GEOSTAT 3 – A European Implementation Guide for the Global Statistical Geospatial Framework

Karin Hedeklint, Statistics Sweden, karin.hedeklint@scb.se

Abstract

Members of the ESS (European Statistical System) are currently seeking to collectively modernise their statistical production systems, to transform their operations and to derive new relevant metrics and indicators. This modernisation process involves an ambition to level up integration of geospatial information in statistical production.

One of the main goals of the GEOSTAT 3 project is to develop a European implementation guide for the Global Statistical Geospatial Framework (GSGF). The development of 'GSGF Europe' is believed to be an important step towards a better and more coordinated integration of statistical and geospatial information. The ultimate objective of GSGF Europe is to increase resource efficiency, to obtain a higher degree of harmonisation between countries and to support creation of a more flexible statistical production.

Whereas the Global framework is a high-level framework consisting of five generic principles, broad enough to apply to any region around the globe, the European guide aims to be a tighter framework reflecting the specific European situation. It will build on the major achievement that have already been done for the availability of geospatial information through the INSPIRE directive and the Spatial Data Infrastructures set up by Member States.

The GEOSTAT 3 project runs until January 2019. For the 2018 Quality conference, the presentation describes the key-elements of the framework as proposed by the project.

Keywords: Geospatial, Framework, Data Integration, Infrastructure, Accessible

1. The GEOSTAT 3 project

GEOSTAT 3 is an ESSnet Eurostat grants project that started in February 2017. The final report will be delivered in January 2019. The general objective of the project is to foster a better integration of statistics and geospatial information in order for the statistical community to provide more qualified descriptions and analyses. The main

task is to develop a European implementation guide for the Global Statistical Geospatial Framework (GSGF), taking into account existing conditions and initiatives, including European and national frameworks.

The project is also supporting the European Forum for Geography and Statistics (EFGS), in its networking and information sharing, by maintaining the EFGS website and organising the yearly EFGS conferences.

The GEOSTAT 3 project consists of a network of eight NSI's and three sub-contractors, where Statistics Sweden is the coordinator[1]. By sub-contracting the German NMCA BKG, which is chairing the UN GGIM: Europe working group on data integration, the project has ensured an operational interaction with UN GGIM: Europe.

2. Drivers for integration of statistical and geospatial information

There are a number of drivers for a closer integration of statistical and geospatial information. Within the National and International Statistical Systems, there is a move towards an increased use of administrative data and registers for census purposes. In parallel, many countries have launched national geospatial strategies to geocode administrative records in order to support data integration. In the ESS context, the 2021 population and housing census is the first driver for geo-enabling statistics and integrate geospatial information into statistical production.

European and global sustainable development programs increasingly require reliable and relevant information in terms of higher spatial resolution and increased abilities for spatial and thematic disaggregation. Hence, Agenda 2030 and its Sustainable Development Goals (SDGs) is another driver that is pushing for a closer integration of statistical and geospatial information.

There is a growing need to add the "where" dimension in public information and statistics, both in general and in the monitoring of the SDGs in particular. Thus, the statistical and geospatial community have a common task to build frameworks and workflows that can enable integration of data sources in a flexible way. Using data from a range of sources and for multiple purposes requires their integration into a common reference system of harmonised concepts.

---

[1] The other NSI's are from the Netherlands, Finland, Norway, Poland, Austria, Estonia and Portugal. The sub-contractors are the German and the Norwegian Mapping and Cadastral Agencies and MDmapping.

3. The Global Statistical Geospatial Framework (GSGF)

The international statistical and geospatial communities recognised this challenge and responded by establishing a UN Expert Group (UN EG-ISGI) to develop a Global Statistical Geospatial Framework (GSGF)[2]. The GSGF provides the international statistical and geospatial community with a common approach to connecting socio-economic and environmental data to appropriate locations, and improves the accessibility and usability of this geospatially-enabled data. The framework should act as a bridge between statistics and geospatial information, between statistical institutes and geospatial agencies, and between statistical and geospatial standards, methods, workflows and tools.

The framework is not intended to provide a detailed implementation plan or design for its implementation, but rather provide guidance on what should be available in countries leaving a lot of flexibility on the "how".

4. Why a European implementation guide?

The European adoption of the GSGF is believed to be an important step towards a better and more coordinated integration of statistical and geospatial information. The outcome of the GEOSTAT 3 project will be an implementation guide that does not aim to introduce new organisational structures, but rather to streamline the existing ones and provide bridges between initiatives.

The relative regional homogeneity between countries within Europe, concerning National Spatial Data Infrastructures and National Statistical Systems, allows for a coordinated implementation of the GSGF in Europe. In order to do this, the implementation guide needs to add flesh to the bones of the GSGF and go beyond the generic global principles, by providing recommendations building on the specific European situation. In particular, the implementation guide rests on two major cornerstones as foundation for the statistical geospatial integration in Europe:

- The achievement undertaken for the availability and harmonisation of geospatial information through the implementation of the INSPIRE directive

---

[2] United Nations Expert Group on the Integration of Statistical and Geospatial Information 2016. Background Document on Proposal for a Global Statistical Geospatial Framework (Advanced Draft as of 28/07/2016).

and the National Spatial Data Infrastructures (NSDIs) set up by EU Member States.

- The existing and well-established structure for collaboration and harmonisation of European official statistical provided through the European Statistical System (ESS).

In essence, INSPIRE and the ESS constitutes the institutional and geographical scope of the recommendations as they provide rational infrastructures and mechanisms (legal, technical, collaborative and financial) to harmonise the implementation of GSGF. Accordingly, the EU/EFTA countries form the core area for the implementation guide. However, Europe is wider than the EU. Apart from the fact that non-EU/EFTA countries cannot rely on INSPIRE and the ESS, the GEOSTAT 3 project suggests that most recommendations are relevant also for these countries.

5. GSGF Europe – principles

The GSGF is a high-level framework that consists of five broad principles that are considered essential for integrating geospatial and statistical information. Each of the principles is defined by a set of objectives, and is supported by international, regional and, where applicable, domestic standards and good practices.

Figure 1. The Global Statistical Geospatial Framework (GSGF)
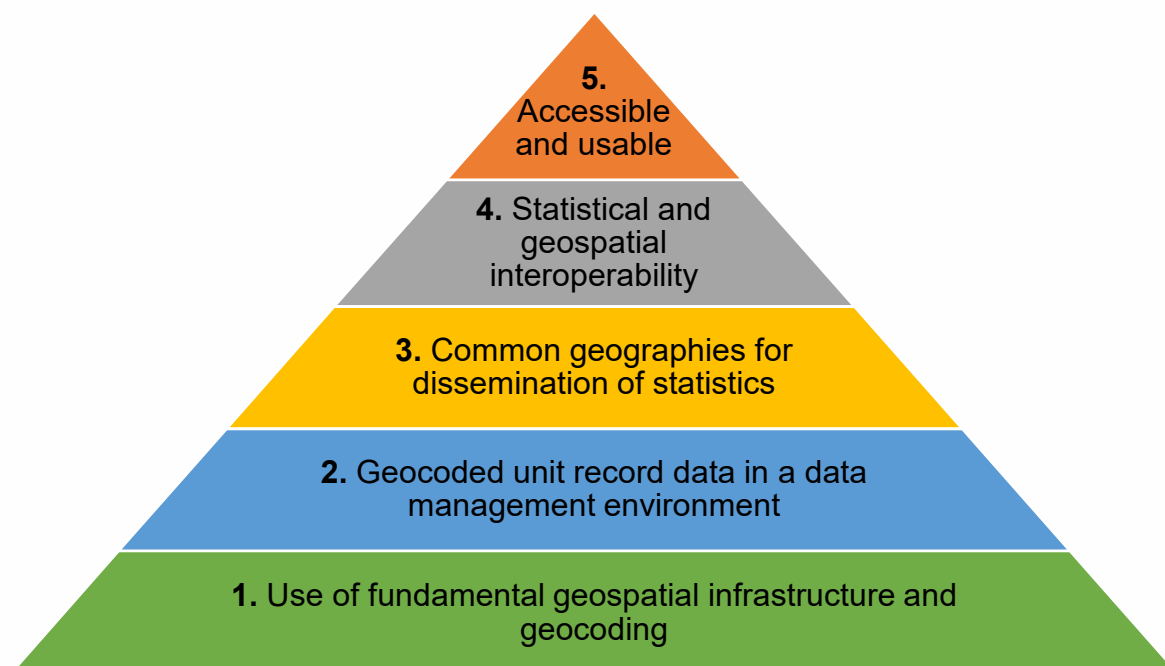
5.1. Principle 1 – Use of fundamental geospatial infrastructure and geocoding of statistical information

Principle 1 describes the geospatial infrastructure needed to geocode information, whereas the actual process of geocoding is occurring in principle 2. The goal is to create an infrastructure that will support the provision of high quality and standardised location attributes. The data should be existing geospatial information, provided by specialised authorities. It could be standardised physical addresses, cadastral parcels or building identifiers, that can be linked to each unit in a statistical or administrative dataset, such as a person, household, business, building or parcel/unit of land. Time and date stamping these locations will place the unit both in time and in space.

When obtaining locations and geocodes it is recommended to use official, fundamental geospatial data from the National Spatial Data Infrastructures or other nationally agreed sources.

GSGF Europe suggests that the following requirements should be met to fulfil the objectives of principle 1:

1. Use data from National Spatial Data Infrastructures
2. Use point-based location data for geocoding
3. Build formal working relationships on institutional agreements

The INSPIRE directive sets out the framework for this principle, as Member States are obliged to follow specific requirements for the sharing of data and structural harmonisation. On the content-side, and considering other data quality aspects such as scales, density etc., the UN-GGIM: Europe has launched the Core Data Concept[3], that will complement the INSPIRE data specifications.

Principle 1 requires the cooperation between statistical institutes and the national agencies that are responsible for the geocoded reference data. The cooperation should ideally rely on (formal) agreements on roles and responsibilities for organisations and, when applicable, comply with the framework set by the INSPIRE

---

[3] http://un-ggim-europe.org/

directive, taking advantage of the coordinating/steering committees and technical structures created for its implementation.

5.2. Principle 2 – Geocoded unit record data in a data management environment

Whereas principle 1 describes the requirement for the geospatial infrastructure needed for a consistent geocoding, principle 2 is dealing with one specific use of the infrastructure; the actual process of geocoding. The framework recommends that the geocoding occur within a data management environment, where main aspects are the storage of data and the quality of both input and output data.

GSGF Europe suggests that the following requirements should be met to fulfil the objectives of principle 2:

1. Enable data integration through consistent data architecture
2. Store location only once
3. Build an effective and safe data management environment
4. Ensure consistency and quality of geocoding results
5. Develop a consistent approach to manage non-matching data
6. Use point-of-entry validation in collection of administrative or statistical data

The recommendations aim at creating good quality of the geocoding results, and facilitating future use of the data. A clear management of data will enable future changes in geographies over time, or data linking processes that aim to integrate information of varying nature and sources.

5.3. Principle 3 – Common geographies for dissemination of statistics

The goal of principle 3 is to obtain a common and consistent set of geographies for production and dissemination of statistics. Geographies are the spatial representations of the administrative, statistical or functional division of a country, also known as administrative or statistical *units*. Use of common geographies will ensure that all statistical data is consistently enabled and that users can discover, access, integrate, analyse and visualise information seamlessly for geographies of interest.

The following main groups of geographies can be identified:

- *National* administrative, statistical and functional geographies,

- *European* statistical geographies within the TERCET framework (LAU and NUTS areas),
- Global, European and national statistical *grids*.

The capability among NSIs and geospatial agencies to deliver information for geographical areas defined by the user, is increasingly important. However, such geographies cannot be included in the group of common and official geographies maintained by statistical or geospatial agencies.

GSGF Europe suggests that the following requirements should be met to fulfil the objectives of principle 3:

1. Use and maintain a consistent framework of national statistical and administrative geographies
2. Use and maintain a consistent framework of European statistical geographies
3. Consolidate use of existing statistical grid systems and explore the potential of evolving global systems

There are several international standards for the harmonisation of these common geographies, where the INSPIRE regulation contains specifications for the accessibility to data. There are both global and European projects that are working for a harmonised international grid system.

5.4. Principle 4 – Statistical and geospatial interoperability; Data, Standards and Processes

There are clear benefits of greater interoperability between statistical and geospatial data and metadata standards, from cataloguing to data interchange and better communication. Overcoming structural and syntactic barriers between data and metadata from different providers will enhance the efficiency of discovery, access, and use of geospatially enabled data.

Both the statistical and geospatial data communities operate their own general data models and metadata capabilities. Within the statistical community, there is a need to build geospatial processes and standards into statistical business processes in a more consistent manner. Therefore, the framework recommends an approach where geospatial frameworks, standards and processes are incorporated into the Common

Statistical Production Architecture and its components. In particular, the Generic Statistical Business Process Model (GSBPM) needs to refer more to the use of geospatial data and methods.

GSGF Europe suggests that the following requirements should be met to fulfil the objectives of principle 4:

1. Improve geospatial workflows within statistical production
2. Publish data once and leave it at its source, to be reused many times
3. Increase use of automated solutions for merging geography and statistics
4. Explore the potential of Linked Open Data

Within this principle, The GEOSTAT 3 project is making a special effort on promoting a proof of concept for a tool for automated linking of geographies and statistics, in a concept called Table Joining Services. It is well suited for cases where multiple themes of statistical data share the same reference geographies. Instead of publishing the same geometries repeatedly, the geometry is published only once and multiple statistical content can be added on demand.

Figure 2. The Table Joing Service concept
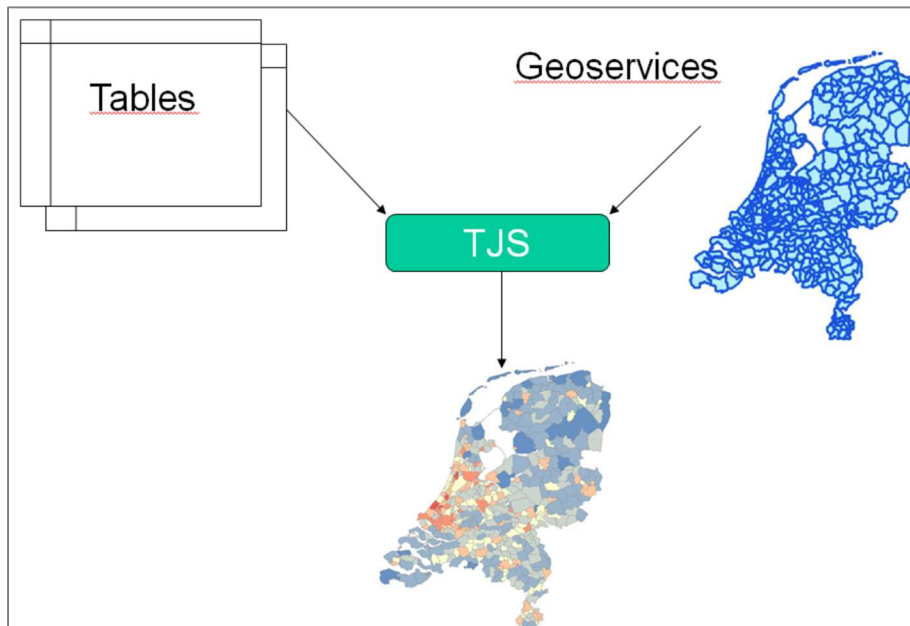
5.5. Principle 5 – Accessible and usable geospatially-enabled statistics

The goal of principle 5 is to make sure that geospatial statistics is accessible and usable in the best possible way. The recommendations includes the making of policies, standards and guidelines, which support the release, access, analysis and

visualisation of information. There is a wide range of legislative and operational issues that organisations need to be aware of when releasing and analysing information about people and businesses in a spatial context. One important aspect is to ensure accessible data, using safe mechanisms that protect privacy and confidentiality. On the other hand, you need to enable open access to data in order to undertake a broad range of analyses to inform decision-making.

Other issues of relevance include data quality in its different dimensions, particularly with regard to reliability, timeliness, and relevance. Analysis, dissemination and visualisation are also included in the work of Principle 5.

GSGF Europe suggests that the following requirements should be met to fulfil the objectives of principle 5:

1. Implement clear and simple data licensing policies
2. Increase use of European services for easy access to pan-European data
3. Use service oriented national data portals supporting dynamic integration of data
4. Define clear national and European rules to ensure protection of privacy
5. Facilitate data search and use through improved guidance and cataloguing

6. On-going work in the GEOSTAT 3 project

The final version of GSGF Europe will be presented to Eurostat in January 2019. As a quality control of the model, the Statistical institutes that are participating in the project will test the framework. Three of the UN Sustainable Development Goal indicators, where you need to combine geospatial and statistical data, have been chosen for the test. The chosen indicators are:

- 11.2.1 Access to public transport,
- 11.3.1 Land consumption by population growth,
- 11.7.1 Built up areas of cities, open space for public use.

The results of the test, together with an updated version of GSGF Europe, will be presented at the EFGS conference in October 2018. In the final version of the guide, there will be an annex with good practise examples.