GeoStat

GSGF Europe: Reference Architecture

GEOSTAT 4

eurostat

Title: GSGF Europe: Reference Architecture

Author: GEOSTAT 4

Project: Eurostat ESSnet grant project GEOSTAT 4

Grant agreement number: 945503 - 2019-FI-GEOSTAT4

Summary

GEOSTAT 4 project has produced the first draft of the GSGF Europe reference architecture. This reference architecture describes actors, roles, processes, services and concepts so that the geospatial statistical community can share the same view of their operating environment, discuss the same concepts and build solutions on the same common base.

The architectural descriptions are based on a Lean Architecture method which means that architecture descriptions are produced only when they are needed and add value. The architectural descriptions comply with the ArchiMate Enterprise Architecture modelling language 3.1.

Content

Summary 2
1. List of Abbreviations 4
2. Introduction 5
3. Why a reference architecture? 5
4. Architectural method 6
5. Roles and Actors 6
6. Processes 7
7. Services 9
8. Conceptual model 9
9. References 10
10. Annexes 11
10.1 Annex I – Roles and Actors Map 11
10.2 Annex II – A List of Roles and Actors 11
10.3 Annex III – An Example of a Roles and Actors Map, FI 11
10.4 Annex IV – Geospatial Activities 11
10.5 Annex V – A List of Generic Services 11
10.6 Annex VI – The Key Concepts 11

1. List of Abbreviations

CSPA – Common Statistical Production Architecture

EA – Enterprise Architecture

EFGS – European Forum for Geography and Statistics

EG-ISGI – Expert Group on Integration of Statistical and Geospatial Information

GAMSO – Generic Activity Model for Statistical Organisations

GeoGSBPM - Geospatial view of Generic Statistical Business Process Model

GSBPM – Generic Statistical Business Process Model

GFGS – Global Forum for Geography and Statistics

GSGF – Global Statistical Geospatial Framework

GSIM – Generic Statistical Information Model

HLG-MOS – High-level Group for the Modernisation of Statistical Production and Services

IT – Information Technology

NGIA – National Geospatial Information Agency

NSI – National Statistical Institute

TOGAF – The Open Group Architecture Framework

2. Introduction

The aim of this GSGF Europe reference architecture is to integrate Enterprise Architecture (EA) into an organisation's development work. GEOSTAT 4 project has produced the first draft of the GSGF Europe reference architecture. This reference architecture describes actors, roles, processes, services and concepts so that the geospatial statistical community can share the same view of their operating environment, discuss the same concepts and build solutions on the same common base. This allows the interoperability of implemented solutions and their sharing between organisations at the national and international level. GSGF Europe reference architecture specifies how the objectives presented in the GSGF Europe strategy map will be achieved.

3. Why a reference architecture?

An enterprise architecture aims to create an environment which can change and support business strategies and goals described in the EFGS strategic map (GEOSTAT 4, 2021). It shows what the business needs are, where the organisation wants to be, and ensures that the IT strategy aligns with this. Enterprise architecture helps to remove silos, improves collaboration within an organisation and ensures that the technology is aligned to the business needs. This work also helps in standardising organisations at the national and international level.

In the fields of enterprise architecture and software architecture a reference architecture provides a template solution for an architecture for a particular domain. It also provides a common vocabulary with which to discuss design and implementations, often with the aim to stress commonality. The international statistical community has some widely accepted reference architectures. The most widely used of these are:

- CSPA (UNECE, 2021a)
- GSBPM (UNECE, 2019a)
- GeoGSBPM (UNECE, 2021b)
- GSIM (UNECE, 2019b)
- GAMSO (UNECE, 2019c)

These reference architectures have been considered in the building of the GSGF Europe reference architecture. Especially GSBPM and GSIM frameworks have strong relations to this work.

The GSGF provides the international statistical and geospatial community with a common framework to connect statistical data, such e.g., socio-economic and environmental data, to appropriate locations and improves the accessibility and usability of this geospatially enabled data. However, it is a high-level framework, meaning it is not intended to provide a detailed implementation instruction but rather guidance on what should be available in countries, leaving a lot of flexibility on the "how".

The GSGF Europe reference architecture aims to act as a template for statistical organisations in the development of their own geospatially enabled Enterprise Architectures. It provides a common framework incorporating geospatial frameworks, standards and processes in a consistent manner, in order to produce geospatially enabled data and statistical services.

4. Architectural method

The architectural descriptions produced in GEOSTAT 4 are based on a Lean Architecture method which means that architecture descriptions are produced only when they are needed and add value. It was identified that the most useful descriptions would be actors, roles, processes, services and conceptual models. These views provide generic insights that can cover various operational use cases and serve high level objectives that are described through strategic maps with drivers and goals, which will be produced in WP6 of the GEOSTAT 4 project.

The architectural descriptions comply with the ArchiMate Enterprise Architecture modelling language 3.1. The ArchiMate modelling language is an open and independent Enterprise Architecture standard that supports the description, analysis and visualisation of architecture within and across business domains. ArchiMate is hosted by The Open Group and is fully aligned with TOGAF. The Specification of the ArchiMate modelling language 3.1 is available on The Open Groups web pages (The Open Group, 2019).

Graphical architectural descriptions are produced with the Archi application. Archi is open-source modelling toolkit for creating ArchiMate models.

5. Roles and Actors

The role and actor documents aim to describe the roles in the production of geospatial statistics and link them to the more concrete actors. The roles and actors are defined at three different levels: national, European and global (Annex I). Descriptions at European and global level identify the main actors in the field of geospatial statistics and link them to generic roles. The national level role diagram focuses on the identification and production of generic roles, which can be recognised across countries and organisations. To these roles, there are linked actors, which may vary between the countries. For instance, at the national level, the role “Provider of Intelligent Geospatial Components” is carried out mainly by National Mapping and Cadastral Authorities but in some countries can be endorsed by other organisations, like public-private consortiums, academic research centres, etc. The key point has been to document generic roles that are the same although the actors responsible for that role may vary. In addition to modelling and linking actors and roles, they are also documented in excel sheet format (Annex 2).

The statistical and geospatial communities play a large role in the successful implementation of the GSGF. In GSGF, it is recognised that each country seeking to implement the GSGF, should first ensure that there is a clear assessment, planning and agreement on priority areas for action and agreement on roles and contributions by key stakeholders and a range of national institutions, particularly NSIs and NGIAs.

Statistics Finland has produced an example from its own operating environment, in which some of the roles are linked to responsible actors (Annex 3). The idea is that each country can produce their own representation of the actor diagram. One significant remark is that users of geospatial statistics are identified only as a role in the generic diagram. It would be a very challenging task to identify and document all users as actors in a generic model. Therefore, they are not documented in this study.

6. Processes

The specifics of statistical-geospatial production processes are that they are very dependent on geospatial production processes. These are usually carried out by National mapping agencies and aim to produce geospatial reference data and tools (Web service or user interface). These can be used for different purposes, and particularly by statistical-geospatial processes, in order to produce and disseminate integrated geospatial and statistical information. These processes must be designed to ensure consistency in terms of the production of geolocated information. The target as defined by the GSGF is expressed in principle 1 (“Use of fundamental geospatial infrastructure and geocoding”). The goal of this principle is « to obtain a high quality, standardised physical address, property or building identifier, or 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 microdata level)».

The statistical-geospatial production is based both on geospatial reference data (boundaries, areas, national address repositories, etc.) and statistical data collection (surveys, administrative data, big data, smart statistics, etc.). The statistical-geospatial processes are usually carried out by statistical organisations and aim to produce geospatially enabled statistics and tools (Web service or user interface) to collect, produce and disseminate at a detailed geographical level. Most statistical-geospatial processes only use geospatial data and services in order to produce their statistical deliverable. These processes are generally referred to as geocoding. Principle 1 of the GSGF stipulates that in this case, the process of obtaining locations and geocodes should use relevant, fundamental geospatial data from National Spatial Data Infrastructures or other nationally agreed sources.

Besides, when the statistical-geospatial process uses data and tools provided by the geospatial process, it is likely that the Review and validate sub-process (GSBPM 5.3) will lead to the detection of errors. In this case, there might be a need to modify geospatial data based on the outputs of the statistical-geospatial process in order to improve the quality. In this case, there should be a feedback from the statistical process to the geospatial one, with a quality bridge between the two infrastructures, geospatial and statistical.

However, the interactions between these processes can be even more complex. Some statistical-geospatial processes may also end-up creating new geospatial objects that should be added or changed in the geospatial reference data. For instance, statistical pipelines dedicated to pendular migrations e.g., home to workplace trips may produce specific areas measuring the geography of local labour markets. The computed boundaries of these local labour markets can be considered as a geospatial object produced through statistical-geospatial processes. While the statistical deliverable could be databases measuring unemployment rates in each of these areas.

Consequently, there is a need for governance and management processes to handle changes, workflows and supportive activities. The description of processes within the GSGF Europe reference architecture emphasises in particular the relationships between the two types of production processes. Defined at an international level it should help to adapt national architectures to local contexts.

The UN Expert Group on the Integration of Statistical and Geospatial Information (EG-ISGI) recognised that a top-down approach is required in order to incorporate geospatial frameworks, standards and processes into the Common Statistical Production Architecture (CSPA) and its components (GEOSTAT 3, 2019). In particular, the Generic Statistical Business Process Model (GSBPM) needs to refer to the use of geospatial data, methods and processes in the statistical production process. Especially data, standards and methods that are described in the GSGF.

Geospatial Task Team under the Supporting Standards Group of the High-Level Group on the Modernisation of Official Statistics (HLG-MOS) has developed a geospatial view of the GSBPM (henceforth GeoGSBPM). The GeoGSBPM describes geospatial-related activities. In particular those activities that are needed to produce geospatially enabled statistics, using the framework of the GSBPM. The result is that GeoGSBPM describes activities that are smaller than subprocesses of GSBPM. The Geostat 4 project takes these activities as modules that can be used as components in national process descriptions. The use of these common components in different countries makes the process diagrams understandable and interoperable between countries. These activities will be linked with roles to show who is responsible for what and to generic services that are needed to perform these activities. Geospatial activities are also documented in an excel list (Annex 4)

7. Services

To carry out efficient statistical-geospatial processes, a set of modular services are necessary. Some of these services would be implemented by the NSIs (statistical services), others made available by the National Spatial Data Infrastructure (NSDIs, geospatial services). These services would potentially be used by both stakeholders, statistical institutes and mapping agencies, but also by external users.

A list of generic services needed to produce geospatial statistics was identified and described by the Geostat 4 project (Annex 5). This list includes previously widely agreed services such as services for geocoding and dissemination of geospatial data and some other services. These services can be linked to the geospatial-statistical activities in process diagrams to perform the necessary tasks. To indicate the ownership and the use of these services they can be linked to the roles and actors presented in corresponding diagrams.

8. Conceptual model

The key concepts of the geospatial statistic are documented by geospatial statistical conceptual model (Annex 6). The model is intended to be a high-level general presentation that is valid in different countries. Due to this, details (e.g., multiplicity) have been omitted from the model. Also dimensions like time might vary according to national practices so they are not part of this model although they are significant from the point of view of database modelling.

The model consists of three types of elements:

- Yellow business object elements represent the main concepts of the model
- Orange business object elements provide practical examples of the most common use of the main concepts
- Green elements represent processes

When a statistical unit is linked to location, it becomes a geospatial object which has a direct link to geospatial terminology and geospatial elements. The aim of the Conceptual model is to describe these key elements for geospatial statistical data. These concepts should be linked to the inputs and outputs of the processes /activities and they should be described at a more detailed level for national needs.

One important aim of the architecture work is that we will have a common terminology and understanding within the statistical and geospatial communities - at least at the overall level described here. In addition to these architecture descriptions a terms and definitions document will be produced. This terms and definitions document describes the concepts shown in the conceptual model and builds a common understanding in this area.

9. References

GEOSTAT 4, 2021. Development Plan of the EFGS. Deliverable 6.2 of the GEOSTAT 4 project. Eurostat ESSnet grant project GEOSTAT 4.

GEOSTAT 3, 2019. GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe. Proposal from the GEOSTAT 3 project. EFGS and Eurostat. (https://www.efgs.info/wp-content/uploads/geostat/3/GEOSTAT3_GSGF_EuropeanImplementationGuide_v1.0.pdf)

The Open Group, 2019. the ArchiMate® 3.1 Specification. (https://pubs.opengroup.org/architecture/archimate3-doc/)

UNECE, 2021a. Common Statistical Production Architecture. CSPA. Version 2.0. (https://statswiki.unece.org/display/CSPA/CSPA+v2.0)

UNECE, 2021b. Geospatial view of Generic Statistical Business Process Model. GeoGSBPM. Version 1.0. (https://statswiki.unece.org/display/GSBPM/GeoGSBPM)

UNECE, 2019a. Generic Statistical Business Process Model. GSBPM. Version 5.1. (https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1)

UNECE, 2019b. Generic Statistical Information Model. GSIM. Version 1.2. (https://statswiki.unece.org/display/gsim/Generic+Statistical+Information+Model)

UNECE, 2019c. Generic Activity Model for Statistical Organisations. GAMSO. Version 1.2. (https://statswiki.unece.org/display/GAMSO/GAMSO+v1.2)

10. Annexes

Annex I – Roles and Actors Map

Roles and actors are described at three levels: national, European, and global.

Annex II – A List of Roles and Actors

Actors and roles are also documented in excel sheet format.

Annex III – An Example of a Roles and Actors Map, FI

Statistics Finland has produced an example of its own operating environment, in which some of the roles are linked to responsible actors.

Annex IV – Geospatial Activities

Geospatial activities identified by UN HLG-MOS are documented in an excel list.

Annex V – A List of Generic Services

A list of generic services needed to produce geospatial statistics was identified and described in an excel sheet format.

Annex VI – The Key Concepts

The key concepts of the geospatial statistic are documented in a conceptual model.