

Geospatial Enhancements in Statistical Production at Statistics Portugal

Rossano Figueiredo¹, Ana Santos², Magda Ribeiro³, Francisco Sardinha⁴

¹Statistics Portugal (Instituto Nacional de Estatística), Portugal

²Statistics Portugal (Instituto Nacional de Estatística), Portugal

³Statistics Portugal (Instituto Nacional de Estatística), Portugal

⁴Statistics Portugal (Instituto Nacional de Estatística), Portugal

Abstract

Better statistical-geospatial data integration is a strategic priority in the future vision of the European Statistical System (ESS) to support evidence-based decision-making and policy lifecycle and forecasting. In this regard, the consistent production of standardised geospatial statistics is a key line of action in modernising official statistics in the European context through the mission to implement the Global Statistical Geospatial Framework attending the regional statistical-geospatial operating environment (GSGF Europe). Throughout the implementation journey of this statistical-geospatial framework, quality is one of the topics that needs to be more deeply addressed and studied by the statistical community, especially when handling geospatial data, processes and services raises confidentiality and privacy concerns.

Since geospatial aspects are not specifically mentioned in the common quality framework for the ESS, future enhancements on this topic are required to tackle this gap and further appropriately embody it within all four levels of quality assurance, from statistical regulations to methods and tools. To achieve this goal, National Statistical Institutes (NSIs) need to monitor, align and streamline geospatial processes and services into the statistical business processes and services that handle geospatial content, namely by adopting a quality management approach underpinned by quality feedback procedures and geospatial quality reporting guidelines, methods and tools. Thus, geospatial quality management needs to be formally embodied in the statistical business production model as an overarching process and high-level corporate activity to produce high-quality geospatial statistics.

Statistics Portugal (SP) currently documents its statistical production process through its Statistical Production Process Manual (MPPE), which is fully aligned with the Generic Statistical Business Process Model (GSBPM, UNECE), version 5.1. In this manual, SP has also added a third, more detailed level of the process, going beyond the phases and sub-processes, as it describes the tasks and responsibilities of the process. Furthermore, SP acknowledged the review of its MPPE as an opportunity to introduce geospatial enhancements by more extensively documenting geospatial-related activities, services, tasks and respective guiding descriptions and drafting a quality management framework for statistical-geospatial data integration. These added geospatial enhancements will ensure the quality management, assessment and improvement of input and output geospatial data, processes and services considering standardisation, harmonisation and other key aspects outlined in GSGF Europe. This work also intends to guide the experts regarding the geospatial dimension in statistical business production, promote a common understanding and build synergies on tasks and responsibilities to produce geospatial statistics in a systematic and consistent manner.

Keywords: data integration, statistical business production model, geospatial statistics, geospatial quality

1. The Statistical Production Process Manual of Statistics Portugal

The Statistical Production Process Manual (MPPE) (INE, 2020) is an in-house reference quality document and a guiding production manual that uses the GSBPM conceptual model to describe the activities supporting the development of a statistical operation. It addresses an internal version of the statistical business production that systematically identifies and describes the phases, sub-processes and main tasks enabling a common language that facilitates communication within the organisation and builds synergies concerning roles and responsibilities. Nowadays this manual is used for many activities within the institution and is a reference in mapping the statistical production process.

The review of the reference document considered the previous version – MPPE 2020 (V 2.0) - following the regular update procedures of the 3rd Edition related to either required adjustments from changes in the production process or internal and external suggestions. This version included changes related to adjustments from the 5.1 version of GSBPM, simplified of schemes and contents by aggregating some tasks, the National Data Infrastructure project, responsibilities in collecting and managing data, aspects of the Information Security Management System and the General Data Protection Regulation and reinforcement of evaluation activities. From the geospatial perspective, new tasks were added to the matrix related to geospatial data, namely regarding the design of requirements and capacity building of the Geographical Information Infrastructure (IIG).

The MPPE breaks down into the Framework and Statistical Production Process Model. Firstly, the Framework is inspired by the UNECE GSBPM principles and structure (version 5.1, 2019) and follows the internal manuals to describe the activities supporting the production of official statistics. Secondly, the conceptual model outlines the description of the statistical production process adopted by SP breaking it down into three hierarchical action levels: i) phases; ii) sub-processes; and iii) tasks (120 in total). The application detail at the task level under each sub-process comprises the main difference with GSBPM addressing a more operational level of the model and overall structure. Moreover, this model systematises and describes the tasks of a statistical operation in an orderly manner (adjustable to any operation and not including cross-cutting activities) – into a production process matrix that outlines responsibilities,

stakeholders (responsible and intervening units), applicability dimensions according to the type of statistical operation and information sources as well as additional notes and remarks.

Although the geospatial contribution in the production process has been extended addressing tasks related to IT systems planning, the IIG and geospatial visualisation (e.g., Geolnq¹), new geospatial-related tasks and activities supporting the development of a statistical operation need to be extensively documented and described. The current version of MPPE outlines the functional requirements for geospatial development in several tasks (e.g., production needs of the IIG), however more detailed geospatial considerations need to be formally mapped in the next version of MPPE towards the implementation of the GSGF Europe (GEOSTAT 4, 2021) at the institutional level. In this regard, SP acknowledged the review opportunity of this reference in-house documentation to include new geospatial contents and enhance existing ones in the updated production process matrix at the task level considering the GSGF Europe Principles and supporting Requirements and Recommendations (GEOSTAT 4, 2022). This work aimed to support and improve the quality management of SP from the geospatial perspective in producing standardised geospatial statistics based on common best practices and internally established quality assurance approaches that handle geospatial data, processes and services.

In the current version of MPPE (2.0 version) minor geospatial considerations were outlined in the description of the following sub-processes: i) 2.3 – Design collection; ii) 2.4 – Design frame and sample; iii) 3.3 – Reuse or build dissemination components; iv) 4.3 – Run collection; v) 5.1 – Integrate data; and vi) 6.1 – Prepare draft outputs. In sub-process 2.3 (Design collection) geospatial data is included in the activities related to planning the integration of the statistical operation into the specified system supporting the collecting and reception of input data and management of the collection instruments in which technology systems handling this type of data are required (e.g., Geolnq). The sub-process 2.4 (Design frame and sample) acknowledges the importance of the geospatial component, specifically the spatial relationships of data, for defining the frame and specifying the conditions for selecting the sampling frame and the sample. The technical requirements supporting geosampling were addressed in more detail in sub-process 2.6 (Design production systems and workflow). The sub-process 3.3 (Reuse or build dissemination components) recognises geospatial statistics and maps as privileged dissemination means of statistical outputs in SP's Portal. Mapping web services, GIS applications and dynamic infographic handling geospatial data, services and

¹ Geographic visualisation application for the sample buildings.

related capabilities are other examples of the Portal's geospatial components. Concerning the sub-process 4.3 (Run collection) the geocoding was referred to in assigning geographical codes (geocodes) to collected data during the survey collection depending on the geographical basis/geospatial datasets used and the metadata territorial classification. In sub-process 5.1 (Integrate data), alongside statistical and administrative data from internal and external sources and providers, geospatial data is explicitly outlined as an infrastructure data type for data integration to derive statistics or for validation and processing purposes. Lastly, sub-process 6.1 (Prepare draft outputs) highlights the construction of web-based GIS applications and geospatial services (geographical bases, maps, among others) to support statistical dissemination.

At the task level, some functional requirements for geospatial development were assigned to a set of tasks in the statistical production matrix. Thus, the following 8 out of the mapped 120 tasks (6,67%) included notes and remarks embodying geospatial aspects, namely related to the IIG capabilities for sampling, collection, processing, analysis and dissemination needs:

- Task no. 11: Plan the statistical operation in the supporting IT systems (1.6 - Prepare and submit business case)
- Task no. 15: Identify the methods and modes used, the timing, as well as the functional requirements for collecting and receiving micro-data (2.3 – Design collection)
 - Task related to the sub-process 2.6 – Design production systems and workflow and 3.1 – Reuse or build collection instruments)
- Task no. 17: Specify the functional requirements for the integration of micro-data in the National Data Infrastructure, including Business Intelligence Systems (2.3 – Design collection)
 - Task related to the sub-process 2.6 – Design production systems and 3.2 – Reuse or build processing and analysis components)
- Task no. 30: Specify the functional requirements for integrating the data in the Data Warehouse and for supporting data processing and analysis (2.5 – Design processing and analysis)
 - Task related to the sub-process 2.6 – Design production systems and workflows and the sub-process 3.2 – Reuse or build processing and analysis components)

- Task no. 31: Specify the functional requirements for disseminating statistical information (2.1 - Design outputs)
 - Task related to the sub-process 3.3 – Reuse or build dissemination components)
- Task no. 32: Specify the requirements of the Geographical Information Infrastructure (IIG) to meet production needs (2.6 – Design production systems and workflows)
 - Task related to the sub-process 3.4 – Configure workflows on capacity building)
 - This task identifies the types of location data that make up the IIG and that are relevant for geocoding, geospatially enabled statistical/administrative (micro)data and for statistical content (e.g., spatial analysis): i) addresses; ii) buildings; iii) small statistical areas (sections and subsections); and iv) road network (roads and streets)
- Task no. 44: Enable the IIG according to the specified requirements (3.4 – Configure workflows)
 - Task related to the sub-process 2.6 – Design production systems and workflows.
- Task no. 51: Select the sampling frame from the specified reference frame and analyse its quality (4.1 – Create frame and select sample)
 - Task related to the sub-process 2.4 – Design frame and sample and the sub-process 3.4 – Configure workflows.
 - This task encompasses the sampling plan and includes the availability of the IIG capabilities to respond to the identified methodological needs.

This analysis from the geospatial perspective of the MPPE last version demonstrated that there is a set of tasks that require geospatial improvements to be documented and integrated into the next statistical production matrix of the SP to increase the capability to integrate statistical and geospatial data across different phases of statistical operations and produce geospatial statistics in a standardised manner.

2. Preliminary results from the geospatial enhancements in the MPPE update at task level

Under the 2023-PT-GEOS2023 Grant (Enhanced location data and quality framework for geospatial statistics), SP has committed to documenting geospatial-related activities and tasks in the MPPE towards the implementation of GSGF Europe. The objective of updating and

enhancing the new version of the MPPE (updating the level of the main tasks of the statistical production process (3rd level) with geospatial enhancements and respective guiding descriptions in producing geospatial statistics is one of the specific objectives of the WP4 (Geospatial enhancements in statistical production) of the same project. In addition, the WP4 aims to improve the quality of geospatial data and processes used to support statistical business production by developing and documenting geospatial quality methods and tools supporting overarching quality management processes.

These objectives were defined due to the lack of internal guiding documentation related to the geospatial components within statistical production, either from the input and output perspective, i.e. on geospatial data quality assessment, or from the processing perspective, i.e. evaluation of the geospatial processes, statistical processes handling geospatial content and geospatial services. For this paper, SP will present the preliminary results of the first exercise and analysis round on updating the tasks outlined in the new version of the MPPE matrix (129 in total), with geospatial considerations. The geospatial considerations included generic descriptions related to the geospatial-related activities outlined in the GeoGSBPM (UNECE, 2021) and GSGF Europe for the tasks involving geospatial data, processes and services. The Principles and the list of the Requirements and Recommendations of the regional version of the statistical-geospatial global framework (outcomes of the GEOSTAT 3 and 4 projects) were also assigned to each task of this group insofar as they have a direct or indirect functional relationship or represent a production/capability need to effectively carry out the task.

Geospatial considerations were attached to 45 of the 129 tasks mapped in the new version of MPPE based on their description and ability to accommodate operational issues related to geospatial data, processes and services, which could be the input for an aspect-oriented version of the MPPE related to geospatial data integration. The 45 tasks are incorporated in all the 8 GSBPM phases and 27 sub-processes, from sub-process 1.1 (Identify needs) to sub-process 8.2 (Conduct evaluation). 7 tasks under the following five sub-processes are related to all 5 Principles of GSGF Europe since they all support the development of the task and are cross-cutting to the production process for geospatial statistics: i) sub-process 1.1 (Identify needs); ii) sub-process 1.6 (Prepare and submit business case); iii) sub-process 3.6 (Test statistical business process); iv) sub-process 3.7 (Finalise production systems); and v) sub-process 6.2 (Validate outputs). These tasks are required to support the development of the statistical operation covering production elements related to the legal framework,

data/information and user needs, methodological description, feasibility study, pilot-study technical specifications, preliminary methodology document, technical certification, executive memorandum and quality control. Moreover, 23 tasks were assigned to Principle 1 (Use of fundamental geospatial infrastructure and geocoding), 26 tasks were assigned to Principle 2 (Geocoded unit record data in a data management environment), 9 tasks were assigned to Principle 3 (Common geographies for the dissemination of statistics), 6 tasks were assigned to Principle 4 (Statistical and geospatial interoperability) and 7 tasks were assigned to Principle 5 (Accessible and usable geospatially enabled statistics). 32 tasks were assigned to more than one Principle, including the above-mentioned 5 tasks involving all 5 Principles. In general, the tasks assigned to the Principles and respective Requirements and Recommendations address the IIG technical requirements to support the production process, geosampling data collection and management, data integration and dissemination.

3. Conclusions

From this first exercise and preliminary results, it was evident that more extensive and detailed geospatial enhancements in the MPPE matrix need to be further exploited and streamlined to support the production of geospatial statistics in the context of implementing GSGF Europe in SP. In addition, there is an agreed acknowledgement of the need to develop and document considerations on the geospatial (data) requirements and technical specifications, especially regarding administrative data and emerging data sources, and overarching quality management activities handling the geospatial components across the statistical production process. It is also important to transpose the Requirements and Recommendations and the supporting materials of GSGF Europe to the national case, for instance in calculating the geospatial indicators and assessing the statistical-geospatial maturity level. In this regard, future work involves the ongoing development of the update and enhancement of the MPPE and overall quality management of SP with formal geospatial considerations across the different statistical operations and production of guiding documentation and related materials supporting the standardised production of geospatial statistics and moving forward with the GSGF Europe in SP.

Acknowledgment

This work was carried out under the ongoing 2023-PT-GEOS Grant (Enhanced location data and quality framework for geospatial statistics), an EU funded project in the following topic 'SMP-ESS-2023-GEOS-IBA — Geospatial Statistics'. In this regard, Statistics Portugal is grateful for the financial support received from the EU to improve the production of geospatial statistics.

References

UNECE (2021). Geospatial view of Generic Statistical Business Process Model. GeoGSBPM. Version 1.0.

GEOSTAT 4 (2021). GSGF Europe. Eurostat ESSnet grant project GEOSTAT 4, 2019-2022.

GEOSTAT 4 (2022). GSGF Europe: Requirements and Recommendations. Eurostat ESSnet grant project GEOSTAT 4, 2019-2022.

Statistics Portugal, INE (2020). Statistical Production Process Manual (MPPE). 3rd Edition.