



Economic Commission for Europe

Conference of European Statisticians

Sixty-eighth plenary session

Geneva, 22-24 June 2020

Item 5 (b) of the provisional agenda

**Session 2: Experiences and results of concrete steps already taken by NSOs
and the geospatial communities to modernize their role****Realising the potential of geospatial and statistical data
integration at the time of expanded possibilities¹****Note prepared by Statistics Poland***Summary*

This document is presented to the Conference of European Statisticians seminar on “New roles for national statistical agencies and geospatial agencies in emerging national data ecosystems” for Session 2: “Experiences and results of concrete steps already taken by NSOs and the geospatial communities to modernize their role” for discussion.

¹ This document was scheduled for publication after the standard publication date owing to circumstances beyond the submitter's control.



I. Introduction

1. For several years, public statistics in Poland is creating and using spatial data. Taking into consideration modern technologies and information standards, and the technological needs of the censuses and surveys, Statistics Poland had undertaken work to ensure spatial identification of objects on the basis of spatial reference and GIS technologies. Geospatial information is becoming increasingly important in statistical production process and strengthens the role of statistical offices in the geospatial and statistical data integration.

II. Statistical Production Process Model

2. Currently, Statistics Poland has been developing *Statistical Production Process Model (Model Procesu Produkcji Statystycznej – MPPS)* on the basis of the *Generic Statistical Business Process Model (GSBPM)*² which has been enriched with geospatial components. To make a precise analysis of the processes implemented in Statistics Poland in terms of spatial data compliance with the processes recommended by the generic model, the related business processes were mapped to the GSBPM model. This model represents a generic reference standard of statistical business process.

3. The analysis was based on comparison of business functions in each phase of GSBPM with actual and planned processes and business functions implemented in statistical production, with particular emphasis on spatial data.

4. Compliance analysis was made by mapping various activities carried out in Statistics Poland to phases, processes and sub-processes recommended in GSBPM. The practical implementation of relevant processes relating to spatial data and mapping them with GSBPM model showed some important areas are not included in the model. By indicating these areas, the potential shortcomings of the GSBPM model were identified. The shortcomings essentially concerned statistical data “spatialization” aspects from the stage of designing the data collection through geocoding, analysis and dissemination of statistical products with spatial characteristics. These identified geospatial subprocesses that should be implemented in the architecture of statistical production process are:

- Phase 2: sub process “2.5a Design geocoding frame, sample & data collection”
- Phase 4: sub processes “4.1a Geocode frame & sample” and “4.3a Geocode collection”
- Phase 6: sub process “6.2a Prepare spatial analyses & maps”
- Phase 7: sub process “7.2a Manage spatial analyses & maps using GIS”.

5. Introducing the sub-processes listed above enriches the statistical production process model with a spatial component. This will allow to better understand spatial data and its role in the statistical production process, and to standardize the methodologies merging statistical data with spatial data. These new geospatial subprocesses (the yellow blocks on the graph below) can be found in the *Statistical Production Process Model* supporting Polish statistical production and integration of statistical and geospatial components for better and more comprehensive geospatial analysis.

² <https://statswiki.unecce.org/display/GSBPM/IV.+Levels+1+and+2+of+the+GSBPM>

Table
Statistical Production Process Model

1	2	3	4	5	6	7	8
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update outputs systems	8.1 Gather evaluation inputs
1.1a Prepare Statistical Programme (SSPPS)	2.2 Design variable descriptions	3.2 Reuse or build processing & analysis components	4.1a Geocode frame & sample	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.2 Consult & confirm needs	2.3 Design collection	3.3 Reuse or build dissemination components	4.2 Set up collection	5.3 Review & validate	6.2a Prepare spatial analyses & maps	7.2a Manage spatial analyses & maps using GIS	8.3 Agree an action plan
1.3 Establish output objectives	2.4 Design frame & sample	3.4 Configure workflows	4.3 Run collection	5.4 Edit & input	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	
1.4 Identify concepts	2.5 Design processing & analysis	3.5 Test production system	4.3a Geocode collection	5.5 Derive new variables & units	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5a Design geocoding frame, sample & data collection	3.6 Test statistical business process	4.4 Finalise collection	5.6 Calculate weights	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production system & workflow	3.7 Finalise production system	4.5 Verify plans	5.7 Calculate aggregates		7.6 Satisfaction research	
1.7 Prepare a general plan for the implementation of the production process	2.7 Design quality indicators	3.8 Verify projects and construction plans		5.8 Finalise data files		7.7 Prepare an improvement plan for the next research cycle	
1.8 Prepare a detailed plans for subsequent phases	2.8 Verify the general plan and detailed plans						
1.9 Verify the concept to improve satisfaction							

III. Geographic information system supporting statistical production process

6. Geospatial components have had more and more significant role in the Polish statistical production process. To conduct analyses which show territorial dimension in changes of different phenomena, statistical data with very precise location information is needed. Only official statistics has been collecting all necessary unit data with reference to points (x, y coordinates) that is essential for further processing and producing aggregates enabling credible and high-quality geospatial analysis. For census purposes it is also essential to have georeferenced data at the level of x, y coordinates.

7. To present the results of the previous round of Censuses: Agricultural Census in 2010 and Population and Housing Census in 2011, Statistics Poland used GIS technology. A data visualization platform was prepared and launched for internal and external users – the Geostatistics Portal³ to facilitate dissemination of georeferenced statistical information.

8. 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

³ <https://geo.stat.gov.pl/imap/?locale=en>

data, by way of processing, is presented in depersonalized form, preserving statistical confidentiality.

9. Currently, at the time of expanded possibilities, Statistics Poland has been working on the extension of the Geostatistics Portal and focusing on a new project for designing innovative advanced new generation tool that will support the statistical office in the role as an office for geospatial statistical data and for more sophisticated detailed data analysis centre.

10. The expansion project's name is "Spatial Statistical Data in the Information System of the State" (called PDS). Its aim is to broaden the scope and increase availability of statistical information and methods of geostatistical analyses using the resources of official statistics. The PDS project is a response to the needs expressed following the implementation of current Geostatistics Portal taking into account new data sources, IT tools and open source technologies. This aim will be achieved by designing a modern IT platform supporting the use of geographic information system in statistical studies and enabling more advanced and sophisticated geospatial analysis.

11. Within the PDS project, the functionalities of the previously provided services in the Geostatistics Portal 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, and also to 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 will be developed, which will facilitate the use of exploratory spatial data analyses using statistical information, analyses in the field of geostatistical modelling and support supplementing users' own data with geostatistical information and analyses.

12. Three completely new external services will emerge as a result of the PDS project:

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

13. The Exploratory geostatistical data analyses using statistical information available in the Portal as well geostatistical modelling services will allow users to:

- Import and supplement the user's own data with statistical data available in the PDS system
- Supplement the data with additional variables (the so-called derived variables) – calculated on the basis of already imported and supplemented user's data with using the available set of mathematical and geostatistical functions
- Explore the collected data (supplemented with derived variables) in geostatistical terms, evaluate their quality, and supplement and modify the data
- Visualise and publish the obtained results of analyses.

14. A wide range of methods of statistical spatial analysis will be made available, including: central tendency statistics, statistical dispersion, cluster analysis and spatial autocorrelation. The geostatistical modelling analyses service will also enable to build and apply a probabilistic model allowing the inference (estimation) of the value of the dependent variable on the basis of the results the probability sample survey (i.e. on the basis of the data collected).

15. The enrichment of user content with geostatistical information and analyses available in the Portal will allow users to:

- Import own content to analyse it
- Conduct content analysis with the use of "text mining" mechanisms to detect keywords referring to categories of objects available in the PDS System and to categorise user's content in an appropriate way

- Select the objects in the PDS System best matching the detected words and key phrases and establish correlations between the objects searched in the PDS System to select analyses available in the PDS System
- 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.

16. Additionally, the PDS project will also introduce new internal service facilitating efficient maintenance of spatial address databases.

IV. Conclusions and recommendations

17. Statistical data with precise geographical reference is essential to facilitate comparative analysis and achieve better statistical geospatial data production. At the time of expanded possibilities, the statistical office can increase its role by creating new generation tools for statistical analysis.

18. The Geostatistics Portal is used for spatial presentation of Poland's largest statistical information resource, enabling the publication of aggregated statistical data in the form of various types of spatial analyses, presented on maps with statistical confidentiality. Created PDS portal will enable very advanced analyses of statistical information, e.g. exploratory geostatistical data analyses or geostatistical modelling. The strength and uniqueness of the current Geostatistical Portal and planned new PDS platform results from the ability to perform reliable spatial analysis on the unit data being only in the possession of official statistics.

19. The PDS system will be developed in line with *Statistical Production Process Model* and will be part of a wider architecture of statistical production process in the Polish official statistics. The portal will present official statistical data and results of geostatistical analyses, critical for the functioning of the state, local government and local communities. The information will be accessible in a convenient graphical form, accelerating the decision-making process. The further goal will be to develop and present the results of censuses of the round 2020-2021.
