EUROPEAN FORUM FOR GEOSTATISTICS

PRODUCTION PROCEDURES FOR A HARMONISED EUROPEAN POPULATION GRID

AGGREGATION METHOD

Project	ESSnet project GEOSTAT 1B
Agreement No.	50502.2009.004-2009.860
WP	WP-1C Geostatistics by aggregation
Task	Task 1C.1: The production of datasets according to the "Bottom up" approach
Deliverable	SUMMARY REPORT OF THE METHODOLOGY FOR GENERATING GEOSTAT 2011 GRID
Date	2013-01-01
Contributors	Geostat 1B Consortium

INTRODUCTION

GEOSTAT action is for developing guidelines for producing harmonised grid statistics and promoting national grid initiatives in National Statistical Institutes. Reasoning for this action comes from two major political definitions of EU: European Statistical System (ESS) strategic goal to integrate spatial information and statistics and INSPIRE initiative's goal to develop interoperable spatial datasets. INSPIRE focus on 34 dataset themes, of which one is population distribution. Population distribution is described as population data by municipalities and grids.

These guidelines are meant for helping National Statistical Institutes to produce harmonised population grid data compatible with INSPIRE specifications. Many elements in these guidelines will also be relevant for the production of other detailed geo-referenced data.

TWO MAIN APPROACHES FOR CONVERTING NATIONAL DATA TO A EUROPEAN GRID

1. CONVERSION AND AGGREGATION OF POINT-BASED SOURCE DATA INTO GRIDS

When national micro data is available and it is or can be georeferenced accurately a production process may start from the point based data. This process is the most straightforward and high quality way to produce European wide harmonised grid data.

2. CONVERSION OF EXISTING NATIONAL GRID DATA INTO EUROPEAN GRID DATA (ETRS89-LAEA)

Existing national grid data can also be converted into a European harmonised grid data, but it is here important to take into the account the differences in coordinate systems. This method allows transformation of coordinate systems considering the differences in location and form. In order to reduce the risk for errors it is always safest to return to approach 1 if possible. There are two processes for handling this transformation:

A. CONVERSION FROM NATIONAL TO EUROPEAN GRID DATA USING CENTROIDS

This process is suitable for converting national grid cells that:

- are smaller than the grid cell size of the European grid.

- have a coordinate reference system with characteristics which are not too different from the LAEA coordinate system

- are part of very large grid data sets (large number of grid cells) and/or when the processing capacity of software and hardware is limited

B. CONVERSION FROM NATIONAL TO EUROPEAN GRID DATA USING INTERSECTION

This process is suitable for converting national grid cells that:

- are of the same size as the grid cell size of the European grid and where national coordinate system is substantially different from ETRS89-LAEA. These differences occur in regions and countries far away from the centre of the LAEA projection (latitude of origin 52ºN and longitude of origin 10ºE, central Germany).

- are not part of very large grid data sets and/or when the processing capacity of software and hardware are not an issue

If neither of the approaches above are possible there are also disaggregation methods for estimating grid data. There is more information about disaggregation methods *in other parts of the EFGS website*.

There are many ways to carry out grid production processes. Each phase can be done by using various software or tools. The production environment as well as available source data may widen or restrict methods or tools that can be used during the process.

THE EUROPEAN GRID ETRS89/LAEA

Eurostat has made available a harmonised grid net of the whole of the Europe. In addition to the European wide grid net there are country clips of the grid net covering only one country at a time. The grid net provides reference grid cells for the harmonised grid data in Europe. It is using the Lambert-Azimuthal-Equal Area projection (ETRS89-LAEA Europe – EPSG:3035) and it is compatible with INSPIRE specifications. More information about ETRS89/LAEA can be found in the report "Map Projections for Europe" (http://www.ec-gis.org/sdi/publist/pdfs/annoni-etal2003eur.pdf)

TERMS USED

Aggregation: Here aggregation is the function for summing data entities into a total sum per individual grid cell

Aggregation method: Producing grids by aggregating geo-referenced micro data (also called bottom-up approach)

Disaggregation method: In the absence of geo-referenced micro data this method produces grids, using statistical data for the lowest available administrative/territorial units in combination with auxiliary spatial data (also called top-down approach)

Hybrid method: Producing grids by combining the aggregation and disaggregation method

Micro data: Collection of data entities in form of registers in National Statistical Institutes

Georeferenced micro data: collection of statistical data entities with assigned geographical coordinates of buildings/addresses

Spatial dataset: Dataset including geographical objects

Tabular dataset: Dataset including tabular data

Populating: Database terminology for filling up database storage space

Grid dataset: A form of spatial dataset that describes space as an array of equally sized grid cells arranged in rows and columns

National grid data: Grid dataset produced using a national coordinate system

Centroid: Central point of a grid cell

Reprojection: Conversion from one coordinate system to another

European grid ETRS89/LAEA: Grid dataset defined in the coordinate reference system ETRS89 – Lambert Azimuthal Equal Area – Europe. For more information see http://spatialreference.org/ref/epsg/3035/ or http://www.ec-gis.org/sdi/publist/pdfs/annoni-etal2003eur.pdf

Country clip: A clipped grid dataset per country; in this case the ETRS89/LAEA grid produced by Eurostat

Grid identifier, GRD_ID: Harmonized grid cell identifier in European LAEA grid net

ArcGIS: A commercial Geographical Information System (GIS) software

gvSIG: An open source GIS software that can be downloaded from gvSIG Portal, http://www.gvsig.org/web/ .

How to convert national data to a European grid:

The process flows for generating a harmonised European grid are similar and have common phases but approaches and source data are different. The process flows are described below with explanations. The explanations are to be found in the guidelines, chapter one to six.

I Conversion and aggregation of point-based source data into grids

This method can be used when national micro data is available. This method produces the most accurate and high quality grid data.

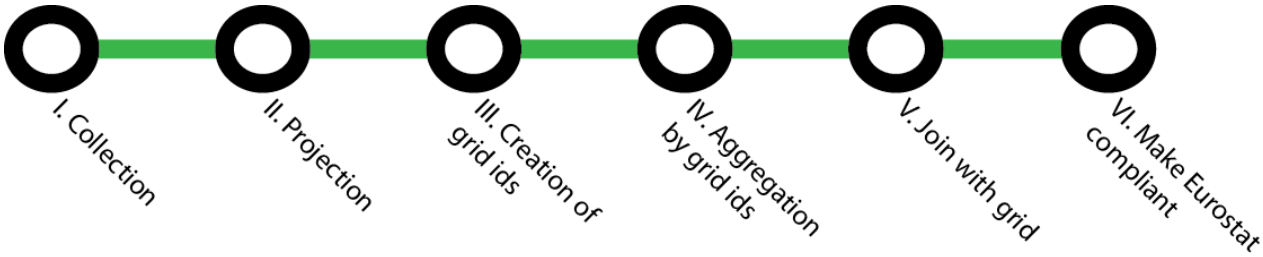


Fig 1. Process flowchart of micro data conversion and aggregation.

There are various ways that NSIs can convert and aggregate point-based micro data. The standard process is described in figure 1 including five phases. Using GIS software enables the reader to go from phase II to phase V directly.

I Collection

The data collection of micro data varies in between NSIs, but also within a NSI where data collection of various themes varies. Chapter 1 describes how the data collection and preparation can be carried out in a NSI. All NSIs will here download the European LAEA grid net and evaluate its spatial extent by comparing it to national borders.

II projection

Chapter 2.1 describes how to convert the coordinate system used for national micro data into the European coordinate system ETRS89-LAEA

III Creation of grid cell identification codes by calculating or by spatial join with the LAEA grid net

In order to generate grid cell identification codes several approaches are possible depending on the accessible tools. Chapter 3 explains how grid cell identification codes can be added to the reprojected grid data using calculations of tabular data or by joining the data spatially with European grid cells collected in phase I.

IV Aggregation of population figures by LAEA grid cells

Chapter 4 describes how the point based micro data can be aggregated (summed up) per grid identifier as generated in phase III.

V Join with the LAEA grid net

Chapter 5 gives a description of how the aggregated data can be joined with the European grid.

VI MAKING THE POPULATED GRID DATA COMPLIANT WITH THE DATA DESCRIPTION OF EUROSTAT

Chapter 7 gives a description of how the generated population data can become compliant with the data description defined by Eurostat.

II CONVERSION OF EXISTING NATIONAL GRID DATA INTO EUROPEAN GRID DATA (ETRS89-LAEA)

This method can be used for converting existing national grids into a European harmonised grid dataset. It is here important to be aware of the differences in coordinate systems in order to reduce the risk for errors (location and form) when reprojecting the data.

A. CONVERSION FROM NATIONAL TO EUROPEAN GRID DATA USING CENTROIDS

This method can be used when national grid data is available and when the grid cells are smaller than the grid cell size of the output grid (in this case the European LAEA grid). Moreover, the method gives good results when the national coordinate reference system has characteristics which are not too different from the LAEA coordinate system, i.e. when nationally defined square grid cells would not be distorted too much when reprojected into LAEA. This process is an alternative for NSIs where the grid data is heavy and the hardware and software capacity is limited. The reason is that working with point data is not as heavy as handling grids.

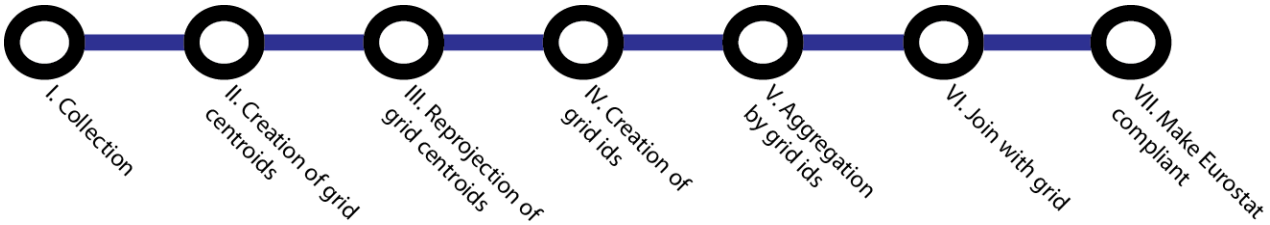


Fig. 2 Process flowchart of recasting grid data

Workflow phases and references to more detailed guidelines:

I COLLECTION

Chapter 1 describes how the data collection and preparation can be carried out in a NSI. All NSIs will here download the European LAEA grid net and evaluate its spatial extent by comparing it to national borders.

II CREATION OF GRID CENTROIDS

Chapter 2.2 describes how to transform grid data to centroids or central points of each grid cell

III REPROJECTION OF CENTROIDS INTO LAEA

Chapter 2.2 describes how to reproject the centroids from the national coordinate system to the LAEA coordinate system.

IV CREATING GRID CELL IDENTIFICATION CODES BY CALCULATING OR BY SPATIAL JOIN WITH THE LAEA GRID NET

In order to generate grid cell identification codes several approaches are possible depending on the accessible tools. Chapter 3 explains how grid cell identification codes can be added to the reprojected grid data using calculations of tabular data or by joining the data spatially with European grid collected in phase 1.

V AGGREGATION OF POPULATION FIGURES BY LAEA GRID CELLS

Chapter 4.2 describes how the data related to the grid cell centroids can be aggregated using GIS software.

VI Join with the LAEA grid net

Chapter 5 gives a description of how the aggregated data can be joined with the European grid.

VII MAKING THE POPULATED GRID DATA COMPLIANT WITH THE DATA DESCRIPTION OF EUROSTAT

Chapter 7 gives a description of how the generated population data can become compliant with the data description defined by Eurostat.

B. Conversion from national to European grid data using intersection

This method can be used when national grid data is available. This is an alternative to the recast method but may give better results if only 1 km national grid data is available comparing to the recast method. However, the amount of polygons to be handled during the process may get huge and that may make data processing heavier than in the case of the recast method. If data processing is an issue, the territory of a country could be split into parts for the sake of processing.

The intersection method is especially appropriate when the national coordinate system substantially differs from the LAEA coordinate system, e.g. for regions/countries located far away from the centre of the LAEA projection. In these cases, the intersection method will result in a more regular population distribution over the territory.

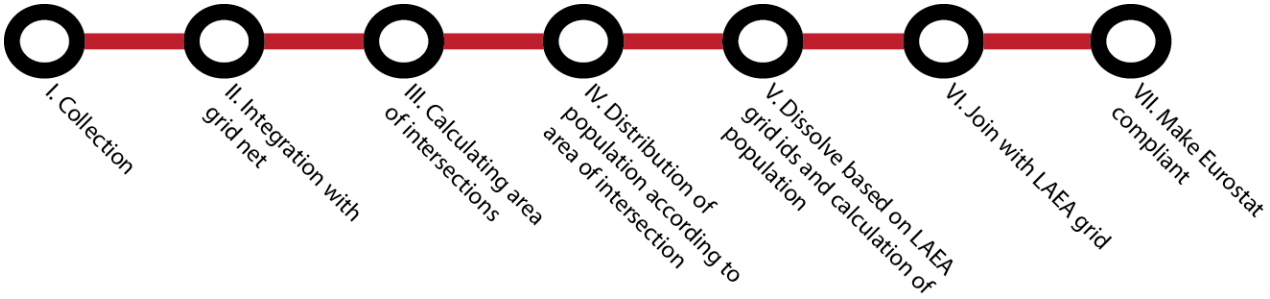


Fig. 3 Process flowchart of grid cell intersection.

Workflow phases and references to more detailed guidelines:

I Collection

Chapter 1 describes how the data collection and preparation can be carried out in a NSI. All NSIs will here download the European LAEA grid net and evaluate its spatial extent by comparing it to national borders.

II Phase II intersection with various grid nets to phase VI joining the data with the LAEA grid net.

Chapter 6 contains an illustrated description making it easy to follow phases II to VI.. From (II) intersection of various grid nets to (III) Calculating the area of intersections, (IV) Distribution of the population according to the area of intersection, (V) Dissolve based on LAEA grid cell identification codes and calculation of population and finally (VI) Joining the data to the LAEA grid net.

VII MAKING THE POPULATED GRID DATA COMPLIANT WITH THE DATA DESCRIPTION OF EUROSTAT

Chapter 7 gives a description of how the generated population data can become compliant with the data description defined by Eurostat.

GUIDELINES:

SOFTWARE USED

Various software has been tested for the various approaches and the aim of this chapter is to give an overview of the alternatives tested in GEOSTAT 1B and leaving it to the reader to choose the approach that suits her/him best. The software used are GIS Software (ArcGIS, gvSIG), Statistical software (SAS), Database Management software (SQL), and Microsoft Office software (MS Excel).

In order to facilitate the work for the reader efforts have been made to make the procedure more operational. The tools and training material are therefore to be placed at the website of www.EFGS.info.

1. COLLECTION

1.1 COLLECTING DATASETS

The first step is to collect the datasets needed for populating (database term for filling up) the LAEA grid net country clip. The datasets needed here are:

- Georeferenced national micro data or populated national grid data
- LAEA grid net country clip: Can be downloaded from http://www.efgs.info/data/european-datasets/eurogrid

1.2 ANALYSING MICRO DATA AND LAEA GRID NET COUNTRY CLIP

The original European grid net consists of grid cells covering the whole of Europe. The grid net has been clipped per country by Eurostat in order to make the dataset lighter and easier to handle when populating the dataset.

Before starting populating the LAEA grid net country clip it is important to validate the spatial extent of this dataset in order to ensure that the whole country is covered and that no data (e.g.addresses or buildings) will be missed out.

Best reference materials for testing are highly detailed country border data sets of each country, but correctly georeferenced micro data or national grid data can also be used.

1.2.1 ANALYSING THE SPATIAL EXTENT OF THE GRID NET GRID NET COUNTRY CLIP

Grid net should cover completely the area covered by reference material as borders, micro data or national grid. Grid cells covering areas outside the country borders do not matter as long as they are not too many and not too heavy to handle. These cells will not be populated in the later process described below. Worst is if grid cells are not covering the whole country, since this may result in the loss of statistical data.

Using spatial analysis it is possible to verify the coverage of the "country clip" with national reference data. Firstly, it is important to make sure that the projections correspond. This is explained in chapter 2. Secondly, verify that the micro data fall inside the national borders. This can be done in ArcGIS and gvSIG:

ArcGIS:	**gvSIG version 1.12.0 Final:**
1. Add national borders and micro data of interest	1. Add national borders and micro data of interest
2. Carry out the spatial selection and verify the data:	2. Select all features in the national boundary data:
A. Go to Selection > Select By Location	A. Activate the data layer of national boundaries in the list of contents.
B. In the dialog box enter:	B. Go to View > Selection > Select all

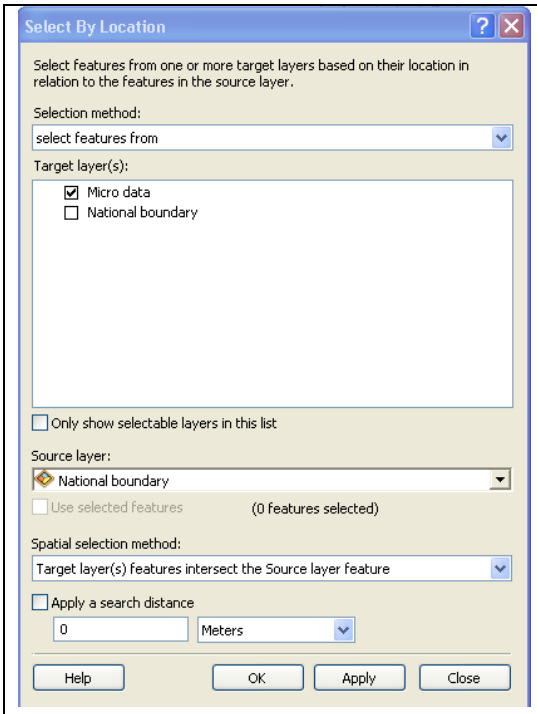


Fig. 4 Select by location tool in ArcGIS

C. In the table (right click on Micro data layer and select "Open attribute table") verify that all data is selected

C. Activate the data layer: micro data of interest

3. Carry out the spatial selection and verify the data:

D. Go to View > Selection > Spatial selection

E. Before clicking on the "New set" button, choose: "Select features from active layers that "Intersect with" selected features in layer: "National boundaries"

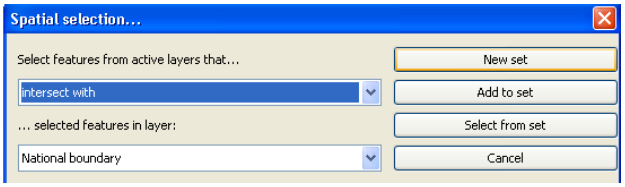


Fig. 5 Spatial selection tool in gvSIG

F. In the Micro data table (View > Show attribute table) verify that all data is selected

Thereafter, add the "country clip" in the GIS software and follow the spatial selection procedure again with following modifications:

ArcGIS:	**gvSIG version 1.12.0 Final:**
1. Add the "country clip" grid 2. Change the Selection method in the "Select By Location" dialog box to be: "Select from currently selected features in". Change "Source Layer" to be the "country clip" 3. In the table (right click on Micro data layer and select "Open attribute table") control the selected data	1. Add the "country clip" grid 2. Select all features in the "country clip": 3. In the "Spatial selection" dialog box set "… selected features in layer:" to be "Country clip" and execute the selection clicking on the "Select from set" button 4. In the Micro data table (View > Show attribute table) control the selected data

In case there are reference data left out, the country clip should have extra cells added. Please contact Eurostat (ekkehard.petri@ec.europa.eu)

2. REPROJECTION

Projection is a method of representing the surface of a sphere or other three-dimensional body on a plane[1]. This transformation results in a distortion and varies depending on the purpose of the map. In order to preserve the sphere-like body properties each country has various ways to project national map data. This is also the case on a European level. In order to populate the European LAEA grid the national point-based microdata or grid data need to be reprojected into ETRS89/LAEA.

Reprojection can be done spatially in GIS applications but also by other coordinate conversion applications (usually launched by national mapping agencies). This chapter presents how to carry out reprojection using GIS software both for point - and grid net data.

2.1 REPROJECTION OF NATIONAL MICRO DATA

For point-based micro data, reprojection can be done in both ArcGIS and gvSIG .

ArcGIS:

1. To reproject data from national projection to ETRS89-LAEA using Project tool. See figure below.

A. In ArcToolbox go to Data Management > Tools > Projections and Transformations > Project

B. Enter the micro data as input and verify that the input coordinate system is correct. Define the output dataset and coordinate system.

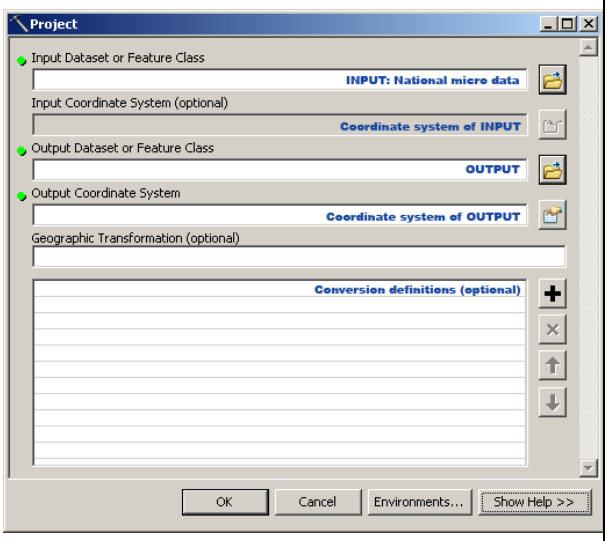


Fig. 4 Projection tool in ArcGIS

2. Once the data are reprojected it is important to identify the next steps in process and what tools to be used. In case a tabular approach (see chapter 3) will be used for generating grid cell identification codes it is necessary to generate x

gvSIG version 1.12.0 Final:

1. To reproject data from national projection to ETRS89-LAEA using Reproject tool. See figure below.

A. Verify that the projection of the micro data set was defined correctly once opened in gvSIG. Right click on the data in the table of contents and choose "Properties". Otherwise, open the data View > Add layer and set current projection.

B. Go to Reproject tool (View > geoprocessing tools > Data conversion > Reproject)

C. Enter the micro data as input and verify that the input coordinate system is correct. Define the output dataset and coordinate system.

D. In order to ensure correct reprojection define the transformation code. In this case EPSG 1149 was used to convert WGS84 to ETRS89.

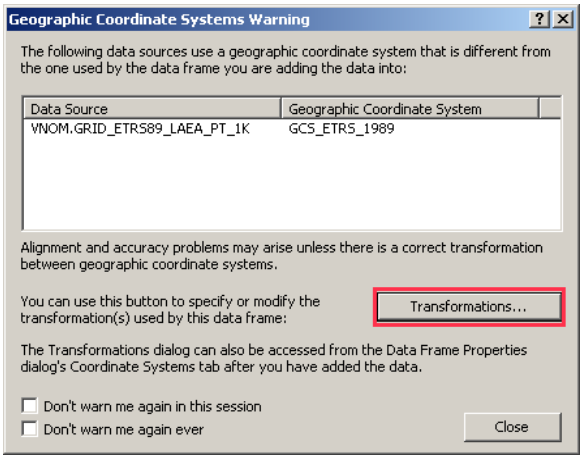
Be aware that gvSIG doesn't generate projection files .prj as ArcGIS, but projection needs to be defined in the gvSIG when adding the dataset to a view. In order to recognize the projection add the EPSG code in the file name e.g. Micro_data_3035.shp

[1] http://en.wikipedia.org/wiki/Map_projection

and y coordinates for the newly reprojected point dataset. With a GIS approach this is not necessary:

A. In the table view of the newly reprojected dataset click on the option button and select "Add field" to the table. Set the first field name to be "X" and the second to be "Y" and set the datatype to "Long integer". Right click thereafter on the field heading and select "Calculate Geometry". Set here "X coordinate" or "Y coordinate" property of the field and "meters" as units.

In ArcGIS it is also possible to make a projection "on the fly" in a dataframe of ArcMap. If the datum of the local projection hasn't datum ETRS89 a transformation message will be shown where it is possible to identify the transformation parameters.



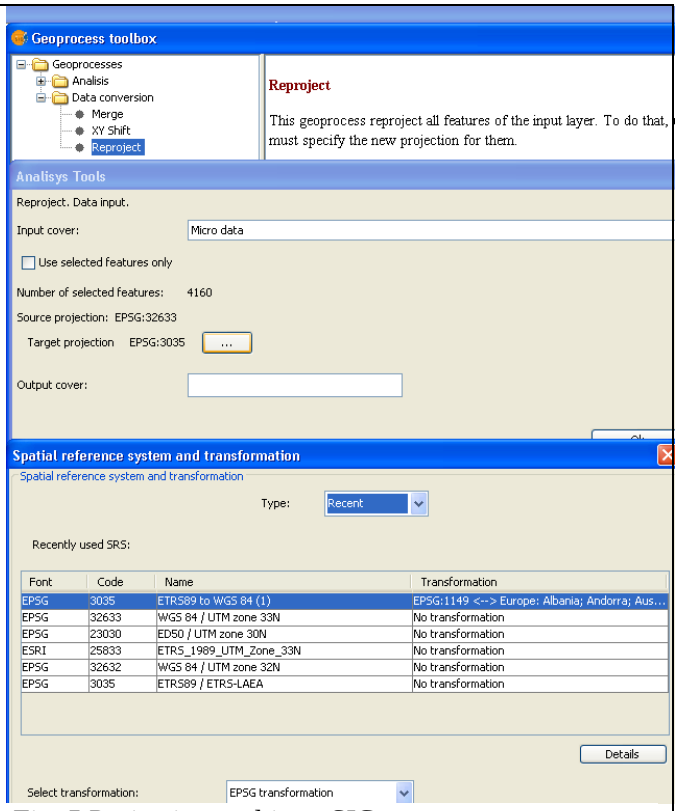


Fig. 5 Projection tool in gvSIG

2. Once the data are reprojected it is important to identify the next steps in process and what tools to be used. In case a tabular approach (see chapter 3) will be used for generating grid cell identification codes it is necessary to generate x and y coordinates for the newly reprojected point dataset. With a GIS approach this is not necessary:

A. In the Sextante Toolbox go to Sextante > Tools for point layers > Add coordinates to points

B. Enter the newly reprojected micro dataset as input and define a new output name.

2.2 Reprojection of national grid cell centroids

In order to have most reliable results, it is recommended that the spatial data layers used, have a same coordinate system and projection.

When the source data is national ready-made grid data, the grids cells have to be transformed into centroids first. Grid cell centroids must be reprojected before further phases.

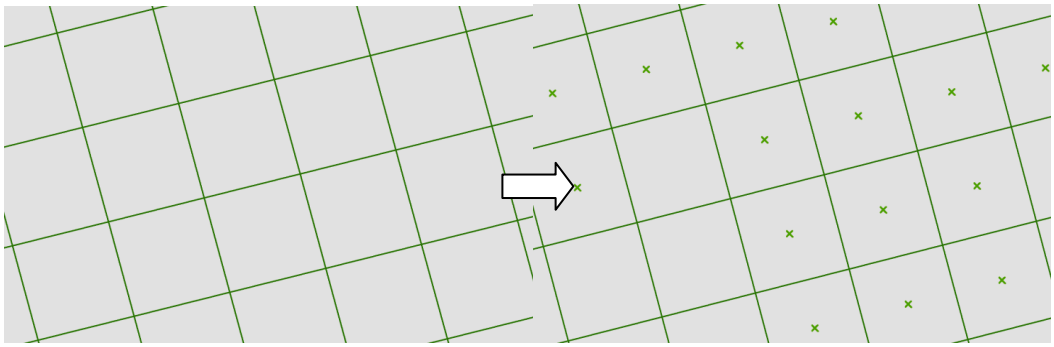
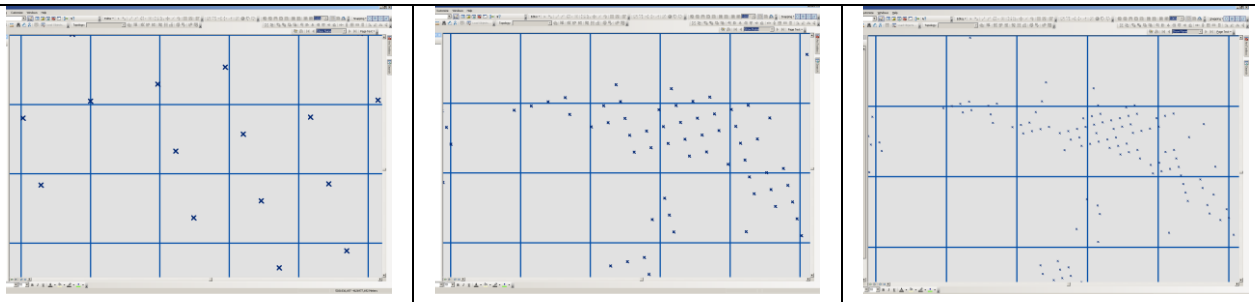


Fig. 6 Grid data have to be transformed into centroids first. The centroids are then reprojected into ETRS89-LAEA.

When ready-made grid data is used as source data, it should be noted that the grid cell size of national grid data determines the accuracy of derived data later. Basic rule is, that the finer the original grid data is the more precise will derived data become. For example, if we are producing 1 km harmonized grid data, having ready-made grid data as source data, 1 km grid cell centroids are fewer and more sparsely distributed, as 125m grid cell centroids. The spatial accuracy improves when we have more source data "spots".



1000 m grid cell size	250 m grid cell size	125 m grid cell size

Fig. 7 The grid cell size of national grid data determines the accuracy of derived grid data.

2.2.1 GRIDS CELLS TO CENTROIDS AND REPROJECTION IN GIS SOFTWARE

The centroid of grid cells can be generated in both ArcGIS and gvSIG:

ArcGIS:	**gvSIG version 1.12.0 Final:**
1. There are various ways to generate centroids in ArcGIS. The easiest way to carry out this work is by using the "Mean Center" tool. See figure below or appendix 1 for other alternatives. A. In ArcToolbox go to Spatial Statistics Tools > Measuring Geographic Distributions > Mean Center B. Enter the national grid data as input and the centroid layer as output and define the Grid identifier as "Case field". In order to produce centroids do not enter any weights.	1. In the toolbox "Sextante" there is a tool named "Centroids" that is easy to use. A. In the Sextante Toolbox go to Sextante > Tools for vector layers > Centroids B. Enter the national grid data as input and the centroid layer as output.

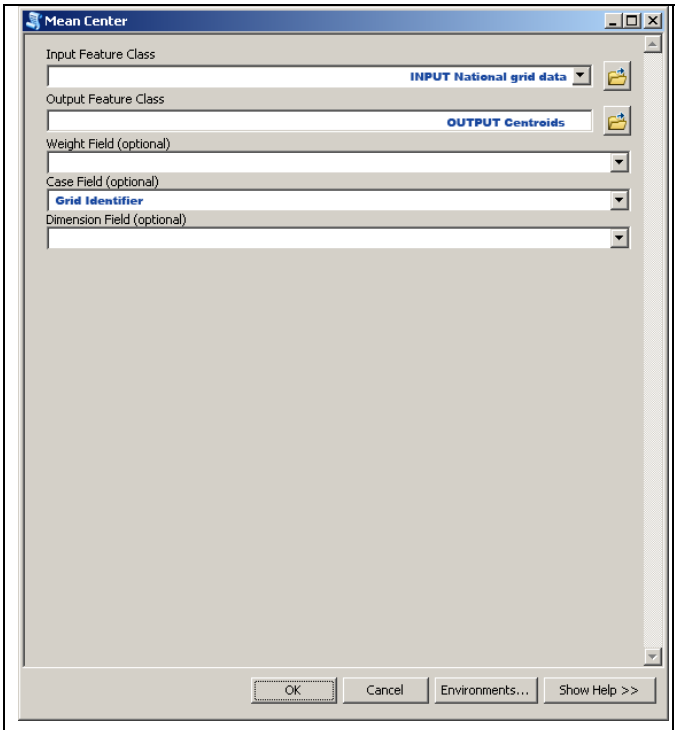


Fig. 8 Mean Center tool in ArcGIS

2. Once the centroid points are generated follow the reprojection instructions under chapter 2.1 above.

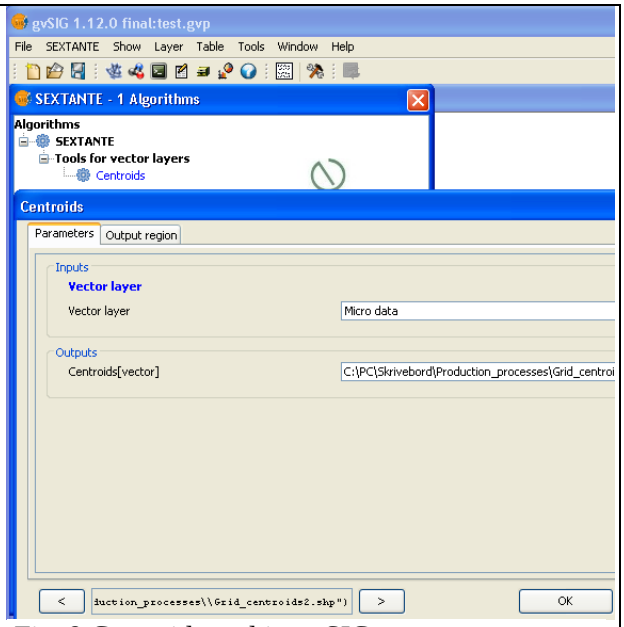


Fig. 9 Centroids tool in gvSIG

2. Once the centroid points are generated follow the reprojection instructions under chapter 2.1 above.

2.3 Reprojection of the country clip of LAEA grid net to national coordinate system

Another alternative is to reproject the LAEA grid to the national coordinate system. For this it is possible to use the same procedures as described in chapter 2.
All the geographical operations will then be carried out using the national projection and it is not necessary to project the different data sources used. Afterwards the results for each GRID cell can be linked using the Grid cell identification code as a common key.

3. Generating grid cell identification codes

Grid cell identification codes are based on grid cell's lower left-hand corner coordinates truncated by grid cell size (e.g. 1kmN4534E5066 is result from coordinates Y=4534672, X=5066332 and the cell size 1000)

This chapter will describe two main approaches:

Tabular data approach: Working with tabular data is recommended when working with large grid nets and/or when the processing capacity of software and hardware is limited.

GIS approach: following the same structure as in the previous phases with the spatial functions in ArcGIS parallel to the corresponding functions in gvSIG.

On micro data level the grid cell identification code can be calculated from coordinates of e.g. building or address points.

In order to create grid cell identification codes for grid data in LAEA (tabular or spatial data), the grid cell's lower left-hand coordinates must be derived first from the data. These coordinates are used as inputs. If data hasn't got coordinates in separate columns, you must create those first.

3.1 TABULAR DATA APPROACH

Tabular data approach generates grid cell codes using calculations and this chapter gives an overview of some tools and languages that can be used, Working with tabular data is also possible in GIS software and these examples are to be found in appendix 2.

3.1.1 GENERATING GRID CELL IDENTIFICATION CODES FOR NATIONAL MICRO DATA USING SAS

When using SAS, you may use this program including the floor function rounding down the coordinates:

```
data micro_data_id(keep=GRD_ID Y_coord X_coord);
     set micro_data ;
     length y  x $4.;
     yn=floor(Y_coord/1000);   /* rounding down coordinates  */
     xn=floor(X_coord/1000);
     x=xn;                     /* converting to characters */
     y=yn;
     GRD_ID = "1kmN" !! strip(y) !! "E" !! x; /* defining grid id */
run;
```

3.1.2 GENERATING GRID CELL IDENTIFICATION CODES FOR NATIONAL MICRO DATA USING SQL

When using e.g. MS SQL Server Management Studio, you may use this expression:

```
SELECT  '1kmN' + CONVERT(varchar(4),CAST(grid_dataset.Y_coord /1000 as integer))
            + 'E' + CONVERT(varchar(4),CAST(grid_dataset.X_coord /1000 as integer)) as
       GRD_ID
FROM     micro_data
```

3.1.3 GENERATING GRID CELL IDENTIFICATION CODES USING EXCEL

Newer versions of Microsoft Office Excel allow working with higher number of rows and opening dbf file in Excel might be an alternative to generate the grid cell identification code in a GIS software. It is here important to change the "Defined name" in order to change the range of columns. This needed to be able to add new columns to the dbf file. In Excel 2010 this can be done by redefining the "Defined Name" formula or by:

a. selecting the range of data that you want to name,

b. on the "Formula tab", in the "Defined Names" group, click "Create from Selection" and

c. In the "Create Names from Selection" dialog box, designate the location that contains the labels by selecting the Top row, Left column, Bottom row or Right column check box[2].

[2] http://office.microsoft.com/en-us/excel-help/define-and-use-names-in-formulas-HA010342417.aspx

Grid cell identification code can be created in Excel by using this macro. Define xField, yField and calcField before running the macro. The x and y coordinates will be pasted in starting on row 2, otherwise change the “i” value.

```
Sub GRID_CODES()
' GRID_CODES Macro
    Dim i, xField, yField, calcField, xRow, yRow, calcRow As Integer

    Application.ScreenUpdating = False  ' Unactivate screenupdate
    Application.DisplayAlerts = False  ' Unactivate alerts

    xField = 1  ' Set the field with x coordinates 1 equals field A
    yField = 2  ' Set the field with y coordinates 2 equals field B
    calcField = 7 ' Set the field for calculations 7 equals field G

    Cells(1, calcField) = "FLOOR_X"
    Cells(1, calcField + 1) = "FLOOR_Y"
    Cells(1, calcField + 2) = "LEFT_X"
    Cells(1, calcField + 3) = "LEFT_Y"
    Cells(1, calcField + 4) = "GRID_ID"

    i = 2
    Cells(i, xField).Select
    Do Until Cells(i, xField) = "" ' do until empty string in xfield

       Cells(i, calcField) = Excel.WorksheetFunction.Floor(Cells(i, xField), 1000)
       Cells(i, calcField + 1) = Excel.WorksheetFunction.Floor(Cells(i, yField), 1000)
       Cells(i, calcField + 2) = VBA.Strings.Left(Cells(i, calcField), 4)
       Cells(i, calcField + 3) = VBA.Strings.Left(Cells(i, calcField + 1), 4)
       Cells(i, calcField + 4) = "1kmN" + Str(Cells(i, calcField + 2)) + "E" + Str(Cells(i, calcField + 3))
       Cells(i, calcField + 4) = VBA.Strings.Replace(Cells(i, calcField + 4) , " ", "")

    i = i + 1
    Loop
End Sub
```

3.2. GENERATING GRID CELL IDENTIFICATION CODES FOR NATIONAL MICRO DATA OR NATIONAL GRID DATA USING GIS SOFTWARE

The GIS software approach is here described with ArcGIS and gvSIG, but there are also other tools and possible alternatives. Some more examples are to be found in appendix 2.

ArcGIS:

1. There are various ways for generating grid cells identification codes in ArcGIS. The easiest way to carry out this work is by using the "Spatial join" tool. See figure below or appendix 2 for other alternatives.

A. In ArcToolbox go to Analysis Tools > Overlay > Spatial Join

B. Enter the point dataset as target features and the "country clip" as join features. Set "Intersect" to be the match option:

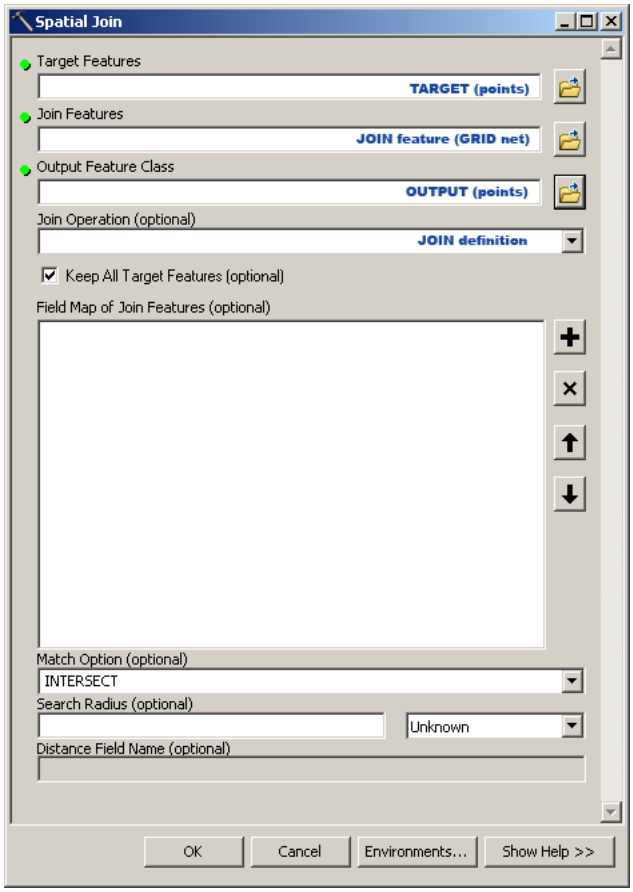


Fig. 10 Spatial join tool in ArcGIS

This command can also be used in a script environment as Python:
arcpy.SpatialJoin_analysis("TARGET "Points", "JOIN feature (GRID net)", "Output (points)", "JOIN_ONE_TO_ONE", "KEEP_ALL", "", "COMPLETELY_WITHIN")

gvSIG version 1.12.0 Final:

1. The easiest way to generate grid cells identification codes in gvSIG is by using the "Spatial join" tool. See figure below. This can be a heavy process and not manageable by gvSIG. Generating grid cell identification code using Excel can here be an alternative.

A. Verify that the projection of the micro data set was defined correctly once opened in gvSIG. Right click on the data in the table of contents and choose "Properties". Otherwise, open the data View > Add layer and set current projection.

B. Go to Spatial join tool (View > geoprocessing tools > Analysis > Spatial join)

C. Enter the point dataset micro data as input cover and the "country clip" as join features. Do <u>not</u> Select the tic box "Use nearest geometry" and Name the output cover. By not selecting "Use nearest geometry" the user selects intersection as transfer criteria.

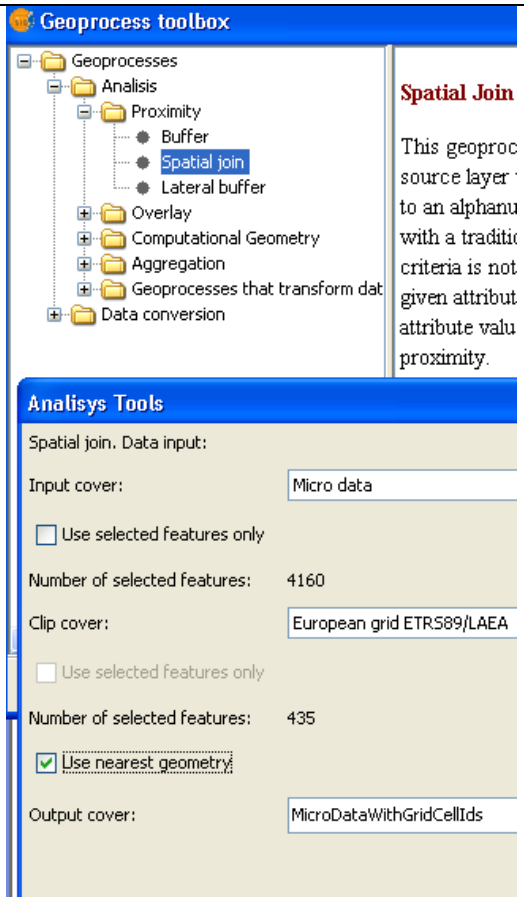


Fig. 11 Spatial join tool in gvSIG

4. AGGREGATION OF POPULATION FIGURES BY LAEA GRID CELLS

By generating the grid cell codes these identifiers can be used for aggregating the micro data by in a grid cell. As in chapter 3 this chapter describes two approaches; tabular data - and GIS software approach.

4.1 AGGREGATING MICRO DATA USING TABULAR DATA APPROACH

This chapter describes how SAS and SQL can be used for aggregating the data by grid cells using the identifiers created in the previous chapter

4.1.1 AGGREGATING MICRO DATA USING SAS

When using SAS you can use SAS procedures (e.g. PROC SUMMARY) or program code as below. Input is the micro data with LAEA grid identifier as generated in chapter 3.1.

```
proc sort data=micro_data_id;
 by GRD_ID;
run;


data grid_dataset_POP_sum(KEEP=GRD_ID POP_TOT);
    set micro_data_id;
    by GRD_ID;
    if first.GRD_ID then do;
```

```
        POP_TOT=0;
    end;
    POP_TOT+1;
    if last.GRD_ID then do;
        format GRD_ID $12. POP_TOT 8.;
        output;
    end;
run;
```

4.1.2 Aggregating micro data using SQL

Based on the micro data with LAEA grid identifier generated in 3.2 aggregation is possible using following query:

```
select GRD_ID, count(*) as POP_TOT
into #grid_dataset_POP_sum
from micro_data_id
group by GRD_ID
order by GRD_ID
```

4.2. Aggregating of grid cell centroids using GIS software

ArcGIS:	**gvSIG version 1.12.0 Final:**
1. Data by grid cell can be aggregated in ArcGIS using the Summary Statistics tool. See tool interface in the figure below. This phase can be avoided by proceeding to chapter 5.2 directly, making use of the Spatial Join tool. A. In ArcToolbox go to Analysis Tools > Statistics > Summary Statistics B. As input table enter the table that is generated with the tabular approach or the point dataset generated using the chapter 3.2. approach. Set the population field to be summed up and the case field to the GRID ID. Define the output table name:	1. Data by grid cell can be aggregated in ArcGIS using the Summary Statistics tool. See tool interface in the figure below. A. Aggregation in gvSIG: Summarize table tool. Open the table using following icon in the menu bar  . Once the table is displayed go to  in the menu bar. Choose firstly the “group by” field and select the “Sum” tick box next to the field that you would like to sum. Define thereafter the output table name. The summarize process will run in the background once the “ok” button is selected and shaded. The process is done once the “Summarize table” display is closed down.

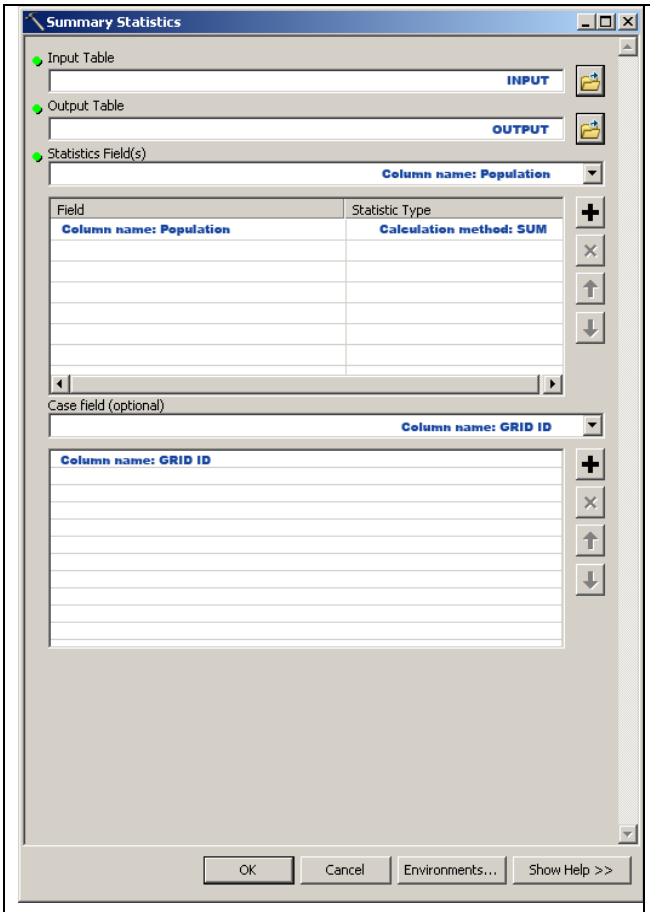


Fig. 12 Summary Statistics tool in ArcGIS

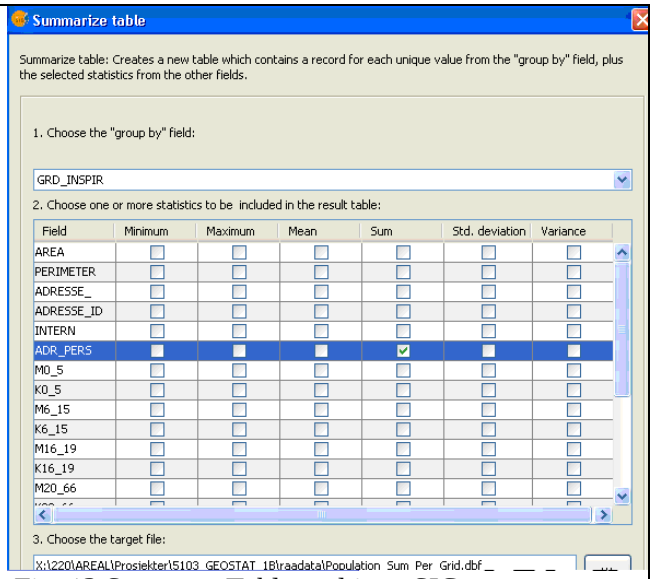


Fig. 13 Summary Table tool in gvSIG

5. JOIN WITH THE "COUNTRY CLIP LAEA GRID NET"

5.1 JOIN GRID CELL CENTROIDS WITH THE "COUNTRY CLIP LAEA GRID NET" USING GIS SOFTWARE

ArcGIS:

1. In order to join the two datasets in ArcGIS use Join fields. Join the attributes in the new aggregated table with the ETRS89-LAEA grid. See tool interface in the figure below.

A. In ArcToolbox go to Data Management Tools > Joins > Join Field

B. As input dataset enter the "country clip" and specify GRID_ID to be the join field. As join table add the aggregated table and the GRID_ID as join field. Select thereafter from the field list the fields to be included in the new joined dataset.

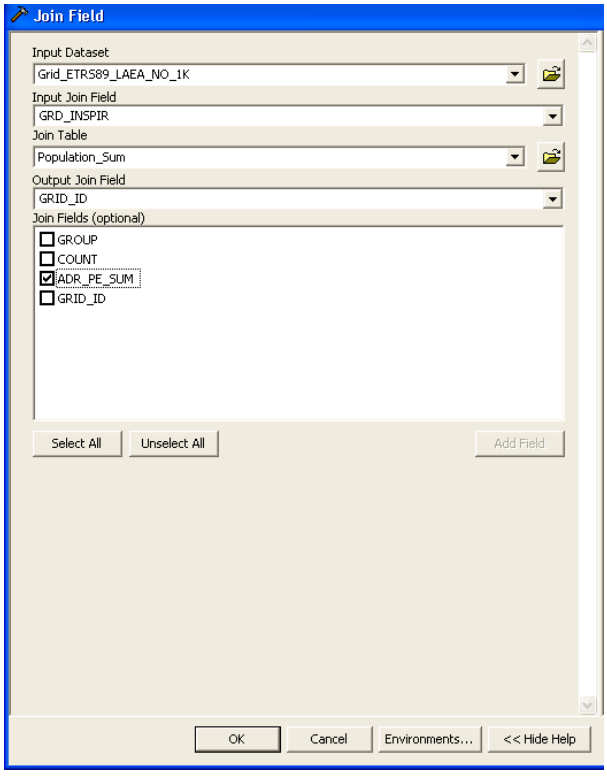


Fig. 14 Join Fields tool in ArcGIS

gvSIG version 1.12.0 Final:

1. In order to join the two datasets in gvSIG use Join fields. Join the attributes in the new aggregated table with the ETRS89-LAEA grid. See tool interface in the figure below.

A. Open the table using following icon in the menu bar  "Attribute Table" and go to Table > Join.

B. When joining the table choose first the ETRS89-LAEA grid and the GRD_ID field to be used in the join. In the next display choose the aggregated table and field to use in the join. The process will run in the background once the "ok" button is selected and shaded. The process is done once the "Join tables" display is closed.

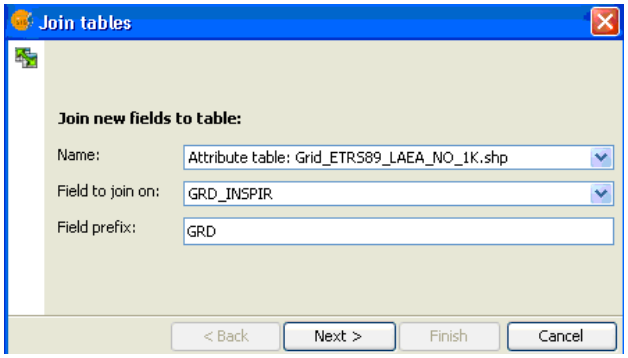


Fig. 15 Join tables tool in gvSIG

5.2 AGGREGATING AND JOINING GRID CELL CENTROIDS USING SPATIAL JOIN IN ARCGIS

It is also possible to make use of the spatial join in ArcGIS for aggregating the centroid data and joining the data with the "Country clip LAEA grid net" in one go.

A. In ArcToolbox go to Analysis Tools > Overlay > Spatial Join

B. Enter the “Country clip LAEA grid net” as target features and the centroid points as join features. Set “join one to one” as join operation and “Intersect” to be the match option:

C. In order to aggregate data with the Spatial Join tool use the Merge Rule by clicking the right mouse button on the field of interest in the field map (see figure below), Select here ”Sum“ as merge rule. As match option choose ”Intersect“ or alternatively ”Contain“. The input for this tool is a grid net with grid codes and a centroid layer with statistical information to be aggregated at the same time.

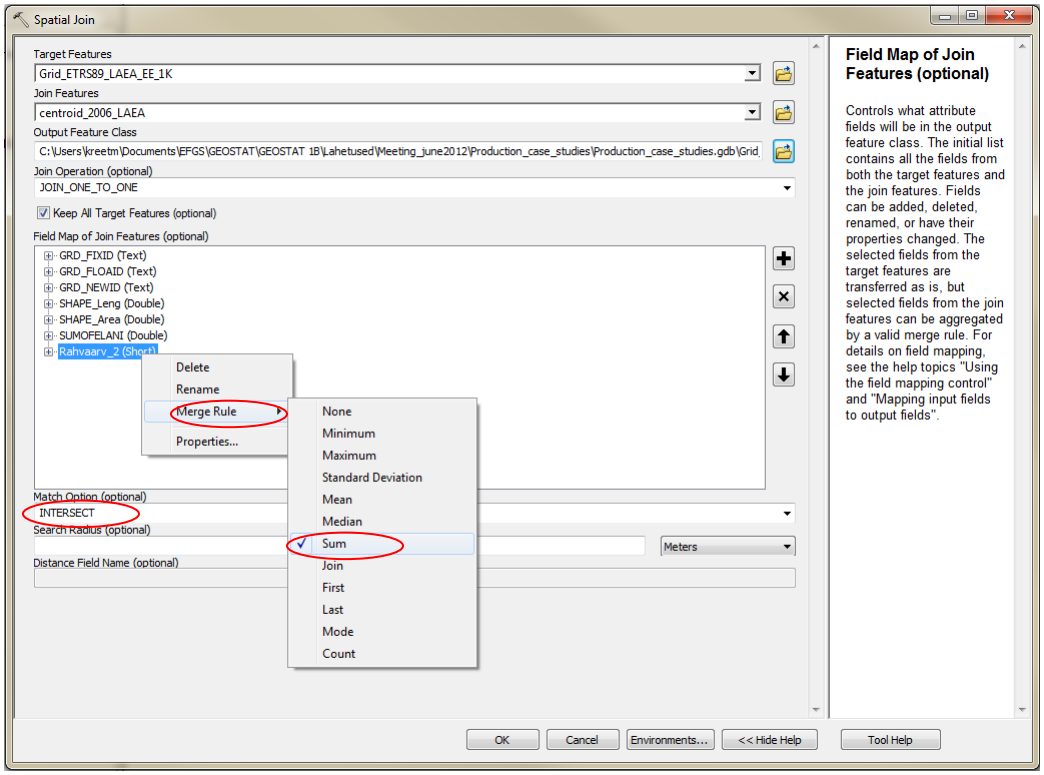


Fig. 15 Spatial Join tool in ArcGIS with “Merge Rule”: “Sum” and “Match Option”: “Intersect“

6. CREATING GRID CELL INTERSECTIONS AND COUNTING POPULATION SUMS USING GIS SOFTWARE

Phases of the method:

ArcGIS:	**gvSIG version 1.12.0 Final:**
1. Reproject the “country clip” (stored in national coordinate system) with associated population into LAEA projection	1. Reprojection of the “country clip” (stored in national coordinate system), follow explanations under chapter 2.1

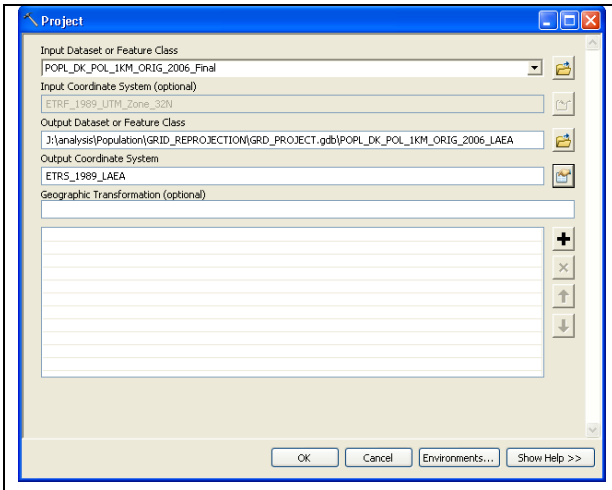


Fig. 16 Reprojection of the "country clip"

ArcGIS:

2 Calculate area in a new field

A. Enter the attribute table of the selected dataset and click on the "Options" button and select "Add field". Enter a name and set type to be "float".

B. Click on the title heading of newly created field. Select "Calculate Geometry". Add thereafter the same inputs as in figure 17.

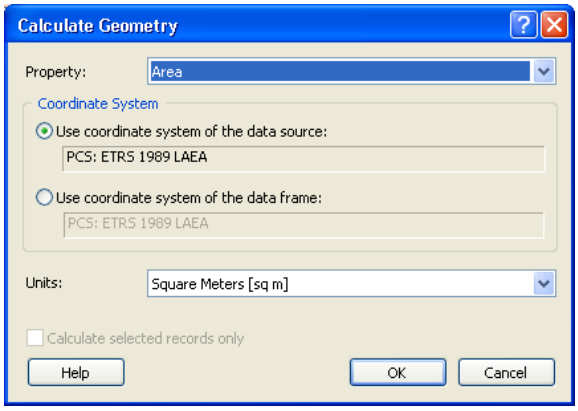


Fig. 17 Calculate geometry tool in ArcGIS

gvSIG version 1.12.0 Final:

2. Calculate area in new field

A. Active the dataset of interest in the table of contents. Go to Layer > "Add geometric info" in the menu bar.

B. Select layer to be calculated and move "Area" over to the right had window as in figure 18.

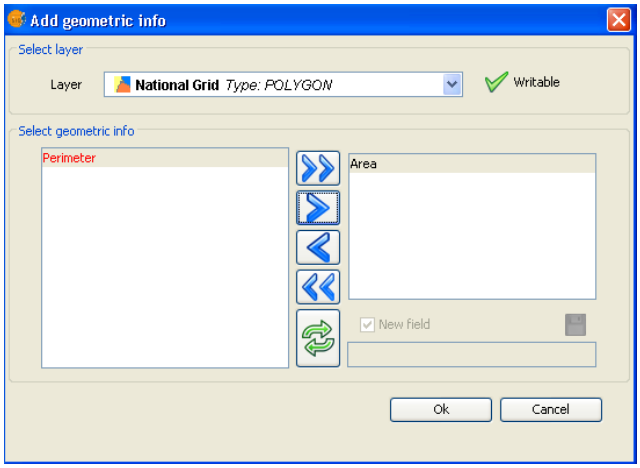


Fig. 18 Add geometric info tool in gvSIG

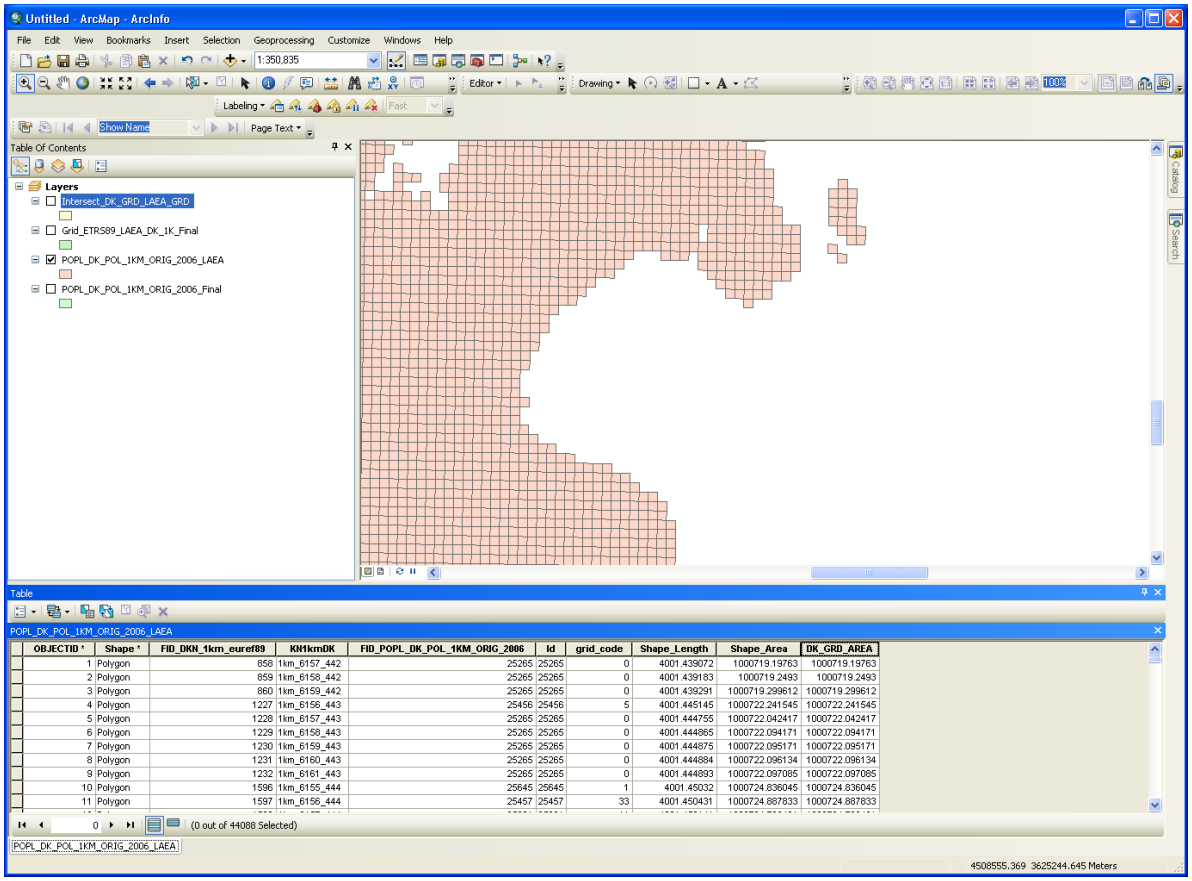


Fig. 19 The view and table including calculated area of the reprojected "country clip"

ArcGIS:

3. Intersect the LAEA reprojected “country clip” with the original “country clip”

A. In ArcToolbox go to Analysis Tools > Overlay > Intersect. Add thereafter the same inputs as in figure 20.

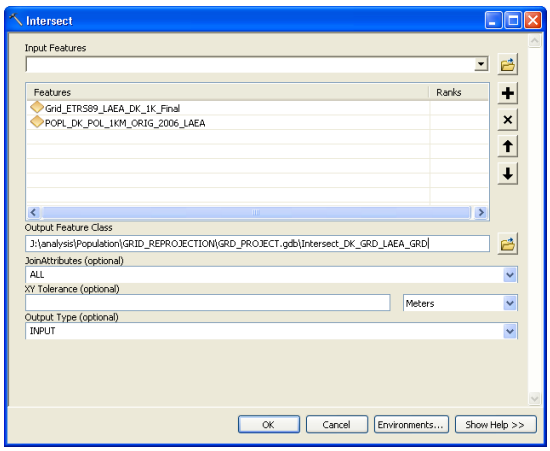


Fig. 20 Intersect tool in ArcGIS

gvSIG version 1.12.0 Final:

3. Intersect the LAEA reprojected “country clip” with the original “country clip”

A. Go to Intersect tool (View > geoprocessing tools > Analysis > Overlay > Intersect). Add thereafter the same inputs as in figure 21.

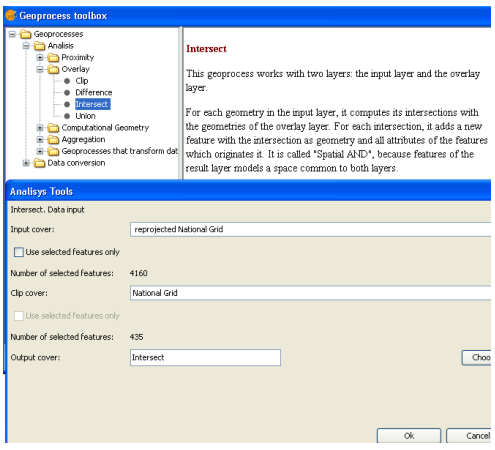


Fig. 21 Intersect tool in gvSIG

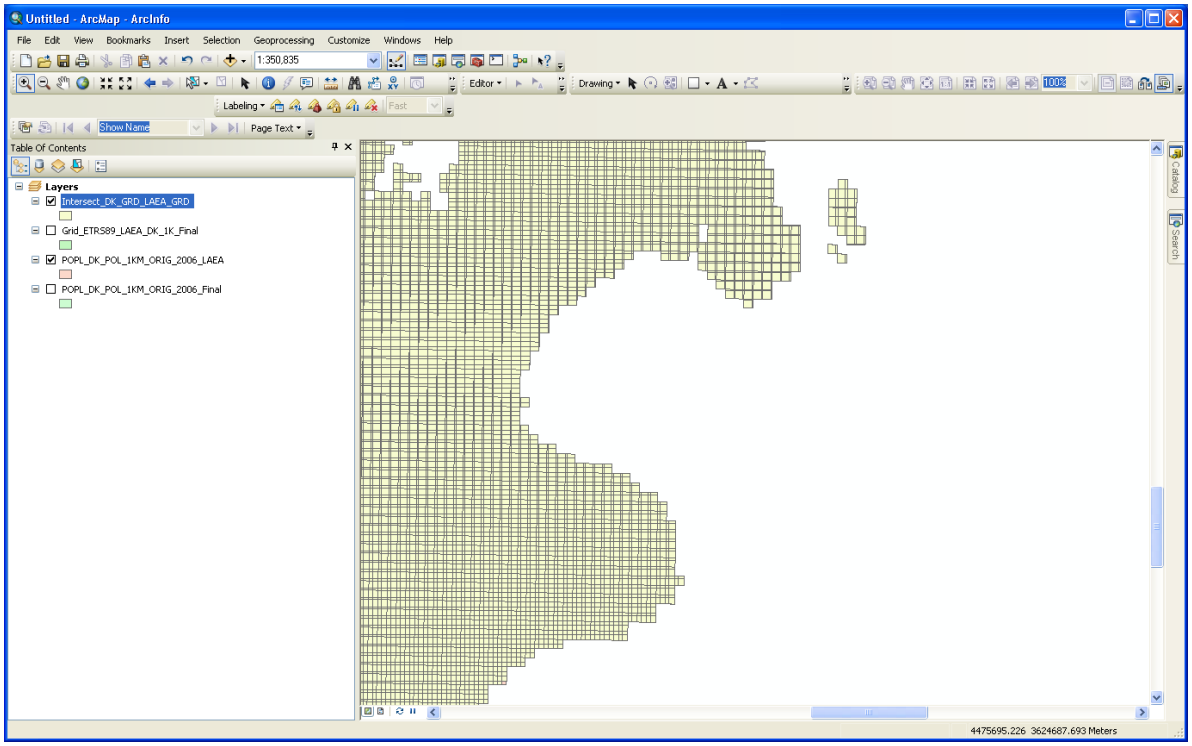
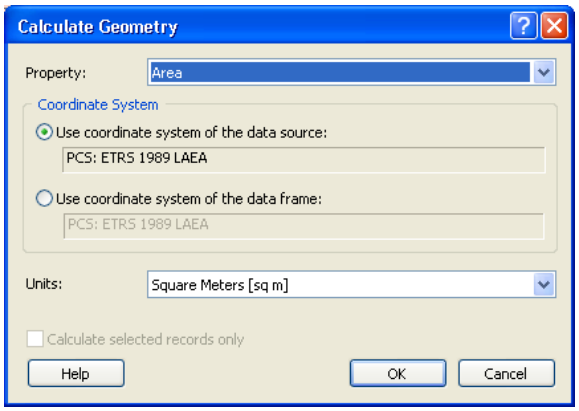
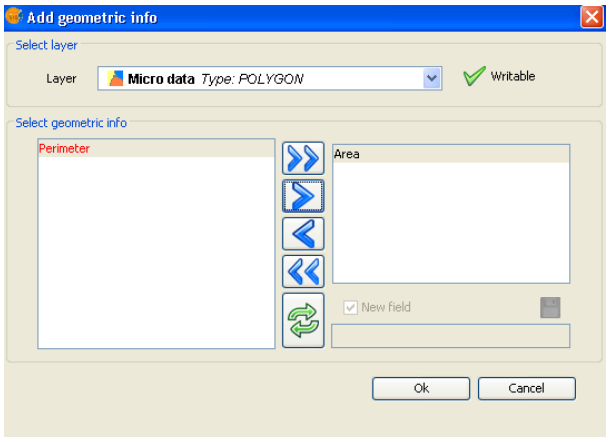
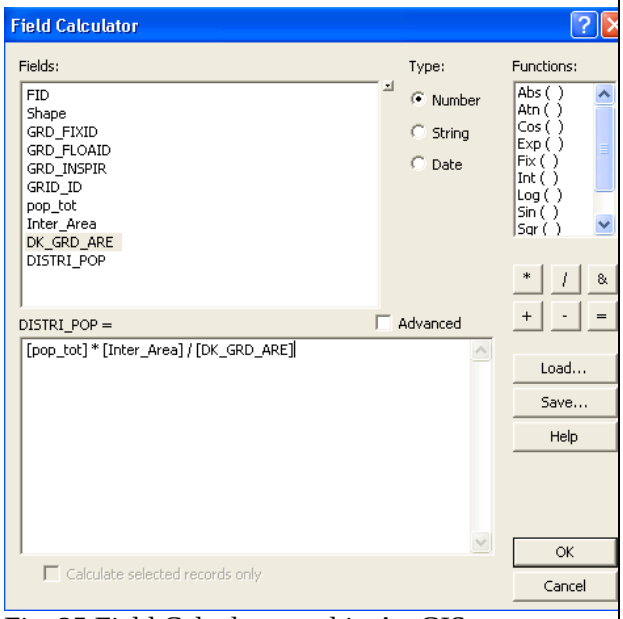
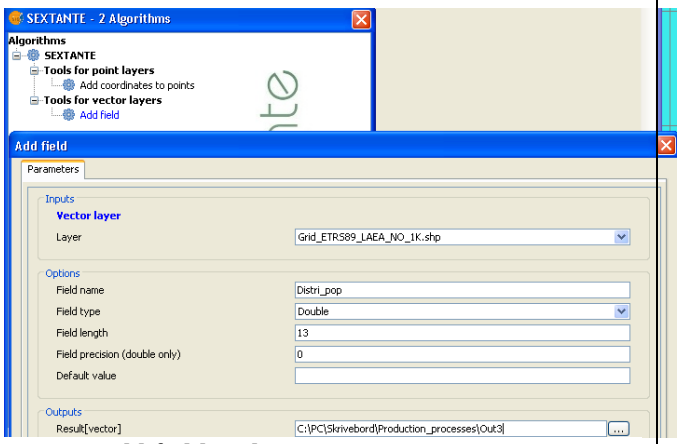
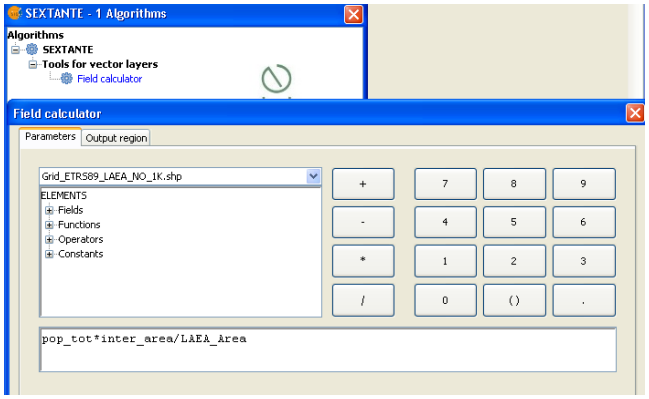


Fig. 22 The view illustrating the result of the intersection

ArcGIS:	**gvSIG version 1.12.0 Final:**
4. Calculate area of the intersected dataset by following description A. Enter the attribute table of the selected dataset and click on the "Options" button and select "Add field". Enter a name and set type to be "float". B. Click on the title heading of newly created field. Select "Calculate Geometry". Add thereafter the same inputs as in figure 23.  Fig. 23 Calculate geometry tool in ArcGIS	4. Calculate area of the intersected dataset by following description A. Active the dataset of interest in the table of contents. Go to Layer > "Add geometric info" in the menu bar. B. Select layer to be calculated and move "Area" over to the right had window as in figure 24.  Fig. 24 Add geometric info tool in gvSIG

ArcGIS:	**gvSIG version 1.12.0 Final:**
5. Create and calculate a new field in order to	5. Create and calculate a new field in order to

distribute the population according to the area of the intersection. This calculation is based on the assumption that the distribution of population inside each original national grid cell is uniform. A. Enter the attribute table of the selected dataset and click on the "Options" button and select "Add field". Enter a name and set type to be double or "float". B. Click on the title heading of newly created field. Select "Field Calculator". Add thereafter the following formula: [original population]*[area of intersection]/[area of the "country clip"national grid cell in LAEA]. See also figure 25.  Fig. 25 Field Calculator tool in ArcGIS	distribute the population according to the area of the intersection. This calculation is based on the assumption that the distribution of population inside each original national grid cell is uniform. A. Add a new field. In the Sextante Toolbox go to Sextante > Tools for vector layers > Add field  Fig. 26 Add field tool in gvSIG B. Calculate field. In the Sextante Toolbox go to Sextante > Tools for vector layers > Field calculator  Fig. 27 Field calculator tool in gvSIG

ArcGIS: 6. Dissolve the intersected features, based on the LAEA grid identifier. While dissolving, calculate the sum of the distributed population. A. In ArcToolbox go to Data Management Tools > Generalization > Dissolve. Select GRID_ID as dissolve field and choose the statistic type "Sum" of the calculated distributed population as in figure 28.	**gvSIG version 1.12.0 Final:** 6. Dissolve the intersected features, based on the LAEA grid identifier. While dissolving, calculate the sum of the distributed population. A. Add a new field. In the Sextante Toolbox go to Sextante > Tools for polygon layers > Dissolve. In the field list select "sum" of the calculated distributed population. See figure 29.

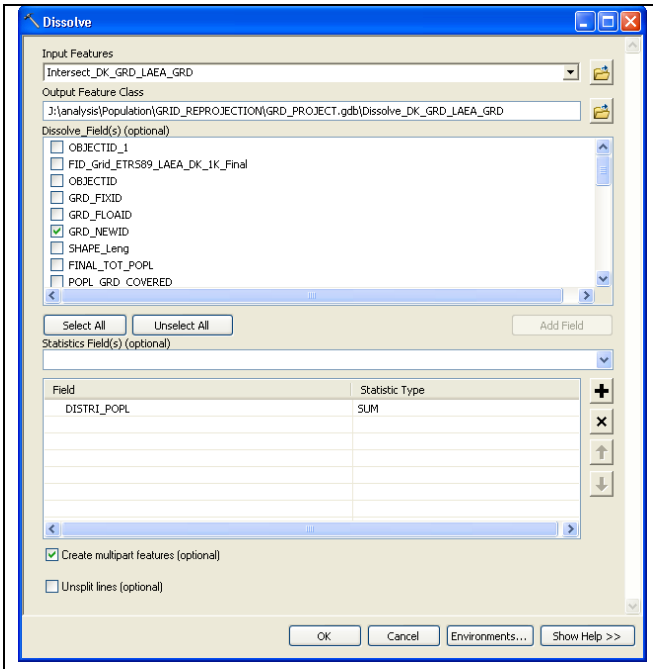


Fig. 28 Dissolve tool in ArcGIS

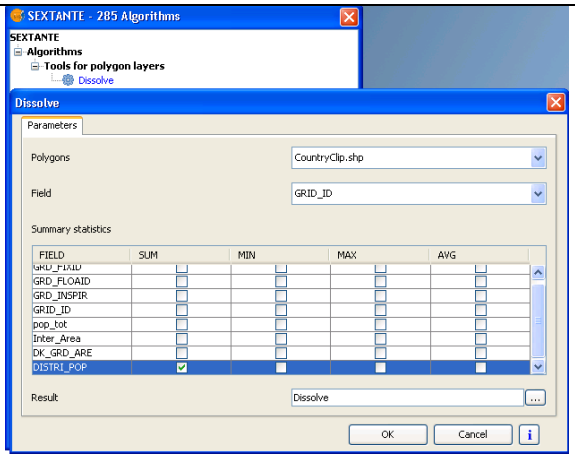


Fig. 29 Dissolve tool in gvSIG

ArcGIS:

7. Create and calculate a new field in order to store the distributed population in integer format.

A. Enter the attribute table of the selected dataset and click on the "Options" button and select "Add field". Enter a name and set type to be double or "float".

B. Click on the title heading of newly created field. Select "Field Calculator". Add thereafter the following formula: round([sum_distributed_population]). See also figure 30.

gvSIG version 1.12.0 Final:

7. Create and calculate a new field in order to store the distributed population in integer format.

A. Add a new field. In the Sextante Toolbox go to Sextante > Tools for vector layers > Add field

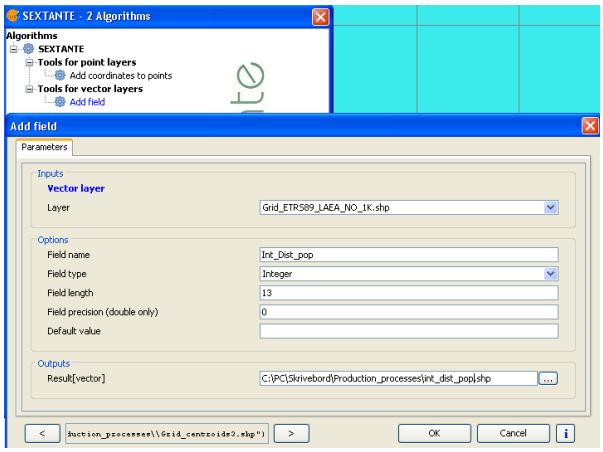


Fig. 31 Add field tool in gvSIG

B. Calculate field. In the Sextante Toolbox go to Sextante > Tools for vector layers > Field calculator

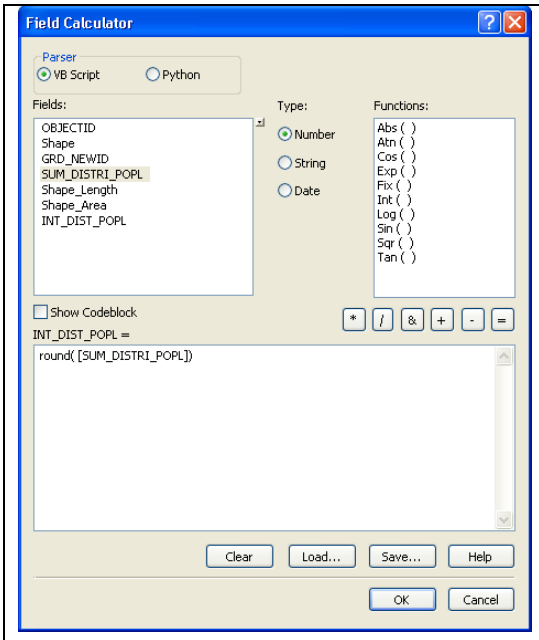


Fig. 30 Field Calculator tool in ArcGIS

7. MAKING THE POPULATED GRID DATA COMPLIANT WITH THE DATA DESCRIPTION OF EUROSTAT

Based on the experiences from GEOSTAT 1A Eurostat has developed data descriptions to be used in the GEOSTAT 1B project. For more information see Appendix 3. This chapter gives an overview how the population grid produced using the guidelines above can become conform to the data description of Eurostat.

The file naming of datasets should follow the following standard:

Name of the file	GEOSTAT_grid_POP_1K_CC_YYYY (CC: country code, YYYY: ref. year, e.g. GEOSTAT_grid_POP_1K_SE_2011

Each grid cell should include following data separated with semicolon:

Column Names	GRD_ID; METHD_CL; YEAR; POP_TOT; CNTR_CODE; DATA_SRC

Where:

GRD_ID	Identification code of the grid cell (lower left-hand corner) according to INSPIRE
METHD_CL	Method used to determine the population of the grid cell; A (aggregated), D (disaggregated) and M (mixed). In the case of border cells calculations have to be made in order to remove double counting of the population.
YEAR	Reference year of the data

POP_TOT	Population count of the grid cell rounded to integers (in the case of border cells the cell contains the share of the population for the country in CNTR_CODE).
CNTR_CODE	ISO code of the country in which the grid cell is located (in the case of border cells the grid cell is reported from all neighboring countries with the same GRD_ID but different CNTR_CODE attribute).
DATA_SRC	For national datasets the country code; for the European disaggregated datasets the data source. For border cells a similar approach as for CNTR_CODE is adopted.

In order to populate the grid according to the data description both tabular and the GIS approach can be used. Below follows the description on how to create and populate new fields using the GIS approach

ArcGIS:	**gvSIG version 1.12.0 Final:**
1. Create and calculate new fields. A. Enter the attribute table of the selected dataset and click on the "Options" button and select "Add field". Enter a name and set type to be "string" or "integer". See also figure 25. B. Click on the title heading of newly created field. Select "Field Calculator" and add thereafter string or year of interest.	1. Create and calculate new fields. A. Add a new field. In the Sextante Toolbox go to Sextante > Tools for vector layers > Add field (see figure 26) B. Calculate field. In the Sextante Toolbox go to Sextante > Tools for vector layers > Field calculator (see figure 27)

APPENDIX 1. MORE ALTERNATIVES FOR GENERATING GRID CELL CENTROIDS

A.1.1 GENERATING GRID CELL CENTROIDS IN ARCGIS USING THE FEATURE TO POINT TOOL

Grid cell centroids can be generated in ArcGIS e.g. by using Feature To Point tool. This requires an application extension license.

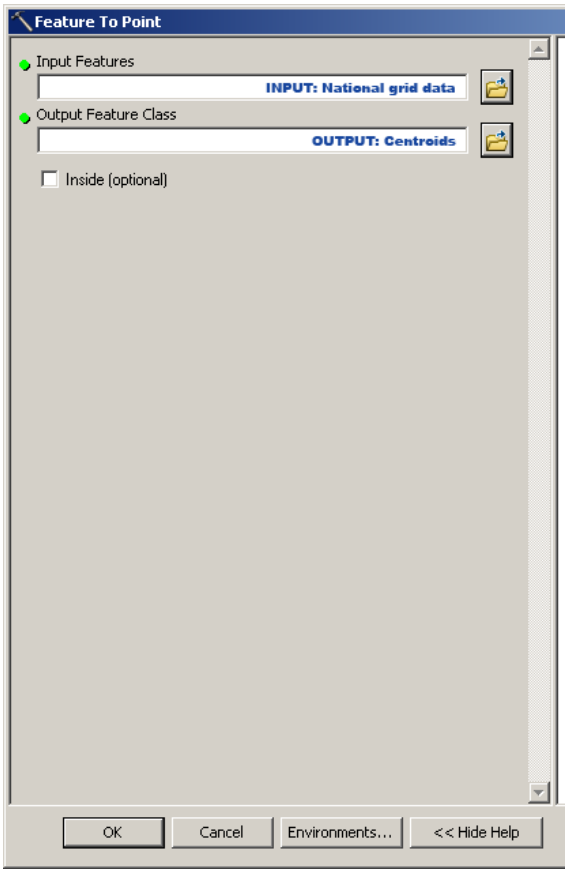
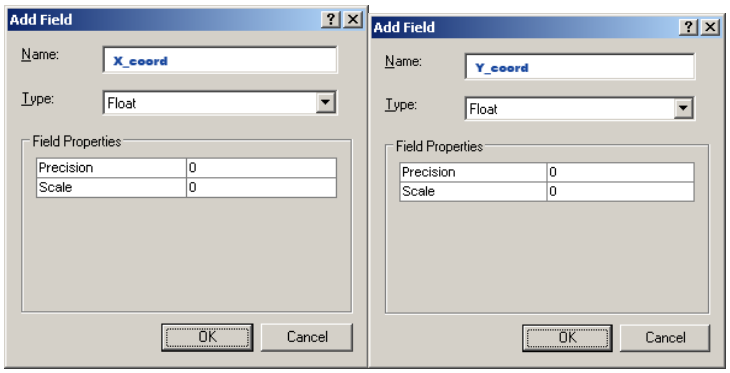


Fig. 31 Feature To Point tool in ArcGis.

A.1.2 Generating grid cell centroids in ArcGIS by calculating coordinates

Grid data centroids can also be generated by calculating grid cell middle coordinates from lower left-hand corner coordinates (that usually stands for the reference point of a grid cell). After creating new columns and using Field Calculator to add 500 meters to coordinates (in case of 1 km grids), you may create a point layer. Mouse right-click on the table (in table of contents), and choose Display X Y data. ArcGIS creates an Events layer that has to be saved first as data layer, before further processing is possible.



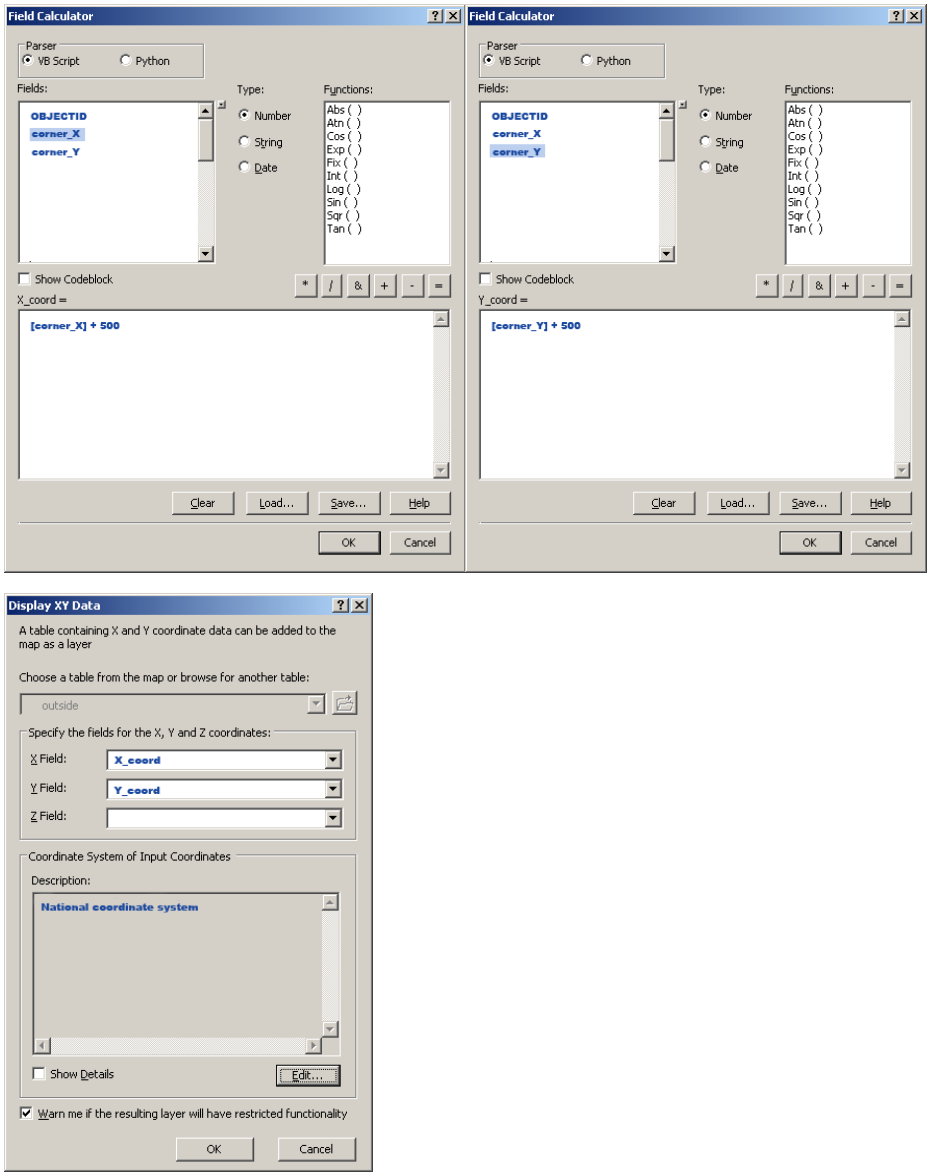


Fig. 32 Creating grid cell centroids from tabular grid data. First new columns are added. Then by using Field Calculator grid cell middle coordinates are calculated. Grid cell centroids are created from coordinates by using Display X Y Data function.

APPENDIX 2. MORE ALTERNATIVES FOR GENERATING GRID CELL CODES

A.2.1 CREATING GRID CELL IDENTIFICATION CODE FOR GRID DATA USING PYTHON SCRIPTING IN ARCGIS

For polygon data, the grid cell identification code can be created by using Python scripts.

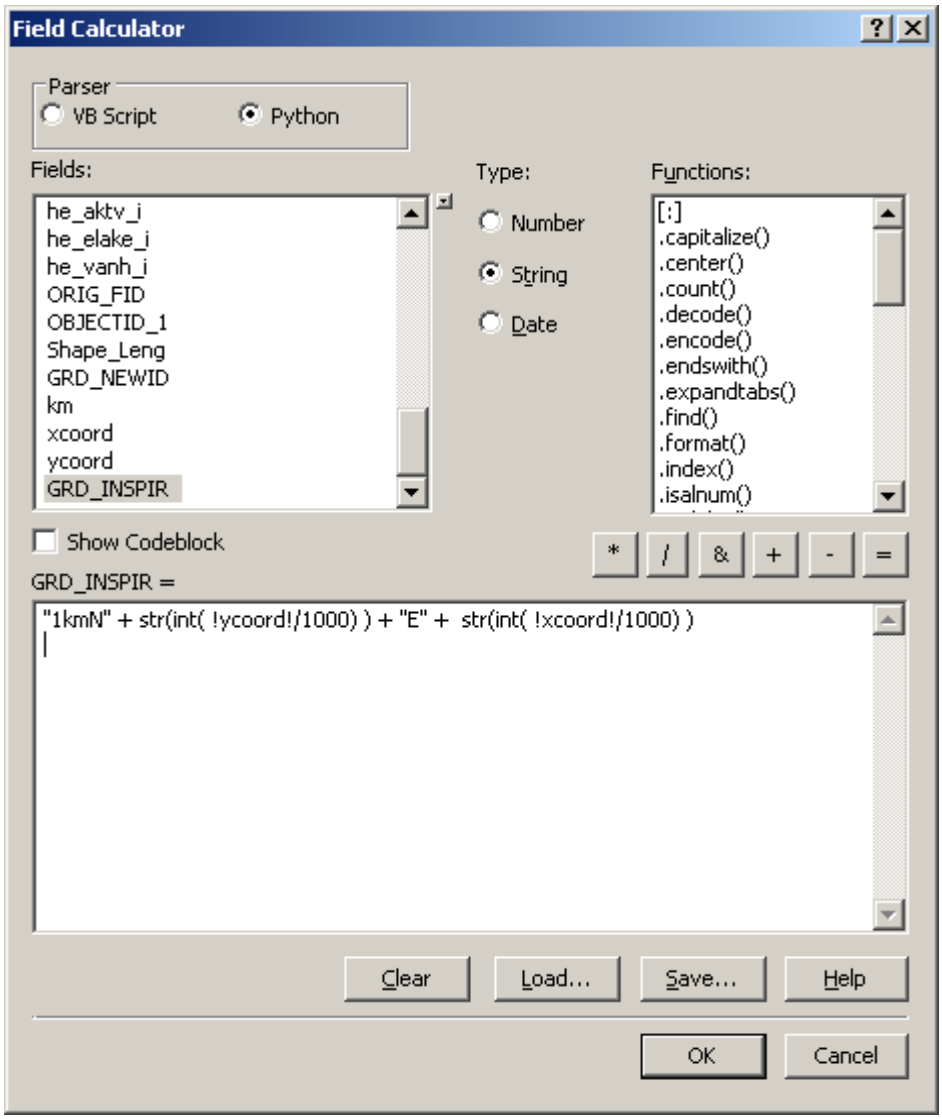


Fig 33. There have to be fields containing the grid cell's lower left corner coordinates in the table and an empty field for the grid cell identifier.

A.2.2 Creating lower left-hand corner coordinates in ArcGIS

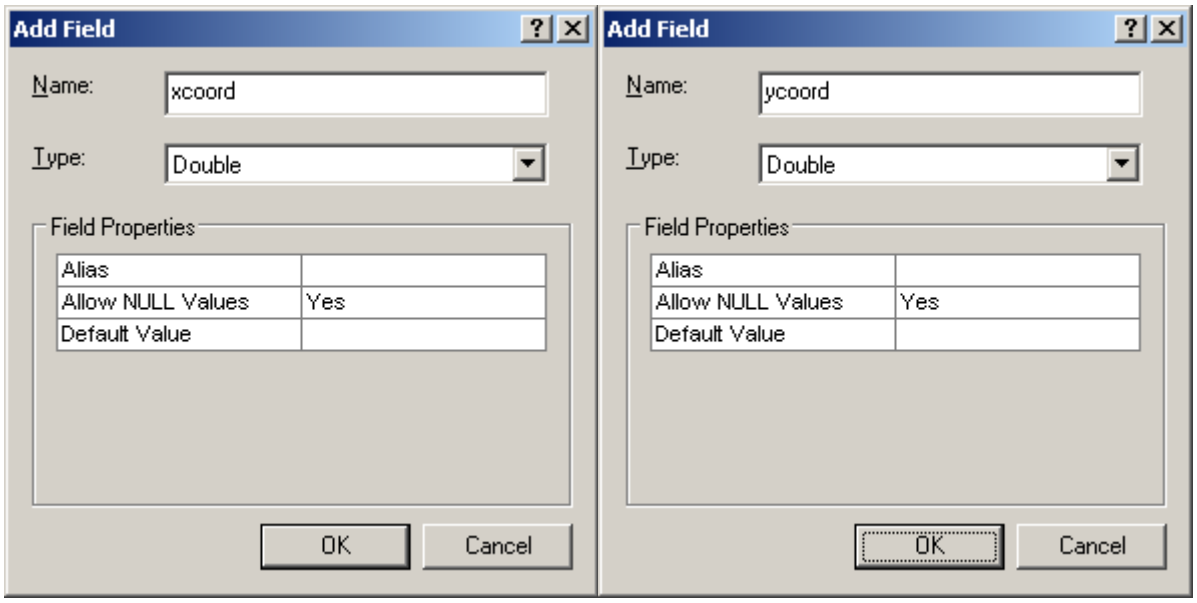


Fig. 33 Adding new fields.

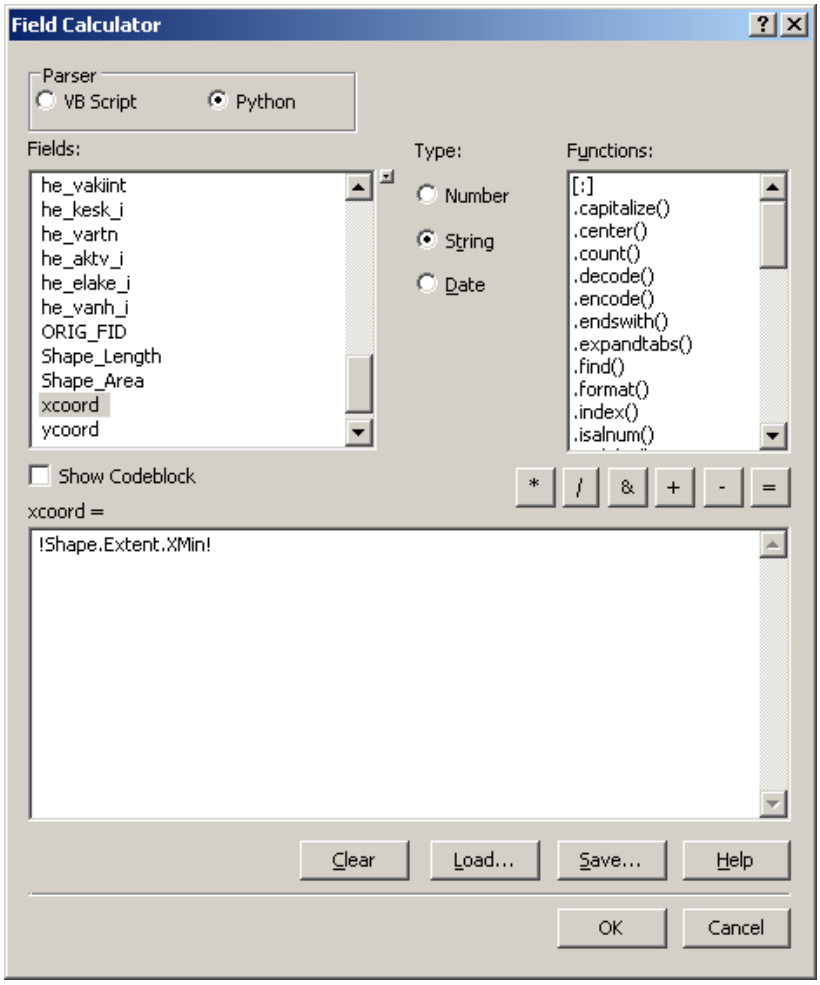


Fig. 34 Counting lower left-hand corner coordinates.
Scripts: !Shape.Extent.YMin! !Shape.Extent.XMin!

APPENDIX 3. DATA DESCRIPTION OF THE GEOSTAT POPULATION DATASET BY EUROSTAT

The proposal is to collect and disseminate data per country. The data will also be disseminated in a European dataset. In order to keep both datasets aligned we propose to make the combination GRD_ID and CNTR_CODE the primary key of the data.

The starting point is a European-wide grid net of 1 km^2 cells corresponding to the INSPIRE specifications[3] and covering EU27 + EFTA countries. The grid net contains all grid cells intersect the landmass of the countries concerned, including all inland waters. The grid net can be divided into country nets allowing for easier data handling. The grid net files will be named Grid_ETRS89_LAEA_1K_CC whereby CC is the ISO country code, so Grid_ETRS89_LAEA_1K_SE for Sweden. The European grid net file will be named Grid_ETRS89_LAEA_1K_EU.

Those grid nets represent the framework for the integration of national grid data. The actual grid dataset consists of .csv text files with the unique INSPIRE grid cell code as reference to the grid net.

The conventions for naming the national grid dataset file and the variables are as follows:

Name of the file	GEOSTAT_grid_POP_1K_CC_YYYY (CC: country code, YYYY: ref. year, e.g. GEOSTAT_grid_POP_1K_SE_2011
Column Names	GRD_ID; METHD_CL; YEAR; POP_TOT; CNTR_CODE; DATA_SRC

Where:

GRD_ID	Identification code of the grid cell (lower left-hand corner) according to INSPIRE
METHD_CL	Method used to determine the population of the grid cell; A (aggregated), D (disaggregated) and M (mixed). In the case of border cells calculations have to be made in order to remove double counting of the population.
YEAR	Reference year of the data
POP_TOT	Population count of the grid cell rounded to integers (in the case of border cells the cell contains the share of the population for the country in CNTR_CODE).
CNTR_CODE	ISO code of the country in which the grid cell is located (in the case of border cells the grid cell is reported from all neighboring countries with the same GRD_ID but different CNTR_CODE attribute).
DATA_SRC	For national datasets the country code; for the European disaggregated datasets the data source. For border cells a similar approach as for

[3] http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SU_v3.0rc2..pdf.

	CNTR_CODE is adopted.

Cells with zero total population are preserved (POP_TOT = 0), for example:

GRD_ID;METH_CL;YEAR;POP_TOT;CNTR_CODE;DATA_SRC
...
1kmN4101E4453;A;2006;**0**;NO;NO
...

Regarding the METHD_CL attribute for each grid cell, there can be various combinations of actual methods combining aggregation, disaggregation and even other estimation methods. For the sake of simplicity the goal is to differentiate cells which have been merely disaggregated using a dasymetric approach ('D'), from those grid cells which are simply 'point in polygon counts' based on detailed georeferenced source data ('A') and from those where various methods and data sources were applied to estimate and model the population figure ('M'). For further details metadata will be used.

As an illustration, a shared border cell between Sweden will be reported by SE as:

GRD_ID	1kmN4101E4453
METHD_CL	A
YEAR	2011
POP_TOT	2
CNTR_CODE	SE
DATA_SRC	SE

and by NO as:

GRD_ID	1kmN4101E4453
METHD_CL	A
YEAR	2011
POP_TOT	4
CNTR_CODE	NO
DATA_SRC	NO

In the integrated European dataset organised in line records per grid cell the data will look as follows:

GRD_ID;METH_CL;YEAR;POP_TOT;CNTR_CODE;DATA_SRC
...
1kmN4101E4453;A;2006;4;NO;NO
1kmN4101E4453;A;2006;2;SE;SE
...

APPENDIX 4. PRODUCTION CASE STUDIES FOR GENERATING A HARMONISED EUROPEAN POPULATION GRID

The methodology paper "Production procedures for a harmonised European population grid – aggregation-approach" is based on the experiences of the various GEOSTAT 1B partners. Below is a short description on how the various GEOSTAT 1B partners are populating the European LAEA grid net. These production cases can clarify for the reader who to appoint in case of need for assistance with data production.

A.4.1 STATISTICS NORWAY (SN)

A.4.1.1 SHORT INTRODUCTION TO REGISTERS AND THE USE OF GIS IN STATISTICS NORWAY:

- Registers in SN

Central register of Establishments and Enterprises, 1956

Statistics Norway Cadastre

Statistical Population Register 1964

- GIS in SN

Used GIS since 1992

10 persons working with GIS

A.4.1.2 SHORT OVERVIEW OF STATISTICS NORWAY'S PRODUCTION PROCEDURE FOR POPULATING THE EUROPEAN LAEA GRID NET

SN converts and aggregates point-based source data into grids similar to the first process flow described in the methodology paper. Below follows a brief description on how SN carries out the various phases in the process.

I COLLECTION

- Primary data:

Statistics Norway (SN) is using the statistical population register and links this data to the georeferenced address points in Statistics Norway Cadastre. 100 % of the address points is registered. However, a few address points may have low geographic accuracy. Linking the two datasets result in a georeferencing percentage of the population of about 99.7 %.

SN is using SAS for linking population data with address points. The data is stored in a SAS dataset before the points are converted to csv-files then generated in ArcGIS.

- "Country clip" of the European LAEA grid net:

The national boundaries of the Kingdom of Norway cover a vast area, from 57° 57' 31" N (Pysen in Mandal municipality) in the South to 80° 49' 44" N (Rossøya on Svalbard) in the North, and from 09° 04' 39" W (Høybergodden on Jan Mayen) in the West to 33° 30' 59" E (Kræmerpynten on Kvitøya, Svalbard) in the East, including Svalbard[4]. In addition to this Norway have several dependencies in the Antarctica, all in all more than 3 million square kilometers.

SN used both the ArcGIS and gvSIG approach in the methodology paper for carrying out the comparisons between the country boundary, the georeferenced micro data and the "Country clip" of the European LAEA grid net.

[4] http://www.ssb.no/english/yearbook/kart/i.html

II Reprojection

The address points are stored in projection ETRS89-UTM33 and the reprojection to ETRS89-LAEA is carried out as a part of the grid data production. SN is currently using the ArcGIS approach under chapter 2.1, but has also tried gvSIG approach. For more complete and efficient data handling SN will try to make use of a tabular approach for reprojecting the data in the future.

III Creation of grid cell identification codes by calculating or by spatial join with the LAEA grid net

SN is currently carrying out this work in SAS, as under chapter 3.1.1. A future tabular approach to the reprojection phase might have consequences for today's use of SAS.

IV Aggregation of population figures by LAEA grid cells

SN is carrying out the aggregation as described under chapter 4.1.1. In this step SN is also verifying the quality of aggregated population dataset. In order to verify the aggregated data the population figures were compared to the reported total population figures.

V Join with the "Country clip LAEA grid net"

Joining of aggregated data with the "Country clip LAEA grid net" has until today been carried out using the ArcGIS method as described under chapter 5.1.

A.4.2 Statistics Estonia (SE)

A.4.2.1 Short introduction the use of GIS in Statistics ESTONIA:

- GIS in Statistics Estonia

GIS in use since 1996

First grid data for 2000 census, update with 2011 census

5 persons working with GIS

A.4.2.2 Short overview of Statistics Estonia's production procedure for populating the European LAEA grid net

Statistics Estonia uses point-based data for creating grid data. Point-based data is used also in production of European LAEA grid data. At the moment the grid data is updated only during censuses.

SE converts and aggregates point-based source data into grids similar to the first process flow described in the methodology paper. See Fig 1. Process flowchart of micro data conversion and aggregation. Below follow a brief description on how SE carries out the various phases in the process:

I Collection

- Primary data:

Statistics Estonia is using the population and housing census microdata to create grid data. In census conduction the Estonian Land Board building and address data were used. As these data were not 100% timely and accurate, in census 2011 also different map applications (web map application in e-census and desktop map application with GPS device in field work) were used in data collection. That allowed address search and marking the location of new building on the map. In census database every building has a unique identification code that is a key to linking people with buildings. Persons, whose addresses are not known on building level, are joined to the virtual points. Such virtual points are created to the mean centre of the buildings of every settlement.

The census microdata is stored in an Oracle database with building identification codes. The spatial data is stored in ArcGIS geodatabase.

- "Country clip" of the European LAEA grid net:

Statistics Estonia is using ArcGIS software (as described in chapter 1.2.1) for carrying out the comparisons between the country boundary, the georeferenced micro data and the "Country clip" of the European LAEA grid net. The country clip covers the Estonian territory and there is no need for widening the country clip.

II Reprojection

The census spatial point data is stored in Estonian National Projection L-EST 1997 (Lambert Conformal Conical projection). For producing the European LAEA grid the primary point data is projected to European LAEA before joining it with LAEA grid net. For that Statistics Estonia is using ArcGIS software as described under chapter 2.1.

III Creation of grid cell identification codes by calculating or by spatial join with the LAEA grid net

Until now we were using Eurostat grid or "country clips" where they already had grid cell identification codes calculated, so we didn't have to calculate these by ourselves. For joining LAEA grid with microdata we used ArcGIS's spatial join tool as described in chapter 3.2. So basically we were able to join and aggregate the data to LAEA grid net in just one step.

But we tested also several other methods like generating grid codes in Excel and we are considering using grid cell calculations in ArcGIS like described in appendix 2 chapter 2.1 (this is useful for cases where we don't have suitable grid codes calculated in Eurostat grid).

IV Aggregation of population figures by LAEA grid cells

Statistics Estonia stores census micro data with building identification codes in Oracle database. In Oracle database the total number of population by buildings is calculated using the building unique identification codes. Using ArcGIS spatial join tool the point based (building based) population data are joined and aggregated to grids in one step. This is described in chapter 5.2.

In this step SE is also verifying the quality of aggregated population dataset. In order to verify the aggregated data the population figures were compared to the reported total population figures.

V Join with the "Country clip LAEA grid net"

Joining of aggregated data with the "Country clip LAEA grid net" has until today been carried out using ArcGIS as described under chapter 5.2.

A.4.3 Czech Statistical Office (CZSO)

A.4.3.1 Short introduction to registers and the use of GIS in CZSO:

- Registers in CZSO

Register of Census Districts and Buildings, 1997

Business Register,

Database of Population, 2007

- GIS in CZSO

Used GIS since 1998

40 persons working with GIS

A.4.3.2 Short overview of CZSO's production procedure for populating the European LAEA grid net

CZSO converts and aggregates point-based source data into grids similar to the first process flow described in the methodology paper. Below follow a brief description on how CZSO carries out the various phases in the process.

I Collection

- Primary data:

Czech Statistical Office collects primary population data at the level of buildings only during the census and then links this data to the georeferenced building points in Register of Census Districts and Buildings. Approximately 99.5 % of the building points in this register are georeferenced. CZSO is using ArcGIS for linking population census data with building points. Linking the two datasets result in a percentage of georeferenced population at about 99.5 %.

- "Country clip" of the European LAEA grid net:

CZSO used ArcGIS approach in the methodology paper for carrying out the comparisons between the country boundary, the georeferenced micro data and the "Country clip" of the European LAEA grid net.

The quality of the "Czech country clip" is very high, because all micro data and grid cells fall inside the country boundary, so there are no buildings and even no grid cells outside the Czech boundaries.

II Reprojection

The building points are stored in national projection S-JTSK and the reprojection to ETRS89-LAEA is carried out as a part of the grid data production using reference transformation key (algorithm) provided by National Mapping Agency. CZSO is currently using the ArcGIS approach described under chapter 2.1.

III Creation of grid cell identification codes by calculating or by spatial join with the LAEA grid net

CZSO is currently adding grid cell identification codes to reprojected point-based data in ArcGIS, as described under chapter 3.2.

IV Aggregation of population figures by LAEA grid cells

CZSO is carrying out the aggregation as described under chapter 4.2. In this step CZSO also verifies the quality of aggregated population dataset comparing it with the reported total population number.

V Join with the "Country clip LAEA grid net"

Joining of aggregated data with the "Country clip LAEA grid net" is CZSO carrying out using the ArcGIS as described under chapter 5.1.

A.4.4 Bulgarian National Statistical Institute (BNSI)

A.4.4.1 Short introduction to registers and the use of GIS in BNSI:

- Registers in BNSI

- Census 2001, 2011
- National register of populated places
- Information System Demography, 2007 (including statistical population data from 1992)
- Short-term statistics on residential buildings, 2004

- GIS in BNSI

Used GIS since 2000

4 persons working with GIS

A.4.4.2 SHORT OVERVIEW OF PRODUCTION PROCEDURE FOR POPULATING THE EUROPEAN LAEA GRID NET

BNSI is carrying out aggregation for the part of the national territory that has georeferenced address points. Dissagregation takes place where there is no georeferenced micro data or where xx% of population havn't found their spatial representation. All the phases are done in GIS software environment (ArcGIS).

I COLLECTION

- Primary data:

- Census 2011 population data
- Georeferenced parcel address points (Cadastre)
- LAEA grid net country clip, downloaded from http://www.efgs.info/data/european-datasets/eurogrid
- Additional data (ancillary information) for the dissagregation process – Classified populated areas, Urban Atlas, OSM roads…

Census 2011 population data is linked to the georeferenced parcel points with address text field. Additional processing of Cadastre address data is carried out for more precise results and for greater percentage of linking. Linking the two datasets result in a georeferencing percentage of the population of xx %. After analysing primary data the national territory is divided into smaller parts where three types of scenario for gridding the population are determined – purely aggregated areas, purely dissagregated, and combined.

- "Country clip" of the European LAEA grid net:

No national data outside the country clip. BNSI used ArcGIS approach in the methodology paper to check if there is national data outside the country clip. The grid net covers completely the reference materials – micro data, borders, ancillary data that is used in dissagregation.

II REPROJECTION

The address points with population data are stored in WGS84 and the projection to ETRS89-LAEA is carried out using the ArcGIS approach under chapter 2.1.

Or we use on-the-fly projection of ArcGIS, by first opening ETRS89-LAEA grid net. The software automatically integrates layers whose coordinate systems are well defined by projecting the data on the fly. Then the coordinates can be calculated if planned table aggregation. But we carry on with GIS approach.

III CREATION OF GRID CELL IDENTIFICATION CODES BY CALCULATING OR BY SPATIAL JOIN WITH THE LAEA GRID NET

BNSI is using Geostat grid net with grid-ids generated. No national grid system used in statistics yet.

IV AGGREGATION OF POPULATION FIGURES BY LAEA GRID CELLS

Population figures are directly aggregated.

In order to verify the aggregated data the population figures were compared to the reported total population figures.

V JOIN WITH THE "COUNTRY CLIP LAEA GRID NET"

Joining of aggregated data with the "Country clip LAEA grid net" is carried out using the ArcGIS method.

A.4.5 STATISTICS FINLAND (SF)

4.5.1 SHORT INTRODUCTION TO REGISTERS AND THE USE OF GIS IN SF

Registers in SF

Register based census -type statistics annually since 1987

The Register of Buildings and Dwellings 1980

GIS in SF

Used GIS since 1990

about 10 persons working with GIS

SF produces EU harmonised grid data similar to national grid data production. The production infrastructure has been developed to simplify and accelerate the production process including EU harmonised data. SF has published EU harmonised population grid data also as WMS and WFS services following Inspire specifications:

http://geo.stat.fi/geoserver/laearuutu/laearuutu:laea_1km_2011/wms?

http://geo.stat.fi/geoserver/laearuutu/laearuutu:laea_1km_2011/wfs?

A. 4.5.2. SHORT OVERVIEW OF STATISTICS FINLAND'S PRODUCTION PROCEDURE FOR POPULATING THE EUROPEAN LAEA GRID NETI COLLECTION

PRIMARY DATA:

SF links population data to centroids of buildings. Statistics Finland receives the data from the register holder (Population Register Centre) and refines the data for statistical purposes.

SF maintains the spatial information of buildings and joins it to other data in order to produce spatial statistics.

"Country clip" of the European LAEA grid net:

The LAEA grid net country clip is evaluated in order to ensure that the whole country is covered and that no data will be missed out of the grid net. In case there are grids missing extra grids will be added to the grid net.

II REPROJECTION

A reprojection into ETRS89/LAEA is made for the national point-based microdata. Centroids of buildings are converted by using ArcGIS Project tool. Converted coordinates are maintained in the same data warehouse as the national coordinates of buildings (in ETRS89-TM35FIN).

44

III CREATION OF GRID CELL IDENTIFICATION CODES BY CALCULATING OR BY SPATIAL JOIN WITH THE LAEA GRID

NET

The maintenance of building points in the data warehouse includes also grid codes, which are calculated for each building. Beside the national coordinate system, grid id codes are also produced from the LAEA coordinates.

Grid identification codes are generated by using SQL:

```
SELECT   '1kmN' + CONVERT(varchar(4),CAST(grid_dataset.Y_coord /1000 as integer))
            + 'E' + CONVERT(varchar(4),CAST(grid_dataset.X_coord /1000 as integer)) as GRD_ID
FROM    micro_data
```

IV AGGREGATION OF POPULATION BY LAEA GRID CELLS

Data aggregation is made by using SAS.

```
proc sort data=micro_data_id;
by GRD_ID;
run;
data grid_dataset_POP_sum(KEEP=GRD_ID POP_TOT);
set micro_data_id;
by GRD_ID;
if first.GRD_ID then do;
POP_TOT=0;
end;
POP_TOT+1;
if last.GRD_ID then do;
format GRD_ID $12. POP_TOT 8.;
output;
end;
run;
```

V JOIN WITH THE LAEA GRID NET

In order to get tabular population grid data into spatial grids, table and national country clip of the grid net are joined in ArcGIS by using Join Table tool.