„Harmonization of statistical and geodetic divisions in the context of 10 level model, to develop a common framework as a standard of geospatial data production”

Janusz Dygaszewicz, Anna Sławińska
Central Statistical Office of Poland, j.dygaszewicz@stat.gov.pl, a.slawinska@stat.gov.pl

According to adopted rules, in Poland 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 statistics should consider the possibility of harmonizing statistical division (statistical regions and census enumeration areas) with the cadastral division (cadastral units), taking into account the needs of official statistics.

In reference to the above, Polish proposal of the 10 Level Model will be used to better understand and develop statistical and geodetic reference framework as a standard of geospatial data production.

The inventory the level of integration spatial objects used in statistics and geodesy is presented and described below in „*The 10 Level Model*” for harmonisation of statistical and geodetic reference framework:

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 level 6	Statistical regions Enumaration areas
+	POLYGON level 7	?
?	GRID level 8	+
+	LINE level 9	?
+	POINT level 10	+

ADMINISTRATIVE LEVELS (level 1-5)

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.

INDIVIDUAL UNITS FOR INTERIOR PUPRPOSES (level 6)

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 don't like such division of space, prefer cadaster system. These fields need for further discussions. One of the solution 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 work. The main problem arises in case of phenomena which relate to the other ranges than the one mentioned above.

POLYGON (level 7)

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 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.

GRID (level 8)

Some kind of compromise that leads to a good solution in that field is the idea to use 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 - gives 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.

LINE (level 9)

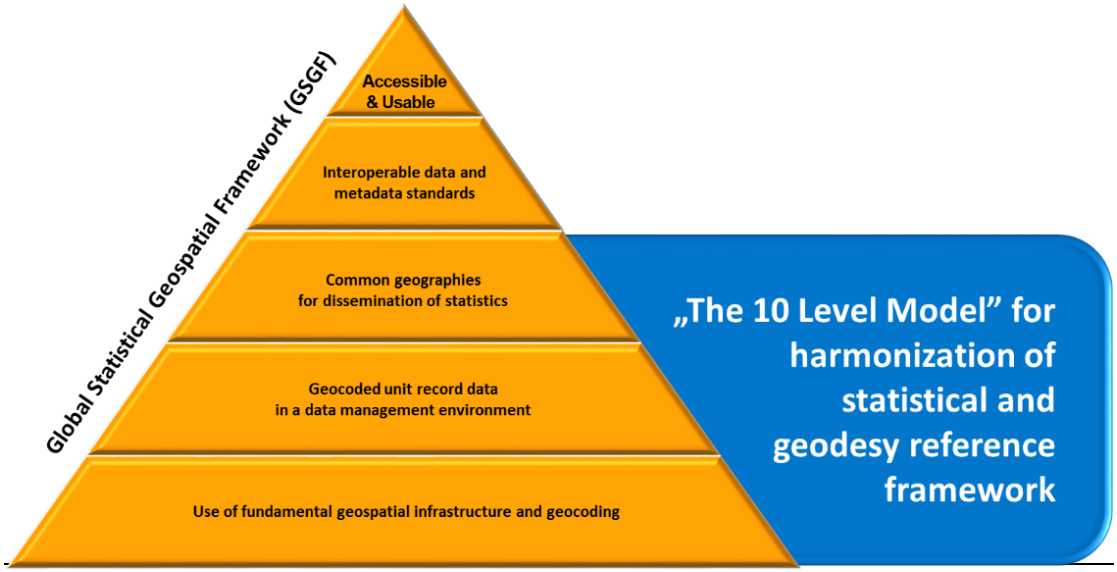
Geodetic data are also presented using linear objects. Currently in statistics there are no surveys that could be presented using this type objects. It is a trend that can gain recognition in the future with the dissemination of statistical data related to e.g. transport, waterways or linear investments. 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).

POINT (level 10)

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).

In Polish opinion the proposed „The 10 Level Model" should be the subject of practical activity in order to overcome existing barriers and as a starting point to make practical progress in the methodology of combining spatial data with statistical data.

Nevertheless the geospatial and statistical data integration landscape is very complex, with many players and initiatives The challenge, however, is to understand how best to achieve this integration in an effective and consistent way. Developing a coherent and systematic approach to linking statistical and geospatial data is likely to take some time. The best way to achieve consistent integration is through having a common method of geospatially enabling statistical and administrative data and integrating this with geospatial information through an internationally agreed the Global Statistical Geospatial Framework (GSGF), which enables comparisons within and between countries. This framework consists of five principles that are considered essential to integrating statistical and geospatial information. Implementation of "*The 10 Level Model*" into three lowest layers of GSGF could support practical activities of integrating statistics with geospatial data. "*The 10 Level Model*" could help to communicate and understand the geospatial capability requirements for statistical information. This could support data sharing between institutions, enhancing the interoperability of geospatial and statistical information and building institutional collaboration between geospatial and statistical communities. As a result it is expected to enable new, better and more integrated information for analysis and decision-making process.



Currently Poland is planning to carry out an experiment to develop a statistical division framework for official statistics with respect to the geodetic division of the country. It will be one of the main tasks in new Eurostat project *Merging statistics and geospatial information in Members States* which will start soon.

It is planned to conduct a feasibility analysis for harmonisation of the statistical division (statistical regions, census enumeration areas) with the geodetic division (cadastral units), considering the needs of official statistics and external stakeholders, which are using the statistical division. Poland will analyse size and diversity of spatial statistical units in terms of population limits as a criterion for delimitation of statistical regions and census enumeration areas. Additionally a comparative analysis of statistical and geodetic division of the country with determination of the size of statistical division units for use on the first stage of survey sampling will be performed. The final stage will encompass the proposal of a statistical division framework for official statistics.

Development of statistical division based on geodetic division should provide better interoperability of sets of data and raise possibilities of statistical geospatial analyses. However, it has to be remembered, that such harmonisation will cause that quality of statistical geospatial data and conducted analyses will depend on the quality of input spatial data from external administrative registers. That is why assessment of the overall quality of external data sets, and especially the quality of data which they include is essential.

Recently Polish official statistics worked on the project which aim was the improvement of the use of administrative sources. As a result, on the basis of Polish experience, the methodology of assessing the usability of administrative data sources has been elaborated. Nevertheless quality assessment should be conducted separately for each register, taking into account its possible use. The methodology of assessing register quality will include a few sections. For purpose of statistical division based on geodetic division the section regarding information about the quality of spatial data register will be the most important. The issues included under this section will enable assessing the overall quality of data sets, and the quality of data which they include. Within this area two criteria has been distinguished, i.e. accuracy and comparability which will be measured by specific indicators.

Polish methodology of assessing spatial data sets could be a proposal of standard for other countries and NSIs which want to harmonise statistical and geodetic divisions to receive better quality of statistical geospatial data and analysis.

In terms of GSGF, quality assessment of spatial registers, as a part of integration of statistical and geospatial information, could take place in lower layers of global framework, whereas the methodology of quality assessment could be implemented to fourth layer of GSGF (*Interoperable data and metadata standards*).