United Nations ECE/CES/ GE.41/2016/13

Economic and Social Council

Distr.: General
19 August 2016

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Eighteenth Meeting
Geneva, 28 - 30 September 2016
Item 6 of the provisional agenda
Innovations in census methodology and technology, and results of testing

Standardization of the geocoding as the basis of linkage the census data with the spatial data and flexible dissemination of final census statistics

Note by the Central Statistical Office of Poland

Summary

Before the 2010 census round the digitization of all statistical division boundaries was carried out within Statistical Offices. According to the rules adopted, boundaries of statistical units are aligned with those of cadastral units where applicable. In connection with the need to preserve the limits of housing and people in the statistical units and taking into account the diversity of terrain and population density, statistical regions and census enumeration areas (EAs) in rural and urban areas have different areas. For urban areas the census EAs can be one block or city block. On the other hand, in rural areas, where the population density is much smaller, the census EAs can cover a whole village. This means that there can be many statistical units inside the boundaries of cadastral units and vice versa, and that statistical division must be maintained on a regular basis by the statistical service. To preserve collinearity in case of any changes made by the cadastral service, appropriate changes must be made on the statistical side. Therefore the possibility should be considered of using the boundaries of cadastral units as territorial division units for statistics. At the same time the current trend is to move from traditional census to register-based census. In that case it is necessary to re-examine certain population and buildings limits and legitimacy of their maintenance. The possibility of harmonizing statistical division (statistical regions and census enumeration areas) with the cadastral division (cadastral units) should be pursued, taking into account the needs of official statistics.

I. Introduction

1. Nowadays, the development process of obtaining and providing statistical information is, to a large extent, based on its spatial reference. In Poland, in the 2010 Census Round, a combination of data coming from administrative sources and register containing spatial data was used for the first time. The application of digital maps and the GPS technologies brought a revolutionary change when it comes to the possibilities of planning and managing census operations, both prior to and during the census.

2. For that purpose, the data obtained from the State geodetic and cartographic resources, as well as ortophotos (processed aerial photographs), were used. With the use of the materials obtained, both from geodetic and statistical resources, it was possible to develop sampling frames for censuses, comprising statistical address points and their spatial reference.

3. Digital maps were an indispensable tool facilitating the work of census enumerators (when it comes to moving around the area, verifying the sampling frame, etc.), gmina (commune) leaders, and voivodship (province) and central dispatchers who could verify on a map the progress of the census and, for example, the route or location of an enumerator, using a dispatching or GIS application, facilitating the work of a gmina leader.

4. In accordance with the organisational principles adopted, a pre-census round was made prior to the census commencement. In pre-census, a census enumerator verified the existence of buildings and supplemented the register with missing address points. An additional aspect of the round involved examining the area where the census was to take place, and possibly resolving any ambiguities which had arisen during data revision on the gmina level.

5. Mobile terminals were equipped with the GIS application, which enabled revisions and showed on the map, among other things, the current location of the census enumerator (GPS) and address points assigned to him. Using the mobile application, the enumerator could change the location of an address point, delete an address point, or add an address point not included in the register, with the help of GPS device. During the pre-census stage, the enumerator was also responsible for controlling the entire areas of the census districts assigned to him. It was particularly important when the revision in the gmina district was performed only on the basis of the registers kept, and the census enumerator was the first and in many cases the only person directly involved in field work in the framework of the census operations. The GIS application was also actively applied during the census – to manage its course. It enabled the monitoring and control of the enumerator's work, as well as the tracing of his movement in the area (among other things, to ensure his safety).

6. The introduction of x,y coordinates and address points in statistical 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 of data collected in public statistics for even the smallest areas. It also facilitated the creation of a spatially-oriented micro database, enabling the conduction of geo-statistical analyses.

7. Currently Polish statistics presents a range of grid-based presentations of population density at statistical data visualization platform - Geostatistics Portal - (geo.stat.gov.pl). 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).

8. The 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 analysis, 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 serves the following functions:

- storing,
- presenting,
- sharing information for a broad group of recipients.

9. The Portal functions on two levels: for internal (official statistics) 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.

10. The interface of the Geostatistics Portal allows its users quick and easy access to resulting statistical information. Data are presented using such cartographical presentation methods as cartograms (choropleth map) and various cartodiagrams. It is also possible to set one’s own parameters for the visualisation of a thematic area for a given cartogram. These include measure, aggregation level (territorial division unit), the number of intervals, 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.

11. Classification of the analyses conducted by 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.

12. Before the 2010/2011 census round the digitization of all statistical division boundaries was carried out within Statistical Offices. According to adopted rules, boundaries of statistical units are aligned with boundaries of cadastral units where applicable. In connection with the need to preserve the limits of housing and people in the statistical units, and taking into account the diversity of terrain and population density, statistical regions and census enumeration areas in rural and urban areas have different areas. For urban areas the census enumeration areas can be one block or city block. On the other hand, in rural areas, where the population density is much smaller, the census enumeration area can cover a whole village. This means that there are many statistical units inside the boundaries of cadastral units and vice versa and that statistical division must be maintained on a regular basis by the statistical service. To preserve collinearity in case of any changes made by the cadastral service, appropriate changes must be made on the statistical side. Therefore, we should consider the possibility of using the boundaries of cadastral units as territorial division units for statistics. At the same time, the current trend is to resign from the traditional censuses to register based census. In that case it is necessary to re-examine certain population and buildings limits and legitimacy of their maintenance. We should seek for the possibility of harmonizing the statistical division (statistical regions and census enumeration areas) with the cadastral division (cadastral units), taking into account the needs of official statistics.

13. In reference to the above polish experience proposal to inventory the level of integration of spatial objects used in statistics and geodesy – "The 10 Level Model" for harmonization of statistical and geodesy reference framework is presented below.

Geodetic System	Layers (suitable for geocoding)	Statistical System
+	NUTS1 - Administrative level 1	+
+	NUTS2 - Administrative level 2	+
+	NUTS3 - Administrative level 3	+
+	LAU1 - Administrative level 4	+
+	LAU2 - Administrative level 5	+
Cadastral units Cadastral parcels	INDIVIDUAL UNITS	Statistical regions Enumaration areas
+	POLYGON	?
?	GRID	+
+	LINE	?
+	POINT	+

II. Administrative levels (level 1-5)

14. In general, on both systems (geodetic and statistical) function five reference layers relating to the administrative division of the country (in Europe NUTS 1, 2, 3 and LAU 1, 2) and from the point of view of data synchronization those layers are treated equally by both systems. That means that data collected in geodesy and through statistical service are referenced to the same geometry that is already established usually by Mapping Agency (MA). That makes it is possible to use this geometry for the process of geocoding statistics. However this process takes place at the high level of aggregation which is more often not satisfactory for users of the statistics.

III. Individual units for interior purposes (level 6)

15. There are cadastral units and cadastral parcels in geodesy, the statistical regions and enumeration areas are in statistics. Geometries of these divisions have a special destiny in those two systems - they are independent. Point of this proposal is to conduct some work for trial to harmonize it and to create a mechanism for combining spatial data with statistical data. Harmonization causing some problems because statistics used statistical units so commonly and, unfortunately, geodesy do not like such division of space, prefer cadaster system. These fields need for further discussions. One of the solutions could be recode of primary coded phenomena (eg. Statistical units or cadaster) to one of a good agreed harmonize layer of the proposed model, but these is not an easy task. The main problem arises in cases of phenomena which relate to the other ranges than the one mentioned above.

IV. Polygon (level 7)

16. In geodesy the polygonal layer is commonly used. In case of environmental phenomena their polygonal ranges are quite problematic to identify due to difficulties in determining the location of its phenomena in space. To obtain the data collected with the required accuracy a new polygonal layer should be developed – dedicated for statistical purposes. But such a badly standardized layer would be characterized by a huge variability and also diversity of surveyed polygons. Consequently, for statistical purposes it would become confusing over time and useless for statistical analysis and comparisons.

V. GRID (level 8)

17. Some kind of compromise which leads to a good solution in that field is the idea to use a grid as a special type of the polygon. Mapping the obtain results to grid of squares - both in research carried out by statistics and observations conducted by geodesy - increases chances to improve coherence of these two systems, particularly with regard to environmental phenomena requiring spatial localization only in the form of a polygon. Such standardization of a polygon ensures grid with appropriately selected mesh. The problem is that the GRID objects should be generally introduced into the existing geodetic system, which is not so popular in mapping agency and may be not an easy task. But this step guarantees the proper development of the correct geocoding environmental phenomena presented in statistics. One kilometer grid is currently used in statistics mainly for the population data presentation and publication. Considering that geodesy is conducting surveys of phenomena with spatial ranges that are also difficult to define, use good fitted size of squares grid could, in the matter of fact, also solve this problem.
Not without significance is the fact that the one kilometer grid is so flexible that it allows for virtual connection of any data at any level of spatial division.

VI. LINE (level 9)

18. Geodetic data are also presented using linear objects. Currently in statistics there are no surveys that could be presented using this types of objects. It is a trend that can gain recognition in the future with the dissemination of statistical data related to e.g. transport, waterways or linear investments. The possibility of creating linear statistics will appear in the near future and it will allow for simple connection between linear statistical data with geometry offered by geodesy (linear geocoding).

VII. POINT (level 10)

19. At the lowest level of geocoding (the highest accuracy), in both systems, points reflecting the spatial position are functioning, usually in the form of x,y coordinates. Currently in this area it is observed the fastest progress in the field of cooperation between statistical and geodesy services. The reason is that in the last census most countries successfully used geometry of the address points and science that time it become an important link between statistical and spatial data (precise point geocoding).

VIII. Expected benefits

20. The following list contains the expected benefits derived from the “10 Level Model” presented above:

- An inventory of the levels of integration spatial objects used in statistics and geodesy for harmonization purposes;
- Supporting Statistical Geospatial Framework (SGF);
- Increase added value of statistical and spatial data;
- The benefits from the harmonization both systems as a result of methodological work around this issue;
- Geocoded unit record data in a data management environment;
- Use of fundamental geospatial infrastructure and geocoding;
- Better quality of geospatial data, integrated statistical and spatial data;
- Better quality of spatial analysis;
- Better interoperability of sets of data, possibility of analyses;
- Establish solid collaboration mapping agency with statistical institution with maintenance timeliness both systems;
- Possibility to lighted need and start enhance the collaboration between mapping agency with statistical institutions.

21. In conclusion, the question marks in the proposed model (lack of grid on the geodesy side and lack of linear objects on the statistical side, develop a mechanism or a key transition the statistical division into the geodesy division) should be the subject of intensive works in order to break down existing barriers and, as a starting point, to make practical progress in the methodology of combining spatial data with statistical data.