GeoStat

GSGF Europe: GSGF and Frameworks Environment

GEOSTAT4

eurostat

Title: GSGF Europe: GSGF and Frameworks Environment

Author: GEOSTAT 4

Project: Eurostat ESSnet grant project GEOSTAT 4

Grant agreement number: 945503 - 2019-FI-GEOSTAT4

Summary

In this document, nine other frameworks are reviewed and linked to the GSGF. The aim is that geospatial information and related standards and aspects will be a natural part of the ESS statistical production. This is done by identifying links and roles between the selected frameworks and the GSGF. The frameworks seem to fit well together. The frameworks are not developed in isolation and must be seen in the context of a whole suite of standards and frameworks.

Often, the frameworks can be viewed as a “cookbook” to make sure that necessary aspects have been identified, e.g. geospatial aspects within all statistical business process steps or information objects. Generally, the frameworks also aim to provide a common vocabulary and a structure to support collaboration activities, particularly in the field of modernisation.

The Global Statistical Geospatial Framework, GSGF is a high-level framework which facilitates consistent production and integration approaches for geo-statistical information. It is generic and permits application of the framework principles to the local circumstances of individual countries. The GSGF acts as a bridge between statistical and geospatial professional domains, between NSIs and NGIAs, and between statistical and geospatial standards, methods, workflows and tools.

Content

1. Abbreviations

API – Application Programming Interface

DCAT – Data Catalog Vocabulary

DDI – Data Documentation Initiative

DWG – Domain Working Group

GSGF – Global Statistical Geospatial Framework

ESS – European Statistical System

EU – European Union

FAIR – Findable, Accessible, Interoperable, and Reusable

GAMSO – Generic Activity Model for Statistical Organisations

GeoGSBPM - Geospatial view of Generic Statistical Business Process Model

GSBPM – Generic Statistical Business Process Model

GSIM – Generic Statistical Information Model

CSDA – Common Statistical Data Architecture

CSPA – Common Statistical Production Architecture

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

IETF – Internet Engineering Task Force

IGIF - Integrated Geospatial Information Framework

INSPIRE - Infrastructure for Spatial Information in the European Community

IR – Implementing Rules

ISO – International Organization for Standardization

IT – Information Technology

NGIA – National Geospatial Information Agency

NSI – National Statistical Institute

OGC – Open Geospatial Consortium

SDMX – Statistical Data and Metadata eXchange

SOA – Service Oriented Architecture

SP – Strategic Pathways

TOGAF – The Open Group Architecture Framework

UNECE – United Nations Economic Commission for Europe

UN-GGIM – United Nations Committee of Experts on Global Geospatial Information Management

UNSC – United Nations Statistical Commission

W3C – World Wide Web Consortium

WFS – Web Feature Service

WMS – Web Map Service

2. Introduction

In this document, nine other frameworks are reviewed and linked to the Global Statistical Geospatial Framework (GSGF). The aim is that geospatial information and related standards and aspects will be a natural part of the ESS statistical production. This is done by identifying links and roles between the selected frameworks and the GSGF.

This document aims to help a reader to:

- understand the frameworks by highlighting the key points of each framework, but not going into details,
- recognise the role of each framework in statistical and/or geospatial fields, and
- distinguish how they are linked to the GSGF.

One European and eight global frameworks are reviewed in this document:

- The Generic Statistical Business Process Model, GSBPM
- The Geospatial view of Generic Statistical Business Process Model, GeoGSBPM
- The Generic Activity Model for Statistical Organisations, GAMSO
- The Generic Statistical Information Model, GSIM
- Common Statistical Data Architecture, CSDA
- Common Statistical Production Architecture, CSPA
- The Integrated Geospatial Information Framework, IGIF
- Infrastructure for spatial information in Europe, the INSPIRE
- Open Geospatial Consortium, OGC

3. The Frameworks Environment

3.1 GSBPM and GeoGSBPM to Identify Activities and Tasks to Convert Data into Statistical Information

The Generic Statistical Business Process Model, GSBPM (UNECE, 2019a) is a collection of related and structured activities and tasks to convert input data into statistical information. In the context of the GSBPM, organisations or groups of organisations perform statistical business processes to create official statistics to satisfy the needs of the users. Essential work is done by UNECE HLG-MOS task team which has also made a document describing geospatial related activities under each sub-process of GSBPM to primarily produce geospatially enabled statistics (UNECE, 2021a).

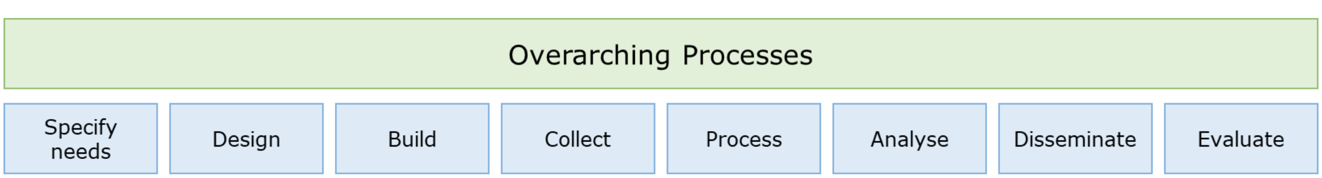

Figure 1. The eight phases of the statistical business process.

The GSBPM also recognises several overarching processes with a strong statistical component that apply throughout the eight phases. These are quality management, metadata management, data management, process data management, knowledge management, and provider management. Greater value will be obtained from the GSBPM if it is applied in conjunction with the GSIM (see 3.3).

3.2 GAMSO to Define Capabilities Needed to Support Statistical Production

The Generic Activity Model for Statistical Organisations, GAMSO (UNECE, 2019b) extends and complements GSBPM by identifying additional activities needed to support statistical production. While individual collaboration typically focuses on modernising a particular aspect of production, as described by the GSBPM, statistical production occurs within a broader context of corporate strategies, capabilities and support. Greater value will be obtained from the GAMSO if it is applied in conjunction with the GSBPM.

The GAMSO is composed of four activity areas: strategy and leadership, capability development, corporate support and production. Each activity area, except production, is further broken down into activities and examples of the activities are provided in the text. For production GAMSO does not seek to redefine these because the GSBPM phases represent the activities in this activity area.

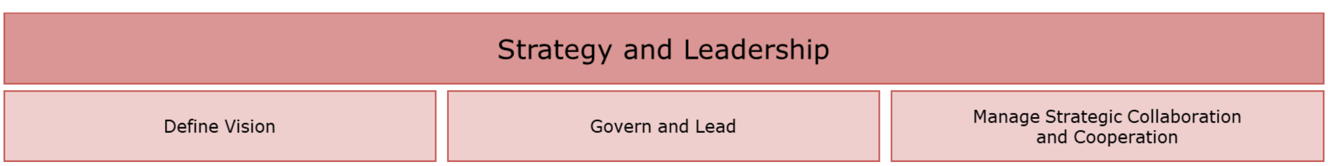

Figure 2. High-level strategic activities that enable statistical organisations to deliver the products and services needed by governments and communities nationally and internationally.

Figure 3. This activity area includes research, development and innovation activities i.e., the development of capabilities that enable the organisation to undertake new activities, or to improve the efficiency of existing ones. When a new capability or a capability improvement is fully integrated in Production, its support is transferred to one or more activities of Corporate Support.

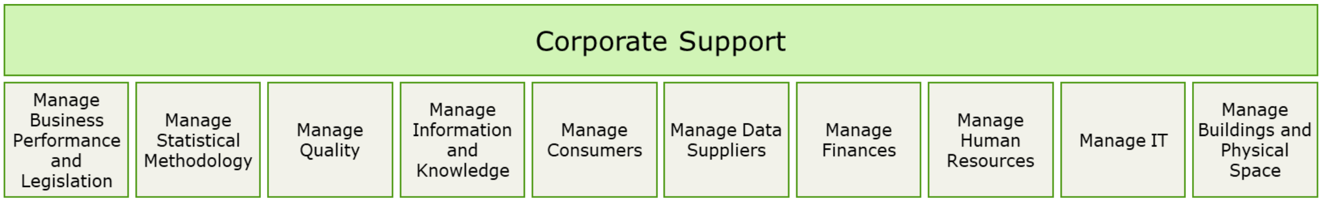

Figure 4. The Corporate Support activity area covers the cross-cutting functions required by the organisation to deliver its work programme efficiently and effectively.

3.3 GSIM to Define and Describe Information Objects in a Harmonized Way

The Generic Statistical Information Model, GSIM (UNECE, 2019c) is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information used in the production of official statistics i.e., information objects. GSIM also enables generic descriptions of the management and use of data and metadata throughout the statistical production process. GSIM identifies more than 100 information objects, which are grouped into broad categories, and explained in more detail in the documentation.

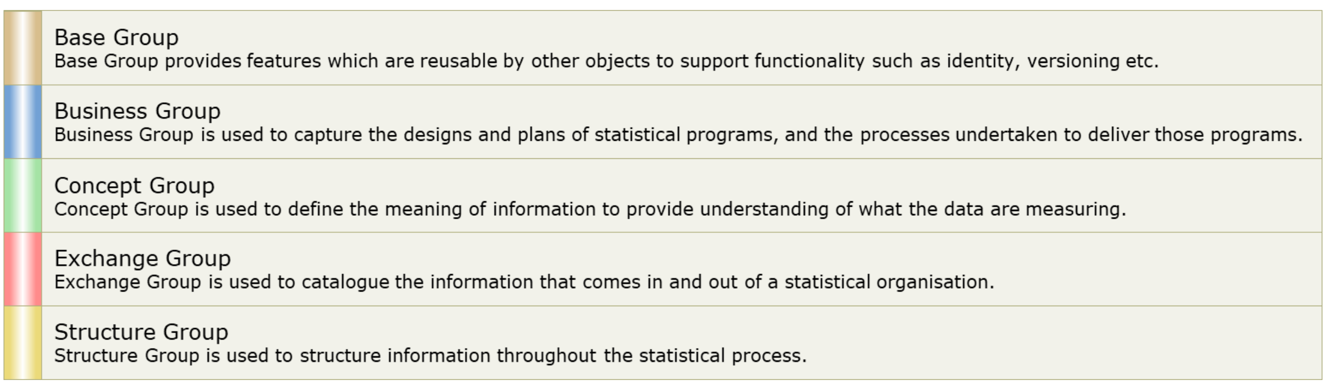

Figure 5. Top-level information object groups in GSIM.

GSBPM and GSIM are recognised as key standards for the modernization of official statistics. They aim to provide common terminology, improving communication about the production of statistics, within and between organizations. This, in turn, facilitates collaboration and reuse and sharing of good practices, methods, components and processes leading to greater efficiency. Greater value will be obtained from the GSIM if it is applied in conjunction with the GSBPM (see 3.1).

By defining objects common to all statistical production, regardless of subject matter, GSIM enables statistical organizations to rethink how their business could be more efficiently organized. The GSIM helps to explain significant relationships among the entities involved in statistical production and can be used to guide the development and use of consistent implementation standards or specifications.

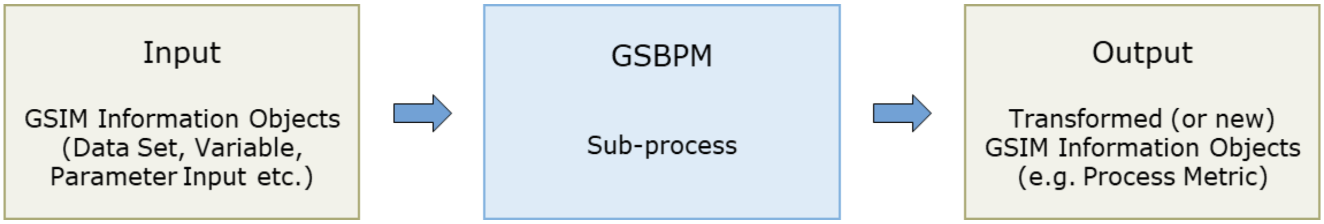

Figure 6. The GSIM and the GSBPM are complementary models for the production and management of statistical information. GSIM provides a set of standardized, consistently described information objects, which are the inputs and outputs in the design and production.

3.4 CSDA to Identify Capabilities That Are Needed to Manage Data

Common statistical data architecture, CSDA (UNECE, 2018) shows organisations how to organise and structure their processes and systems for efficient and effective management of data and metadata; starting from the external sources through the internal storage and processing up to the dissemination of the statistical end-products. The scope of the CSDA includes all GSBPM phases, specifically the designing, building and use of processes and systems in statistical data collection, production, analysis and dissemination, based on external needs, in statistical organisations. There are no restrictions to the types of data. In particular in order to, help organisations to modernise themselves, CSDA shows how to deal with the newer types of data sources such as Big Data, Scanner data, Web Scraping, etc.

CSDA is about data, and as there is no data without metadata, CSDA considers both integrally as information. Although CSDA is loosely based on TOGAF, it should be stressed that "data" to statistical organisations means something different from what is understood by most industries. "Data", to statistical organisations, is the raw material, the parts and components and the finished products, rather than the information needed to support and execute the organisation's primary processes. Although, also in statistical organisations, there is data that plays that role. Thus, CSDA is not what TOGAF calls a Data Architecture. According to TOGAF, a Data Architecture is part of the Application Architecture.

CSDA provides statistical organisations with a few key principles aiming to direct the way capabilities are identified and defined. The main content of the current version of the CSDA is the description of the Capabilities and Building Blocks. Capabilities are abilities, typically expressed in general and high-level terms that an organisation needs or possesses. Capabilities typically require a combination of organisation, people, processes, and technology. Building Blocks represent (potentially re-usable) components of business, IT, or architectural capability that can be combined with other Building Blocks to deliver architectures and solutions. CSDA is restricted to conceptual and logical descriptions of components. Non-statistical, business processes such as HR and Finance, are out of scope for the CSDA.

The capabilities are divided in core capabilities and cross-cutting capabilities. The Core Capabilities is where the actual work i.e., statistical production, happens. Besides the operational phases, the CSDA addresses cross cutting issues: Data Governance, Traceability, Quality and Security.

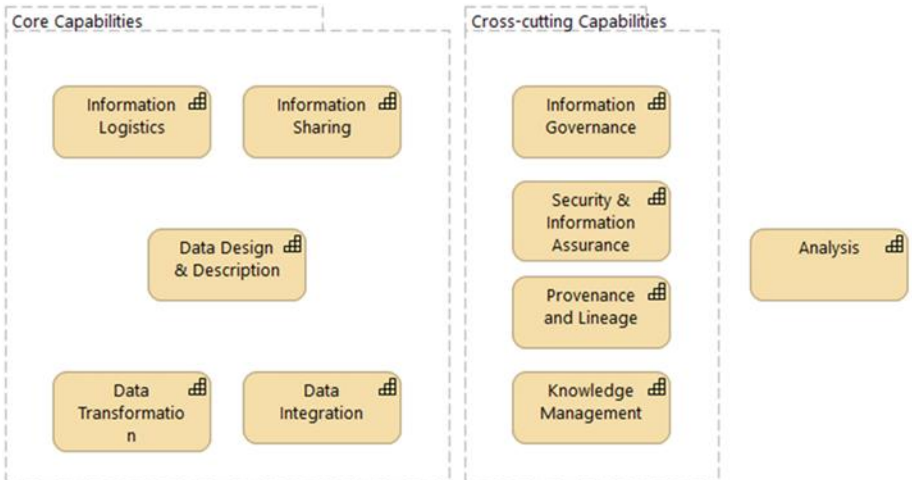

Figure 7. A glance at the CSDA model.

The CSDA key principles:

- Information is managed as an asset throughout its lifecycle: information includes both the data and the metadata describing that data, and all objects that describe the context, content, controls and structure of data and metadata.
- Information is accessible: information is discoverable and usable, and it is available to all unless there is good reason for withholding it.
- Data is described to enable reuse: data must have sufficient metadata so it can be understood outside its original context, and connections between data objects as well as restrictions to data usage must be documented.
- Information is captured and recorded at the point of creation/receipt: information should be captured and recorded at the earliest point in the business process to ensure it can be used by subsequent processes, and subsequent changes to information should be documented at the time of action.
- Use an authoritative source: within a business process, there should be an authoritative source from which information should be sourced and updated, and the existing information should be reused instead of recreated or duplication.
- Use agreed models and standards: key information should be described using common, business-oriented, models and standards, agreed by the organisation.
- Information is secured appropriately: information should be classified according to its level of sensitivity, secured and managed accordingly.

These principles are compatible with FAIR data principles: Findable, Accessible, Interoperable, and Reusable.

3.5 CSPA to Develop, Share and Reuse Joint Components in Statistical Production

Common Statistical Production Architecture, CSPA (UNECE, 2021b) is an industry architecture for the official statistics industry. An industry architecture being defined as a set of agreed common principles and definitions designed to promote greater interoperability within and between different stakeholders that make up an industry. The CSPA Business Architecture defines the need to identify business functions and organise them into business processes operationalised by business services, also applied in the geospatial domain where web services are the bases (Ariza-López et al., 2021). CSPA aims to reduce the cost of developing and maintaining processes and systems.

CSPA builds on and uses GAMSO, GSBPM and GSIM, so as to improve the common understanding and alignment necessary for joint development, sharing and reuse of components. It gives users an understanding of the different statistical production elements that make up a statistical organization i.e., processes, information, applications, and services, and how those elements relate to each other. It also provides a common vocabulary to discuss implementations, with the aim to stress commonality. Furthermore, CSPA enables the vision and strategy of the statistical industry, by providing a clear, cohesive, and achievable picture of what is required to get there.

CSPA is separated into four "perspectives". These are business architecture, information architecture, application architecture, and technology architecture.

CSPA, a modernisation blueprint includes five steps:

1. standardisation of statistical pipelines, facilitated by common description (and then, common design) using GSBPM and GSIM frameworks
2. demise of specific and non-shareable technical implementations of statistical processes
3. coordinated development of statistical services using international standards (sdmx, ddi, etc.) and based on SOA
4. adoption of generic and shared services as described in CSPA framework.
5. re-building statistical pipelines as a plug-and-play assembly of shared or reused services.

One key aspect of CSPA is that it promotes solutions based on an architectural style called Service Oriented Architecture (SOA). This style focuses on services (statistical services in this case).

Design principles of CSPA:

- Re-use existing before designing new
- Design new for re-use and easy assembly
- Processes are metadata driven
- Adopt available standards

- Designs are output driven
- Enable discoverability and accessibility
- Statistical Services have service-level agreements

Information principles of CSPA:

- Manage information as an asset
- Manage the information lifecycle
- Protect information appropriately
- Use agreed models and standards
- Capture information as early as possible
- Describe to ensure reuse
- Ensure there is an authoritative source
- Preserve information input into Statistical Services
- Describe information by metadata

The CSPA catalogue is a proposal to facilitate the access to the set of shared services proposed by the Eurostat and other INS members. Within the catalogue, a CSPA service is split between the definition of the service, the specification and the implementation (Figure 8). A service can have different definitions. Each definition can have several specifications and each specification is related to zero or more implementations.

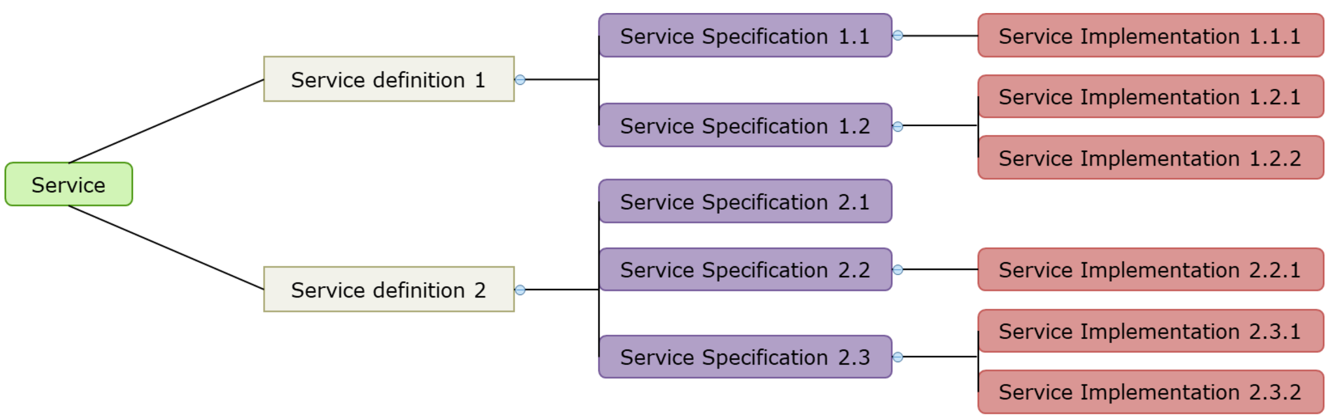

Figure 8. A CSPA service is split between the definition of the service, the specification and the implementation.

CSPA provides a practical link between the conceptual standards of GSBPM, GSIM, and statistical production. The shareable services of the CSPA catalogue use a common breakdown of supported statistical functionalities, through the use of GSBPM phases and GSIM concepts to describe the way the service can be used i.e., which GSBPM phase and what kind of input and output data are considered.

CSPA services respect the following characteristics:

- "Plug and play" architecture: replacing Statistical Services should be as easy as pulling out a component and plugging another one in.

- "Loose coupling": CSPA Services are independent, that is, they do not talk directly to each other. Organizations will need a technology solution to support communication between CSPA Services.
- "Granularity and embedding": An atomic or fine-grained statistical service encapsulates a small piece of functionality whereas, a coarse-grained statistical service will encapsulate a larger piece of functionality, for example, a whole GSBPM sub process.

3.6 IGIF to Develop and Strengthen National Geospatial Information Management

The Integrated Geospatial Information Framework, IGIF (UN-GGIM, 2020) is a strategic guide to enable coordinating, developing, strengthening and promoting effective sharing and use of geospatial information for policy formulation, decision-making and innovation. IGIF has been developed in collaboration between the United Nations and the World Bank.

Geospatial information provides an integrative platform for all digital data that has a location dimension to it. IGIF is focused on concepts, methods, standards and guides to address global goals as well as the national needs of all countries. IGIF aims to translate all these concepts to more practical implementation guidance. It does this by leveraging seven underpinning principles, providing eight goals and nine strategic pathways as a mean for governments to establish more effective geospatial information management practices and policies. Structurally the IGIF comprises a strategic-level overarching guide, a practical-level implementation guide focused on the nine strategic pathways (Figure 9), and a country-level action plan.

Figure 9. The structure of IGIF.

NSDIs have historically focused on the collection of data and the implementation of technologies. In contrast, the IGIF additionally focuses on the governance, policy, financial, capacity and engagement processes necessary to collect, maintain, integrate and share geospatial information, through all levels of government and society, in a modern and enabling technology environment.

3.7 INSPIRE to Create a Common Spatial Data Infrastructure

Infrastructure for spatial information in Europe, the INSPIRE directive (EU, 2007), creates an EU spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. INSPIRE aims to enable the sharing of environmental spatial information among public sector organisations, to facilitate public access to spatial information across Europe and to assist in policy-making across boundaries. INSPIRE defines a consistent set of web services dealing with interoperability actions, including data download and metadata exchange.

It is based on five common principles, establishes a set of common Implementing Rules which guarantee compatibility and interoperability in a community and transboundary context, and addresses 34 Spatial Data Themes needed for environmental applications (Figure 10).

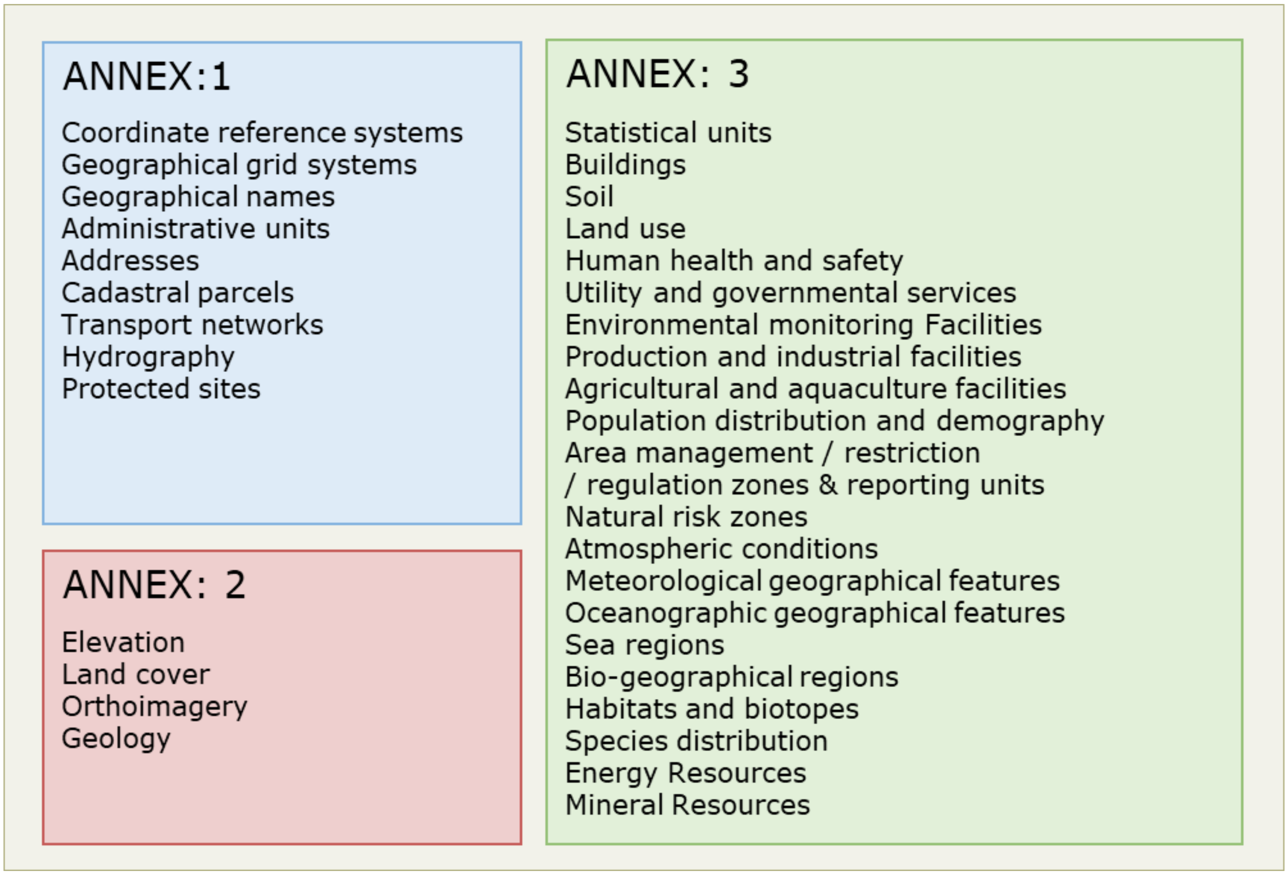

Figure 10. 34 spatial data themes of INSPIRE.

INSPIRE has set out the framework for sharing interoperable geospatial datasets and services and for deployment of a network of web services to document, view and download the geospatial datasets. The main goal is full discovery, accessibility and sharing of the national geospatial datasets owned by the public sector. Member States are obliged to establish NSDIs, to be built on international and national standards. The interoperability of geospatial data and services within a NSDI relies on the internationally agreed ISO 19100 family and the respective OGC standards and processes supported with feasible standards by W3C, IETF and other international standardisation bodies.

INSPIRE is based on five common principles:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.

- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

Furthermore, common Implementing Rules (IR) were adopted in a number of specific areas in order to ensure that the spatial data infrastructures of the Member States are compatible and usable in a Community and transboundary context:

- Metadata
- Data Specifications
- Network Services
- Data and Service Sharing
- Spatial Data Services
- Monitoring and Reporting

3.8 OGC to Improve Access to Geospatial or Location Information

Open Geospatial Consortium, OGC (2021) is a worldwide community committed to improving access to geospatial, or location information. OGC creates free, publicly available geospatial standards that enable new technologies. OGC also manages an agile and collaborative research & development process - the OGC Innovation Program - that anticipates and solves real-world geospatial challenges experienced by OGC members. The mission of OGC is to make location information more Findable, Accessible, Interoperable, and Reusable (FAIR).

OGC Statistical Domain Working Group (Statistical DWG) identifies requirements and use cases of how geospatial and statistical standards can support the integration of geospatial information into the statistical system and for the purposes of broad discovery, analysis and use. The establishment of the Statistical DWG sets ground for the premise that the statistical domain is evolving to move from the traditional capture of data through surveys such as census, to a more dynamic integration of data from across a wide range of domains that can make statistics more timely, more accurate and better understood. It states that geospatial information is a key data source in transforming the way that statistics are produced as it has the potential to fit into every component of the Generic Statistical Business Process Model (GSBPM). The complexity of bringing together survey, administrative and big data with geospatial information and earth observation is driving an increased need for standardization to support better statistical outputs.

The main focus areas of Statistical DWG are

- Data integration: how can semantic web technology be used to improve the integration of geospatial and statistical datasets?
- Data Capture: how can geospatial data be used to include the capture and georeferencing of survey, administrative and big datasets used in the production of statistics?

- Data processing: as geospatial objects do not fit within traditional database structures these are currently handled externally to the rest of the statistical process through dedicated geospatial systems.
- Data exchange: given the lack of interoperability between statistical and geospatial data architectures, frameworks, metadata and standards, how can data be shared through a service orientated approach?
- Data analysis: a lack of comparability between geographic datasets at a national, regional, continental and global level makes international analysis difficult.
- Knowledge: there is a lack of understanding within the statistical community of the geospatial tools and standards that can support the statistical system, and a corresponding situation with the geospatial community.

A set of tasks are defined for the Statistical DWG which include:

- Reviewing the GSGF.
- Identifying existing OGC standards that could have a statistical application within the GSGF.
- Identifying cases studies of existing OGC work being used to integrate statistics and geography, such as the Discrete Global Grid Systems and the Table Joining Service.
- Reviewing the relationship between the Standards pathway of the IGIF and the Statistical DWG.

4. Connections to the GSGF

4.1 The Global Statistical Geospatial framework, GSGF

The Global Statistical Geospatial Framework, GSGF (UNSC and UN-GGIM, 2019) is a high-level framework which facilitates consistent production and integration approaches for geo-statistical information. It is generic and permits application of the framework principles to the local circumstances of individual countries. In addition to the principles GSGF features Inputs, Key Elements, and Outputs. Through all these components, the GSGF acts as a bridge between statistical and geospatial professional domains, between NSIs and NGIAs, and between statistical and geospatial standards, methods, workflows, and tools.

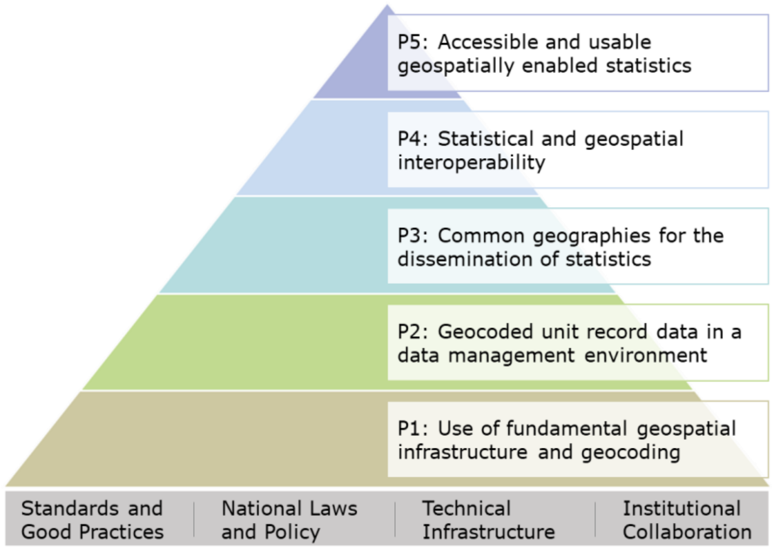

Figure 11. The five principles and four key elements of the GSGF.

4.2 Connections between the GSGF and Other Frameworks

Often, a framework can be viewed as a "cookbook" to make sure that necessary geospatial aspects have been identified, e.g. within all statistical business process steps (GSBPM) or information objects (GSIM). Generally, the frameworks also aim to provide a common vocabulary and a structure to support collaboration activities, particularly in the field of modernisation.

A rough division between the studied frameworks can be made; namely GSBPM, GAMSO, GSIM, CSDA, and CSPA have been developed and peer reviewed by the international statistical community, whereas IGIF, INSPIRE, and OGC have a geospatial focus. GeoGSBPM can be considered as an integration of these two fields.

However, the frameworks seem to fit well together. The frameworks are not developed isolated but must be seen in the context of a whole suite of standards and frameworks, e.g. CSDA links to GSBPM, GSIM, and CSPA. Where applicable, the CSDA also links to other international standards such as TOGAF, DDI, SDMX, etc. Different frameworks support the work from different aspects, and they complete each other, but should also be used differently.

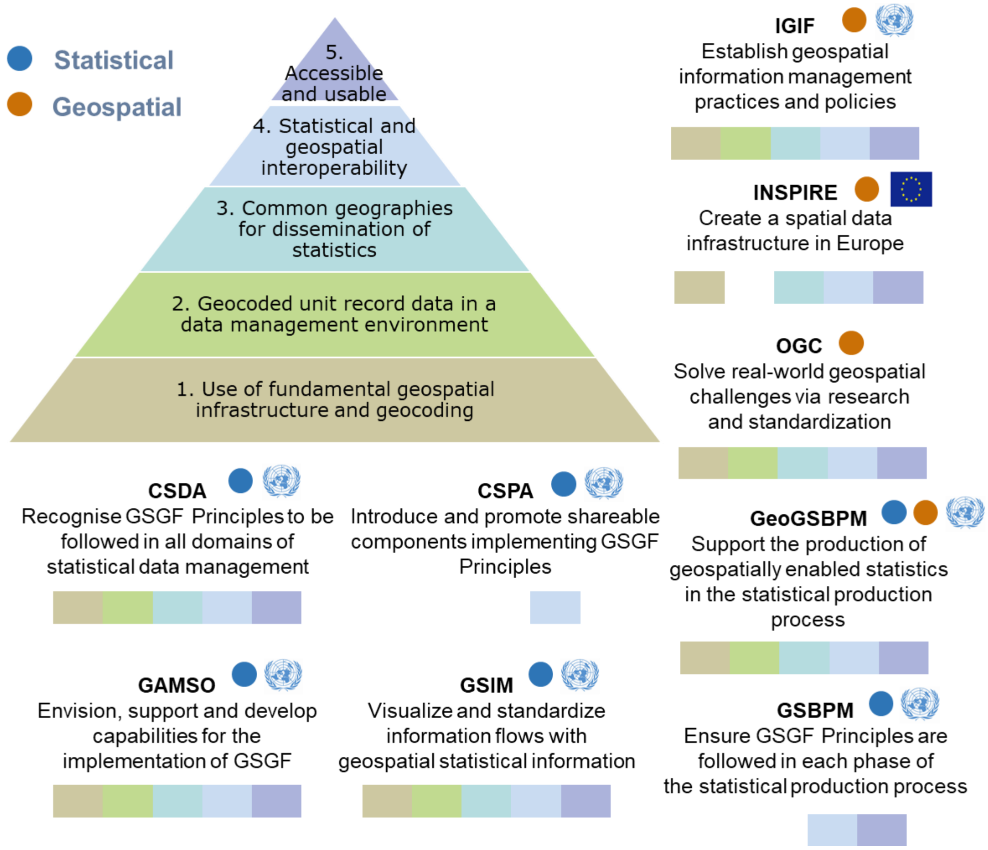

Figure 12. The role of each framework in statistical and/or geospatial fields and how they are linked to the GSGF principles.

4.2.1 GSBPM and GeoGSBPM

GSBPM together with the results of the HLG-MOS task team, GeoGSBPM, is a good tool to describe the geospatially enabled statistical business process. They provide a steady starting point to describe the interactions between the statistical and geospatial pipelines and thereby to build consistent flows between these two pipelines according to the GSGF principles.

GSBPM and GeoGSBPM have an immediate connection to the Principle 4. Especially, recommendations related to the requirement 4.1 "Improve geospatial workflows within statistical production are important". Furthermore, recommendations that are in the meeting point of the statistical and geospatial production processes are to be noted.

Furthermore, GSBPM enables an identification of the sub-process phases for which the quality issues are important, in order to produce accessible and usable geospatially enabled statistics (Principle 5).

4.2.2 GAMSO

There is identified an overlap between activities in GSGF and GAMSO. In Europe, the implementation of the GSGF is much more about policies, responsibilities, resources, communication, collaboration and commitment to common vision and strategic goals than for instance, availability of IT systems and technology. Thus, changing a system like statistics is a slow process with many people involved. GAMSO focuses on widescale assets that make implementation of GSGF possible on the whole. For instance, studying GAMSO shows that there is a need for support from management level and high-level strategic activities. A lot of cooperation between different pipelines is required inside and between organisations too. Thus, paying attention to all three activity areas identified in GAMSO can support the work on the way to the goals defined by GSGF. The fourth activity area in GAMSO is Production. The Production activities in GAMSO are identical with those in GSBPM. GAMSO is by construction fully consistent with GSBPM.

Furthermore, studying GSGF from the point of view of the GAMSO, capability development supports implementation of GSGF by providing a common approach to help to compare organisation's own capabilities to the required ones and identify and plan capability improvements. In addition to development of capability elements there, is a challenge of maintenance too. Thus, in the implementation of GSGF, GAMSO looks at long term goals, transformation and also cultural questions: How to get organisations more used to using geospatial data.

4.2.3 GSIM

In GSIM, many information objects are directly relevant for GSGF principles and recommendations. GSIM could support visualisation and standardisation of information flows with geospatial information. This can facilitate collaboration and reuse and sharing of good practices, methods, components and processes and thereby, lead to greater efficiency.

GSIM Information objects could be used to tell a story relevant to the geospatial community (Figure 13).

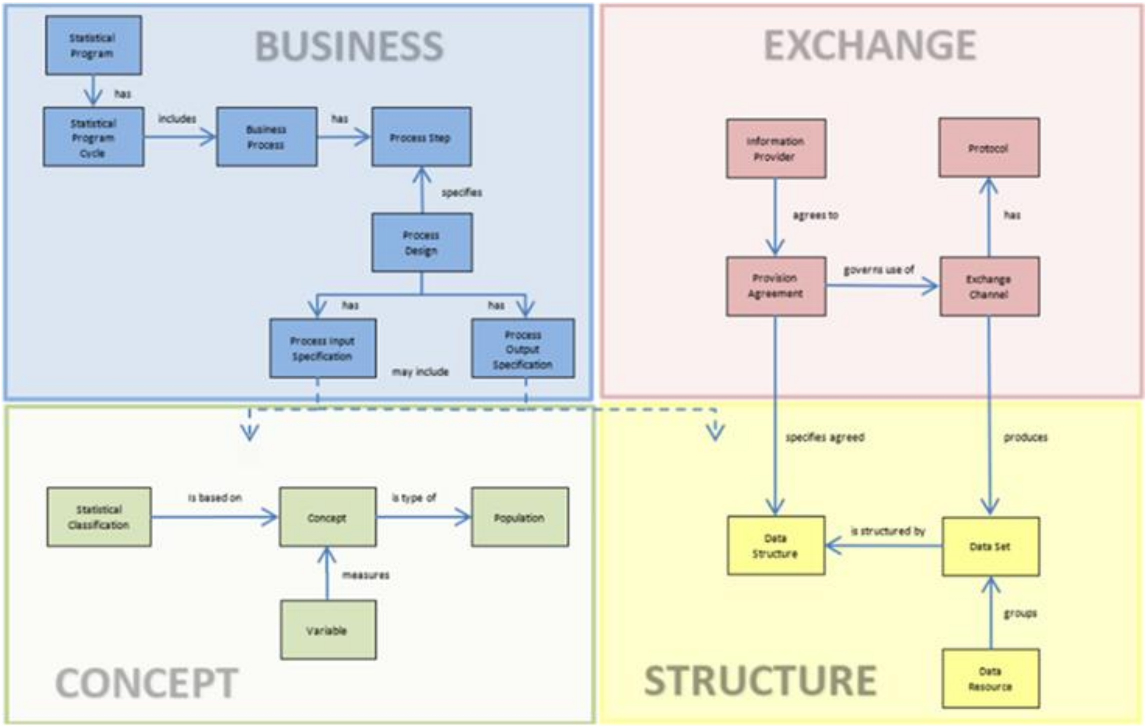

Figure 13. GSIM information objects tell a story, an example.

Clarifying concepts in GSGF could also benefit from GSIM (Principle 5) and statistical and geospatial interoperability could be enabled by using GSIM (Principle 4). Furthermore, identifying links between GSIM and GSGF, can produce a combined set of metadata. This could support a standard way to collect metadata giving support to implementation. statDCAT-API, geoDCAT-API can be relevant for catalogues of statistical and geospatial information to share in production.

In GSIM, the Identifiable Artefact and Administrative Details in particular support implementation. Identifiable Artefact is an abstract class that comprises the basic attributes and associations needed for identification, naming and other documentation.

4.2.4 CSDA

CSDA provides key principles which overlap and support those of the GSGF. Similar to process phases in GSBPM, CSDA can be used as a tool to recognise GSGF principles to be followed in each domain/capability of statistical data management.

The capability model of CSDA can help to design business and application layers. Capabilities represent stable, self-contained business functions, decomposed to the level at which they are useful. This breakdown will help to describe in detail the business layer of the reference architecture, and to identify the application services required, with a view to harmonizing processes at international level and sharing tools. Now, this is continued by formalizing in an Excel file the links between the elements of the three frameworks (GEOSTAT 4, 2021):

- CSDA Key Principles with the GSGF principles (already partially made, requires further discussion and validation) – this could translate a high-level analysis: principles analysis of both frameworks
- CSDA Capabilities with GSBPM stages/processes

4.2.5 CSPA

In the context of the nine studied frameworks and GSGF, focus is on the application architecture layer within CSPA. CSPA has an immediate connection to the Principle 4. CSPA gives, for statistical pipelines, a generic way to describe services implementation, based on interoperability and standardisation, like what is done by INSPIRE on the geospatial side. Both CSPA and INSPIRE are supporting standardized architecture. It is notable that according to GSGF, there are four different layers of interoperability: legal, organisational, semantic and technical. To some extent in statistical side, legal and organisational sides are covered by GAMSO (capabilities) and GSBPM (processes), while semantic and technical sides are partly covered by GSIM and CSPA.

In addition to Principle 4, three complementary GSGF principles can be supported by CSPA:

- Principle 1 is aligned with CSPA values, for avoiding data duplication and silos.
- Principle 2: CSPA facilitates the building and delivering of geo services in a managed and crosscutting collection of statistical pipelines
- Principle 5: Having CSPA compliant service-oriented dissemination is a way to ensure accessible dissemination.

The CSPA catalogue is a key tool to support implementation of GSGF principles on standardization of statistical geospatial pipelines. The CSPA catalogue is a reference tool to promote shareable services using a common breakdown of supported statistical functionalities, whereas geospatial services are included in INSPIRE catalogue. The services could be divided into two different kinds of services:

- Data services – their main goal is to disseminate data without changing their content. These services could support implementation of the GSGF principles 1 and 5
- Functional services – their main goal is to deliver transformational functions so as to implement specific treatments and methodologies. These services could support more particular requirements of the GSGF principles

4.2.6 IGIF

There is still a lack of awareness and understanding of the vital and integrative cross-disciplinary role of geospatial information and related enabling architectures. Considerable promotional effort is needed also at the policy and decision-making levels. More extensive institutional collaboration, enhanced interoperability and technical integration across various national information systems and platforms is required. Geospatial information has typically, been collected, processed and distributed in organisational silos; resulting in data duplication, and use of non-standard solutions, proprietary formats and incoherent classifications. As a result, data harmonisation, maintenance and integration remain problematic.

Where GAMSO promotes the supporting actions within an organisation, IGIF is more useful at interorganisational level and helps to identify tasks that require a shared responsibility between organisations. Most of the IGIF strategic pathways (SPs) coincide, align with and complement GSGF key elements and can be considered potentially helpful for implementation of the GSGF. In particular, in SP 4 Data, GSGF is mentioned as the primary tool for integration issues.

GSGF mostly considers integration of fundamental, authoritative data sources from governmental data providers and custodians with clearly agreed roles and responsibilities defined by legislation. New emerging data domains, e.g. health and other non-authoritative data sources and users, e.g. communities and citizens, with no formally agreed roles and responsibilities pose a challenge for GSGF implementation. Here, the IGIF provides guidance on actions, methodology and tools to cover and capture the potential of non-authoritative data collection, management and delivery, in particular in SP 5 Innovation and SP 7 Partnerships.

The IGIF is primarily focused on providing guidance for developing countries with weak existing digital infrastructures and poorly developed governance and policy-related structures for geospatial data management. Hence, the strategic pathways comprise many elements and tools that are not directly applicable in the European context. This fact calls for delicate selection of potentially useful elements within each strategic pathway that could be utilised for GSGF Europe implementation.

As a result of a lightweight analysis of the interlinkages between IGIF and GSGF at a holistic level, the strategic pathway SP 4 Data is found to be an obvious elemental and integral component for GSGF implementation. On the other hand, SP 3 Financial, SP 8 Capacity and Education and SP 9 Communication and Engagement do not provide any straightforward new insights to GSGF implementation issues in the European context. The role and potential of the remaining strategic pathways of IGIF, i.e. SP 1 Governance and institutions, SP 2 Policy and Legal, SP 5 Innovations, SP 6 Standards and SP 7 Partnerships need to be explored and, together with SP 4 Data, reflected against GSGF Europe's requirements and recommendations.

Another perspective for further work would be to identify whether the GSGF Europe model is missing any obvious elements which could be complemented by IGIF. This work, however, was decided to be excluded from the scope of GEOSTAT 4 and can be considered to be useful follow-up work after GEOSTAT 4.

However, we took some issues stemming from our analysis related to Task 1.1.2.2 "Organisational, legal and other nontechnical aspects around the implementation of the GSGF" and discussed the implications in the non-technical context in more detail. In particular the role of innovation (SP 5) which is currently missing from GSGF Europe was explored further.

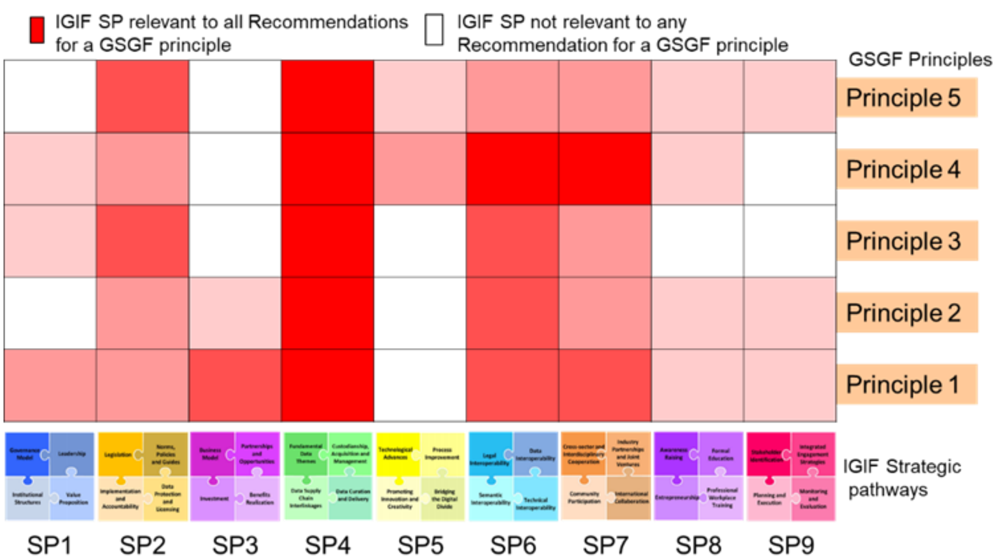

Figure 14. IGIF Strategic Pathways vs. the GSGF principles: Relevance of Strategic Pathways to the Recommendations for the principles provided in GEOSTAT 3 project report "GSGF Europe – Implementation guide for GSGF in Europe".

4.2.7 INSPIRE

INSPIRE is recognised as an important framework under the GSGF Europe, as a legal framework that implements a structure supported by NSDIs and an organisational schema for the whole EU. Thanks to INSPIRE, NSDIs have been put in place in EU Member States and it is rational to use data from these NSDIs as a source for geocoding of statistical information. INSPIRE constitutes the institutional pillar of the recommendations as they provide rational infrastructure and mechanisms (legal, technical, collaborative and financial) to deploy a harmonized implementation of the GSGF Europe, both for ESS Member States and non-members. INSPIRE is already recognised as an important framework in several recommendations under the GSGF principles 1, 3, 4 and 5:

- The spatial data themes of INSPIRE (Annexes 1 and 3) provide common standards to describe and share spatial data so that it can be found, shared and used more easily and is readily combined and integrated (Principle 1).

- Data standards of INSPIRE, e.g. enumeration areas and grids, support integration of data from different sources (Principle 3).
- In Principle 4, there is the greatest shortage of INSPIRE which requires only the latest/updated available data.
- Geospatial information is available through an INSPIRE on-line GeoPortal, e.g. WMS and WFS services and metadata (Principle 5).

Furthermore, INSPIRE provides training material and INSPIRE library. Thus, INSPIRE provides important support for interoperability, but also lacks some aspects, e.g. versioning and linkages of historic data and developing geographies.

4.2.8 OGC

For the GSGF, OGC is mainly a standards and technological framework being aware of the potential that can be achieved by the integration of geospatial and statistical information or, the other way around, by spatially enabled statistics.

The OGC framework supports the goals and recommendations of the GSGF by crossing all five GSGF principles. A more detailed analysis down to the recommendations level also confirms this, being relevant to many of them, identifiable directly or through other instances or frameworks that are also interrelated with the OGC framework, such as the Infrastructure for Spatial Information in the European Community (INSPIRE), the GSGF itself or the International Organization for Standardization (ISO).

Being transversal to the GSGF, it is possible to identify several of its recommendations for which components of the OGC framework are directly relevant, such as standard for geocoding API (1.2.4), geocoding standard (2.5.3), compliant view and download services aligned with INSPIRE (3.1.4), Table Joining Service (4.1.1 and 4.4.2), Statistical DWG (4.1.2), role on semantic interoperability issues, such as ontologies for addresses and buildings (4.2.1), use of compliant services and non-proprietary formats such as Geopackage for file deliveries (5.2.1), publication of simple features as also defined in ISO 19125 along with the complex data models defined by INSPIRE (5.2.5).

Given that OGC is mostly concerned with technological aspects and more specifically targeted to the standards domain, it is considered useful to broaden the scope of the OGC framework to deal with other relevant geospatial standardization issues, such as ISO, INSPIRE related components and industry driven dynamics, as well as components of statistical processes that are already spatially enabled.

Given that the INSPIRE framework is broader than OGC, more focused
on content and although its standards component envisages compliance with OGC and ISO, it is much more difficult to implement as, for instance, data models are often too complex and not all of them are easy to be used. On the other hand, some of the components dealt with by OGC, such as for instance geocoding, downloading and grid are already being considered by the statistical community in an autonomous way, providing compliance that is assured with established standards and practices within a fitness to purpose approach.

Although driven by different reasons and approaches, the OGC and ISO are closely related and in the long term must achieve some kind of convergence, being the first more suitable for the GSGF itself, as for Europe there are many more issues to consider. In what concerns pros and cons, it is worthwhile to highlight that while OGC implies a costly participation and a free access to standards, ISO is the other way around, based on free participation and costly access, which arises from their respective industry or institutional driven nature.

5. References

Ariza-López, F.J., Rodríguez-Pascual, A., Lopez-Pellicer, F.J., Vilches-Blázquez, L.M., Villar-Iglesias, A., Masó, J., Díaz-Díaz, E., Ureña-Cámara, M.A., González-Yanes, A., 2021. An Analysis of Existing Production Frameworks for Statistical and Geographical Information: Synergies, Gaps and Integration. Academic Editors: Kainz, W. and Wu, H. ISPRS Int. J. Geo-Inf. 2021, 10(6), 374 (https://doi.org/10.3390/ijgi10060374)

EU, 2007. Infrastructure for spatial information in Europe. INSPIRE. EU Directive (https://inspire.ec.europa.eu/)

GEOSTAT 4, 2021. GSGF and Frameworks Environment: the CSDA, GSGF, GSBPM matrix. An Excel file. Eurostat ESSnet grant project GEOSTAT 4.

OGC, 2021. Open Geospatial Consortium. (https://www.ogc.org/)

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

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

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

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

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

UNECE, 2018. Common Statistical Data Architecture. CSDA. Version 2.0. (https://statswiki.unece.org/display/DA/CSDA+2.0)

UN-GGIM, 2020. Integrated Geospatial Information Framework. IGIF. (https://ggim.un.org/IGIF/overview/)

UNSC and UN-GGIM, 2019. The Global Statistical Geospatial Framework. GSGF. (https://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf)