



Economic and Social Council

Distr.: General
8 September 2015

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Seventeenth Meeting

Geneva, 30 September to 2 October 2015

Item 5 of the provisional agenda

Grid statistics: opportunities and challenges

Grid statistics: opportunities and challenges

Note by the Central Statistical Office of Poland¹

Summary

The introduction of x, y coordinates and address points in statistical data enabled significant change of the previous spatial identification system and resulted in shifting from area assignment (census districts) to point assignment. It had a fundamental consequence for the application of geoinformatics in official statistics. The change of the assignment mode allowed for a more flexible grouping of data collected in public statistics even for the smallest areas. It also facilitated the creation of a spatially-oriented micro database, enabling the conduction of geo-statistical analyses.

Currently Polish statistics presents a range of grid based presentations of population density at statistical data visualization platform - Geostatistics Portal. Experience gained during the preparation of these visualizations shows that data presentation in grid cells is basically very accurate, allows an easy comparison as all cells have the same size and are stable over time. Moreover grids integrate easily with other scientific data (e.g. meteorological information) and grid systems can be constructed hierarchically in terms of cell size thus matching the study area. Grid cells can also be assembled to form areas reflecting a specific purpose and covering the study area (mountain region, water catchment).

¹ Prepared by Janusz Dygaszewicz, Director of Programming and Coordination of Statistical Surveys Department, President of European Forum for Geography and Statistics (EFGS).

I. Overview

1. For several years public statistics in Poland is creating and using spatial data. In order to guarantee the technical conditions required by INSPIRE directive with regard to modern technologies and information standards, and taking into consideration the technological needs of the census and current surveys, the Central Statistical Office had undertaken work in order to ensure the spatial identification of the included objects on the basis of digital maps and GIS technologies. Among employees of official statistics about 210 people were trained in GIS tools. With the use of reference materials in the form of spatial data obtained from the national geodetic and cartographic fund (orthophotomap, national register of boundaries and areas of country territorial divisions units, data from the land and buildings register) and own materials in the public statistics (maps with the statistical distribution boundaries and situational sketches of the buildings location in the field) following sets were created:

- statistical address points (the x, y location of all buildings from the National Official Register of Territorial Division of the Country – TERYT, means all buildings containing dwellings),
- statistical regions,
- census enumeration areas.

Figure 1

Spatial address database

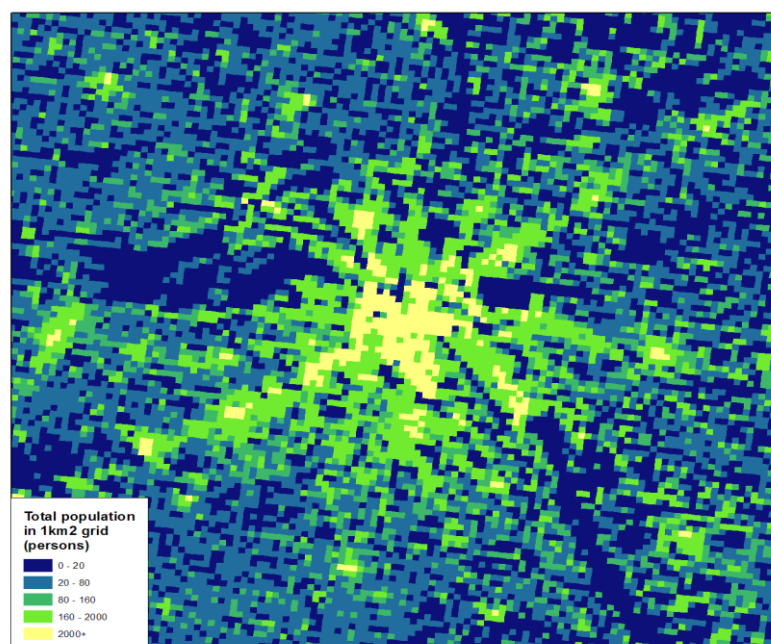


2. These sets are part of the so-called spatial address databases. The data from these databases were used to carry out the Agricultural Census 2010 and Census of Population and Housing 2011. It was used in three ways. First, to prepare field works concerning census enumeration areas allocation to enumerators. Second, during the census, to manage the census enumerators in field works using GPS. And third, after the census, for spatial reference statistics data - all data were collected with reference to a particular statistical address point (or sometimes the plot centroids in the agricultural census) so it was possible to make statistical multidimensional spatial analysis.

3. In order to disseminate the results of Agricultural Census 2010 and Population and Housing Census 2011 on maps Central Statistical Office created the Goestatistics Portal. For the purposes of spatial visualization, data aggregated to various administrative and statistical units (e.g. municipalities, counties, NUTS regions, grid) were prepared maintaining statistical data confidentiality. The portal is publishing predefined thematic cartograms based on the aggregated data. In addition, users are able to edit thematic maps and print maps produced by themselves. For users within official statistics the portal is providing various tools for advanced spatial analysis on microdata (e.g. analysis in a user defined area). Results of such analysis can be published in the portal if it meets the requirements of statistical data confidentiality.

Figure 2

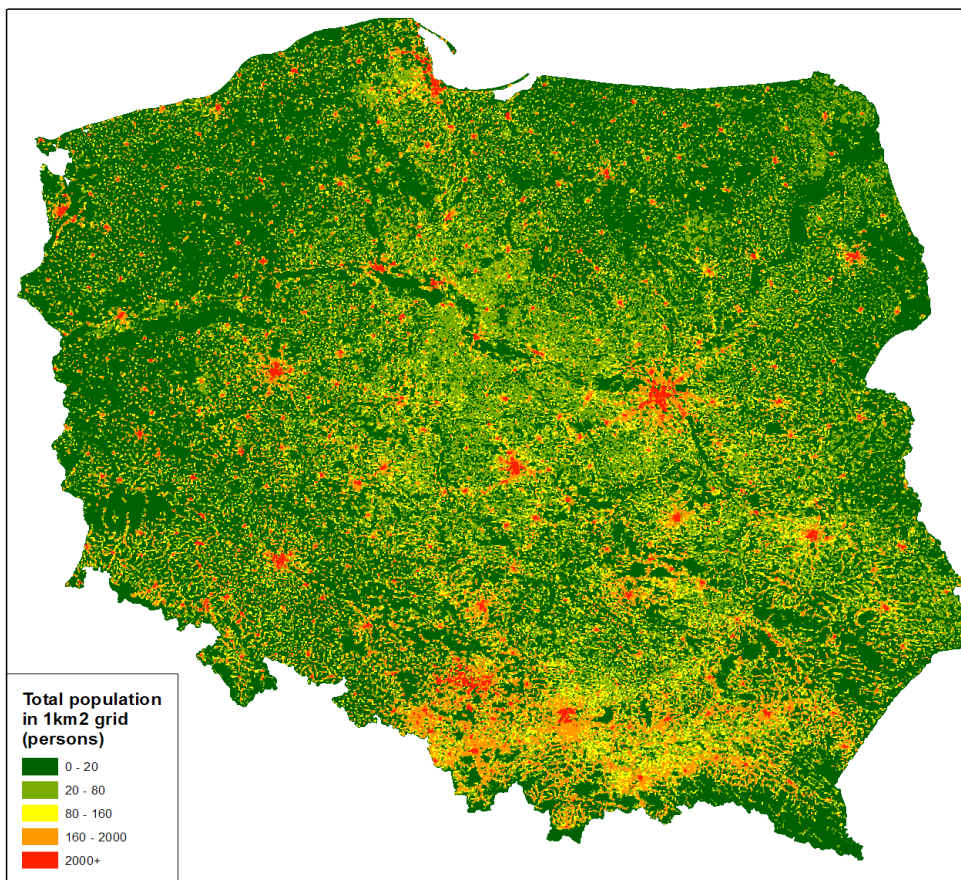
Warsaw population density by 1km² grid



4. Introducing address points coordinates (x , y) for unit statistical data at the data collection stage made it possible to change the previous system of spatial identification of these data and to move from area classification (census districts) to point classification. This had a key, even revolutionary, meaning for the application of geostatistics. Changing the classification allowed for more flexible grouping and dissemination of data in national censuses - for the smallest areas. It also made it possible to create a base of microdata of a spatial nature enabling the carrying out of spatial analysis of various phenomena, concerning, for instance:

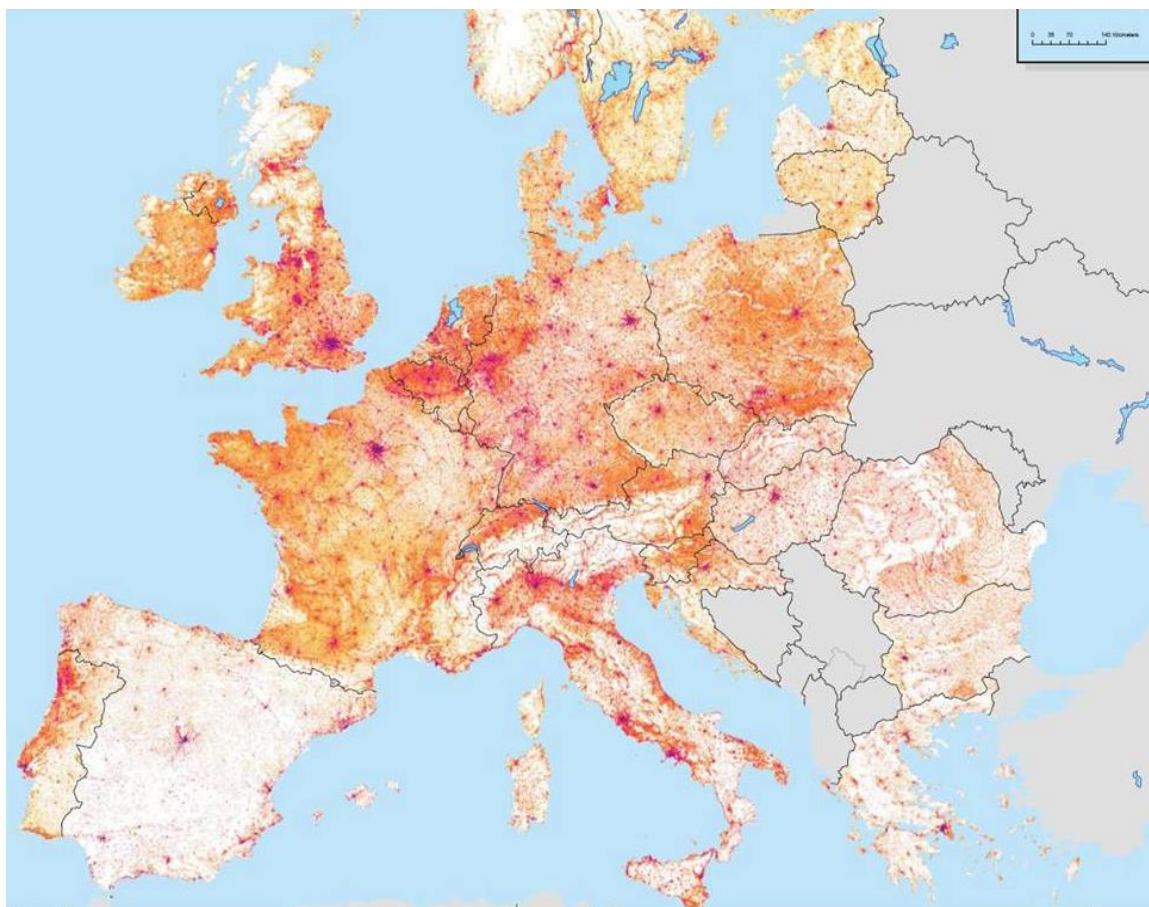
- demography (e.g. the average distance between children's and parents' residence, commuting to work, school, distance to hospital etc.),
- urbanisation and planning (e.g. useful in determining the boundaries of urban agglomerations, metropolis, and the drawing up of land development plans),
- agriculture and environment (analysing the structure of crops, environmental pollution),
- the economy (e.g. analysing the effects of burdensome road and industry investments).

Figure 3
Spatial visualisation of demographic data in Poland by 1 km² grid



5. Classification of the analysis conducted by points with coordinates x, y also made it possible to become independent of boundaries changes in the regional division of the country, usually resulting in changes in census districts and laborious recalculations. This facilitated a comparative analysis of time series, regardless of the changes taking place in this division. An additional advantage was the possibility of the aggregation of data both in the structure of the NUTS administrative division and the GRID divisions prepared within the GEOSTAT projects (Eurostat). It shows that with this approach, it is possible to compare different phenomena not only at the national level but also at the European level regardless of the local administrative divisions.

Figure 4
The grid contains the total residential population on the date of the census 2011 (Eurostat)



6. Currently Polish statistics presents a range of grid based presentations of population density at statistical data visualization platform - Geostatistics Portal. Aside from census and other surveys results the Geostatistics Portal, as part of the national spatial data infrastructure, is currently publishing datasets, services and metadata regarding INSPIRE Annex III data themes: statistical units & population distribution (demography). Experience gained during the preparation of these visualizations shows that data presentation in grid cells is basically very accurate, allows an easy comparison as all cells have the same size and are stable over time. Moreover grids integrate easily with other scientific data (e.g. meteorological information) and grid systems can be constructed hierarchically in terms of cell size thus matching the study area. Grid cells can also be assembled from areas reflecting a specific purpose and covering the study area (mountain region, water catchment).

II. Summary

7. Works towards the establishment of a pan-European grid has started and are currently being carried out by EFGS (European Forum for Geography and Statistics) with a great success. EFGS started as a voluntary cooperation between National Statistical Institutions (NSIs) in the Nordic countries in 1998, on use of geographic information systems (GIS) and statistics. Today EFGS is mainly concentrated on the development of the

best practices in the production of grid-based-statistics in Europe within GEOSTAT projects. It is also worth mentioning that grid was listed as one of the standards in INSPIRE directive.

8. Grid is currently gaining international recognition and relevance in terms of the international comparison of phenomena, possibly, as well as a tool to support the SDGs (Sustainable Developments Goals).

9. Talking about grids - the biggest challenge is to collect the data with the best possible accuracy (preferably to an address point). Later on there are only advantages!
