Methods for generating population grid statistics

'A grid for representing thematic information is a system of regular and georeferenced of cells, with a specified shape and size, and an associated property' (1).

The production of gridded statistics usually consists of two elements: a grid net and a statistical table. A grid net is a vector layer of regular equal-sized grid cells covering a certain spatial extent or study area. It can be made with almost any GIS environments or software that supports vector data. The grid net represents a steady and interoperable location of different statistical objects enabling data comparability regarding differently scaling events over, also ensuring time stability of time series data. Above all, it can provide a basis for the integration of different thematic data, covering different attributes from various domains

Grid-based statistics are territorial statistics from the same analysis unit which the first precondition for generating them is that thematic data can be georeferenced, preferably using direct x and y geographic coordinates as a location code. This means that the data units have direct or indirect links to geographic coordinates.

In short, grid-based statistics can be computed using *aggregation* or *disaggregation* methods. If there is a point-based data model (or point-based object foundation) available, for instance including geographic coordinates of buildings, address points or real estates, the method to be used is called the aggregation method. In GIS terms, it uses a point-in-polygon procedure to aggregate the data. Aggregation is based on accurate locations of points where the data points can be added up inside each grid cell. This is the most accurate approach and provides the best quality. The most common aggregation technique among several alternatives is the simple aggregation from point sources presenting an easy and straightforward approach.

If no direct aggregation from point-based data sources is possible and the only available data is area-based (e.g., enumeration areas) the precondition for grid generation by aggregation is that the boundaries of the areas, or at least the centroids of the polygons, are available as an input.

Both aggregation and disaggregation methods are highly dependent on the tools used for the analysis. It is also generally assumed that there should be as much producer control over the processes executed by the hardware and software tools used in the process.

Aggregation method

The overall workflow of data aggregation is illustrated in the figure below.

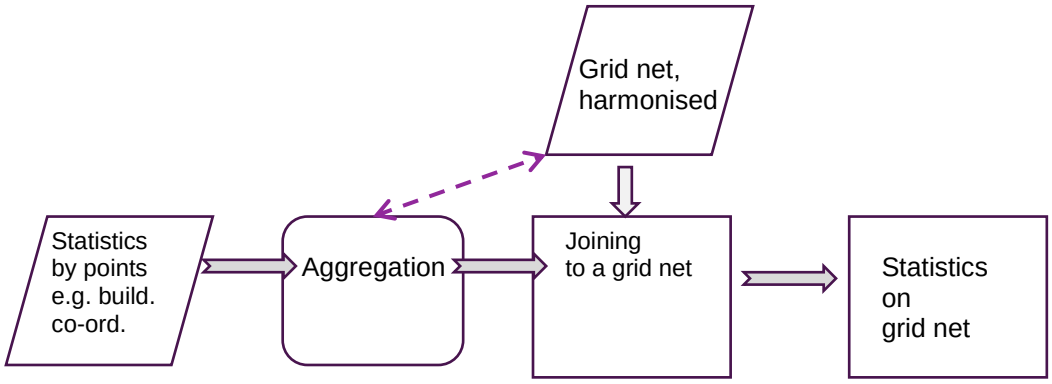

Figure 1. Illustration of a data aggregation workflow for gridded statistics

There are several alternatives for conducting an aggregation process. The whole process may be done using statistical software or with the help of GIS software, or a combination of both.

In terms of data quality, the coordinate system and projection system used should be the same for the grid net (target grid) and the data georeferenced (point-based data source) and should remain so during the production phases. This is the preferred option to ensure the best quality.

If required, a data transformation stage has to be carried out first, to avoid data quality concerns. The conversion from one projection system to another, if needed, is best done with the source data. However, data which is already gridded can be converted to another coordinate system, although this will result in small changes to the data. In this case it is recommended to recast the data from the smallest available grid size to minimise the impact (2).

Disaggregation method

The overall workflow of the disaggregation approach is described in detail in the literature (3)(4). Although refining a European disaggregated grid was outside the scope of this project, such a grid will be fundamental to producing a grid map for the entire European territory for a period of some years, until such time as national grid initiatives are fully established.

The disaggregation method is particularly useful when large territories are covered in a multinational to global context requiring cross-border harmonised data posing some data representation and interpretation concerns. The choice of source data depends on data availability and the size of the grids in the final output. Spatial statistics by census enumeration areas are usually the best starting point. If the area of the statistical geography (enumeration area) is greater than the target grid cell, ancillary data should be used to predict the distribution of statistical observations according to corresponding grids. For population grids, various ancillary data are available to predict indirectly where there are populated areas and where there is not. A disaggregation model is needed to estimate, for each grid cell or part of grid cell, different density categories defined by ancillary data. The calculation of density categories and weights depends on the ancillary data available (3)(4)(5).

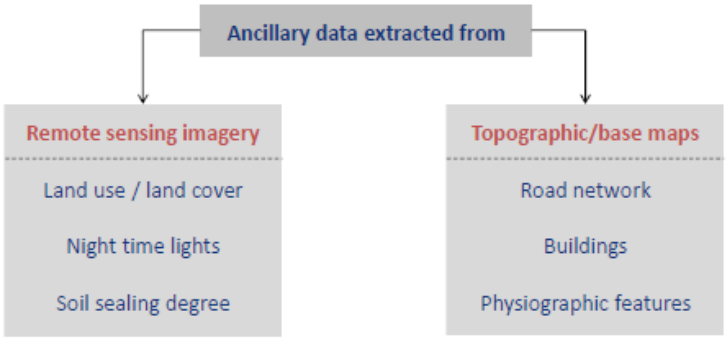

Figure 2. Examples of ancillary data for estimation of population in grid statistics.

A specific case of disaggregation is where enumeration areas are so small that they fit inside the target grids. In these cases, the aggregation method can be applied instead (6) for more details.

National disaggregation initiatives often have access to high-resolution ancillary datasets that are not available with comparable quality throughout Europe (5). Some countries have used information about the location of buildings to predict the distribution of population inside the grids (6)(7). This means that when the enumeration area (with a known population) is split across grid cells, then the population of a grid cell depends on the percentage breakdown of address points within each enumeration area.

References

(1) European reference grid 2003. 1st Workshop on European Reference Grids. Short Proceedings. Ispra 27-29 October 2003. JRC ESDI Action 2142
(2) Koivula et al., 2011 in The European Forum for GeoStatistics — ESSnet project GEOSTAT — Representing Census data in a European population grid
(3) Batista e Silva, F., Gallego, J., Lavalle, C. (2011). The effect of ancillary data in population dasymetric mapping. European Forum for Geostatistics, Lisbon, October 2011.
(4) Gallego, Javier (2010). A population density grid of the European Union. Population and Environment. Springer Science+ Business Media, LLC.
(5) Goerlich, Fransisco and Cantarino, I. (2011). Downscaling Population with a High Resolution Land Cover Data Set for Spain. EFGS Conference. Lisbon, October 2011.

(6) Corcoran, Dermot (2011). A population grid for the Republic of Ireland: Making use of national databases and local geography. European Forum for Geostatistics, Lisbon, October 2011.
(7) Schoenmakers, B-J. 2011. Creating 2001 to 2011 population grids using Census Geography. Statistics Portugal. European Forum for Geostatistics, Lisbon, October 2011.