United Nations ECE/CES/2019/24

Economic and Social Council

Distr.: General
9 April 2019

Original: English

Economic Commission for Europe

Conference of European Statisticians

67th plenary session
Paris, 26-28 June 2019
Item 2 (b) of the provisional agenda
New data sources – accessibility and use
Session 2: Skills needed to use new data sources

Improving data integration with the help of the Global Statistical Geospatial Framework

Note by Statistics Sweden

Summary

Geospatial information and its integration in statistical production is becoming increasingly important, as national statistical offices respond to user needs for statistics on accessibility, on the urban/rural dimension or statistics at the local level in general.

Statistics Sweden has done a national assessment of the Global Statistical Geospatial Framework, a framework developed under the United Nations Global Geospatial Information Management and endorsed by the United Nations Statistical Commission. The assessment focuses on the five principles of the framework from a capability point of view, where the capability to integrate statistical and geospatial information is supported by people, systems, processes, information, methods, standards and frameworks as well as institutional arrangements.

The document will focus on how countries can improve their data integration capability by using a global framework in combination with a global model, the Common Statistical Production Architecture. The Swedish assessment has helped Statistics Sweden to identify areas for improvements related to the different capability elements. It has also helped to show where the status is good enough.

This document is presented to the 2019 Conference of European Statisticians seminar on "New data sources – accessibility and use", session 2 "Skills needed to use new data sources" for discussion.

GE.19-05944(E)

1905944

I. Introduction

1. Territorial policies at the national and European levels, together with the 2030 Agenda for sustainable development[1], have put pressure on European national statistical offices (NSOs) to produce relevant and timely statistics[2]. The 2030 Agenda aims to leave no one behind; from a geographical perspective this means, for example, taking the urban-rural dimension into account, as well as issues of accessibility.

2. It is more important than ever to relate official statistics to the realities people are facing in, for example, deprived neighbourhoods. Furthermore, disaster risk management or measures to improve accessibility to public transport, schools, health clinics, or green areas requires local-level statistics. The European Statistical System[3] has recognised this necessity by funding the GEOSTAT projects[4], such as GEOSTAT 2, thereby providing guidance to NSOs on the fundamentals of a more flexible statistical infrastructure that is point-based[5].

3. Regional statistics are regularly produced by all the countries in European Union, typically using the Nomenclature of Territorial Units for Statistics (NUTS) classification of administrative units in all member States[6]. The Eurostat database provides harmonised regional statistics, but for users looking for more geographic detail, the situation is far from harmonised. An initial step to improve this has been to introduce the use of statistical grids, degree of urbanisation, and so on, together with classifying Local Administrative Units as the new geographic outputs as per the revised EU regulation on territorial typologies, Tercet[7].

4. Statistical grids are set to become one of the dissemination areas in the next European round of population and housing censuses in 2021[8]. Production of grid-based statistics calls for the modernisation of the way countries conduct their censuses and leads the way to a register-based statistical system.

II. Linking people and businesses to place

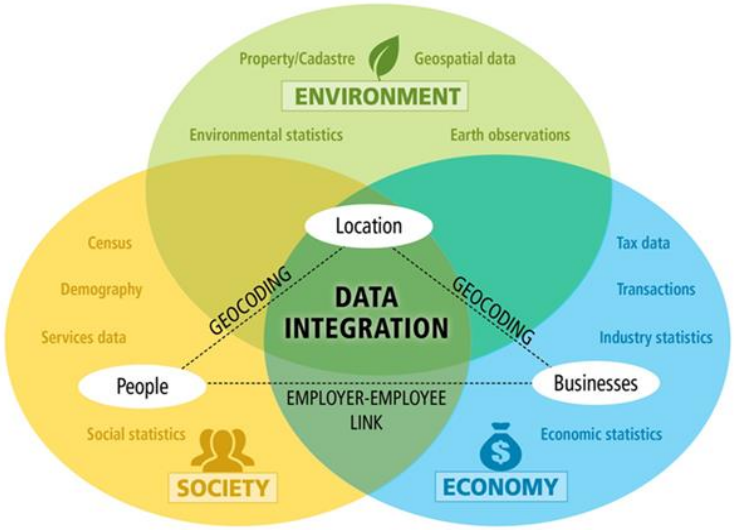
A. Benefits of statistical geospatial integration

5. The benefits of statistical geospatial integration are starting to be widely recognised and the UN Committee on Global Geospatial Information Management has created an Expert Group on Integration of Statistical and Geospatial Information (EG–ISGI)[9] [9]. Statistics Sweden is taking an active part in the work of the expert group and has also coordinated the ESSnet project GEOSTAT 3, which focused on developing a European implementation of the Global Statistical Geospatial Framework (GSGF)[10]. An ESSnet project is a network of several ESS organisations aimed at providing results that will be beneficial to the whole European Statistical System (ESS).

6. The GSGF was adopted in 2016 and directly supports the integration of statistical and geospatial information[11]. Statistics Sweden has started implementing the GSGF nationally and it has proved to be a useful tool for planning and identifying improvements in this aspect of statistics, as described later in this paper.

7. The GSGF helps add location to statistical data that describes individuals and businesses. It also helps organise the geospatial data management and the processes needed to get high-quality statistical output.

Figure 1
Linking society, economy, and environment through data integration[12]



8. Furthermore, the GSGF provides a bridge between the official statistical agency and the geospatial profession. Statistical units are geocoded using geospatial identifiers, such as cadastral parcel, address, or building. Data are then aggregated and released for the geographic boundaries in the boundary (or geography) layer. This supports integration of data across diverse sources and provides the bridge between statistical and more traditional geospatial datasets; through use of a common geography and greater flexibility in geographic outputs.

B. The five principles of the Global Statistical Geospatial Framework

9. The GSGF is a high-level framework, which means that it is not intended to provide a detailed implementation plan or design; rather, it provides guidance on what should be available in countries, providing a lot of flexibility on how it is done.

10. To help countries implement the framework, supporting documents were developed in 2018, both at the global level (by the EG–ISGI) and at the European level (through the GEOSTAT 3 project). The GEOSTAT 3 implementation guide[13] gives a good overview of how to understand the principles from a European perspective.

1. Principle 1: Use of fundamental geospatial infrastructure and geocoding

11. The goal of this principle is to obtain a high-quality and standardised location attribute or reference. The framework recommends using existing geospatial information provided by specialised geospatial authorities (e.g. physical address, cadastral parcel, building, or some other location description) in order to assign accurate coordinates and/or a small geographic area or standard grid reference to each statistical unit (i.e. at the micro-data or unit record level).

2. Principle 2: Geocoded unit record data in a data management environment

12. The GSGF recommends that the linking of a geocode to each statistical unit record in a data set (i.e. a person, household, business, building, or parcel/unit of land) occurs within a data management environment. Persistent and version-controlled storage of high precision geocodes enables any geographic context to be applied when preparing the data to be released in the future (i.e. in aggregating data into a variety of larger geographic units or to adapt to changes in geographies over time).

3. Principle 3: Common geographies for dissemination of statistics

13. To enable comparisons across data sets from different sources, the GSGF recommends that a common set of geographies be used for the display, reporting, and analysis of social, economic, and environmental information.

4. Principle 4: Statistical and geospatial interoperability – data, standards and processes

14. Both the statistical and geospatial data communities operate their own general data models and metadata capabilities; however, these are often not applied universally. The statistical community uses the Generic Statistical Information Model (GSIM), the Statistical Data and Metadata Exchange (SDMX), and the Data Documentation Initiative mechanisms. The geospatial community, meanwhile, makes use of other standards [14].

15. Within the statistical community, there is a need to build geospatial processes and standards into statistical business processes in a more consistent manner. UNECE has revised the Generic Statistical Business Process Model (GSBPM)[15] during 2018, including better descriptions of the use of geospatial data and methods in the statistical production process.

5. Principle 5: Accessible and usable geospatially enabled statistics

16. This principle of the GSGF emphasises the need to identify or, where required, develop policies, standards, and guidelines that support the release, access, analysis, and visualisation of geospatially enabled information. One important aspect of this principle is to ensure that data can be accessed using safe mechanisms that not only protect privacy and confidentiality, but also enable access to data to undertake various analyses that foster decision making.

III. The capability to integrate statistical and geospatial information

A. Statistics Sweden's capability assessment

17. A capability enables a statistical organisation to undertake one or more activities. The general elements that are required for a capability, according to the Common Statistical Production Architecture, are people, processes, methods, technology, standards, and frameworks[16]. As part of the GEOSTAT 3 project, working on how to apply the GSGF in Europe, Statistics Sweden has performed a national assessment of each of the framework's principles.

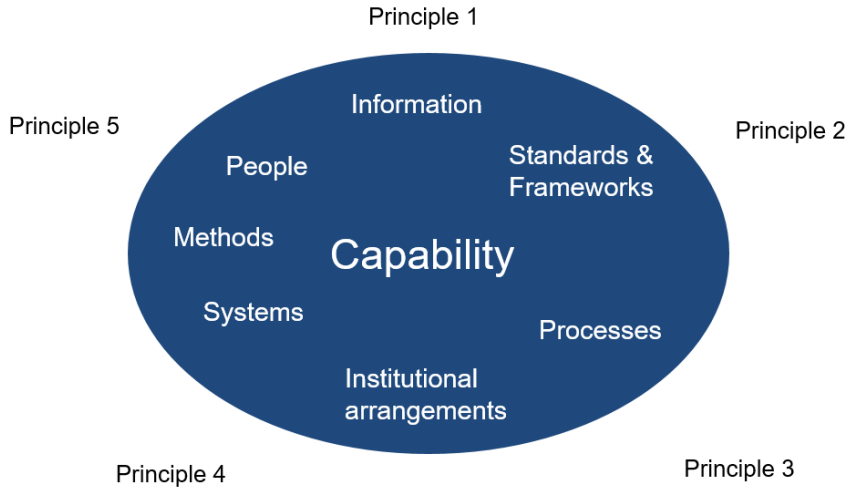
Table 1
Capability elements

Capability elements	
People	The skills, knowledge and aptitudes required from people to undertake the activity.
Systems	The ICT applications, hardware and platforms required by the organisation to undertake the activity.
Processes	The preferred set of steps or tasks undertaken by the business, with reference to methods, standards and systems where necessary, to perform the activity in an efficient and effective manner.
Methods	The set of specific techniques or algorithms required to undertake the activity.
Standards and frameworks	The standards, frameworks, guides and policies required to undertake the activity.

Capability elements	
Information	The data needed to perform statistical and geospatial integration activities.
Institutional arrangements	Legal and institutional infrastructure required to conduct and support the activity.

18. Looking at the framework from a capability point of view proved helpful, as the framework principles need to be underpinned by certain skills, processes, methods, and so on. The assessment showed that, in the case of statistical and geospatial integration, the elements underpinning the capability may be expanded to include information and institutional arrangements too[17].

Figure 2
Elements needed for an NSI to have the capability to integrate statistical and geospatial information



19. The assessment exercise can easily be conducted by NSIs who have recently started to work on statistical–geospatial integration, as well as those looking for a way to improve it. The steps proposed in the Swedish report are simple and start with user needs.

20. Looking at each of the elements underpinning a capability for each principle in the framework revealed the strengths and weaknesses of Statistics Sweden. Setting goals and deciding which activities to prioritise makes it easier to include actions in yearly plans and to follow up. The actions can be categorised as either new activities that need to be initiated or as those that need to be improved. The assessment has also shown the activities that need to be maintained or reduced.

B. Need to start new activities

21. Statistics Sweden identified nine actions that will lead to improved capability in statistical and geospatial integration. For example, providing guidance on geospatial data and geocoding practices, similar to the ABS documentation online[18]. This will help streamline the internal processes at Statistics Sweden as well as help other authorities responsible for official statistics. We also need a systematic approach to the geospatial integration of all registers with improved geospatial accuracy as a result. This will save time later in the processes, as no additional quality improvements are necessary.

22. Providing all statistical geographies as open data is an important action. Another is to have geospatial data and metadata embedded into Statistics Sweden's statistical production architecture. This will allow better integration between statistical business process, data management, metadata and infrastructure.

23. Furthermore, we need to setup concrete targets for desirable geographical quality in registers (e.g. the share of workplaces in the business register to be successfully geocoded). We also need to launch interactive map services on scb.se, and a number of core indicators on small area statistics geographies should be released as open data.

C. Need for improvements

24. Fourteen actions were put down in the "improvement category", targeting areas where we are working already – but where more can be done to improve our capability in statistical and geospatial integration.

25. We need to improve our feedback from Statistics Sweden to Lantmäteriet, the National Geospatial Agency, on quality issues in the data deliveries. We should also improve the use of agreed terminology and data quality declarations. We would benefit from increased use of SQL server spatial to conduct spatial operations. The address information in the Business register needs improvements, so that all workplaces can be properly georeferenced.

26. The use of geospatial information through the whole production process needs to be improved along with increased support by the GIS-experts to the rest of the organisation. Organised training courses and workshops in how to use geospatial data for your statistical analysis is one example.

27. Furthermore, we should increase the availability of all administrative and statistical geographies used by any agency within the national statistical system as open data through a web service. As far as possible, also historical data should be provided to ensure that historical statistics can be properly mapped onto accurate geographies.

28. We also need to improve our confidentiality methods to take into account the more flexible way that users download and combine various datasets (geographic differencing). We should also provide a geographical map entry to the Statistical Database (as a complement to the tabular entry that currently exists).

D Maintaining what works well

29. The assessment also shows what is already working well. Statistics Sweden has a good cooperation with Lantmäteriet and other geospatial data providers. The register-based statistical system with administrative data sources is the foundation for integration.

30. The REGINA service[19] is a valuable improvement in accessibility to the historical changes in administrative geographies. It has been improved with view services but will need to be managed and improved to include more historical administrative geographies.

31. Expertise in search, view and download services required by INSPIRE is valuable when working on how to increase our web services.

32. There are automatic confidentiality processes working well, making production of small area statistics efficient reducing disclosure risk. Statistics Sweden has a Delivery Portal for statistics on demand, which is a safe delivery system for commissioned services.

E. Reducing inefficiencies

33. When assessing capability elements, it was evident that a lot could be gained by reducing inefficient processes regarding quality checks and duplication of work. Cumbersome ad hoc correction of data and quality controls late in the production process should be avoided.

34. There is also inefficient data storage and duplications of databases. Statistics Sweden no longer stores historical versions of geospatial information. They are stored in a national data storage that makes the production more complex. There is also unnecessary confusion about different versions of geographies (scale, boundary changes etc) which need to be reduced.

35. Moving towards more open data will reduce the need for chargeable services, which will make small area statistics more accessible and more used in society.

III. Conclusions

36. Geospatial information is still considered something that mostly concern a small group of GIS-experts at Statistics Sweden. This assessment is targeting where new initiatives and improvements are necessary to get an increased uptake of geospatial data and processes in the whole organisation.

37. For each principle, a number of capability elements have been assessed resulting in proposals on what we need to start, improve, maintain or remove. Activities identified in the assessment has since then been incorporated in the yearly work plans and important step towards more open data has been taken. However, more work is needed to reach the goal of having geospatial data and processes included and utilised by the whole organisation.

38. The assessment exercise is warmly recommended to any NSO wanting to improve its geospatial capability. By implementing the Statistical Geospatial Framework with a focus on relevant capability elements the NSO can prioritise where to invest in skills, processes and tools.

IV. References

[1] United Nations General Assembly. Resolution 70/1. Transforming our world: the 2030 Agenda for Sustainable Development.
[2] Haldorson M., Zaccheddu P., Fohgrub B., Petri E. (2016). Geospatial information management in Europe – responding to the user needs, Statistical Journal of the IAOS 32
[3] European Statistical System, http://ec.europa.eu/eurostat/web/european-statistical-system/overview?locale=fr
[4] GEOSTAT Project Descriptions, http://www.efgs.info/geostat/
[5] Moström J. (editor) et al (2017). A Point-based Foundation for Statistics, Final report from the GEOSTAT 2 project
[6] NUTS Regulation (2003). Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003, on the establishment of a common classification of territorial units for statistics (NUTS), OJ L 154, 21.6.2003
[7] Tercet, Territorial Typologies of the European Union. http://ec.europa.eu/eurostat/web/nuts/tercet-territorial-typologies
[8] European Census, https://ec.europa.eu/eurostat/web/population-and-housing-census/overview
[9] UN-GGIM Expert Group on Integration of Statistical and Geospatial Information, http://ggim.un.org/UNGGIM-expert-group/
[10] The GEOSTAT 3 Project, http://www.efgs.info/geostat/geostat-3/
[11] UN EG-ISGI Background Document (2016). United Nations Expert Group on the Integration of Statistical and Geospatial Information: Global Statistical Geospatial Framework: Linking Statistics and Place, Current status and plans for development, July 2018
[12] Illustration developed by Statistics Australia and Statistics New Zealand
[13] GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe, Proposal from the GEOSTAT 3 project, Version 1.0, 28 February 2019.
[14] Documentation from UNECE Workshop on Statistical and Geospatial Standards, Stockholm November 2017, https://www.unece.org/index.php?id=45404
[15] Generic Statistical Business Process Model. GSBPM, http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0
[16] Common Statistical Production Architecture, CSPA, https://statswiki.unece.org/display/CSPA/Common+Statistical+Production+Architecture
[17] Haldorson M., Moström J. (2018): Implementing the Statistical Geospatial Framework at Statistics Sweden, National Report as part of the GEOSTAT 3 project
[18] The ABS Website on Statistical Geography, http://www.abs.gov.au/geography
[19] Statistics Sweden's REGINA Web Service, Information available in English in a pdf document: http://regina.scb.se/