ESPON
EUROPEAN SPATIAL PLANNING OBSERVATION NETWORK

NUTS 5 Database User Manual

December 2004

Erik Gløersen, Christian Lindner (IRPUD)

Compilation of initial
DG REGIO Mountain Study database:
Jörg Grimm (IRPUD), Alexandra Hill (IRPUD),
Michael Löchl (IRPUD),
Carsten Schürmann (RRG)

NORDREGIO
Nordic Centre for Spatial Development

(a) Table of contents

(b) Figures

(c) Tables

1. Introduction

1.1 Background

As part of the DG REGIO Mountain Study, a NUTS 5 database was established, covering all municipalities (i.e. both mountainous and non-mountainous ones) in the 23 European countries where mountain areas had been identified. After the delivery of the Mountain study, ESPON was granted permission to use the resulting datasets, which belong to the European Commission – DG REGIO. ESPON then financed an extension of this dataset to the remaining, non-mountainous countries for a few core indicators (concerning demography and employment see list below). These non-mountainous countries specifically concerned by the present delivery are

- Estonia,
- Latvia,
- Lithuania,
- Denmark,
- the Netherlands,
- Malta.

All other ESPON countries were covered by the original Mountain study database.

The present delivery contains a pan-European socio-economic database at NUTS 5-level. While it demonstrates that NUTS 5 data are available in a wide range of countries, there are still considerable efforts to be accomplished comparing the indicator definitions in different countries, increasing the comparability of data from country to country, and testing the database in order to identify errors and omissions. The database must be seen as the first stage within the process of compiling NUTS 5 data for the ESPON space.

The original Mountain study data gathering exercise was coordinated by Erik Gløersen (Nordregio) and Michael Löchl, Christian Lindner, Jörg Grimm, Alexandra Hill and Carsten Schürmann (IRPUD). It was carried out with the help of a wide network of national correspondents:

- Federal Institute for Less-Favoured and Mountain Regions, BABF (Austria, Germany);
- Société d'études géographiques, économiques et sociologiques appliquées – SEGESA (France);
- Ph de Boe consultants (Belgium, Luxembourg, the Netherlands);
- Nordregio (Denmark, Sweden, Finland, Estonia, Latvia, Lithuania, Malta);
- National Agricultural Research Foundation, NAGREF (Greece);

- Scottish Agricultural College (SAC) (Ireland, United Kingdom);
- CENSIS (Italy);
- Quaternaire (Portugal);
- Multicriteria Consulting, MCRIT (Spain);
- National Centre for Regional Development (Bulgaria);
- Agricultural Research Institute (Cyprus);
- University of South Bohemia (Czech Republic);
- Regional Environmental Centre for Central and Eastern Europe (Hungary);
- Polish Society of the Development of Mountain Areas (Poland);
- Training and Innovation Centre for Development in the Carpathians (Romania);
- Slovak Agricultural University (Slovakia);
- Oikos Ltd (Slovenia);
- West Norway Research Institute (Norway);
- L'Ecole Polytechnique Fédérale de Lausanne (Switzerland).

The present extension of the database was coordinated by Erik Gløersen (Nordregio) and Christian Lindner (IRPUD), and carried out with the help of:

- Saviour Formosa (Malta Environment and Planning Authority - Malta)
- Thérèse Hanquet and Philippe de Boe (PhdB Consultants – the Netherlands)
- Loreta Grizickaite (Lithuania)
- Iveta Straume (CSB – Latvia)
- Maria-Kristiina Kontus (Statistical Office – Estonia)
- Alexandre Dubois (Nordregio – Denmark)

This is explanatory note provides an overview of the data gathered as part of the DG REGIO Mountain Study, and of the data collected as part of the present extension of the database for ESPON.

Concrete information about which indicators are or are not available at NUTS 5 scale in each country will not be found in this note, but in the Access metadatabase described below.

1.2 Contents of the Database User Manual

The main objective of this Database User Manual is to give advice on technical issues related to the data delivered as part of the Mountain study and the extension of the database financed by ESPON.

Following this introduction, Chapter 2 provides information on the contents of the CD-ROM.

Chapter 3 describes the methodological steps performed in order to settle the list of indicators to be collected for the geodatabase, differentiated by statistical and GIS indicators, and briefly explains how they were processed for analysis and to fit into the common geodatabase format.

The next two chapters provide in-depth descriptions of the developed metadata database (Chapter 4) and of the geodatabase (Chapter 5), constituting the heart of the *Database User Manual*.

Chapter 6 summarises the availability of statistical and GIS indicators, and outlines some problems encountered when setting up the database.

2 The CD-ROM

2.1 Content of the CD-ROM

The CD-Rom contains four main directories:

Geodatabase
The directory contains the Info Tables produced, together forming the comprehensive *GIS Database*. Chapter 5 of this *Database User Manual* provides comprehensive information on the structure, contains and sub-directories of this geodatabase.

Metadata
This directory contains the associated metadata in form of an MS Access database. This file also includes tools to edit, browse and query the metadata. Chapter 4 describes the contents of the Access file, and explains the usage of these tools.

Boundaries
The directory contains the ArcInfo layers with the administrative boundaries.

Infrastructure facility layers
The directory contains the ArcInfo layers with the geographical positions of a selection of European airports, hospitals and universities.

2.2 Software Requirements

The geodatabase is stored as separate compressed zip files in the **/GEODATABASE** directory of the CD-ROM. WinZip can be used to decompress these files.

The unzipped database files can be imported into ArcInfo, ArcView or ArcGIS.

2.3 Installation of metadatabase

The data and metadatabase are delivered on a CD-ROM.

The metadata database as described in Chapter 4 is delivered as a self-extracting setup routine. After copying all files of the directory **/METADATA** from CD-ROM onto hard disc, the installation routine is launched clicking the SETUP.EXE file, which guides you through the rest of the installation. After finish, the metadata database is available from the **START** menu as in the folder **MOUNTAIN/METADATA**.

3 Indicator Collection

3.1 Principles of the indicator collection

The implementation of the DG REGIO Mountain Study required the use of a wide range of data sources, some developed specifically for the study, others already available. The basic principle was to use existing national and supra-national statistical sources as primary input, supplemented by a number of GIS indicators. The reference year for all indicators is 2001. Data for 1991 has been collected for some core indicators.

For the collection of the statistical indicators, an iterative approach was developed. The starting point was a list of indicators proposed by DG REGIO. Based on this list, a dialogue was established, involving DG REGIO, the coordination team and the national correspondents. Only indicators considered relevant with regards to analytical purpose of the study and which were likely to be available in a reasonable proportion of countries were kept in the final list.

The next step then was to establish rules and technical requirements for the data collection, in order to give as much advice as possible to the National correspondents, which were in charge of the actual data collection. For instance, these rules referred to the detailed guidelines for metadata collection, or detailed technical advice on how to deliver the data as such. At the end of this step, the National correspondents then collected the actual data from national or supra-national statistical sources.

A similar approach was used to constitute the list of GIS indicators to be calculated, through a dialogue between members of the research consortium and DG REGIO.

In order to analyse the indicators, and set up the database, it was necessary to group the indicators into indicator fields. The following categories were developed for the Mountain study, trying to reflect the greatest challenges of mountain areas:

- Indicators related to geography;
- Demographic indicators;
- Economic indicators;
- Indicators related to agriculture;
- Infrastructure indicators;
- Indicators related to the environment.

The geographical indicators include data concerning land use, climatic and topographic conditions. Another set of indicators refers to demography, including population numbers, age structures and densities, births, deaths and migration. Indicators on the economy included information on employment, unemployment and in-/ out-commuting. Infrastructure indicators mainly concern facilities for tourists, health care, education, and accessibility, especially regarding the transport network provision. Indicators in the environmental category include mainly information about different kinds of land uses and land coverage.

As the main objective of the Mountain study was to develop a comprehensive and harmonised NUTS 5 GIS database for mountain areas, several prerequisites had to be fulfilled.

Issues related to the collection of data from national statistical sources

A key issue in the process of collecting statistical data was to ensure consistency in the definition of the respective indicators, which is essential to harmonise the data from the different countries. For a small number of countries, lack of harmonisation regarding indicators, definitions and methodology between regions also had to be solved.

Wherever possible, data were collected at NUTS 5 (municipality) level. However, a number of indicators were only available at more aggregated NUTS levels. In part, this is due to the small size of the NUTS 5 areas in some countries. Another problem encountered in many countries is the change of NUTS 5 boundaries over time. Whenever numerous boundary changes occurred, the data were fitted to 1997 boundaries, by standardising all indicators per capita or per area, by calculating percent shares, or by calculating index values.

Issues related to the collection of data from GIS sources

The key challenge with respect to the GIS indicators was (a) to find layers that cover the entire study area, (b) with the appropriate spatial resolution necessary for the Mountain study, and (c) with the appropriate attributive resolution (e.g. the type and number and contents of the attributes associated with these layers had to match the study requirements).

In order to enable easy processing, handling and exploring of indicators, unique indicator codes were developed and assigned to

each indicator. These codes comprise a combination of letters and numbers. The codes were constructed a way they indicate the indicator fields, indicate whether they represent statistical or GIS indicator, and last but least indicate a unique identifier. Eventually, the common format of the indicator codes is as follows:

xx-yy_zz

where **xx** represents the indicator field abbreviation (**A** for agriculture, **D** for demography, **E** for economy, **G** for geography, **EN** for the environment, **I** for infrastructure and **IN** for initiatives), **yy** indicates whether it is an statistical indicator (**NS** = National statistics) or an GIS indicator (**GIS**), and **zz** represents a unique indicator number.

3.2 Reference years

All statistical indicators were requested for the year 2001, but in order to be able to have data for more countries, 2000 or 1999 were also considered when data for 2001 were not available in a country.

The same approach was taken as far as historical data are concerned. Detailed information for each indicator can be found in the metadatabase.

3.3 Statistical Indicators gathered in all ESPON countries

The following indicators have been collected both for mountainous and non-mountainous countries.

Demographic indicators

- Total population and by sex (i.e. total, female, male) (D-NS_1)
  - Total population and by sex (total) by municipality (D-NS_1a)
  - Total population and by sex (female) by municipality (D-NS_1b)
  - Total population and by sex (male) by municipality (D-NS_1c)
- Number of inhabitants over 60 and number of inhabitants under 15 (D-NS_2)
  - Number of inhabitants over 60 by municipality (D-NS_2a)
  - Number of inhabitants under 15 by municipality (D-NS_2b)

- Average number of births and deaths per year by municipality (D-NS_3)
- Number of births per year (annual average 81-91 and 91-2001) (D-NS_3x)
- Number of deaths per year (annual average 81-91 and 91-2001)(D-NS_3y)
- Number of immigrants and emigrants by municipality (D-NS_4)
- Number of immigrants (annual average 81-91 & 91-2001) (D-NS_4x)
- Number of emigrants (annual average 81-91 & 91-2001) (D-NS_4y)
- Net migration (for all years from 1981 to 2001) (D-NS_4a)
- Total number of foreign people by municipality (D-NS_5)

Economic indicators

- Total number of active people by sex (i.e. total, female, male) (E-NS_1)
- Total number active people (total) by municipality (E-NS_1a)
- Total number active people (female) by municipality (E-NS_1b)
- Total number active people (male) by municipality (E-NS_1c)
- Total number of employees by sector (agric., industry, services) (E-NS_2)
- Total number of employees in the agricultural sector (E-NS_2a)
- Total number of employees in the industry sector (E-NS_2b)
- Total number of employees in the services sector (E-NS_2c)
- Total number of unemployed people by sex by municipality (E-NS_4)
- Total number of unemployed people (total) (E-NS_4a)
- Total number of unemployed people (female) (E-NS_4b)
- Total number of unemployed male people (male) (E-NS_4c)

3.4 Statistical indicators gathered in countries with mountainous areas only

The following statistical indicators were gathered for the DG REGIO Mountain Study, but not for the extension of the database for the rest of the ESPON space :

Indicators related to agriculture

- Number of farms by municipality (A-NS_1)
- Livestock population (average annual number of animals) (A-NS_3)
- Total livestock (livestock unit equivalents) by municipality (A-NS_3a)
- Cattle by municipality (A-NS_3b)
- Pigs by municipality (A-NS_3c)
- Poultry by municipality (A-NS_3d)
- Sheep by municipality (A-NS_3e)
- Utilisable agricultural area by municipality (in sq. km) (A-NS_8)

Economic indicators

- Total number of unemployed people under –25 years by municipality (E-NS_5)
- Total number of long-term unemployed (i.e. longer than 1 year) (E-NS_6)
- Level of qualification: number of people with highest education (E-NS_11)
- Disposable income per capita (E-NS_18)
- Commuting patterns: In- /out-commuting municipality (E-NS_22)
- Total number of active people (incl. self-employed & job-seekers) (E-NS_23)

Infrastructure indicators

- Total number of overnight stays of tourists by municipality (I-NS_5)
- Number of doctors by municipality (I-NS_12)
- Number of secondary by municipality (I-NS_18)

3.5 GIS Indicators available for mountainous and non-mountainous areas

The availability of GIS indicators is bound to the availability of layers providing basic information necessary to derive indicators. Compared to the statistical indicators, data availability of basic layers is somewhat better, so that quite a number of indicators could be derived. One important exception from that are indicators related to the environment. There, the availability of GIS layers is somewhat poor. Although some indicators could be calculated (e.g. settlement area, open space, climatic contrast index), many could not because of missing or inaccessible data (e.g. NATURA 2000 areas, habitat areas of certain species), or because the resolution of the existing layers was not appropriate enough for this type of study.

Another drawback of the GIS indicators is that except for some of the accessibility and infrastructure indicators, data for the ultra-peripheral regions is not available (Azores, Canary Islands, Madeira, French DOM regions).

The following GIS indicators were calculated:

Indicators related to agriculture

- Total forest area by municipality (A-GIS_5)
- Share of forest area by municipality (A-GIS_6)
- Total arable land by municipality (A-GIS_7)
- Share of arable land by municipality (A-GIS_8)

Indicators related to the environment

- Degree of urbanisation by municipality (EN-GIS_7a)
- Climatic Contrast Index by municipality (EN-GIS_20a)

Indicators related to geography

- Longitude and latitude of municipality centres (G-GIS_1)
  - Longitude of municipality centre (G-GIS_1a)
  - Latitude of municipality centre (G-GIS_1b)
- Minimum, mean and maximum altitudes (G-GIS_2)
  - Mean altitude by municipality (G-GIS_2a)
  - Minimum altitude by municipality (G-GIS_2b)
  - Maximum altitude by municipality (G-GIS_2c)
- Ratio minimum/maximum elevation by municipality (G-GIS_3)
- Standard deviation of elevation by municipality (G-GIS_5)
- Mean and steepest slopes (G-GIS_7)
  - Mean slope by municipality (G-GIS_7a)
  - Steepest slope by municipality (G-GIS_7b)

- Total area (G-GIS_8)
    - Total municipality area (G-GIS_8a)
- Absolute and average minimum / absolute maximum temperature (G-GIS_10)
    - Average minimum temperature by municipality (G-GIS_10b)
    - Average maximum temperature by municipality (G-GIS_10d)
- Land use by 14 land use types by municipality (G-GIS_17)
    - Land use by coniferous forest by municipality (G-GIS_17a)
    - Land use by deciduous forest by municipality (G-GIS_17b)
    - Land use by mixed forest by municipality (G-GIS_17c)
    - Land use by grassland by municipality (G-GIS_17d)
    - Land use by rainfed arable land by municipality (G-GIS_17e)
    - Land use by irrigated arable land by municipality (G-GIS_17f)
    - Land use by permanent crops by municipality (G-GIS_17g)
    - Land use by scrubland by municipality (G-GIS_17h)
    - Land use by barren land by municipality (G-GIS_17i)
    - Land use by permanent ice and snow by municipality (G-GIS_17j)
    - Land use by wetlands by municipality (G-GIS_17k)
    - Land use by inland waters by municipality (G-GIS_17l)
    - Land use by urban areas by municipality (G-GIS_17m)
- Typology according to main land use type (G-GIS_19)
    - Typology of municipalities according to main land use (G-GIS_19a)
- Airline distances from municipality centres to next cities (G-GIS_22)
    - Airline distance from municipality centres to nearest city (G-GIS_22a)
    - Average airline distance to the next three cities (G-GIS_22b)
- Airline distances to capital city from municipality centres (G-GIS_23)
    - Airline distance from municipality centres to capital city (G-GIS_23a)
- Public infrastructure supply: hospitals (G-GIS_30)
    - Airline distance to next hospital by municipality (G-GIS_30a)
    - Airline distance to second next hospital by municipality (G-GIS_30b)
    - Airline distance to third next hospital by municipality (G-GIS_30c)
    - Average airline distance to next 3 hospital by municipality (G-GIS_30d)

- Average airline distance to all hospital by municipality (G-GIS_30e)
- Public infrastructure supply: universities (G-GIS_32)
  - Distance to nearest university by municipality (G-GIS_32b)
  - Potential accessibility to universities by municipality (G-GIS_32d)
  - Travel time to nearest university by municipality (G-GIS_32e)
- Share of land use by 14 land use types by municipality (G-GIS_36)
  - Share of coniferous forests by municipality (G-GIS_36a)
  - Share of deciduous forests by municipality (G-GIS_36b)
  - Share of mixed forests by municipality (G-GIS_36c)
  - Share of grassland by municipality (G-GIS_36d)
  - Share of rainfed arable land by municipality (G-GIS_36e)
  - Share of irrigated arable land by municipality (G-GIS_36f)
  - Share of permanent crops by municipality (G-GIS_36g)
  - Share of scrubland by municipality (G-GIS_36h)
  - Share of barren land by municipality (G-GIS_36i)
  - Share of permanent ice and snow by municipality (G-GIS_36j)
  - Share of wetlands by municipality (G-GIS_36k)
  - Share of inland waters by municipality (G-GIS_36l)
  - Share of urban areas by municipality (G-GIS_36m)

Infrastructure indicators

- Number of ports by municipality (I-GIS_16a)
- Number of airports by municipality (I-GIS_17a)
- Car travel time to nearest airport by municipality (I-GIS_17c)
- Potential accessibility to airports by municipality(I-GIS_17f)
- Peripherality by car to population by municipalities (EU average) (I-GIS_18)
- Peripherality by car to population by municipalities (nat. averages) (I-GIS_20)
- No of people living within 1h car driving time by municipality (I-GIS_30a)
- No of people living within 30 min driving time by municipality (I-GIS_31a)
- Length of road network by type of road (I-GIS_32)
  - Length of motorways by municipality (I-GIS_32a)
  - Length of dual-carriageways by municipality (I-GIS_32b)
  - Length of other trunk roads by municipality (I-GIS_32c)
- Density of road network by type of road (I-GIS_33)
  - Density of motorways by municipality (I-GIS_33a)
  - Density of dual-carriageways by municipality (I-GIS_33b)
- Density of other trunk roads by municipality by municipality (I-GIS_34)

- Length of existing and planned TEN road network by municipality (in km) (I-GIS_35a)
- Length of existing TEN road network by municipality (in km) (I-GIS_36a)
- Length of planned TEN road network by municipality (in km) (I-GIS_37a)
- Density of navigable waterways by municipality (I-GIS_38a)
- Density of existing TEN road network by municipality (I-GIS_39a)
- Density of planned TEN road network by municipality (I-GIS_40a)
- Density of existing & planned TEN road network by municipality (I-GIS_41a)

3.6 Indicator Processing

For each indicator the national experts provided the coordination team with data sheets including municipality ID codes and an additional (see Chapter 4 for more information about the metadata database). After the data were received, several checks were needed concerning NUTS level, data gaps, and data inconsistencies and provided ID codes. Data with obvious inconsistencies have not been included. National experts have been consulted when estimates have been calculated.

Some gaps in the database are due to boundary changes between 1997 (SABE 1997 boundary system) and the provided data for 2001 (referring to a national 2001 boundary system). In these cases, it was not possible to join data for all municipalities of a country into the database.

Finally, in the cases of Slovakia and the UK, the data at municipal level were calculated by overlaying data corresponding to an alternative municipal delimitation. In these two cases, the quality of the data is in general insufficient for analyses at the municipal level. However, analyses based on different types of municipality aggregates may give valid results.

4 The Metadata Database

4.1 Objectives of the Metadata Database

Constructing a database NUTS 5 level for the whole of Europe implies handling a large amount of information. This not only concerns the actual data themselves, but also information about the data, such as data source, actual year, original data format, data gaps, indicator definitions and units. It is therefore of crucial importance to collect and organise of metadata. These metadata should always be taken into consideration before carrying out any analyses.

Metadata information has been collected country by country, first in order to review data availability, and later in order to get information on indicator definitions, available years, units or data gaps, which all of them are important information for analysing the data. Eventually, the metadata information was transferred into a MS Access database.

In order to collect information about indicators, the National Experts were asked to fill in the 2-page metadata review sheets (see Figures 2 and 3). One sheet was completed for each indicator in each country. These sheets were processed electronically by using VBA macros under Access. These macros then also took care of transforming the original table structure into the relational table structure of the Access database (see Figure 4).

The objectives of this metadata database are threefold:

1. To maintain and browse the indicators available to facilitate efficient access and handling of indicator related information.
2. To enable specific queries and to produce summary statistics on data availability by country and indicator.
3. To produce various lists of indicators (list of statistical or GIS indicators available, those that were not available etc.).

With these objectives, the Access database combines the original information stored in the country-specific and indicator-specific Word files that were produced by the National Experts.

In addition to NUTS 5 data, the metadatabase contains references to a number of indicators developed specifically for the DG REGIO mountain study. These generally correspond to NUTS 5 data which were aggregated at the level of a specific mountain range. References to these indicators, which comprise the indication “by massif” in their name, can be disregarded.

Country:	**Indicator code [1]:**	**Category [2]:**
Indicator definition [3]:		
Data source(s):		
Original data format(s):		**Price:** *0*
NUTS levels [4]: **X** NUTS 0 X NUTS 1 X NUTS 2 X NUTS 3 X NUTS 4 X NUTS 5		
Territory covered [5]:		
Years available: 1981 X yes ڤ no if no, possible alternative years: *xxx* 1991 X yes ڤ no if no, possible alternative years: *xxx* 2001 X yes ڤ no if no, possible alternative years: *xxx*		
Completeness / Gaps [6]:		
1981 *Xxx*	**1991** *xxx*	**2001** *xxx*
Estimations, assumptions, proxies and calculations applied to fill gaps [7]:		
1981 *Xxx*	**1991** *xxx*	**2001** *xxx*

Figure 1. Sample metadata sheet (page 1).

Country:	**Indicator code [1]:**	**Category [2]:**
Overall degree of confidence, reliability [8]: ف low ف medium X high		
Regular update intervals / proposals for future updates of this indicator:		
Other institutions / people contacted [9]: *xxx*		
Final comments [10]: *xxx*		

Figure 2. Sample metadata sheet (page 2).

4.2 Structure of the Access database

Creating a relational database in general means splitting up information into separate tables in order to prevent storing information redundantly and for ease of querying the database. The database for the metadata therefore contains the core table named **DATASETS** which stores the main part of the information provided in the metadata sheets (see Figure 3). To relate each dataset to an indicator the table **INDICATORS** is linked with **DATASETS** by the **INDICATORID**. To assign each indicator to a category of the **CATEGORIES** table these are linked by the **CATEGORYID**. Furthermore, the different countries that are considered in the study are represented through the **COUNTRIES** table and are linked with the **DATASETS** table by the **COUNTRYID**. The information contained in the **AVAILABILITY** table provides along with the assigned **YEARS** table the information for which year data in the

DATASETS table is available plus the statements made on completeness and methods applied to fill data gaps for each year available. The rows for different years (e.g. 1981, 1991 and 2001) in the **AVAILABILITY** table are related to one row in the **DATASETS** table using the **DATASETID** attribute. The information on data sources are also kept in a separate table **DATASOURCES**. The following database scheme illustrates the described database.

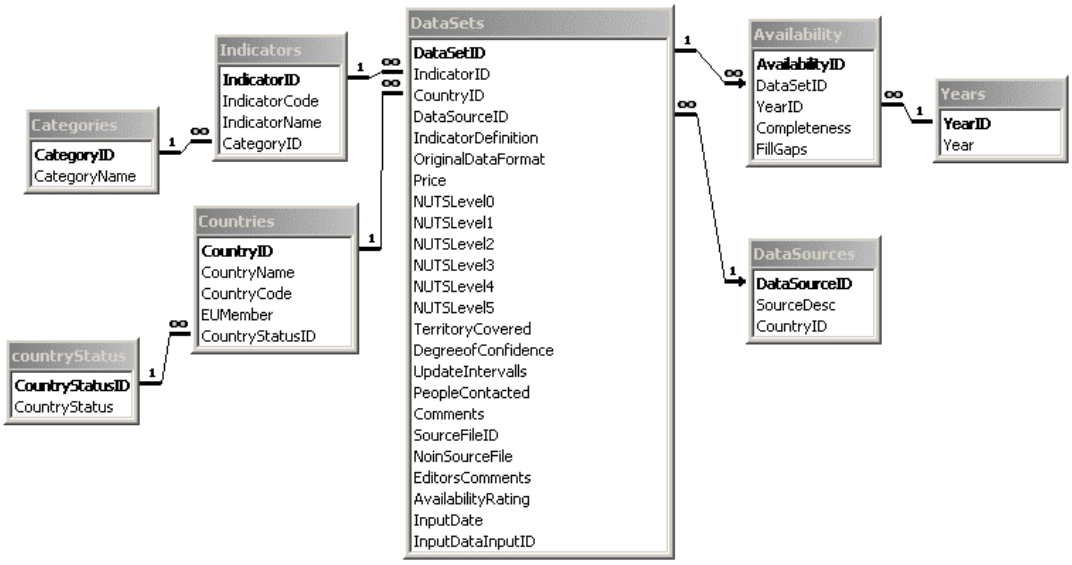

Figure 3. Structure of the Access database.

4.3 Metadata Database Tools

In order to access the huge amount of metadata in a comfortable, convenient and user-friendly way specific metadata database tools were developed on top of the Access database, making use of VBA programming techniques. Each tool was designed for a specific task, as there are separate tools to browse the metadata database, do specific queries and export results into Excel spreadsheet format.

The different tools described in detail in the following paragraphs are accessible via the *Main Menue* of the database which appears when the database is opened. It shows three buttons labelled **Open Database Browser**, **Open Query Builder** and one for closing the application.

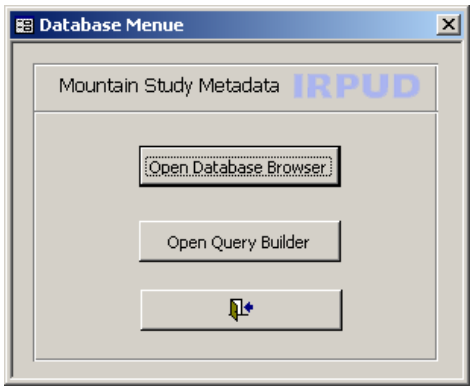

Figure 4. Main menu of the database.

4.4 Browsing Metadata

For this task a database browser on top of Access has been developed which serves as an user interface (Figure 5) allowing the user to directly retrieve the metadata information in question. Therefore, the tool uses queries bringing together the various information stored in the different database tables and presents all metadata information within a single window at a glance, hence there is no need for users to struggle with tabular data. The different fields of the browser's user interface display the different metadata information given per county and indicator analogue to the metadata sheets mentioned above.

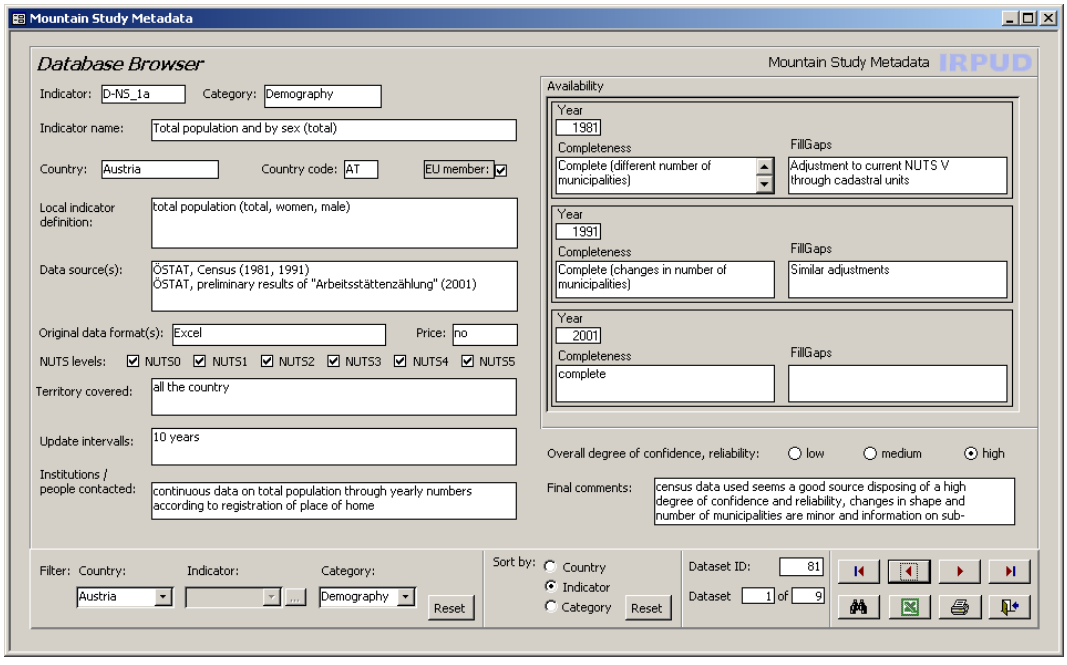

Figure 5. Database browser.

To keep work with the browser handy there is a navigation toolbar located at the bottom of the browser window which provides the user with some useful browsing functions. The filter function enables the user to define and set filter criteria which limits the amount of datasets displayed to the portion matching the filter criteria. The user is able to set a country, indicator or indicator category filter. Contrary to the indicator filter the country and indicator category filter can be applied at once enabling the user to only display datasets related to a certain indicator category for a certain country.

In order to make filtering by indicator easier, there is an advanced filter selection window besides the directly visible dropdown menu, which can be accessed via the small bottom right to it. This was introduced because there may be users not being able to interpret the indicator codes at once. Clicking on this button will open a popup window displaying all indicators with their codes and full definitions. Furthermore the user can set a sort option to display the datasets in question in an appropriate order. Next to the sort option there is an additional display field keeping track of the **ID** number of the dataset currently displayed plus actual information on the number of datasets currently affected by the selection and the number of the current dataset in this amount. Finally the right section of the toolbar is used to navigate through the data. Besides the standard navigation buttons (first record, previous, next and last record) the user is given a search function for text search on every display field plus an export and a print button. The export button enables the user to export the dataset currently displayed to an excel spreadsheet, the print button calls the print function producing a printed version of the current dataset in a preformatted sheet.

4.5 Performing specific Queries

Since the metadata database should allow the user to perform specific queries (i.e. delivering an overview on data availability for the different countries and indicators on comparability of indicators for the different countries, the overall degree of confidence per indicator, data sources etc.), a second tool was developed providing an user interface, the so-called *Query Builder* (Figure 6). This tool helps to find data gaps and provides a fast and consistent overview on analysis potentials. In addition, it enables to easily create lists of indicators of which data are available in existing databases at European, national, or other levels as well as detailed lists of data sources for the indicators identified.

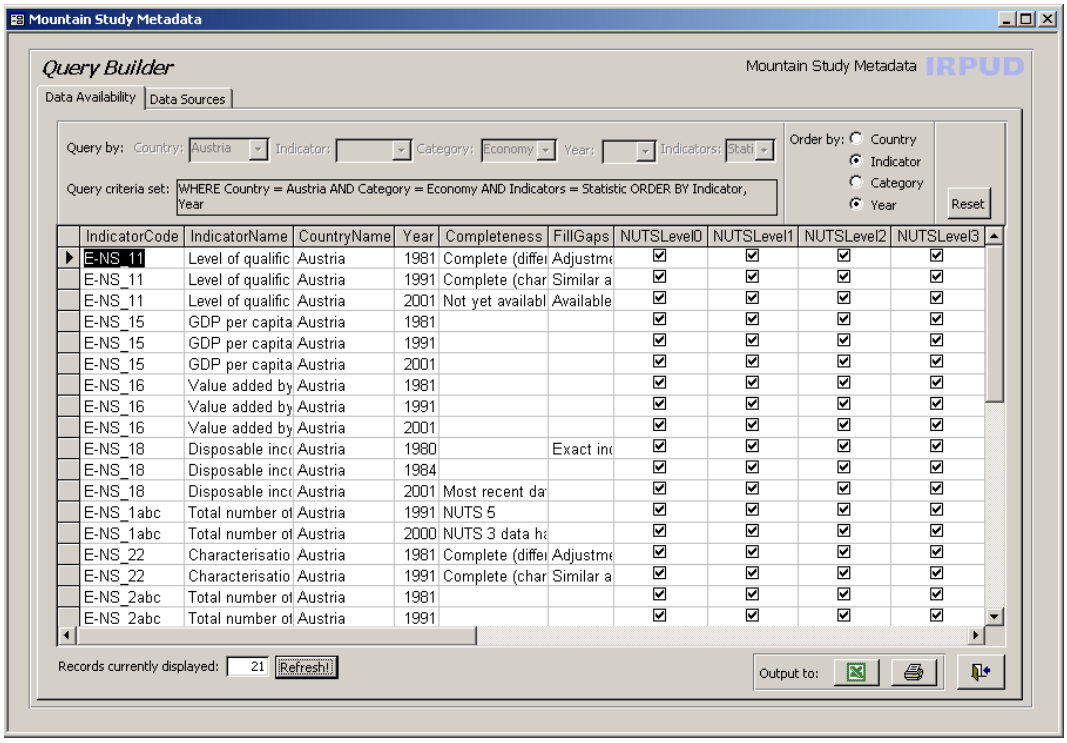

Figure 6. Query builder.

Like the *Database Browser*, the *Query Builder* allows the user to define the amount of relevant datasets by setting criteria like country, indicator and indicator category. In addition it offers specific criteria for specific queries. The means in case of availability, that the user can also set years or indicator types as criteria. The application also has an option to order by these items. The number of datasets affected by the current criteria set is displayed at the bottom of the window on request. To provide export capabilities the lists displayed can be exported to Excel spreadsheets or sent to a printer by using the appropriate buttons at the bottom of the window. For the different queries on data availability and data sources the user only needs to switch the display by clicking on the appropriate tab at the top of the window.

5 Geodatabase Structure

One of the major prerequisites for the GIS database is its compatibility with existing Eurostat/GISCO GIS databases, in particular with respect to the spatial reference system and supported data formats.

5.1 Data formats

The present database uses data formats supported by ArcInfo: ArcInfo Coverages for vector data such as municipality boundaries, and Info Tables for tabular data. These data formats can also be imported into ArcView and ArcGIS.

These data formats are fully compatible with each other, and can be related with each other using standard ArcInfo commands.

5.2 Projection and Co-ordinate System

The municipal map that was compiled by the Mountain Study consortium has been adapted to the standard ESPON projection.

This standard planar projection is a *Lambert Azimuthal Equal Area* projection. It is best suited for large areas, preserving as much as possible the shape of the continent. The projection is characterised by the following parameters:

Units	**meters**
Spheroid	**sphere**
Radius of sphere of reference	**6378388**
Longitude of centre of projection	**15° 00' 00''**
Latitude of centre of projection	**50° 00' 00''**
False easting	**0.0**
False northing	**0.0**

For the French Overseas Territories, alternative individual projection systems were used, defined according to Eurostat specifications.

The generalisation of this map is different from that of the NUTS maps generally used in ESPON. Differences between the two layers („sliver polygons“) will therefore appear when overlaying the NUTS 5 map with these wider scale ESPON layers, if one zooms in on individual regions.

5.3 Info Tables

The data for each municipality are stored in **INFO** tables.

There are two grouped **INFO** tables for demographic and economic indicators collected for the entire ESPON area. Otherwise, each **INFO** table contains one indicator.

Each **INFO** table has two main attributes. The municipality tables contain the **CMRGCD97** attribute (see Table 1), followed by the item storing the indicator value; the name of the column corresponds to the indicator code.

The name of the **INFO** tables corresponds to the indicator code, i.e. the table name directly indicates the contents of the indicator.

In all **INFO** tables, missing values are always represented by ”- 9999”.

5.4 The Structure of the Database

The database is located in the CD-ROM directory **root/GEODATABASE.**

The **INFO** tables included in the database are sorted according to three main categories, with thematic subfolders as described below:

- Data collected in all ESPON countries (folder: **root/Geodatabase/All ESPON countries**):
    - Agriculture
    - Demography
    - Economy
    - Environment
    - Geography
    - Infrastructure
    - Infrastructure maps (geographical positions of airports, hospitals and universities)

- Data collected in countries with mountain areas only (folder: **root / Geodatabase / Mountainous countries**):
    - Agriculture
    - Economy

- Data for French overseas territories (folder: **root / Geodatabase / DOMs** countries):
    - Demography
    - Economy

The **root / Geodatabase / DOMs** folder includes all data available for the French Overseas Territories, which in turn means that the other workspaces mentioned do not store any data for the **DOMs**. Indicators, for which data for the **DOMs** are not available, are not included. Since there are also separate layers representing the municipalities of the **DOMs**, this individual sub-directory provides direct access to the indicators.

This general database structure does not seek to duplicate any existing Eurostat database (e.g. NewCronos, GISCO), but rather provides a comprehensive, GIS-based spatial database. NewCronos Excel sheets (and other data sources) have been transformed into ArcInfo Info Tables or ArcInfo coverage or Grid formats. Aggregates calculated from seamless GISCO layers are linked directly to ArcInfo municipality layers, in order to unlock and enable the full potential of GIS tools to facilitate the analysis of mountain areas.

5.5 The boundary files

The boundary files are located in the CD-ROM directory **root / Boundaries.**

A coverage entitled **MUNIS** representing the 2001 NUTS 5 administrative system for the study area, except for the French Overseas Territories. This layer includes the polygon feature class as being the main feature class. Each NUTS 5 entity is represented by one or several polygons (i.e., in case of island municipalities or in special circumstances, a municipality may consist of several polygons).

This layer is mainly based on the SABE 1997 NUTS 5 layer produced by EuroGeographics (EuroGeographics Association, 1997) and provided through Eurostat/GISCO. However, some adjustments

were made for the UK. In addition, some alternative boundary files were used for some countries:

- Switzerland: boundaries provided by the Swiss Federal Statistical Office),
- Romania: boundaries provided by ESRI Romania,
- Slovenia: boundaries provided by Oikos
- Bugaria: boundaries provided by NCRD). The

The **MUNIS** layer for Europe covers all present EU member states and accession countries, plus Norway and Switzerland. The French Overseas Territories and the other European countries are not covered. The DOM are represented by four individual layers (see below). The **MUNIS** layer includes the following user-defined attributes in the polygon feature class (**PAT**) (internal ArcInfo attributes are not listed here) (Table 1):

Table 1. Attributes available in the municipality layer (PAT).

Attribute	Contents
AREA	Total municipal area (square meters)
PERIMETER	Municipality perimiter
MUNIS_	
MUNIS_ID	Internal municipality ID
CMRGCD97	Municipality code
CMRGCD	Altenrative municipality code
NURGCD	NUTS 3 code (except CH, RO BG)
NURGCDL0	NUTS 0 code
NURGCDL1	NUTS 1 code (except CH, RO BG)
NURGCDL2	NUTS 2 code (except CH, RO BG)
MUNINAME	Municipality name
CMFTTP	Type of polygon C = Condominium L = Lake (area is water only) M = Mainland S = See (area is water only) U = Unknown

The approximate file size of the ArcInfo export file (**E00**) of the **MUNIS** layer is about 263 MBs uncompressed and 82 MBs compressed.

The separate municipality layers for the French DOMs include the same attributes as outlined in Table 1.

The boundary files considers each polygon as a separate entity, and includes sea and lake areas in some countries. The file must therefore be adapted to software used before any analysis can be carried out.

There are also four individual municipality coverages for the French Overseas Territories Guyane (**MUNGUYANE**), Guadeloupe

(**MUNGUADE**), Martinique (**MUNMARTIN**), and Réunion (**MUNREUNION**).

5.6 The Infrastructure Facilities Layers

The **root/Geodatabase/Infrastructure facility layers** folder contains three individual layer providing information on locations of major infrastructure facilities, i.e. location of hospitals (**HOSPITALS**), universities (**UNIVERSITIES**) and airports (**AIRPORTS**).

All the three layers include point feature classes only, i.e. all facility locations are represented by one point in the layer.

The following user-defined attributes are associated with the point attribute table (**PAT**) of the **HOSPITALS** coverage:

Table 2. Attributes available in the hospital layer (PAT).

Attribute	Contents
CITY	Name of the city in which the hospital(s) is (are) located
COUNTRY	ISO country code
HOSPITALS	Total number of hospital in the city
HSPLBEDS	Total number of beds in these hospitals

This layer includes hospitals with more than 300 beds, or with the status of regional or university hospital, only.

The following user-defined attributes are associated with the point attribute table (**PAT**) of the **UNIVERSITIES** coverage:

Table 3. Attributes available in the universities layer (PAT).

Attribute	Contents
CITY	Name of the city in which the university / polytechnique is located
COUNTRY	ISO country code
STATUSUNI	Type of facility available: U = University only P = Polytechnic only UP = University and polytechnic available
STUD_POLY	Number of students at polytechnics in the city
STUD_UNIV	Number of students at universities in the city

This layer includes universities with at least 1000 students only, as well as selected polytechnics.

The following user-defined attributes are associated with the point attribute table (**PAT**) of the **AIRPORTS** coverage:

Table 4. Attributes available in the airport layer (PAT).

Attribute	Contents
IATA_CODE	Official IATA code of the airport
NAME	Airport name
COUNTRY	ISO country code
STATUS	Type of airport C = International airport R = National or regional airport M = Other airport '' = n.a.
FLIGHTS_Y	Number of flights per year
PASS_Y	Number of passengers per year

This layer includes airports with at least one regular weekly flight only. Other (minor) airports dedicated for occasional flights or military use only are not included.

5.7 Exploring the Database

A link between the ArcInfo map layers and the respective **INFO** tables can easily be established using the concept of 'relates' in ArcInfo. The attribute **CMRGCD97** must be used for linking the municipality coverage with the respective municipality **INFO** tables. Once the files have been joined, the **INFO** tables can be browsed, queried, mapped, edited or exported.

6 Data Availability

Data were collected from both national sources with the help of the national correspondents and from various GIS databases, and were classified as either statistical and GIS indicators. As mentioned in Annex 1 of the Final Report, the availability rates among the indicators vary considerably.

6.1 Availability of GIS-indicators

One of the great advantages of using GIS indicators is that once a base layer is available, indicators will also be available for the study area in question. Hence, each GIS indicator is usually based on one data source only, providing information on all countries. Then the same methodology or calculations can be used to derive the indicators that one is interested in, which is a great advantage compared to the statistical indicators, where often attempts have to be made to harmonise different indicator definitions used by different countries.

On the other hand, however, sometimes GIS base layer are not available at all, or are available in a resolution which is too coarse for the purpose of the study, or are available in different formats than required (for example, certain calculations require a polygon representation of a feature, but the layer only includes a point representation), or are available only for parts of the study area, or lastly the base layer is lacking important attribute information.

6.2 Availability of statistical indicators

Table 5 provides an overview of data availability per country. These figures consider 31 spatial entities, as England and Wales, Scotland and Northern Ireland are dealt with separately.

The first four columns deal with the data gathering carried out as part of the DG REGIO Mountain Study. Some data were included directly in the database, and some through the calculation of estimates or other data treatments. Other countries could provide data, but in a format or with a coding system which made it impossible to include them in the database. Finally, some indicators were not available at all in some countries.

All indicators have been collected for the year 2001 if possible. In order to be able to have data for more countries, 2000 or 1999 were also considered when data for 2001 were not available in a specific country.

Table 5. Final data availability per country (notes on next page)

Indicator code[1]	Countries with mountain areas (25 countries considered[2])				Other countries (6 countries considered[6])
	Number of countries with data included	Number of countries with data included through further treatment[3]	Number of countries with data which could not be included[4]	Number of countries with no available data [5]	Number of countries included
A-NS_1	12	2	9	2	-
A-NS_3b	5	0	4	16	-
A-NS_3c	5	0	4	16	-
A-NS_3d	2	0	4	19	-
A-NS_3e	4	0	4	17	-
A-NS_8	12	2	9	2	-
D-NS_1 (1981)[7]	16	0	0	9[8]	-
D-NS_1 (1991)[7]	20	1	0	4[8]	5
D-NS_1 (2001)[7]	21	4	0	0	6
D-NS_2 (1981)[7]	13	1	0	11[8]	-
D-NS_2 (1991)[7]	16	2	0	7	5
D-NS_2 (2001)[7]	19	5	0	1	4
D-NS_3 (1991)[9]	6	8[10]	0	11[8]	2
D-NS_3 (2001)[11]	7	12	0	6[8]	5
D-NS_4 (1991)[9]	3	10	0	12[8]	2
D-NS_4 (2001)[11]	5	13	0	7	3
D-NS_5 (1981)	7	1	0	17[8]	-
D-NS_5 (1991)	9	3	0	13[8]	-
D-NS_5 (2001)	13	5	0	7	3
E-NS_1a	12	12	0	1	-
E-NS_1bc	10	9	0	6	-
E-NS_2(1981)[7]	8	0	0	17[8]	-
E-NS_2(1991)[7]	12	0	0	13[8]	3
E-NS_2(2001)[7]	12	10	1	2	4
E-NS_4[7]	14	10	0	1	4
E-NS_5	11	9	0	5	-
E-NS_6	9	10	0	6	-
E-NS_11	9	7	3	6	-
E-NS_16[7]	0	0	15	10	-
E-NS_18	5	0	13	7	-
E-NS_22	8	0	7	10	-
I-NS_5	8	0	15	2	-
I-NS_12	7	0	15	3	-
I-NS_18	9	0	8	8	-

[1] see Chapter 3.2 for code explanations, indicator is referring to 2001 if not indicated differently
[2] Parts of the UK were counted separately (England and Wales, Scotland, Northern Ireland) as there are differences in terms of availability.
[3] Methodologies like disaggregations and estimations has been used in order to integrate the data into the database.
[4] Data has been collected, but couldn´t be used (i.e. not disaggregateable, wrong year or huge data gaps which are impossible to fill).
[5] as indicated by the national experts.
[6] including those countries without mountain areas: Denmark, The Netherlands, Estonia, Latvia, Lithuania and Malta
[7] information valid for all sub-indicators
[8] including unavailability because of data is based on old boundary system
[9] annual average between 1981 and 1991; information valid for all sub-indicators
[10] includes those countries for which data weren´t based on every requested single year
[11] annual average between 1991 and 2001; information valid for all sub-indicators

6.3 Problems encountered during database construction

Several different procedures have been used in order to overcome problems caused by boundary changes. However, some gaps remain, for example in Eastern Germany. This is because boundary changes occurred between the SABE 1997 boundary system and the boundary system of 2001 to which the provided data is referring.

Moreover, comparisons between countries are sometimes difficult as the definition of some indicators differentiate considerably (e.g., in the case of E-NS_2, some countries only have data for employees, while other supply data on total employment).

As far as the GIS indicators are concerned, three main types problems can be identified:

First, seamless base layers that could be used to construct indicators were lacking or were not available at all (for example, living areas of predators, habitats of species), or were not available in the required data format (for example, there are two GISCO layers providing information on protected areas, however, the feature only point objects, whereas for the Mountain study polygon features were needed). In these cases the construction of the GIS indicator was skipped.

Second, the GIS base layers available did not cover the entire study area, but only parts of it (e.g. CORINE Landcover). In these cases the GIS indicator was only calculated for those covered study area.

Third, in principle GIS base layers were available, but their spatial resolution was too coarse to derive reliable indicators (e.g. mobile phone reception), or the information provided was outdated, incomplete or inconsistent (e.g. energy production layer from Eurostat/GISCO).