



Economic and Social Council

Distr.: General
14 July 2017

Original: English
English and Russian only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Nineteenth Meeting

Geneva, 4–6 October 2017

Item 6 of the provisional agenda

Production and dissemination of geo-referenced coensus data

Creation and dissemination of spatial census data by Central Statistical Office of Poland

Note by the Central Statistical Office of Poland¹

Summary

The Central Statistical Office of Poland (CSO) is creating and using spatial data. In order to present census and survey results on maps and to publish spatial data, CSO prepared and launched a data visualization platform – the Geostatistics Portal. It is a tool for interactive cartographic presentation and publication of aggregated statistical information resulting from various types of spatial data and analyses. In the near future Central Statistical Office (CSO) is going to expand the Portal, which will aim at broadening the scope and availability of statistical information and methods of geostatistical analyses using the resources of official statistics. New functionalities will enable very advanced analyses of statistical information, e.g. exploratory geostatistical data analyses or geostatistical modelling. The Portal will present official statistical data and results of geostatistical analyses, necessary to the functioning of the state, local governments and communities.

¹ Prepared by Janusz Dygaszewicz, Director of Programming and Coordination of Statistical Surveys Department

GE.17-11839(E)



* 1 7 1 1 8 3 9 *

Please recycle



I. Introduction

1. For several years public statistics in Poland have been creating and using spatial data. Regarding modern technologies and information standards, and taking into consideration the technological needs of the census and current surveys, the Central Statistical Office has undertaken work in order to ensure the spatial identification of the included objects on the basis of spatial reference and GIS technologies.

2. In order to present the results of the previous round of Census: *Agricultural Census 2010* and *Population and Housing Census 2011*, Central Statistical Office of Poland prepared and launched for internal and external users a data visualization platform – the Geostatistics Portal.

3. The *Geostatistics Portal* is a tool for interactive cartographic presentation and publication of aggregated statistical information resulting from various types of spatial data and analyses, presented on maps. It is designed to collect, present and provide information to a wide audience, including public administration, entrepreneurs, individual users and research institutions. It is a very helpful tool in making strategic decisions at every management level. The comprehensive solution is tailored to European standards, and all data, by way of processing, is presented in depersonalized form, preserving statistical confidentiality.

II. Point based geocoding

4. The introduction of x, y coordinates and address points in census data enabled changing of the previous system of spatial identification and shifting from area assignment (census districts) to point assignment. It had a fundamental significance for the applications of geoinformatics in official statistics. The change of the assignment mode allowed for more flexible grouping and presenting data collected by public statistic in statistical units smaller than commune, i.e. in statistical regions, census enumeration areas or even in a very small areas, such as a kilometre grid with a cell size of 1 km². It also facilitated the creation of a spatially-oriented micro database, enabling the conduction of geo-statistical analyses on microdata (e.g. analysis in a user chosen area). Results of such analysis can be published in the Portal if it meets the requirements of statistical data confidentiality.

III. The Geostatistics Portal – current functionality

5. The Geostatistics Portal is used for spatial presentation of Poland's largest information resource, enabling the publication of aggregated statistical data in the form of various types of spatial analyses, presented on maps with statistical confidentiality. The Geostatistics Portal is a tool for interactive cartographic presentation and the publication of data acquired in censuses. It facilitates storing, presenting and sharing information for a broad group of recipients.

6. The Portal functions on two levels: for internal (statisticians) and external users, and the scope of presented data is defined through the appropriate roles and authorisations. Internal users have access to both unit and aggregated data, whereas external users only to aggregated data, published taking into account statistical confidentiality.

7. The interface of the Geostatistics Portal provides quick and easy access to resulting statistical information. Data are presented using such cartographical presentation methods as choropleth maps and various kinds of diagram maps. It is also possible to set one's own parameters for the visualisation of a thematic area for a given choropleth map. These

include measure, aggregation level (territorial division unit), the number of classes, classification methods, etc. Aside from the possibility of using ready-made spatial analyses, in the Geostatistics Portal, internal users can draw up custom thematic maps based on a selected feature of the data model, using dynamic spatial analyses, i.e. linear or distance analyses, or object buffering.

8. Classification of the analyses based on points with x, y coordinates gives also the possibility to become independent from boundaries changes (in the regional division of the country), usually resulting in changes of census districts and laborious recalculations. This facilitates a comparative analysis of time series, regardless of the changes taking place in this division. An additional advantage is the possibility of the data aggregation both in the structure of the NUTS administrative division and the grid divisions prepared in CSO during the GEOSTAT projects (Eurostat). With this approach it is possible to compare different phenomena not only on national level but also on European level regardless of the local administrative divisions.

9. Currently Polish statistics presents a range of grid based presentations of population density in the 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).

IV. The Geostatistical Portal - new functionality and services

10. In the near future Central Statistical Office (CSO) is going to expand the Portal, which will aim at broadening the scope and availability of statistical information and methods of geostatistical analyses using the resources of official statistics. It is a response to the needs which were expressed following the implementation of Geostatistics Portal – Phase II project.

11. The functionalities of the previously provided services will be expanded to include i.e. preparing statistical analyses within any spatial division, e.g. defined by the user, downloaded from external spatial data services (WFS), based on “dynamic” grid cells, the possibility of combining statistical data with the user’s own data or geocoding users’ features to use in geostatistical analysis. In addition, new services are to be developed, which will facilitate the use of exploratory spatial data analyses using statistical information, performance of analyses in the field of geostatistical modelling and the support of supplementing users’ own data with geostatistical information and analyses.

12. Three completely new external services will emerge as a result of new developments in the Portal:

- Exploratory geostatistical data analysis using statistical information available in the Portal,
- Geostatistical modelling analyses,
- Supplementation of user content with geostatistical information and analyses available in the Portal.

13. The *Exploratory geostatistical data analysis using statistical information available in the Portal* service will allow users to:

- import and supplement the user's own data with statistical data available in the Portal;
 - supplement the data with additional variables (the so-called derived variables) – calculated on the basis of the already imported and supplemented user's data with the use of the available set of mathematical and geostatistical functions;
 - explore the collected data (supplemented and supplemented with derived variables) in geostatistical terms, evaluation of their quality as well as supplement and modify the data;
 - visualise and publish the obtained results of analyses.
14. A wide range of methods of the statistical spatial analysis will be made available, including:
- central tendency statistics;
 - statistical dispersion;
 - cluster analysis;
 - spatial autocorrelation.
15. The *Geostatistical modelling analyses* service will allow users to:
- import and supplement the user's own data with the statistical data available in the Portal;
 - supplement the data with additional variables (the so-called derived variables) – calculated on the basis of the already imported and supplemented user's data with the use of the available set of mathematical and geostatistical functions;
 - explore the collect data (supplemented and supplemented with derived variables) in geostatistical terms, evaluate their quality and supplement and modify the data;
 - build and apply the probabilistic model allowing the inference (estimation) of the value of the dependent variable on the basis of the results of the probability sample survey (i.e. on the basis of the data collected);
 - visualise and publish the obtained results of analyses.
16. The *Supplementation of user content with geostatistical information and analyses available in the Portal* will allow users to:
- import own content in order to analyse it;
 - conduct content analysis with the use of "text mining" mechanisms in order to detect key words referring to categories of objects available in the Portal and to categorise user's content in an appropriate way;
 - select the objects in the Portal best matching the detected words and key phrases and to establish correlations between the objects searched in the Portal in order to select analyses available in the Portal;
 - configure and generate spatial analyses selected by the user in a graphic form;
 - provide access to the results of selected spatial analyses in a graphic or tabular form.
17. The development of the Portal will also introduce two new internal services facilitating efficient maintenance of spatial address databases and georeferencing survey frames.

II. Conclusion

18. New functionalities will enable very advanced analyses of statistical information, e.g. exploratory geostatistical data analyses or geostatistical modelling. The Portal will present official statistical data and results of geostatistical analyses, necessary to the functioning of the state, local governments and communities. The information will be accessible in a convenient graphical form, accelerating the decision-making processes. The long-term aim will be to develop and present the results of censuses of the round 2020-2021.
