase" and it gathers all the technical and methodological aspects of the work carried out by the UAB in order to develop a method and tools to integrate and analyse heterogeneous datasets to benefit the ESPON Programme, in particular, and territorial and spatial assessment, in general.

The challenge of integrating data which are thematically, spatially and/or temporary heterogeneous was already faced in the former ESPON Programme. The harmonization of datasets in a fixed spatial division (NUTS3) solution, which was chosen by many ESPON 2006 projects, presented some doubts and not always convincing results. Another strategy for data integration was needed.

The Modifiable Areas Unit Problem (MAUP) study (Grasland et al, 2006) already highlighted the integration of heterogeneous databases as one of the most promising application of gridding methods for ESPON. Two potential fields of applications were pointed out: time harmonisation of changing territorial units and thematic harmonisation and combination of heterogeneous spatial sources. Furthermore, the so-called "Eurostat oriented" strategy was suggested to be replaced by another strategy that could be named "EEA oriented", where all the data would be transformed into a grid and integrated on this basis (Grasland et al, 2006).

The terms of reference for the ESPON 2013 DB project already emphasized the need towards a different way of data integration, where socioeconomic information were put into smaller units comparable to those used for environmental or natural data. The analysis of different experiences of data integration worldwide (Deichmann et al, 2001; Gallego, 2001; Nordhaus, 2006) led 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. Moreover, recent territorial cohesion policies derived from the Territorial Agenda 2020 have highlighted the importance for a place-based approach and territorialisation of policies (Zaucha, 2013), which can benefit from spatial data integration like the one applied under the ESPON M4D project by the UAB.

Another important key fact with regard to the methodology described in this technical report was the release of the GEOSTAT 2006 population grid¹. It was an initiative from the European Forum for Geostatistics and Eurostat, through the GEOSTAT project, published in April 2012. It is a one square kilometre gridded dataset containing population figures for the year 2006. It integrates data from national grid initiatives and, wherever this has not been possible, the European disaggregated dataset produced by the Austrian Institute for Technology (AIT) by means of detailed datasets, such as a Soil Sealing layer at 30x30m resolution.

To conclude the presentation section, an introduction to OLAP should be made. Online Analytical Processing is a computer-based technique to answer multi-dimensional analytical queries swiftly. OLAP tools enable users to analyse multidimensional data interactively from multiple perspectives. It experienced a strong growth in the late 90s, but it had been mainly applied to business data in the search for business intelligence. The application of such technique to social and geographical information

¹ http://www.efgs.info/data/GEOSTAT_Grid_POP_2006_1K

as shown in this technical report is rather innovative and provides added value to the integration of data by means of a grid. Moreover, integrating such technique in a web tool, to facilitate querying the database to any kind of user is also an asset of the methodology that has been implemented during the ESPON M4D by the UAB, which has lead to the delivery of different OLAP Cubes and an online webtool to query them. All the technical details with regard to the methodology implemented and results obtained is included in these pages.

2 Background and introductory concepts

The methodology and results achieved under the ESPON M4D project by the UAB started within the former ESPON 2013 Database project, in which the basis for the methodology was set up. Under M4D the methodology has evolved, several OLAP Cubes have been build up at different scales and an online webtool has been developed in order to facilitate the querying of the Cubes in an interactive, more user-friendly manner.

Under the ESPON 2013 DB project the UAB assumed the challenge of “combining socio-economic data measured for administrative zoning (NUTS level) and environmental data defined on a regular grid (like Corine Land cover)”. The ESPON 2006 program had 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 was said in the tender of the ESPON 2013 Database project and by the Modifiable Areal Unit Problem study:

Modifiable Areal Unit Problem (ESPON 3.4.3)

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

Two potential fields of applications were 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, was questioned, and the use of a new strategy was 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).

Existing experiences

At the very beginning, a literature review was made in order to support the best decision on the methodology to be implemented. The following examples show different approaches in which a regular grid structure was used as a common data integrator:

- **"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 density².

- **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 allocation³.

- **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 Framework⁴.

- **"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 cells⁵.

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.

²

http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S14P3_JAVIER_GALLEGO_DO_WNSCALED_POPULATION_DENSITY.pdf

³ "New Metrics for Environmental Economics: Gridded Economic Dats" by William D. Nordhaus.

<http://www.oecd.org/dataoecd/44/7/37117455.pdf>

⁴ www.faro-eu.org

⁵ <http://sedac.ciesin.columbia.edu/gpw-v2/GPWdocumentation.pdf>

The 1 km² European Reference Grid

The 1st Workshop on European Reference Grids⁶ 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.

In this way, the European Reference Grid (ERG) was born, using a standard European projection system (ETRS89-LAEA 52N 10E) and having a spatial resolution of 1 km².

ERG has become an European standard dataset, broadly used by all the European institutions and research centres.

Therefore, taking into account these recommendations and our experience under some EEA projects such as LEAC (Land and Ecosystem Accounting), we proposed to disaggregate socioeconomic data into the **1 km² European Reference Grid**⁷, as it is the way in which valuable data for the ESPON projects, such as the Corine Land Cover datasets, are stored as well.

GEOSTAT Population Grid

In April 2012, an initiative from the European Forum for Geostatistics and Eurostat, through the GEOSTAT project, led to the publication of a gridded population dataset, with a 2006 time stamp. It is a one square kilometre (1 km²) gridded dataset containing population figures for the year 2006. It integrates data from national grid initiatives and, wherever this has not been possible, the European disaggregated dataset produced by the Austrian Institute for Technology (AIT⁸). This results into an integrated single population grid dataset.

⁶ http://eussoils.jrc.ec.europa.eu/projects/alpsis/Docs/ref_grid_sh_proc_draft.pdf

⁷ <http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=760>

⁸ Austrian Institute for Technology

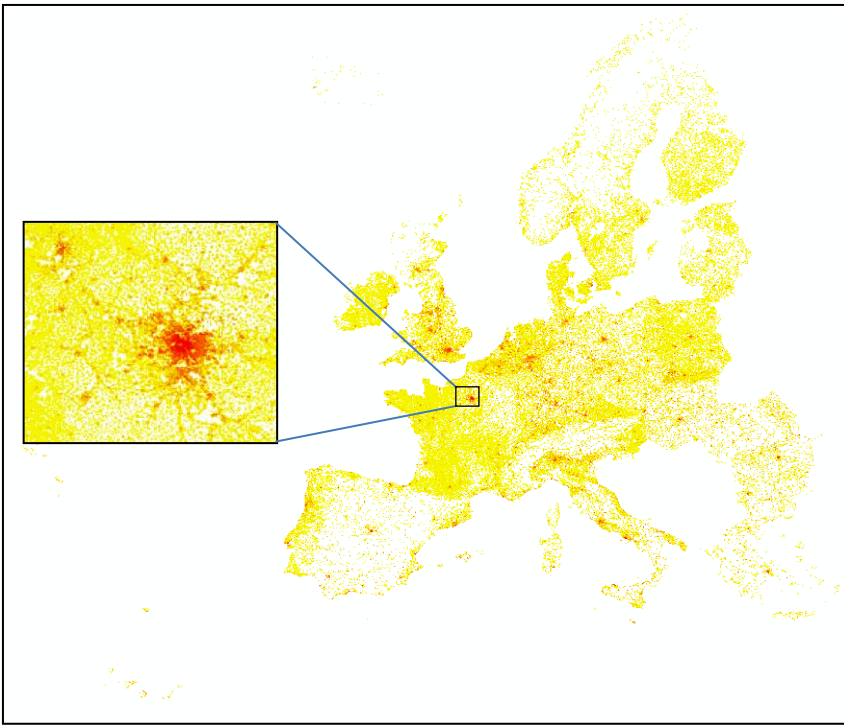


Figure 1. GEOSTAT Population Grid 1 km² (2006)

The added-value of this dataset compared to the former JRC 2001 population grid is the improvement in the population disaggregation phase, made for some countries at national level (AT, FI, FR, SE, DK, PL, EE, ES, SI, NL, NO, UK, PT) and, for the rest, at European level by AIT, by means of detailed datasets, such as the Soil Sealing layer. The GEOSTAT 2006 Population grid can be used by any user within the user's own organisation, but there are some restrictions regarding its use for commercial purposes and its dissemination to third parties.

The GEOSTAT 2006 population grid has become the main ancillary data for the disaggregation process of socioeconomic data, as it will be further detailed in the methodological chapter.

The next section describes the methodology that has been defined and carried out throughout the project, which has resulted in the creation of different ESPON OLAP Cubes, a user manual on how to use them and a webtool aimed at facilitating the querying of OLAP Cubes and obtaining prompt and nice charts and maps out of them.

3 Methodology description

The methodological process in order to build up the ESPON OLAP Cubes can be described in two main phases:

- Data integration within the ERG
- OLAP Cube creation

In figure 4, a general schema of data preparation for integration within an OLAP database is shown.

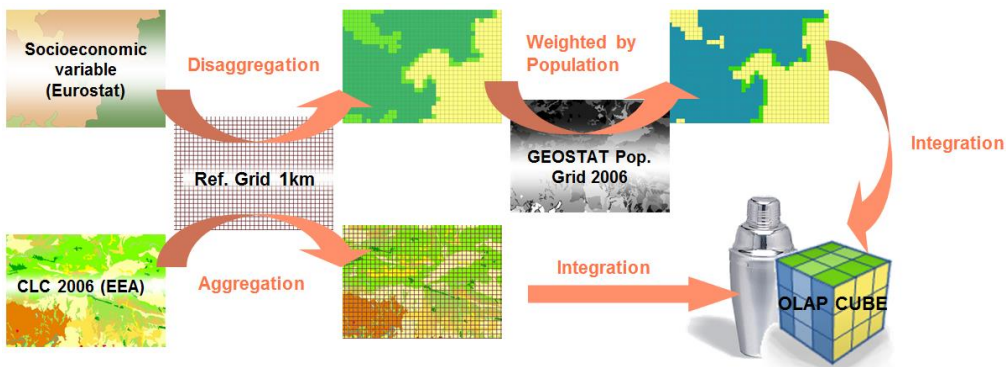


Figure 2. General schema of data processing for OLAP integration

In order to combine socioeconomic data with other types of datasets within an OLAP database, the 1 km² European Reference Grid (ERG), adopted by several European stakeholders at the First Workshop on European Reference Grids (JRC-IES-LMU-ESDI, 2003), is used as the **common integrator layer**. This means, in the end, that each single grid cell must store a single figure for each variable, either numeric or alphanumeric.

As it can be seen in figure 1, in the end the ERG will store all types of information, so that they can be compared on a cell-by-cell basis.

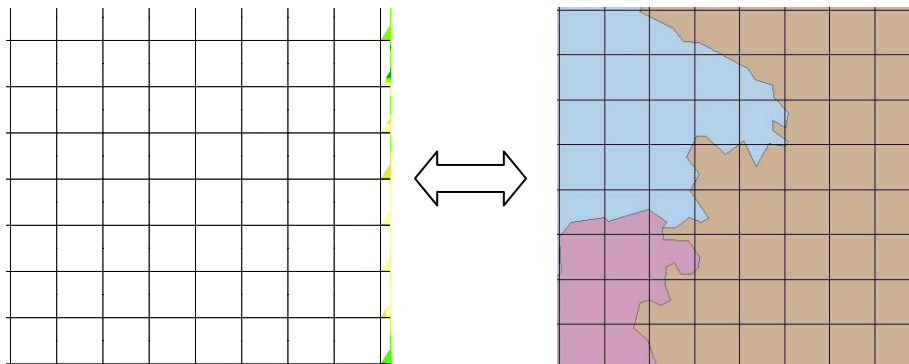


Figure 3. The 1 km² European Reference Grid will hold both environmental and socioeconomic information

Actually, depending on the nature of the dataset or variable, we distinguish:

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

Depending on the type of dataset different procedural steps are applied in order to integrate them within the ERG. These methods are defined in the next section.

3.1 Integration methods

As for **geographic dimensions**, an overlay (intersection) between the dimension and the 1 km² ERG is undertaken, in order to give each grid cell a single feature code (e.g. a NUTS3 code, a LUZ code, etc.). Aiming at avoiding coastal discrepancies, some dimensions, such as NUTS regions, are previously submitted to spatial extension processes generating proximity layers.

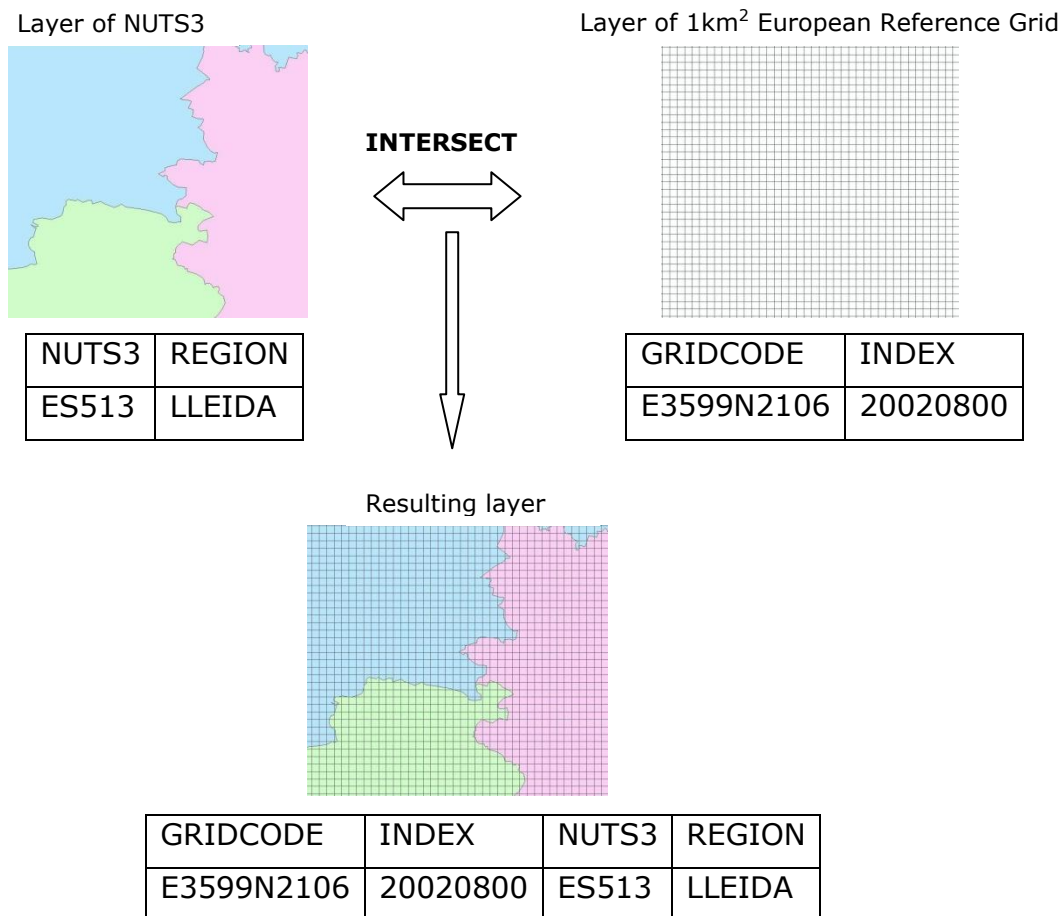
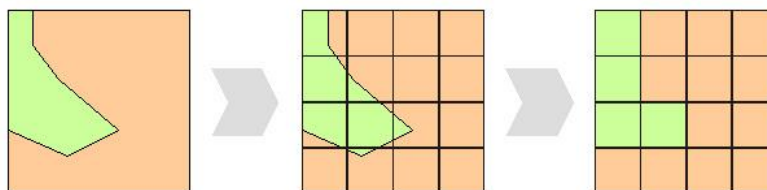


Figure 4. Example of the intersection process

In the case of different geographic units overlaying one cell, a maximum area criterion is applied:

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.



Taking into account that some **thematic layers** have a higher resolution than 1 km² (e.g. Corine Land Cover is available at 100 m² resolution), the combination of such information is carried out at **100 m²**. In this way, it is possible to store, for instance, the different land cover classes and their surfaces for each grid cell. The grid cell index (unique identifier for each grid cell) might appear duplicated in the resulting output table, as many times as different land cover classes exist within the square kilometre. Geographic dimensions will have their code repeated as well:

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

UNIQUE COMBINATION OF UNITS

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

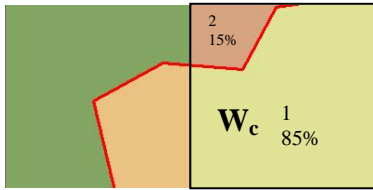
Figure 5. Example of unique combination of units within a grid cell

When it comes to **socioeconomic data**, typically stored by NUTS3, the main challenge is how to disaggregate such figures and get a single value for each cell. In order to do so, population has been used as an **ancillary variable**. With the release of the GEOSTAT 2006 population grid, high quality 2006 population figures are available at 1 km² resolution. That valuable information is used to redistribute population figures at NUTS3 level for the time series 1990-2011 (compiled by ESPON M4D) proportionally to the distribution of population in 2006. When doing so, figures by each grid cell are a proxy of the actual value, but they compensate each other when grid cells are aggregated back to show the result of a query, which always should be done by a bigger reporting unit (NUTS3, LUZ, UMZ...). The rest of socioeconomic variables, such as GDP/capita or unemployment are redistributed in the same way, weighting the figures by population in each grid cell. In this way, those figures can be aggregated back in a query, despite the different NUTS breakdowns or geometries involved (urban delimitations, administrative levels, etc.).

In this case, a proportional and weighted calculation is applied for each cell:

Proportional and weighted calculation: the cell takes a proportionally calculated value by area, but this value is weighted for each cell, according to an external

variable (e.g. population). This method is applied to improve the territorial distribution of a socioeconomic indicator. For instance, a GDP indicator is redistributed by 1 km² grid and weighted by the population figures of each cell (coming from the GEOSTAT database).



$$\text{Cell value} = W_c \sum (V_i * \text{Share}_i)$$

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: $W_c * (V_1 * 0.85 + V_2 * 0.15)$

In the example above, the weighting figure corresponds to the share of population living in that grid cell. In this way, a single figure for each socioeconomic variable is calculated by each unique grid cell.

Once the data dimensions and measures have been overlaid with the ERG and each grid cell has a single figure for each variable (or more, in the case of Corine Land Cover), then it is possible to join all the information in a unique OLAP database. In order to build it up, a star schema has been selected. It is composed by one **fact table** and one **dictionary table** for each geographical or thematic dimension.

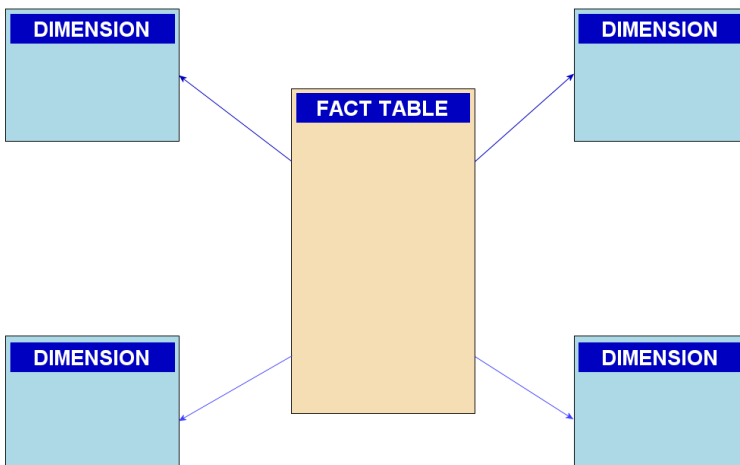


Figure 6. Example of the selected star schema

The **fact table** stores, in the end, all the data dimensions (geographic and thematic) and all the variable measures. They are all referred to a unique identifier for each grid cell. Whenever different land cover classes occur in a single grid cell, the cell code and geographic codes appear as many times as different CLC classes exist. As for measures (numeric values) they are distributed proportionally to the area share of each grid cell piece:

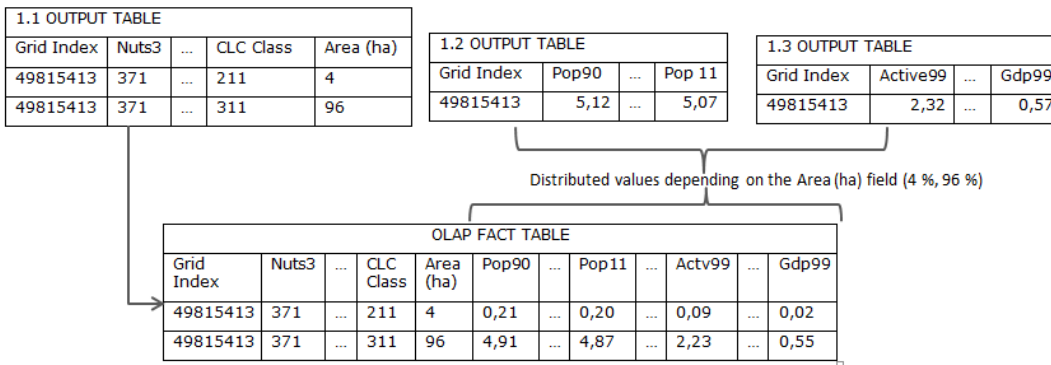


Figure 7. Schema of OLAP fact table built-up

The **dictionary tables** store the description of a specific dimension (e.g. NUTS names). They are put into relation with the fact table by means of the dimension codes.

Once the fact and dictionary tables are built, then the OLAP database (also named OLAP cube, as a shortcut for multidimensional) is ready to be created.

3.2 OLAP Cubes creation

Once the fact and dictionary tables are ready, they can be integrated within a single **OLAP (Online Analytical Processing) database or cube**.

The OLAP technology⁹ 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.

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

The OLAP database (also called **Cube**, as for being multidimensional) is created by means of Microsoft SQL Analysis Services 2000. The resulting product is a .CUB file that can be queried using Microsoft Excel or an online connection to a remote server. A User Manual (annex x) has also been produced to allow users to connect to and query the OLAP database.

⁹Some 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

In this way, users are able to ask some questions taking into account the different socioeconomic, thematic and geographic variables integrated within the database. 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.

As mentioned, those OLAP Cubes can be queried by means of Microsoft Excel through a dynamic table, allowing the user to choose what dimensions appear as rows or columns and which measures will be shown as values after the query. In this way, the user can swiftly get results like the example in figure 4.

	A	B	C	D	E	F	G
1	NUTS3 (2010)	Population 1990 thou	Population 1998 thou	Population 2001 thou	Population 2005 thou	Population 2010 thou	GDP 2006 Million Euro
19	IS	253,79	272,38	283,36	293,58	317,63	,00
20	IT	56.618,42	56.832,68	56.890,15	58.381,92	60.254,84	1.440.002,47
21	ITC	14.933,24	14.851,57	14.884,49	15.399,88	15.975,45	468.376,67
22	ITC1	4.310,28	4.224,99	4.205,74	4.316,11	4.431,88	118.838,66
23	ITC11	2.257,67	2.187,64	2.169,09	2.236,30	2.296,93	63.428,96
24	ITC12	185,14	179,31	177,73	177,57	180,10	4.858,60
25	ITC13	191,90	188,24	187,60	188,19	186,72	4.935,15
26	ITC14	154,36	152,23	151,52	153,69	155,17	3.436,16
27	ITC15	328,95	333,36	337,02	348,93	363,84	9.867,97
28	ITC16	543,93	551,01	554,92	569,49	589,07	16.369,90
29	ITC17	206,96	207,27	207,94	213,06	220,82	4.933,67
30	ITC18	441,37	425,92	419,93	428,87	439,23	11.008,25
31	ITC2	115,23	118,87	120,04	123,68	128,70	3.690,50
32	ITC20	115,23	118,87	120,04	123,68	128,70	3.690,50
33	ITC3	1.699,93	1.613,74	1.584,52	1.597,82	1.621,60	36.948,34
34	ITC31	214,15	208,44	205,76	215,37	221,66	4.367,01
35	ITC32	286,88	277,50	274,30	282,71	288,43	6.329,12
36	ITC33	965,85	903,15	883,25	875,81	883,26	21.102,23
37	ITC34	233,05	224,65	221,21	223,93	228,25	5.149,98
38	ITC4	8.807,80	8.893,97	8.974,19	9.362,26	9.793,27	308.899,17
39	ITF	13.887,83	14.017,39	13.938,78	14.085,78	14.167,64	227.661,38
40	ITG	6.607,85	6.657,64	6.614,65	6.663,13	6.715,40	107.719,64
41	ITH	10.346,29	10.451,05	10.588,16	11.026,92	11.547,94	325.847,11
42	ITI	10.843,22	10.855,03	10.864,06	11.206,21	11.848,42	310.397,67
43	LI	39,20	42,80	33,77	46,42	48,21	41,62
44	LT	3.693,05	3.561,58	3.486,33	3.424,67	3.328,42	23.876,93
45	LU	377,81	420,22	437,02	459,05	499,51	33.163,85
46	LV	2.668,35	2.420,98	2.364,44	2.306,63	2.248,59	15.902,80
47	MK	176,87	183,13	183,42	181,31	179,85	,00
48	MT	352,43	376,51	391,42	402,67	416,96	4.602,48
49	NL	14.860,29	15.622,33	15.954,64	16.273,11	16.542,57	536.973,42
50	NL1	1.590,79	1.637,84	1.668,47	1.698,32	1.710,81	54.343,44
51	NL11	548,31	552,56	560,98	569,58	571,26	24.901,60
52	NL111	145,73	146,76	148,24	148,08	146,24	2.770,09

Figure 8. Sample result of an OLAP Cube query within MS Excel

Section 4 shows the results derived from the implementation of the OLAP technology, including the different OLAP Cubes created under ESPON M4D, plus an advanced web tool to query those OLAP Cubes through a simple web browser.

3.2.1 Specificities of the Urban OLAP Cube

This section details some of the aspects that concerned the creation of the Urban ESPON OLAP Cube. In this case, it was decided to build up an OLAP Cube for the urban areas at a higher resolution than the regular ESPON OLAP Cube for Europe. This fact means that the integration grid has a 100m*100m (1 hectare) cell size instead of the "classical" 1 km².

Input data

The data to be processed and included in the ESPON Urban OLAP Cube was mainly EU based datasets that characterizes cities by its morphology, uses, cover and structure, among others. Those include:

- Urban Morphological Zones (UMZ) 2000
- Morphological Urban Areas (MUAs)
- Functional Urban Areas (FUAs)
- Large Urban Zones (LUZ) 2004
- Large Urban Zones (LUZ) 2012
- Urban Atlas¹⁰ 2006 and update 2012

Additionally, it was decided to include administrative boundaries as LAU2 (Local Administrative Units 2 - GISCO) and population data using GEOSTAT 2006 Population grid.

The input data have different extents and resolution, some covers the Atlantic Portuguese and Spanish islands and others not, most layers have 100 meters resolutions, except for Urban Atlas raster (10 meters resolution) and GEOSTAT Population (1km²). For that reason different quality checks, firstly, and data harmonization, afterwards, have been carried out in order to be able to combine the different datasets correctly.

Quality check and harmonization

Data have been treated in a way that no pixels are skipped: The first step consists on harmonizing data by grouping it by resolution. All 100 meters resolution layers were snapped into the main layer Urban Atlas and then 'No Data' zones were converted to zero, so that when applying the "combine" tool no pixels are left behind. Data cover different extents and zones and when 'no data' is present in a shared/overlapping area no combine is done, so a value must be present.

Population raster at 1km² resolution was integrated in the "combine" as it is so no harmonization is needed as the pixel borders were correctly aligned with all other input rasters.

The struggling point comes when trying to harmonize Urban Atlas due to its resolution (10 m) which increased the processing time of each step and sometimes the results were not fully correct as the software used (ArcGIS 10.1) got some outputs corrupted. Although the big effort and huge time spent treating 10m resolution data, Urban Atlas was integrated in two steps. First, UA 2006 was included in the "combine" and,

¹⁰ <http://land.copernicus.eu/local/urban-atlas>

afterwards, another “combine” was done with the UA 2012 updates. The urban zones outside urban atlas were included afterwards, as it was impossible to deal with Urban Atlas ‘No Data’ areas and convert them into zero.

In order to pre-process the data to be included in the final “combine”, the third one, it was needed to create a mask with presence/absence of urban atlas classes 2006 or 2012. The first step was reclassifying both UA 2006 and UA 2012 ‘Nodata’ to 1 and UA classes to ‘Nodata’. Afterwards, mosaicking both masks into a single one to have a final presence/absence raster was needed. The time spent was huge due to the extent to be covered.

Combination of input data

As mentioned before, the “combine” process had to be done in three steps, first by combining all urban layers and LAU2 at 100m with UA 2006 and Population grid; the second “combine” the same, but using the UA 2012 update instead; and, finally, the third and last one using all urban layers, LAU2, Population raster and the 10m mask with the zones outside UA 2006 and 2012. The “combines” were done always at 10m resolution. The 1km EEA reference grid was included as well because each population pixel should have an ID in order to be distributed within the correct area of urban fabric within its zone.

To merge all “combines” in a final table the tool “append” (access/sql server) was executed twice, using the first “combine” table as the main one, and adding the second and third ones to the end.

Population was distributed among the continuous and discontinuous urban fabric Urban Atlas classes present within that 1 km² grid cell. This process assumes that population is located in urban fabric zones. In areas outside Urban Atlas the population was distributed according to the amount of area covered.

To reduce the amount of rows and enhance the performance avoiding replicated values, the output table was summarized by all columns using a “group by” in MS Access, except for the population column that had to be added.

The final table was inserted into MS SQL Server 2003 database and then the Cube was built in MS SQL Analysis Services Projects database with the following characteristics:

Dimensions:	Measures:	Calculated Members:
LAU2	Population UMZ 2000	Area in Hectares
UMZ 2000	Area	
MUA	Population 2006	
FUA		
LUZ 2004		
LUZ 2012		
Urban Atlas 2006-12		

For each dimension a definition table was constructed with the respective code and a description to be displayed further on.

A final quality check was done comparing the real areas with the areas calculated. It is very important to check that all the definition tables have their corresponding values in the main fact table. Otherwise, they would not be correctly added.

Difficulties

During the processing steps several difficulties were found, mainly related to the size of the input datasets. Both 2006 and 2012 Urban Atlas rasters at 10 meters resolution are extremely big datasets that take a lot of time for each of the steps, when it is needed to harmonize them and create a sub-product. When making a data mosaic, needed to have both UA together, or even to create the mask including the zones outside UA, or reclassifying UA and exporting to other raster; each of these processes always took too much time. Additionally, due to an ArcGIS bug, pyramids are always created when using "mosaic" or "mosaic to new raster", and it was very difficult to find a workaround within ArcGIS 10.1.

During quality check of outputs, when rasters were at 10 m resolution, ArcGIS spent between 3-5 minutes each, only to complete the display of the raster.

The next table shows approximate processing times, to get an idea of how costly the processing was:

Display of rasters 10 m	3-5 minutes each time
Reclassify of 10 m rasters at EU extent	1h 30 to 2h
Mosaic to new raster (masks)	2h to 3h
Mosaic or reclassify with pyramids	More than 2 days
Combine with 8/9 layers	5h

3.2.2 Specificities of the Neighbourhood OLAP Cube

In this chapter the source data used and the different steps of the methodology applied to prepare the Neighbourhood OLAP Cube are detailed.

Data sources

Globcover

Data source: ESA (<http://due.esrin.esa.int/globcover/>)

- Globcover2009_V2.3
- Globcover 200412 200606 V2.2

GDP

Reference year: 2010

Data source: PREVIEW

<http://preview.grid.unep.ch/index.php?preview=data&events=socec&evcat=1&lang=eng>

Population

Reference year: 2000

Data source: GRUMPv1 (<http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-count>)

Recommended Citation(s)*:

Center for International Earth Science Information Network - CIESIN - Columbia University, International Food Policy Research Institute - IFPRI, The World Bank, and Centro Internacional de Agricultura Tropical - CIAT. 2011. Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Count Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://dx.doi.org/10.7927/H4VT1Q1H>. Accessed DAY MONTH YEAR.

Employment

Reference year: 2007

Data source: (ILO : International Labor Organisation, LABORSTA.

<http://laborsta.ilo.org/> and UNSD : UN Statistical Division.

<http://unstats.un.org/unsd/environment/qindicators.htm>).

Step 1. Generation of the reference Grids (RefGrids)

The Geospatial Modelling Environment Tool (<http://www.spatialecology.com/gme/>) has been used to generate the RefGrids in ETRS89.

The command '**genvecgrid**' allows you to generate a regular grid in vector format (<http://www.spatialecology.com/gme/genvecgrid.htm>)

The attribute table contains information on coordinates for each cell.

	FID	Shape	Id	CELLID	MinX	MaxX	MinY	MaxY	CenterX	CenterY
▶	0	Polygon	0	1	3993000	3994000	2760000	2759000	3993500	2759500
	1	Polygon	0	2	3994000	3995000	2760000	2759000	3994500	2759500
	2	Polygon	0	3	3995000	3996000	2760000	2759000	3995500	2759500
	3	Polygon	0	4	3996000	3997000	2760000	2759000	3996500	2759500
	4	Polygon	0	5	3997000	3998000	2760000	2759000	3997500	2759500
	5	Polygon	0	6	3998000	3999000	2760000	2759000	3998500	2759500
	6	Polygon	0	7	3999000	4000000	2760000	2759000	3999500	2759500

Figure 9. Attribute table of a vector grid generated with the command 'genvecgrid' of GME tool

Step 2. Field "CELLCODE"

Note that GME tool generates an error in the attribute table. The field name "MinY" and "MaxY" should be inverted.

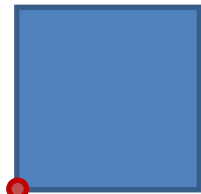
An identify field "CELLCODE" in string format was added to each cell of the RefGrids. The name contains the cell resolution (1km) and Xmin and Ymin coordinates divided by thousand.

Example : CELLCODE= 1kmE4009N2565

"1kmE" & ([MinX]/1000) & "N" & ([MinY]/1000)

"1kmE" & ([MinX]/1000) & "S" & (-[MinY]/1000)

The AOI is too large to be process at once; therefore the RefGrids have been generated separately for 10 different zones.



Xmin ;
Ymin

RefGrid (zone)	Xmin (left)	Xmax (right)	Ymin (bottom)	Ymax (top)
Europe1_GME	1547000	3848000	3241000	5541000
Europe2_GME	3847000	6148000	3241000	5541000
Europe3_GME	1066000	3847000	926000	3241000
Europe4_GME	3847000	6147000	941000	3241000
Europe5_GME	6147000	8447000	941000	3241000
AfricaE_GME	4255000	7129000	-615000	2092000
AfricaW_GME	1390000	4265000	-615000	2092000
Russia_GME	4936000	8210000	2472000	6830000
Greenland_GME	1549000	4937000	5159000	7157000
Feroe_GME	3400000	3489000	4355000	4473000

Step 3. Buffer of 5 km

In order to alleviate the processing, RefGrids have been clipped to an area of interest (AOI) defined by a buffer of 5 km around a country layer "GRID_GADM_levels23lam.shp".

Note that information falling outside this border has not be considered.

Step 4. Disaggregating socioeconomic data

The methodology applied consists in:

- 1) converting wgs84 raster data (GlobCover, GDP, POP) into feature
- 2) project the vector file into etrs89
- 3) union it with the RefGrid in etrs89

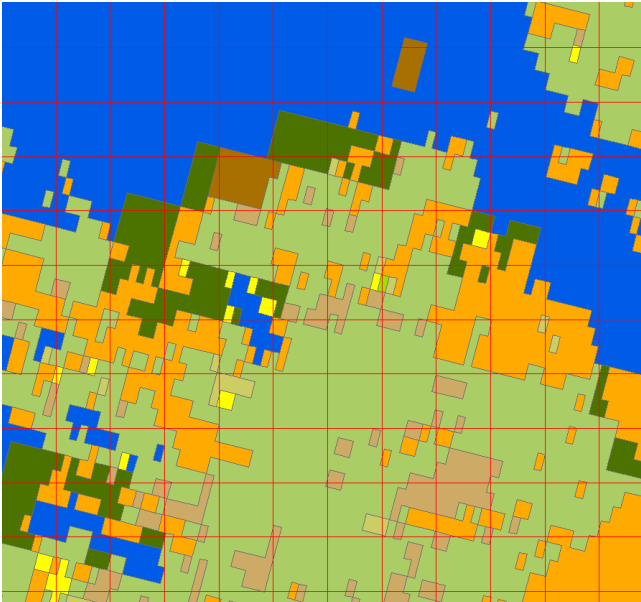


Figure 10. Union of GlobCover (vector) projected into ETRS89 with the RefGrid (1km)

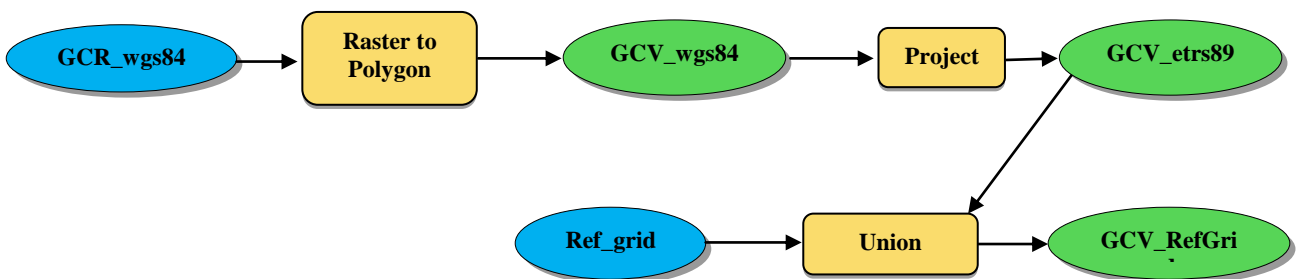


Figure 11. Simplified method schema, example with GlobCover (GlobCover Raster: GCR and GlobCover Vector: GCV)

Step 5. Attribute tables

Tables processing was done in PostgreSQL software.

Globcover

Percent area of each GlobCover class was calculated for each cell, as well as the majority class ("the_class").

GDP

Sum of GDP values by cell of reference.

POP

Sum of population values by cell of reference.

Employment

Percentage of workers by cell of reference.

4 Results

4.1 ESPON OLAP Cubes

Different OLAP Cubes can be created according to the user needs, including chosen data dimensions and measures, if they have been prepared for OLAP integration as explained in the previous section.

In the context of the ESPON M4D, three types of OLAP Cubes have been created.

- **ESPON OLAP Cube** : it covers the entire ESPON area, holding different geographic dimensions plus Corine Land Cover and main socioeconomic variables. This Cube has been updated several times during the project. The working resolution is 1 km².
- **ESPON Urban OLAP Cube** : it covers urban areas (i.e. LUZ coverage) and it holds all the different urban delineations, plus population and Urban Atlas as land cover information. The working resolution is 100 m².
- **ESPON Neighbourhood OLAP Cube** : it covers the Neighbouring areas of the ESPON space (those of the ESPON ITAN project). It includes SNUTS geometries, Globcover information, population, GDP and unemployment. The working resolution is 1 km².

These are the contents of each of the OLAP Cubes developed under ESPON M4D :

ESPON OLAP Cube (version 6.0)

The last update of the ESPON OLAP Cube was delivered in June 2013 and focused on urban geographical dimensions and the newest population statistics. Its content is:

Measures:

- Area in hectares
- GDP 2000 Million Euros
- GDP 2003 Million Euros
- GDP 2006 Million Euros
- GDP 2009 Million Euros
- Population 2000 thousand inhabitants
- Population 2003 thousand inhabitants
- Population 2006 thousand inhabitants
- Population 2009 thousand inhabitants

Thematic dimensions:

- CLC00 Hierarchical
- CLC06 Hierarchical
- CLC90 Hierarchical
- Land Cover Flows 1990-2000

- Land Cover Flows 2000-2006
- Land Cover Flows 2000-2006

Geographic dimensions:

- Nuts 1999 code
- Nuts 1999 name
- Nuts 2003 code
- Nuts 2003 name
- Nuts 2006 code
- Nuts 2006 name
- Nuts 2010 code
- Nuts 2010 name
- Functional Urban Areas (FUA)
- Morphological Urban Areas (MUA)
- Urban Morphological Zones (UMZ)
- Large Urban Zones (LUZ)

ESPON Urban OLAP Cube (version 2.0)

The last update of the Urban OLAP Cube was delivered in June 2014. It includes the following content:

Measures:

- Area in hectares
- Population 2006

Thematic dimensions:

- Urban Atlas 2006
- Urban Atlas 2012

Geographic dimensions:

- LAU2 2008
- LUZ 2004
- LUZ 2012
- FUA
- MUA
- UMZ 2000

ESPON Neighbourhood OLAP Cube (version 1.0)

This OLAP Cube was delivered in June 2014, with the input from UNEP/GRID (University of Geneva), who has processed Globcover and the socioeconomic variables for the entire neighbourhood space. Its contents are:

Measures:

- Population 2000
- GDP 2010
- Employment 2007

Thematic dimensions:

- GLobcover 2004
- Globcover 2009

Geographic dimensions:

- SNUTS 2012

Next table summarises the contents of the three updated OLAP Cubes, as main result for data integration through the grid, under ESPON M4D:

	Resolution	Geographic dimensions	Thematic dimensions	Measures
ESPON OLAP Cube (v6)	1 km ²	NUTS 1999, NUTS 2003, NUTS 2006, NUTS 2010, LUZ, FUA, MUA, UMZ	CLC00 Hierarchical, CLC06 Hierarchical, CLC90 Hierarchical, Land Cover Flows 1990-2000, Land Cover Flows 2000-2006, Land Cover Flows 2000-2006	Area in hectares, GDP 2000, GDP 2003, GDP 2006, GDP 2009, Population 2000, Population 2003, Population 2006, Population 2009
ESPON Urban OLAP Cube (v2)	100 m ²	LAU2 2008, LUZ 2004, LUZ 2012, FUA, MUA, UMZ 2000	Urban Atlas 2006	Area in hectares, Population 2006
ESPON Neighbourhood OLAP Cube (v1)	1 km ²	SNUTS 2012	GLobcover 2004, Globcover 2009	Population 2000, GDP 2010, Employment 2007

4.1 ESPON OLAP Cubes user-oriented scenarios

The objective of this section is showing some examples of particular applications or questions that can be solved by means of one of the ESPON OLAP Cubes that have been prepared under ESPON M4D.

User case #1 : Can we build up a time series between 2000 and 2009 on population and GDP reported by NUTS level 1 (2010) ?

Database to be used : ESPON OLAP Cube

Advantages : Using the Cube we overcome the problem of different NUTS breakdowns when dealing with long time series.

In this first case, we are going to use the OLAP Webtool. We make the following query :

OLAP CUBE VIEWER BETA

Spatial dimensions
NUTS 2010 LEVEL 1 ▾
All members

Thematic dimensions
+ Add new dimension ▾

Measures

Population 2000 (1000 inh.)	⊖
Population 2003 (1000 inh.)	⊖
Population 2006 (1000 inh.)	⊖
Population 2009 (1000 inh.)	⊖
GDP 2000 (in PPS)	⊖
GDP 2003 (in PPS)	⊖
GDP 2006 (in PPS)	⊖
GDP 2009 (in PPS)	⊖

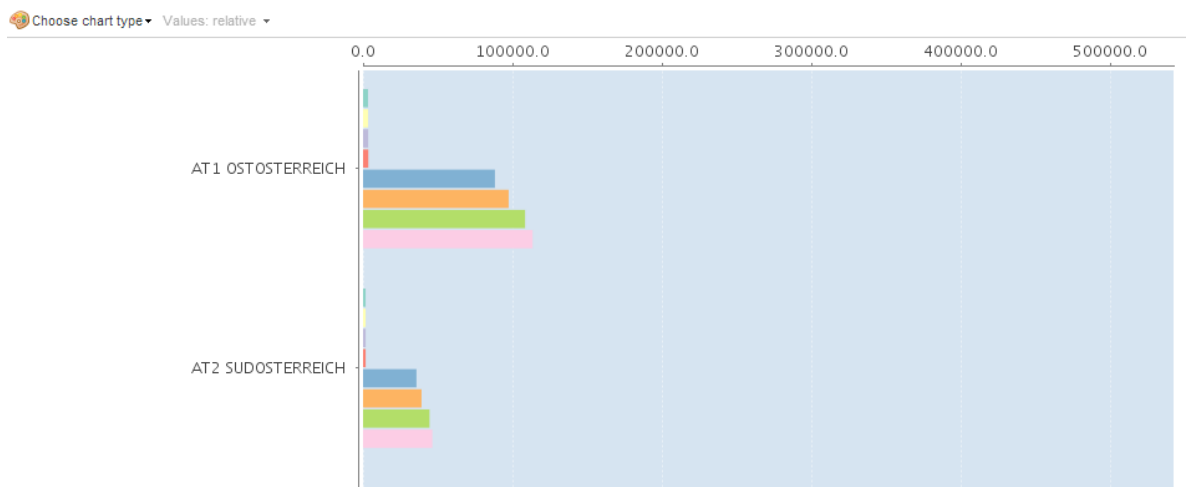
+ Add new measure ▾
 Relative values (%)

Show map
 Show table
 Show charts

Execute query

We have selected to get table and charts:

NUTS 2010 LEVEL 1	MeasuresLevel					GDP
	Population 2000 (1000 inh.)	Population 2003 (1000 inh.)	Population 2006 (1000 inh.)	Population 2009 (1000 inh.)		
AT1 OSTOSTERREICH	3304.0	3358.0	3452.0	3519.0		
AT2 SUDOSTERREICH	1702.0	1710.0	1725.0	1729.0		
AT3 WESTOSTERREICH	2831.0	2875.0	2916.0	2948.0		
BE1 REGION DE BRUXELLES-CAPITALE / BRUSSELS HOOFD*	956.0	987.0	1015.0	1064.0		
BE2 VLAAMS GEWEST	5708.0	5774.0	5857.0	5982.0		
BE3 REGION WALLONNE	3203.0	3228.0	3278.0	3333.0		
BG3 SEVERNA I IZTOCHNA BULGARIA	4304.0	4091.0	3975.0	3880.0		
BG4 YUGOZAPADNA I YUZHNA CENTRALNA BULGARIA	3777.0	3659.0	3634.0	3614.0		
CH0 SCHWEIZ/SUISSE/SVIZZERA	6936.0	7087.0	7239.0	7473.0		
CY0 KYPROS / KIBRIS	0.0	0.0	0.0	0.0		
CZ0 CESKA REPUBLIKA	10029.0	9959.0	10001.0	10200.0		
DE1 BADEN-WURTTENBERG	10243.0	10396.0	10490.0	10512.0		
DE2 BAYERN	11798.0	12013.0	12116.0	12155.0		
DE3 BERLIN	3399.0	3399.0	3404.0	3442.0		
DE4 BRANDENBURG	2465.0	2458.0	2430.0	2391.0		
DE5 BREMEN	653.0	651.0	653.0	653.0		



All the time series are complete:

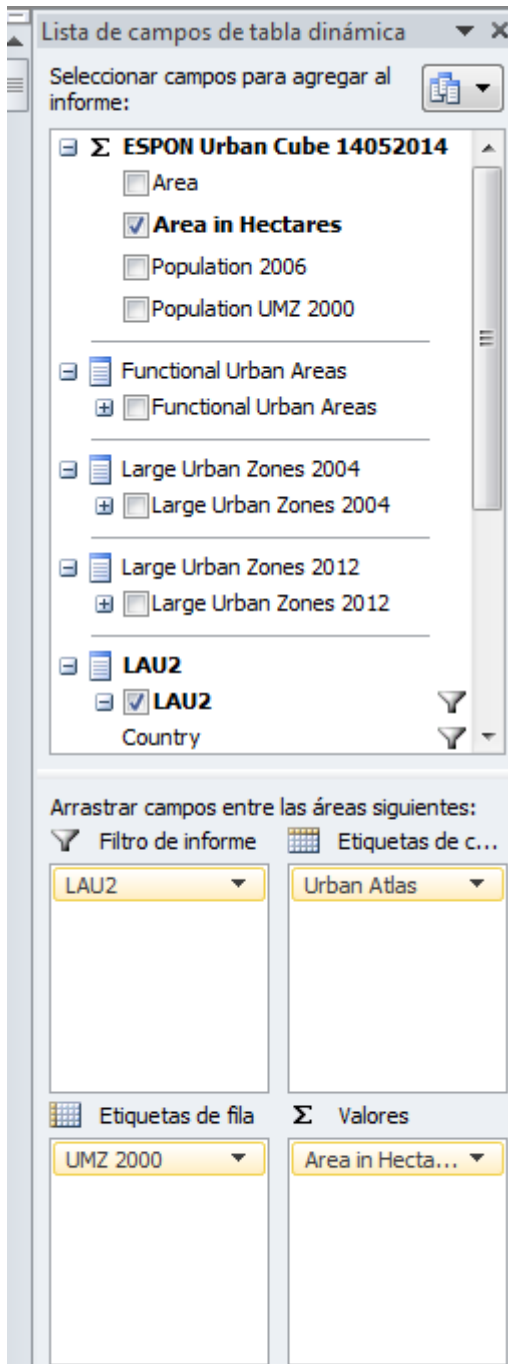
MeasuresLevel							
Population 2003 (1000 inh.)	Population 2006 (1000 inh.)	Population 2009 (1000 inh.)	GDP 2000 (in PPS)	GDP 2003 (in PPS)	GDP 2006 (in PPS)	GDP 2009 (in PPS)	
3358.0	3452.0	3519.0	88189.0	97368.0	108381.0	113380.0	
1710.0	1725.0	1729.0	35703.0	39081.0	44428.0	46344.0	
2875.0	2916.0	2948.0	71248.0	78813.0	90002.0	95476.0	
987.0	1015.0	1064.0	45674.0	51308.0	55195.0	57784.0	
5774.0	5857.0	5982.0	137984.0	153652.0	169816.0	177327.0	
3228.0	3278.0	3333.0	56408.0	62330.0	68530.0	72916.0	
4091.0	3975.0	3880.0	20938.0	24392.0	28775.0	30698.0	
3659.0	3634.0	3614.0	22395.0	29729.0	40751.0	49463.0	
7087.0	7239.0	7473.0	190498.0	206573.0	236734.0	278183.0	
0.0	0.0	0.0	0.0	0.0	0.0	0.0	
9959.0	10001.0	10200.0	133896.0	155536.0	188552.0	206024.0	
10396.0	10490.0	10512.0	267060.0	290974.0	330631.0	335366.0	
12013.0	12116.0	12155.0	319247.0	353627.0	403619.0	399188.0	
3399.0	3404.0	3442.0	70278.0	72305.0	80641.0	91984.0	
2458.0	2430.0	2391.0	39767.0	43185.0	49193.0	49772.0	
651.0	653.0	653.0	19657.0	21831.0	24982.0	24964.0	

User case #2 : Which is the Land Cover (Urban Atlas) of the UMZ in France ?

Database to be used : ESPON Urban OLAP Cube

Advantages : OLAP Cubes facilitate interoperability between the different urban delineations.

By means of Excel, we make the following query :



Result of the query :

	A	B	C	D	E	F	
1	LAU2	FR					
2							
3	Area in Hectares	Etiquetas d					
4	Etiquetas de fila	Agricultural +	Airports	Construction	Continuous Urban F.	Discontinuous Dense Urban F.	Disconti
5	Agen						
6	Aix En Provence	227,29	60	26,57	108,44	628,07	
7	Albi						
8	Alès						
9	Amiens	277,62	81,14	23,09	321,03	637,54	
10	Angers						
11	Angoulême						
12	Annecy						
13	Armentières	88,95		8,37	100,67	269,52	
14	Arras						
15	Avignon						
16	Basel French Part						
17	Bayonne // Anglet // Biarritz						
18	Beauvais						
19	Belfort						
20	Besançon	409,53		37,02	364,14	920,13	
21	Bethune // Noeux Les Mines	3,54				0,55	
22	Beziers						
23	Blois						
24	Bordeaux	1611,67	676,46	52,96	2189,8	7542,85	
25	Boulogne Sur Mer						
26	Bourg En Bresse						
27	Bourges						
28	Brest						
29	Brive La Gaillarde						
30	Bruay La Buissière // Auchel						
31	Caen	820,24	220,66	51,36	1206,62	1327,81	
32	Calais						
33	Chalon Sur Saone						

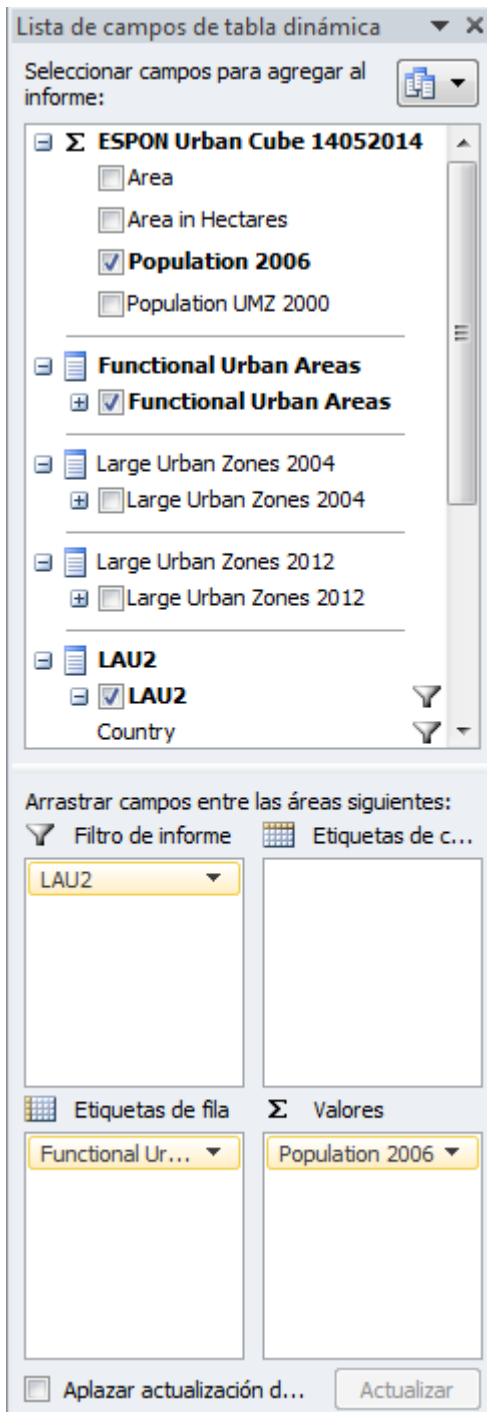
We get the area of Urban Atlas land cover classes in hectares for each French UMZ (blank lines mean that UA is not covering such UMZs).

User case #3 : Which was the population in 2006 for each Functional Urban Area in the UK?

Database to be used : ESPON Urban OLAP Cube

Advantages : OLAP Cubes facilitates populating databases with external datasets (in this case, population is added to FUAs)

By means of Excel, we make the following query :



Result of the query:

	A	B
1	LAU2	UK
2		
3	Etiquetas de fila	Population 2006
4	Aberdeen	305564
5	Alloa - Stirling	74931
6	Ashford	100341
7	Aylesbury	112099
8	Ayr	55120
9	Ballymena	68799
10	Banbury	95262
11	Bangor	49418
12	Barrow-in-Furness	31741
13	Bedford	176077
14	Belfast	483341
15	Birmingham metropolitan area	1958639
16	Blackburn/Burnley	426383
17	Blackpool	292972
18	Boston	64826
19	Bournemouth/Poole	504427
20	Braintree	62892
21	Bridlington	10895
22	Brighton/Worthing/Littlehampton	540465
23	Bristol metropolitan area	642070
24	Burton on Trent	130519
25	Bury St.Edmonds	85294
26	Cambridge	208732
27	Canterbury	149637
28	Cardiff and South Wales valleys metropolitan area	762836
29	Carlisle	126306
30	Cheltenham	159607
31	Chichester	73488
32	Clacton	82100
33	Colchester	206978
34	Coleraine	72273
35	Colwyn Bay	15830
36	Crewe/Nantwich	121976

We get the population figures (2006) for each FUA in the UK.

User case #4 : Which is the population and the area of artificial surfaces in 2009 by each SNUTS?

Database to be used : Neighbourhood OLAP Cube

Advantages : OLAP Cubes facilitate interoperability between land cover (in this case Globcover) and socioeconomic data.

By means of Excel, we make the following query :

The screenshot displays a query builder interface with the following sections:

- Land Cover Categories:** A list of categories with checkboxes. The selected category is **GC2009 - Artificial surfaces and associated areas - Urban areas more 50 percent**.
- Seleccionar campos para agregar al informe:** A pane for selecting fields. The selected fields are **Population in inhabitants** (under NUTS) and **SNUTS Code** (under SNUTS).
- Arrastrar campos entre las áreas siguientes:** A pivot table configuration area with the following settings:
 - Filtro de informe:** Empty.
 - Etiquetas de columna:** Σ Valores.
 - Etiquetas de fila:** SNUTS Code.
 - Valores:** Population in inhabitants, GC2009 - Artificial surfaces and associated are...

Result of the query:

	A	B	C
1	Etiquetas de fila	Population in inhabitants	GC2009 - Artificial surfaces and associated areas - Urban areas more 50 percent
2	AL001	201742	363,7661062
3	AL002	196914	42,86723
4	AL003	255957	206,217485
5	AL004	365508	1502,189265
6	AL005	361422	1801,847787
7	AL006	124428	1502,975624
8	AL007	271148	1431,787134
9	AL008	116578	675,1033823
10	AL009	142346	258,8282174
11	AL00A	254495	2248,225246
12	AL00B	559570	441,7417422
13	AL00C	191830	933,2587528
14	AM00	3079160	36046,57109
15	AZ00	8070218	226151,1403
16	BA011	57468	12921,41746
17	BA021	19656	4369,987095
18	BA022	320999	122155,2033
19	BA023	323571	27978,19929
20	BA024	89791	20645,95621
21	BA025	348815	48126,67331
22	BA026	242796	76004,07153
23	BA027	359839	28252,86521
24	BA028	303836	63132,97513
25	BA029	243427	81069,17963
26	BA02A	42753	19806,0567
27	BA031	603981	211485,7917
28	BA032	135176	50262,04329
29	BA033	89290	20043,15616
30	BA034	87416	58331,65727
31	BA035	218801	30284,51158
32	BA036	144359	66390,96151
33	BA037	202071	57056,22933
34	BY10	1783992	1740,868284
35	BY11	1450936	27076,83659
36	BY20	1524207	3579,522675
37	BY21	1179405	5499,882493
38	BY22	1479095	2408,710093
39	BY23	1379950	5061,200389
40	BY24	1196596	4705,348692
41	DZ101	1126524	1857,545841
42	DZ102	1345391	3851,62425

We quickly get the population figures and hectares of artificial surfaces.

5 ESPON OLAP Webtool

Although querying OLAP Cubes with MS Excel is not complicated, once you get the result of the query producing charts or maps out of it is not straight-forward and specific knowledge is needed. Therefore, the ESPON OLAP Webtool was created with the aim of facilitating the querying of OLAP Cubes for the different users and producing maps and charts in an easy way.

Currently, the tool can be used through this link : <http://158.109.174.100/webbi/>

The webtool has had different improvements since it was launched. Although it currently allows the user to query only the ESPON OLAP Cube (ESPON space one), it is prepared for the possibility to choose amongst different potential OLAP Cubes :

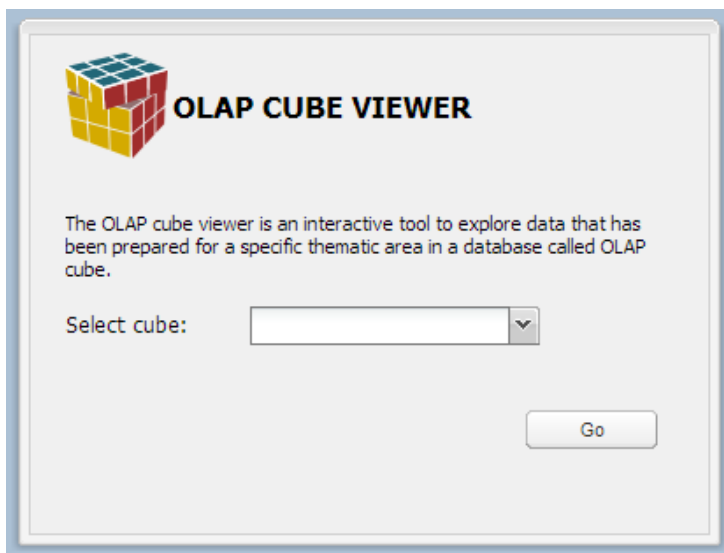


Figure 12. Window to select the OLAP Cube within the ESPON OLAP Cube Viewer

The web tool is used to query an OLAP Cube through a simple web interface, by choosing the data dimensions and measures the user wants to see:

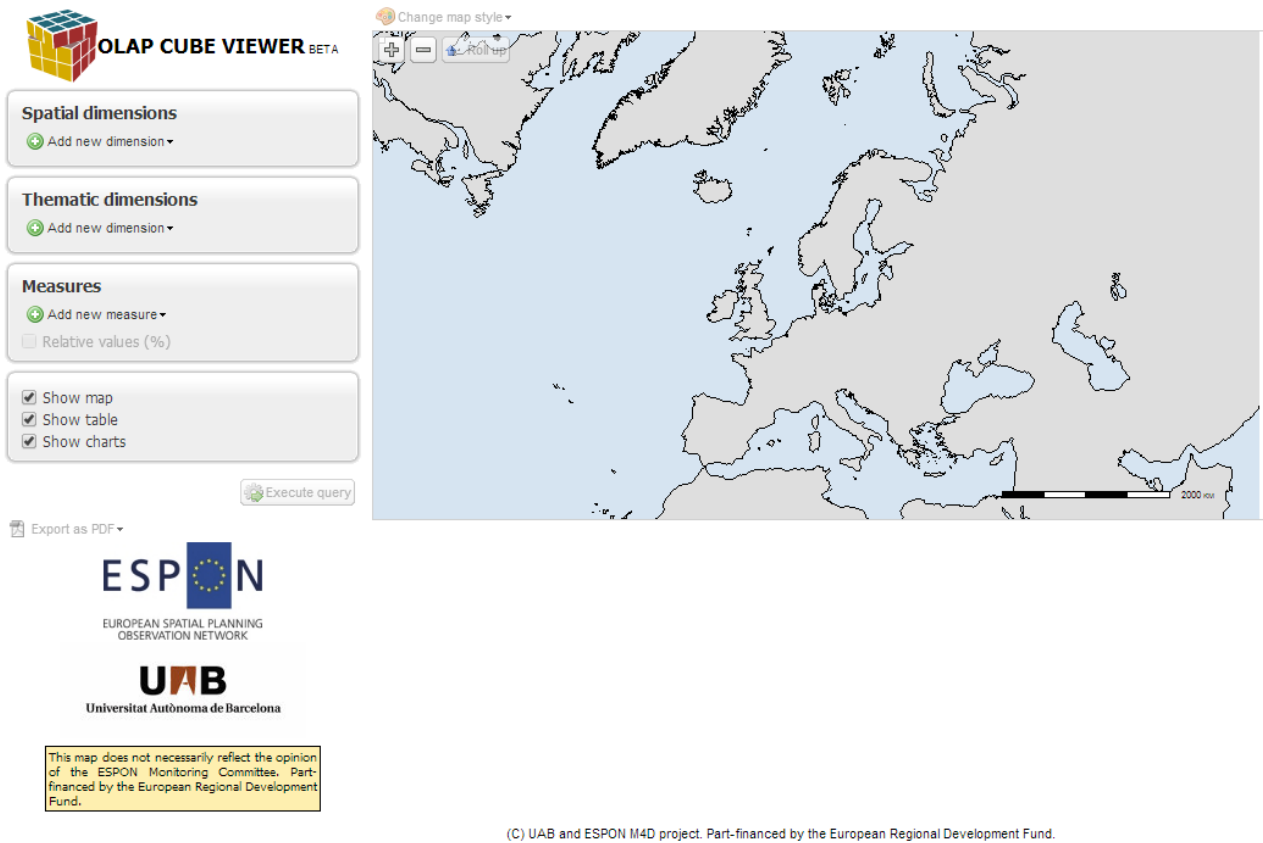


Figure 13. Screenshot of the OLAP web tool homepage

The user only has to select a spatial dimension, a thematic one and a measure (socioeconomic variable) to produce a query. He can also choose which outputs will be obtained after the query is launched (map, table and/or charts).

The selection of a spatial dimension can be refined by selecting just some elements :

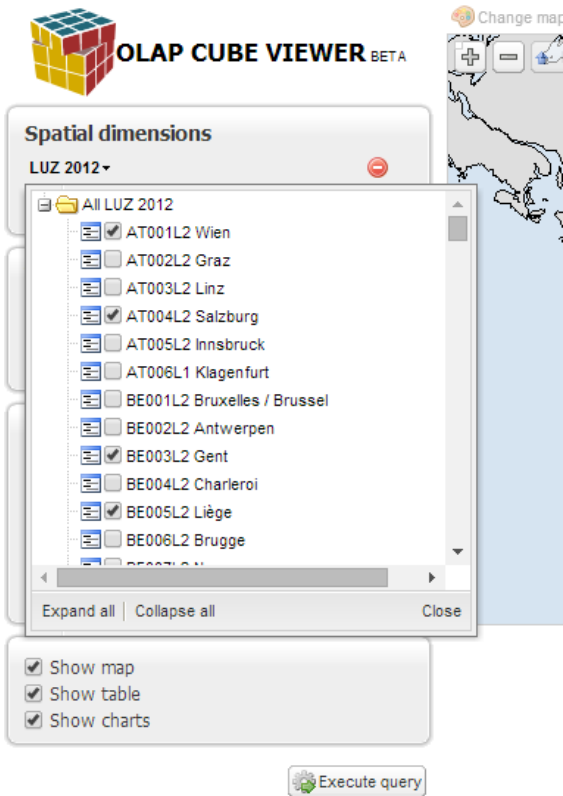


Figure 14. Subselection of spatial dimensions in the webtool

Results are shown at the same time in a table, chart and/or map, depending on user's selection:

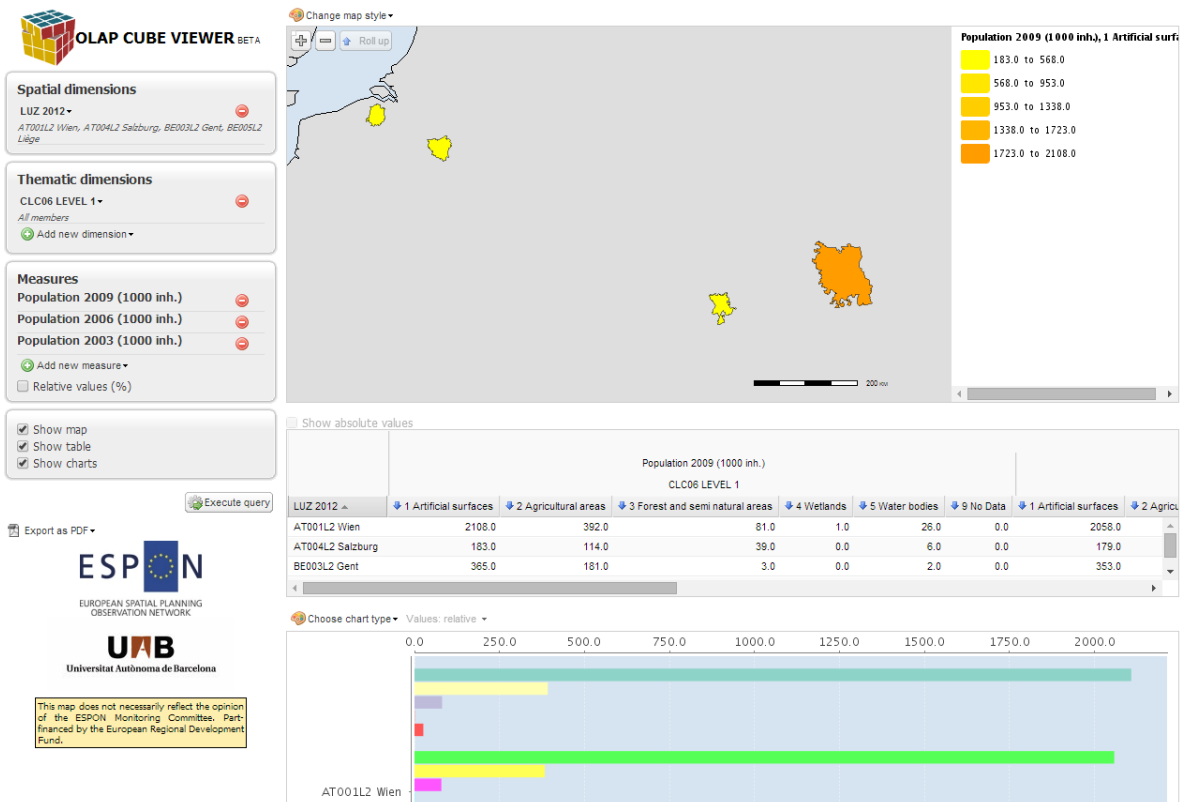


Figure 15. Map, table and charts displayed after a query

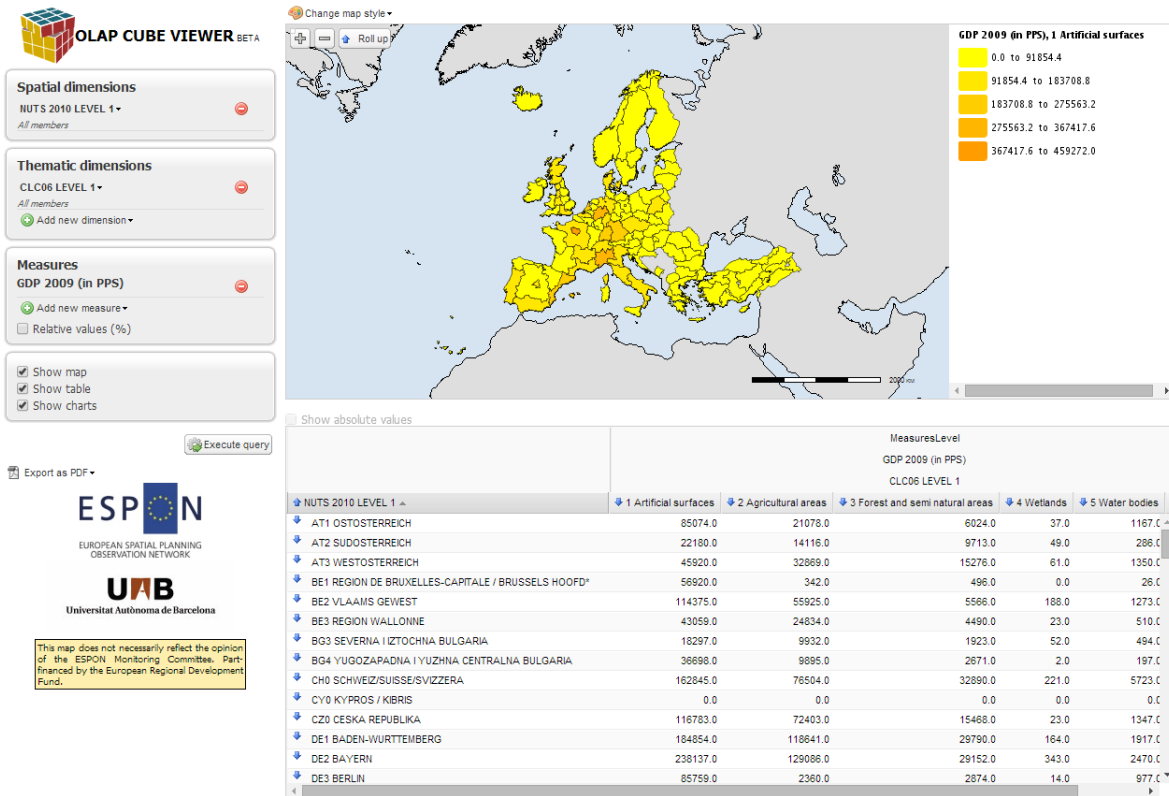


Figure 16. Screenshots of the OLAP web tool results after a query



Figure 17. Example of output from the webtool without map

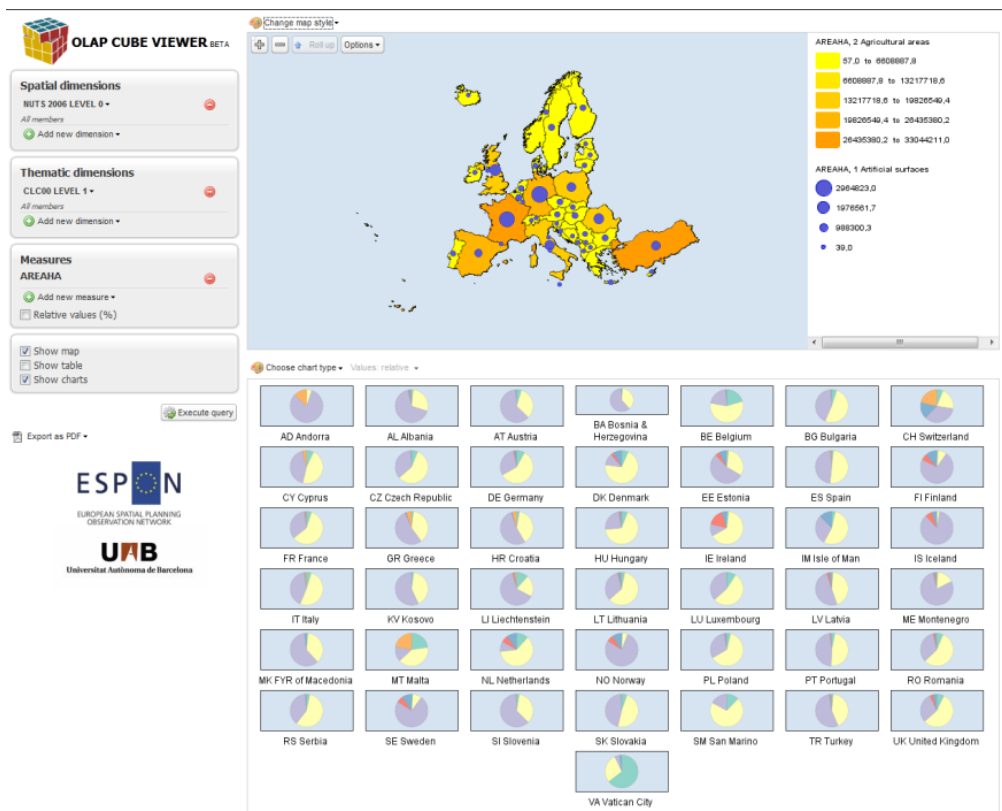


Figure 18. Example of output from the webtool without table

The results obtained via web tool can be exported as PDF files.

Technology of the OLAP webtool

The cube viewer has been built using a number of free software technologies, making possible to re-use the application for different thematic areas. This approach also encourages improvements and customizations from any interested party. In this way, the tool has filled an important gap in the software field, as no comparable open source tool is currently available in the market.

The main technologies used on the project are:

- PostgreSQL + PostGIS, as spatially-enabled relational database management system.
- GeoMondrian and Spatial-OLAP server.
- OpenLayers javascript library in order to provide map visualization and navigation.
- ExtJS and GeoExt in order to build the user interface.
- GeoTools and JTS for geographic data processing and delivering.
- Olap4j, which provides a uniform programming access (API) to different OLAP servers
- JFreeChart, used to generate charts.

Note that all those applications and libraries are open source technologies.

6 Conclusion and further developments

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

Furthermore, this methodology overcomes several classical problems, like working with different NUTS breakdowns or the combination of thematic with administrative and socioeconomic data. It should be stated that the purpose of this method it is not getting detailed disaggregated datasets, but using disaggregation as an intermediate process to combine data, and being able to query them to get sound results in a fast way that might help to analyse trends or facts regarding territorial and social development. Nevertheless, the better disaggregated datasets are built up in Europe, or the better disaggregation techniques can be applied, the better will be the OLAP database produced.

Some key conclusions are summarised in the next bullet points:

- 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:
 - It has an European coverage
 - It follows Inspire specifications
 - It is used for several institutions as the reference grid
 - 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".
- The "proportional and weighted" aggregation method is the one that gives better results, plus some added value to the downscaling. Using disaggregated population as ancillary variable is a good approach to disaggregate socioeconomic figures.
- The different methods are independent from the source data format and can be applied to vector and raster format.
- The OLAP methodology helps to overcome the problem of managing different NUTS breakdowns, as everything is referred to the grid, which never changes over time.
- This methodology allows the integration of socio-economic, thematic and geographic data in a unique database, which facilitates the querying and analysis of such data together.

- 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.
- The ESPON OLAP Webtool facilitates the usage of OLAP Cubes for non-expert users.

There are several improvements and future implementations which are possible in order to make this methodological approach and derived products and tools improve and 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 divisions and Corine Land Cover (e.g. different coastlines).
 - Improvement of the disaggregation performance in terms of time and manageability of the final layer.
 - General improvements on performance, especially for large layers when working at high resolution.
 - Calculation of standard errors introduced by the disaggregation methods.
- b) Follow-up of the EFGS: It might be very useful to keep following up the outcomes of the European Forum for Geostatistics¹¹ in order to contrast the proposed method and upcoming gridded datasets. For instance, a new population grid 2012 is expected to be available.
- c) Usage of new datasets: the recent availability of high quality datasets, such as the population historical time-series database at LAU2 level, produced by DG Regio will definitely be a step forward on data integration at local level, where the OLAP methods defined in this technical report may play an important role.
- d) Integration within the ESPON Database: In the future, a major integration of the different ESPON tools is expected. It should be well analysed the most suitable solution to integrate grid data in the ESPON toolbox and to make the ESPON OLAP Cube available to users through a common ESPON portal or interface. The possible integration of the ESPON OLAP Webtool within the ESPON toolbox should also be discussed and further developed in the future.

¹¹ <http://www.efgs.info/>

Annex 1 – ESPON OLAP Cube User Manual

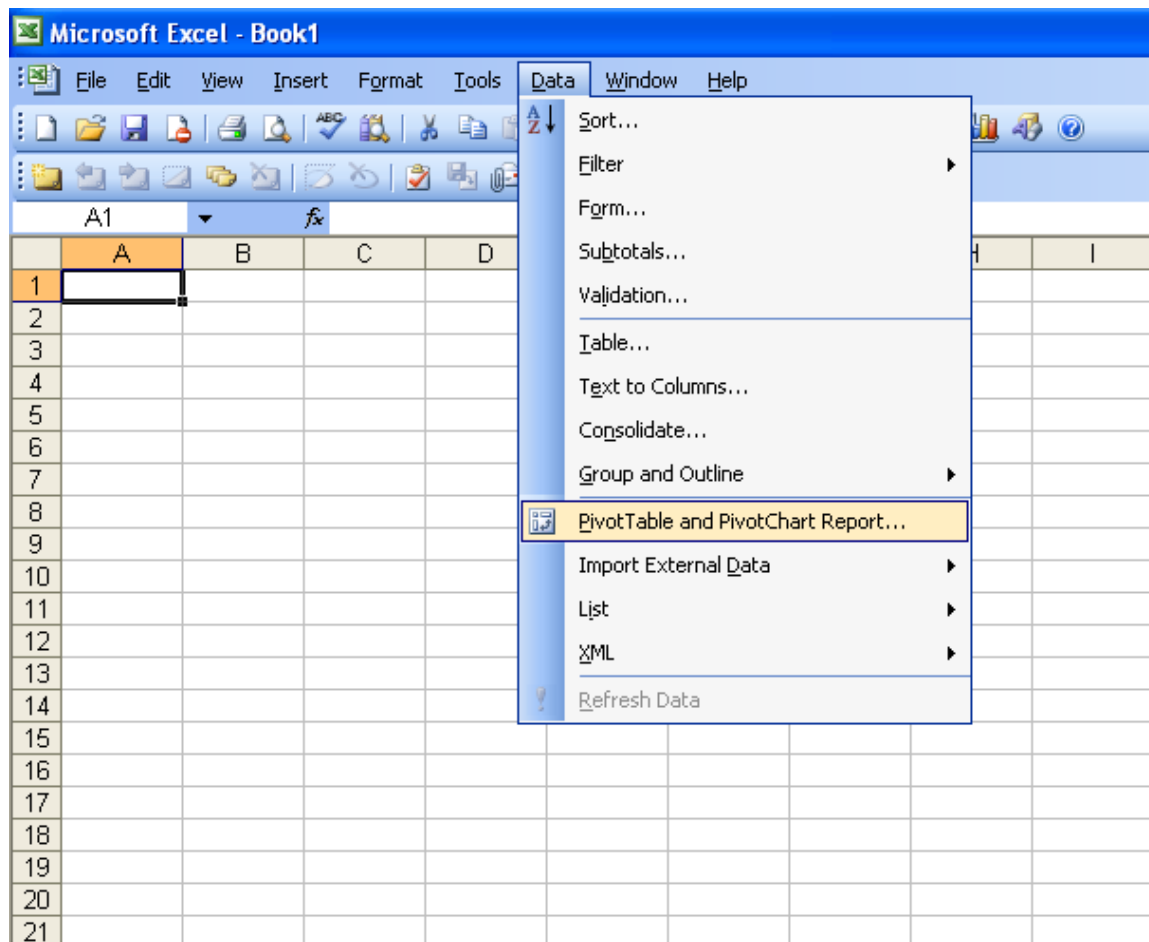
This Annex is aimed at explaining how to query the ESPON OLAP Cube file by means of the default PivotTable application included in MS Excel 2003 and 2010.

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:

MICROSOFT EXCEL 2003

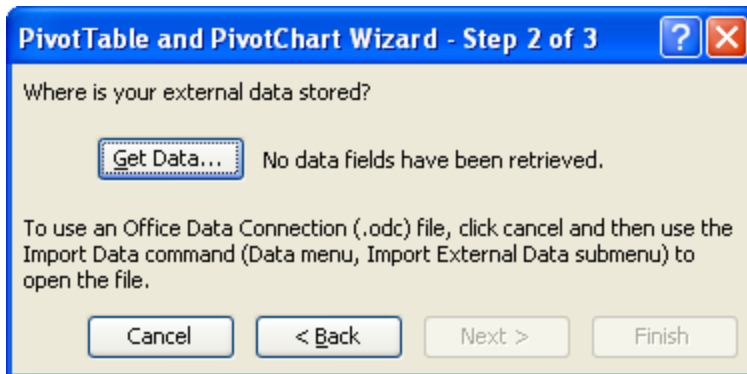
- Select "Pivot Table..." in the Data Menu



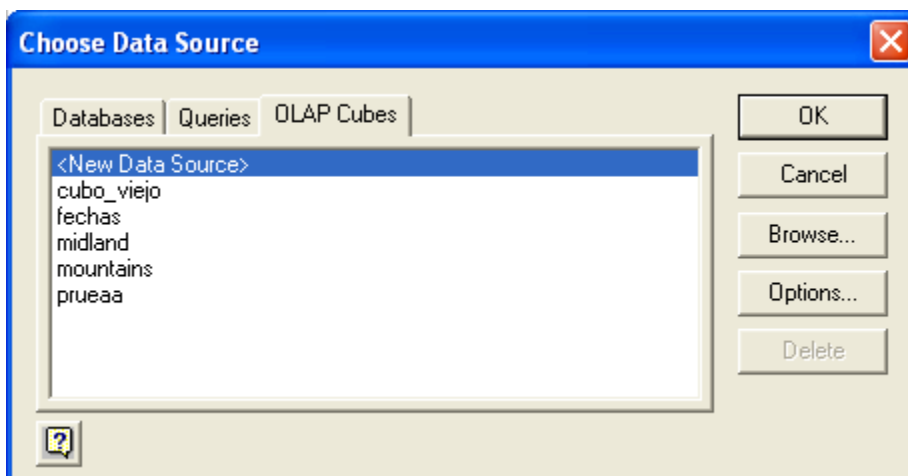
- Select “External data source”¹² and Pivot Table as report type



- Click on “Get Data...”

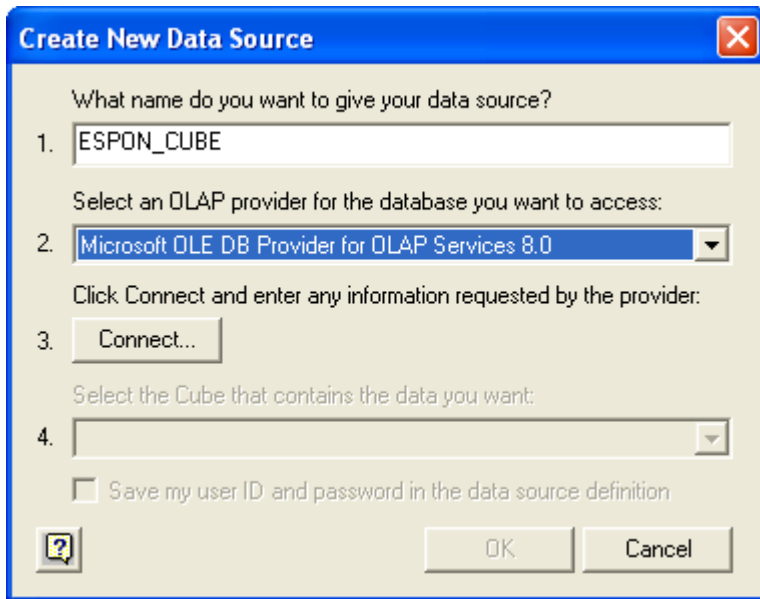


- Go to “OLAP Cubes” Tab. Choose <New Data Source>. Click OK

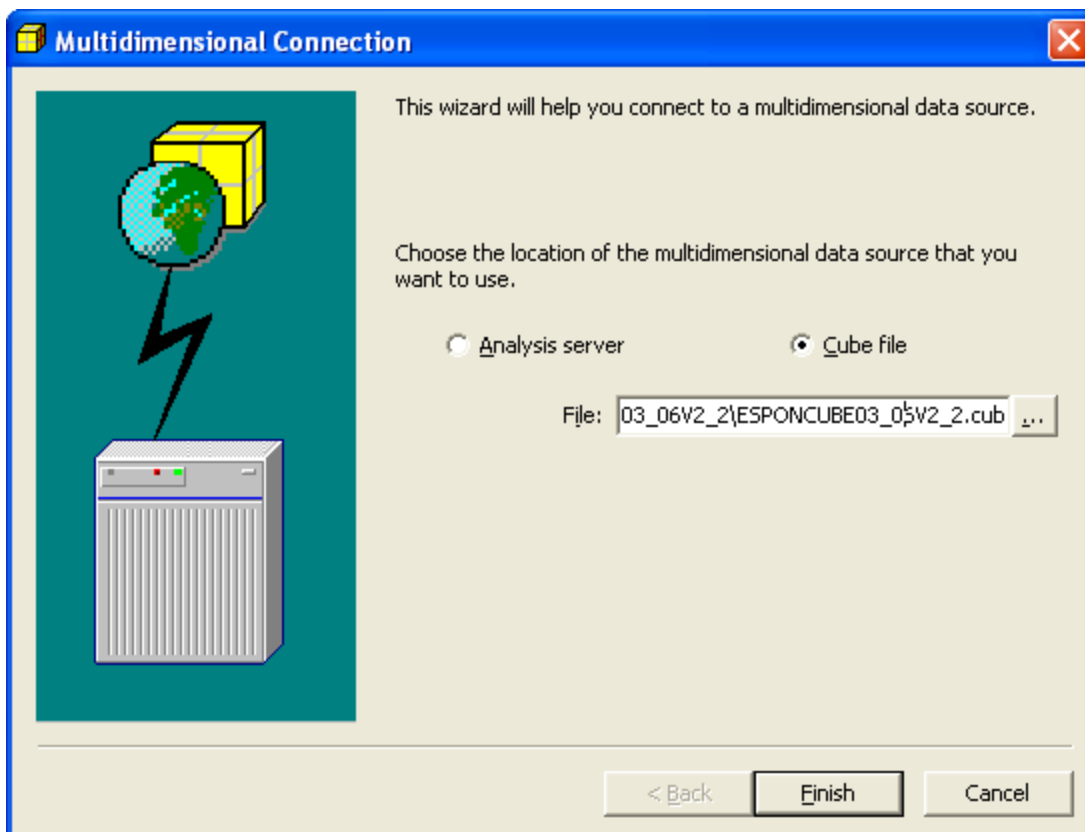


¹² In this case, “external data source” means that the data we are going to build the pivot table from is not coming from an Excel sheet, but from an “external” data source. In our case, this external data source is the .CUB file, i.e. the ESPON OLAP Cube.

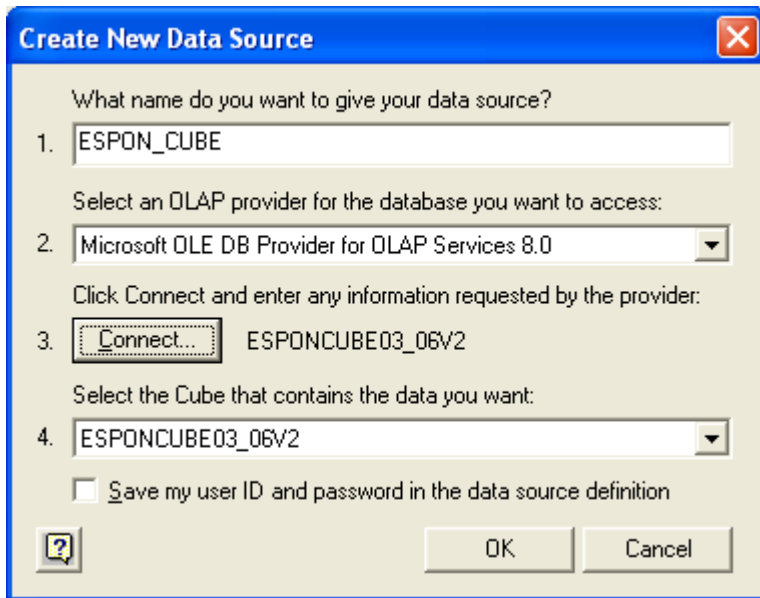
- Write down a name for your connection. Choose "MS OLE DB Provider for OLAP Services 8.0 (note: this component should be installed in order to connect to an OLAP Cube). Click Connect... button.



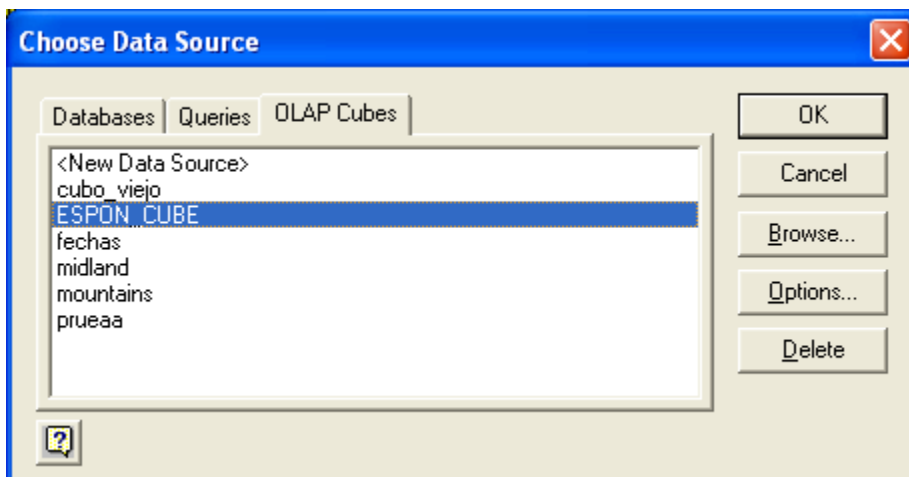
- Choose Cube file, and browse and choose the .cub file in your computer. Click Finish.



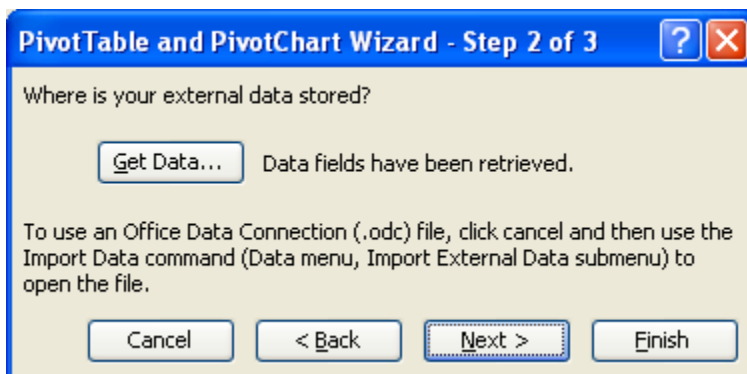
- Click OK



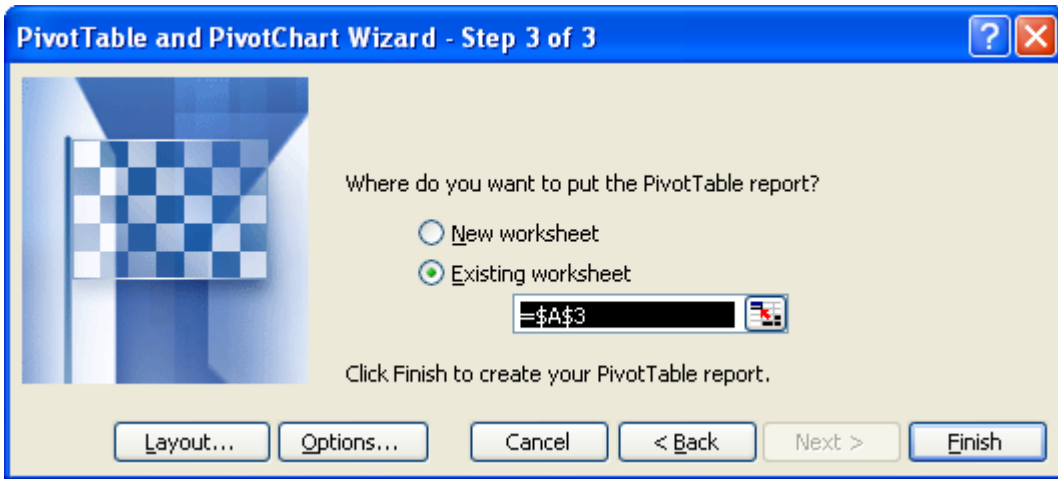
- Choose the connection just created. Click OK



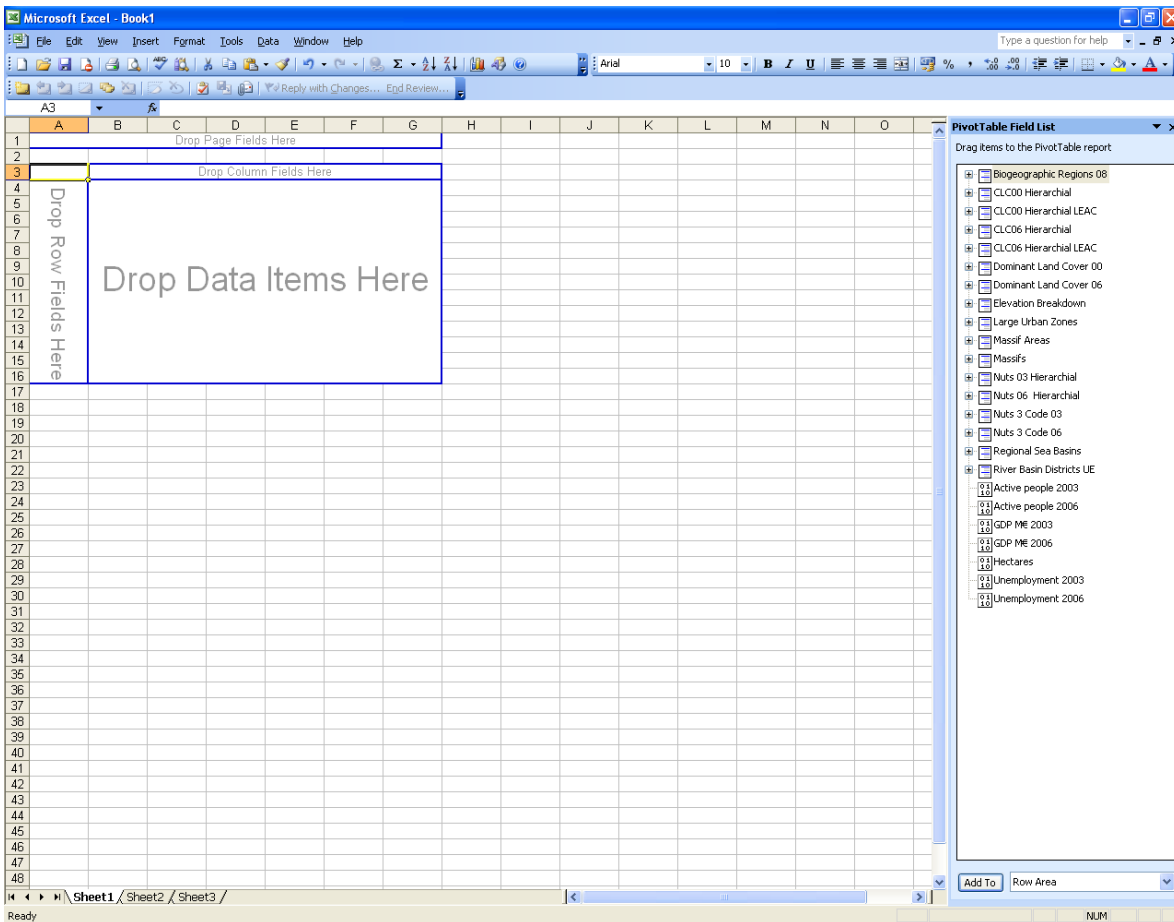
- Click "Next >"



- Choose either a new or existing worksheet. Click Finish.

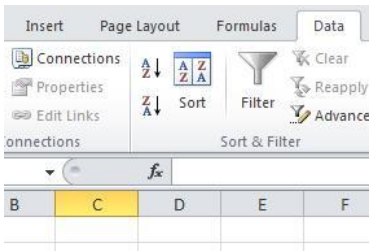


- The Cube is ready!

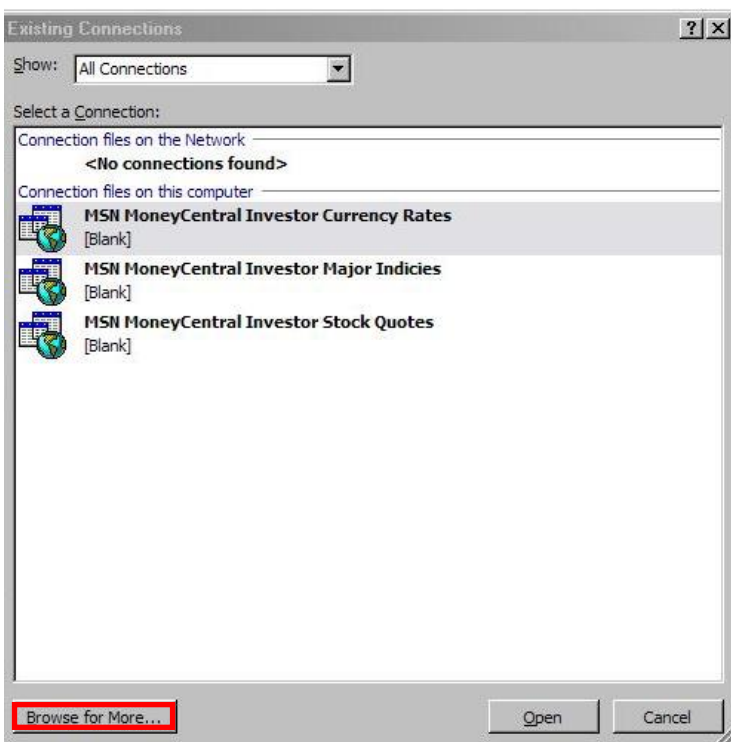
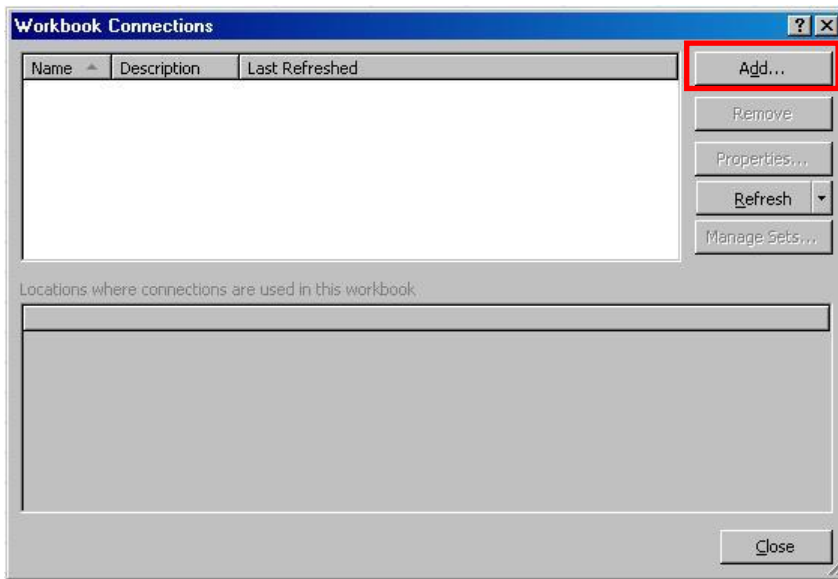


MICROSOFT EXCEL 2010

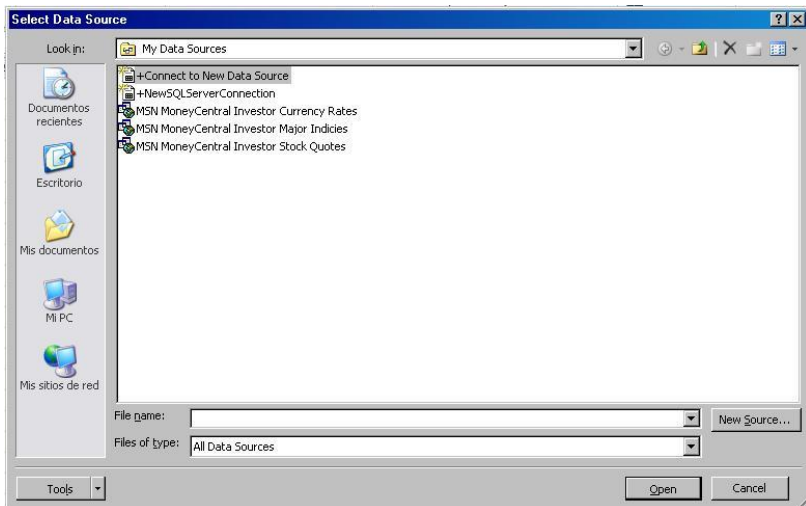
Select "Connections" in the Data Menu



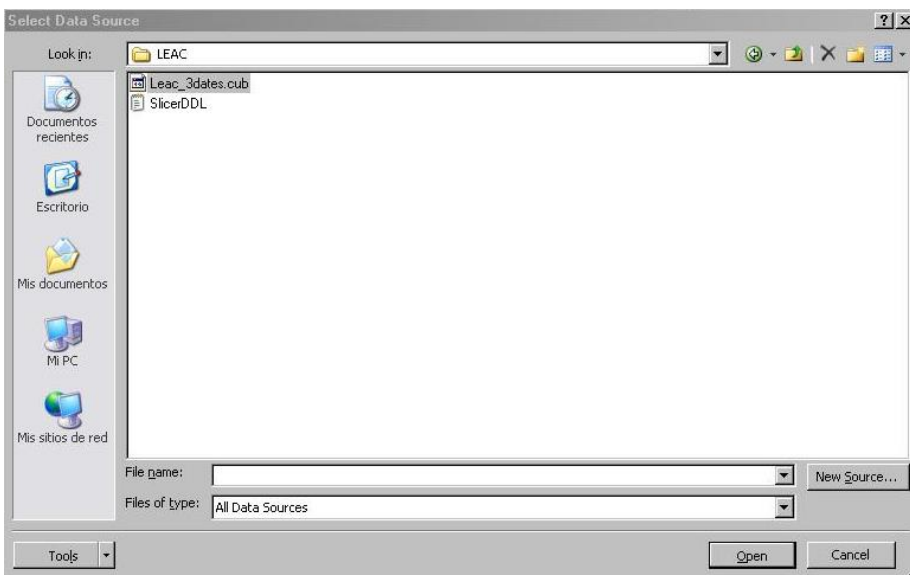
Then, select "Add" and in the next window "Browse for More..."



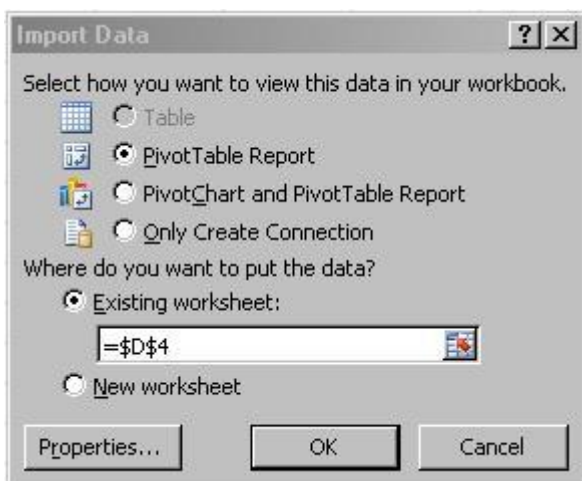
In the window search for the folder where you have your .cub file...



Select it and click "Open"



Select how you want to view the data.
The default is as Pivot Table Report



The Cube is ready to be used!

The screenshot displays an Excel spreadsheet with a PivotTable field list on the right side. The spreadsheet grid shows columns A through N and rows 1 through 40. A callout box is positioned over the PivotTable area, containing the text "PivotTable2" and "To build a report, choose fields from the PivotTable Field List". The field list on the right is titled "PivotTable Field List" and includes a section "Choose fields to add to report:" with a list of fields: Values (Areaah), Biogeographic Regions 08 (Biogeographic Regions 08), Change 1990_2000 (Change 1990_2000), Change 1990_2006 (Change 1990_2006), Change 2000_2006 (Change 2000_2006), City Names (City Names), CLC00 Hierachial LEAC (CLC00 Hierachial LEAC), and CLC00 Hierachial (CLC00 Hierachial). Below this is a section "Drag fields between areas below:" with four empty boxes for Report Filter, Column Labels, Row Labels, and Values. At the bottom of the field list are checkboxes for "Defer Layout Update" and an "Update" button.

How to query the Cube

- Drag & Drop dimensions (e.g. Nuts 3 Code 06) in the Row Area

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Type a question for help

Drop Page Fields Here

Drop Column Fields Here

Drop Data Items Here

PivotTable Field List

Drag items to the PivotTable report

- Biogeographic Regions 08
- CLC00 Hierarchical
- CLC00 Hierarchical LEAC
- CLC06 Hierarchical
- CLC06 Hierarchical LEAC
- Dominant Land Cover 00
- Dominant Land Cover 06
- Elevation Breakdown
- Large Urban Zones
- Massif Areas
- Massifs
- Nuts 03 Hierarchical
- Nuts 06 Hierarchical
- Nuts 3 Code 03
- Nuts 3 Code 06**
- Regional Sea Basins
- River Basin Districts LE
- Active people 2003
- Active people 2006
- GDP ME 2003
- GDP ME 2006
- Hectares
- Unemployment 2003
- Unemployment 2006

Add To Row Area

Ready

- Drag & Drop measures (e.g. GDP ME 06) in the Data Area

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Type a question for help

GDP ME 2006

Drop Page Fields Here

Drop Column Fields Here

Drop Data Items Here

PivotTable Field List

Drag items to the PivotTable report

- Biogeographic Regions 08
- CLC00 Hierarchical
- CLC00 Hierarchical LEAC
- CLC06 Hierarchical
- CLC06 Hierarchical LEAC
- Dominant Land Cover 00
- Dominant Land Cover 06
- Elevation Breakdown
- Large Urban Zones
- Massif Areas
- Massifs
- Nuts 03 Hierarchical
- Nuts 06 Hierarchical
- Nuts 3 Code 03
- Nuts 3 Code 06**
- Regional Sea Basins
- River Basin Districts LE
- Active people 2003
- Active people 2006
- GDP ME 2003
- GDP ME 2006**
- Hectares
- Unemployment 2003
- Unemployment 2006

Add To Data Area

Nuts 3 Code 06	Total
AD???	0,247751848
AL???	0,465693633
AT111	728,1931807
AT112	3193,773684
AT113	1770,642809
AT121	5713,890087
AT122	5451,836716
AT123	4243,001456
AT124	4732,672501
AT125	1896,205201
AT126	6112,046373
AT127	11524,58537
AT130	63483,5594
AT1211	8336,686231
AT1212	2819,554547
AT1213	3623,136297
AT1221	13133,397441
AT1222	2040,074413
AT1223	4728,079903
AT1224	5342,382776
AT1225	3869,352016
AT1226	2434,436218
AT1311	6298,746306
AT1312	21018,23391
AT1313	3483,482771
AT1314	4646,092557
AT1315	6150,905421
AT1321	493,405396
AT1322	4655,104042
AT1323	12759,38914
AT1331	1117,082467
AT1332	9166,609866
AT1333	1153,767175
AT1334	2950,55961
AT1335	7492,388255
AT1341	2947,880575
AT1342	8252,260815
BA???	0
BE100	56130,90852
BE211	36432,06288
BE212	9928,114352
BE213	12205,411349
BE221	11591,85324
BE222	4739,696783

Ready

- Drag & Drop other measures (e.g. CLC06 hierarchical Level 1) in the Column Area

The screenshot shows an Excel 2006 window with a PivotTable report. The PivotTable is located in the 'Drop Page Fields Here' area and displays data for 'GDP ME 2006'. The PivotTable has five columns representing different categories: 1. Artificial surfaces, 2. Agricultural areas, 3. Forest and semi natural areas, 4. Wetlands, and 5. Water bodies. The PivotTable Field List on the right shows the hierarchy of fields, including 'Level 1' under 'CLC06 Hierarchial'.

In Excel 2010, you drag to the areas below the fields list:

The screenshot shows an Excel 2010 window with a PivotTable report. The PivotTable is located in the 'Report Filter' area and displays data for 'Areaha' across various categories. The PivotTable Field List on the right shows the hierarchy of fields, including 'Areaha' under 'Change 2000_2006'. The 'Report Filter' and 'Column Labels' sections are highlighted with a red circle.

In case of using Microsoft Excel 2010 64 bits, you should use the specific Cube created for MS Excel 2010, and simply drag & drop the .cub file to an Excel spreadsheet.

Bibliography

- 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

- 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. <http://www.efgs.info/>

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 Commission:

<http://ec.europa.eu/dgs/jrc/index.cfm>

- OLAP

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

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