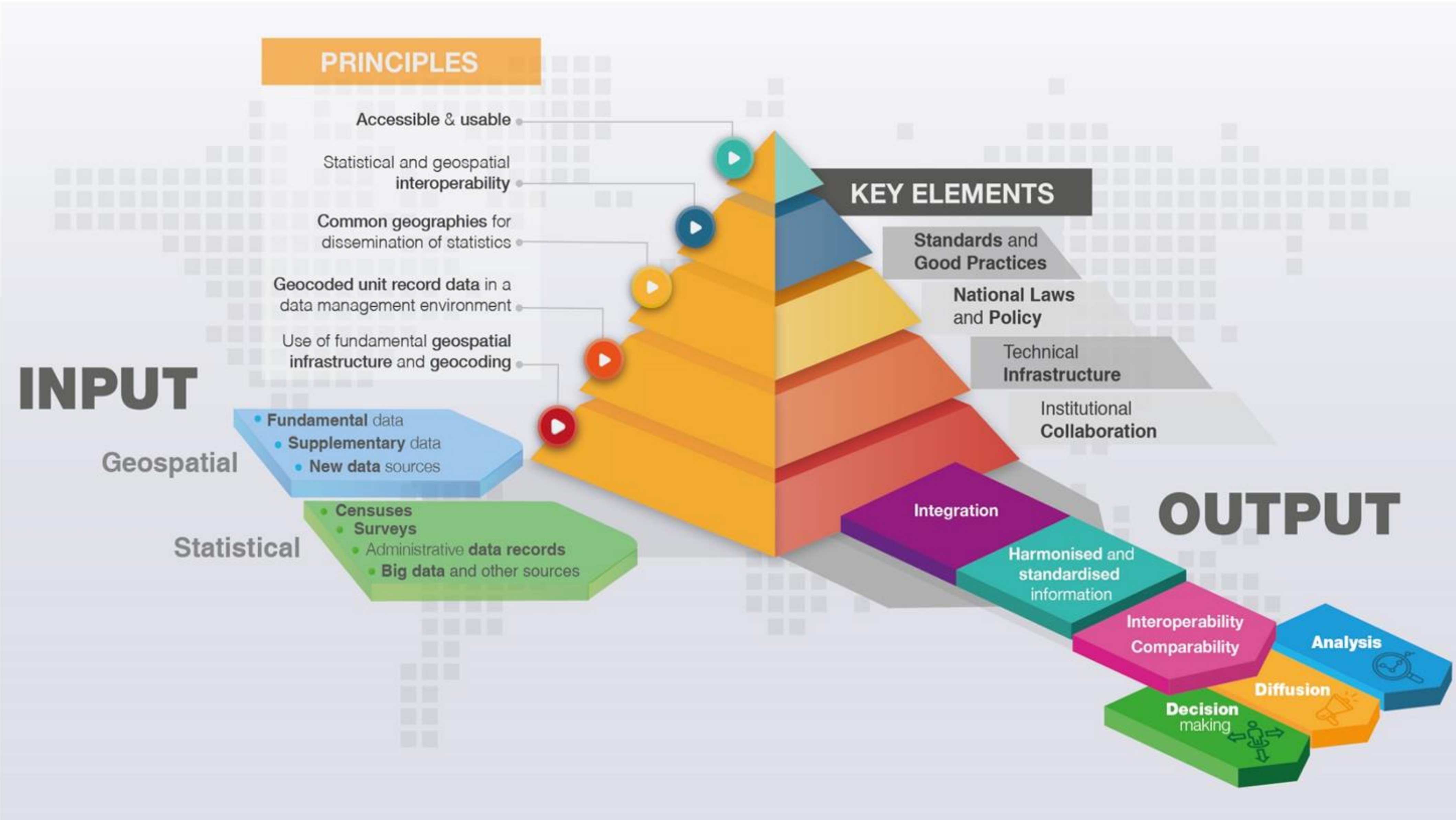
The Global Statistical Geospatial Framework


PRINCIPLES
Accessible & usable
Statistical and geospatial interoperability
Common geographies for dissemination of statistics
Geocoded unit record data in a data management environment
Use of fundamental geospatial infrastructure and geocoding
KEY ELEMENTS
Standards and Good Practices
National Laws and Policy
Technical Infrastructure
Institutional Collaboration
INPUT
Geospatial
Fundamental data
Supplementary data
New data sources
Statistical
Censuses
Surveys
Administrative data records
Big data and other sources
OUTPUT
Integration
Harmonised and standardised information
Interoperability Comparability
Analysis
Diffusion
Decision making

Department of Economic and Social Affairs
Statistics Division

Global Geospatial Information Management Secretariat

The Global Statistical Geospatial Framework

United Nations
New York, 2019

Department of Economic and Social Affairs

The Department of Economic and Social Affairs of the United Nations Secretariat is a vital interface between global policies in the economic, social and environmental spheres and national action. The Department works in three main interlinked areas: (i) it compiles, generates and analyses a wide range of economic, social and environmental data and information on which Member States of the United Nations draw to review common problems and to take stock of policy options; (ii) it facilitates the negotiations of Member States in many intergovernmental bodies on joint courses of action to address ongoing or emerging global challenges; and (iii) it advises interested Governments on the ways and means of translating policy frameworks developed in United Nations conferences and summits into programmes at the country level and, through technical assistance, helps build national capacities.

Statistics Division

The United Nations Statistics Division (UNSD) is committed to the advancement of the global statistical and geospatial systems. It compiles and disseminates global statistical information, develops standards and norms for statistical and geospatial activities, and supports countries' efforts to strengthen their national statistical and geospatial information systems. UNSD facilitates the coordination of international statistical and geospatial activities and supports the functioning of the United Nations Statistical Commission, the United Nations Committee of Experts on Global Geospatial Information Management, and the United Nations Group of Experts on Geographic Names as the apex entity of the global statistical and geospatial systems.

United Nations Committee of Experts on Global Geospatial Information Management

The United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM) is the apex intergovernmental body to discuss, enhance and coordinate global geospatial information management activities by involving Member States at the highest level, to work with Governments to make joint decisions and set directions on the use of geospatial information within national and global policy frameworks, and to develop effective strategies to build geospatial capacity in developing countries.

United Nations Statistical Commission

The United Nations Statistical Commission, established in 1947, is the highest body of the global statistical system. It brings together the Chief Statisticians from member states from around the world. It is the highest decision making body for international statistical activities especially the setting of statistical standards, the development of concepts and methods and their implementation at the national and international level. The Statistical Commission oversees the work of the United Nations Statistics Division (UNSD), and is a Functional Commission of the UN Economic and Social Council.

Notes

The designations used and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries. The term “country” as used in this publication also refers, as appropriate, to territories or areas. The designations “developed regions” and “developing regions” are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process.

[This page intentionally left blank]

Executive Summary

The Global Statistical Geospatial Framework (GSGF) facilitates the integration of statistical and geospatial information. A Framework for the world, the GSGF enables a range of data to be integrated from both statistical and geospatial communities and, through the application of its five Principles and supporting key elements, permits the production of harmonised and standardised geospatially enabled statistical data. The resulting data can then be integrated with statistical, geospatial, and other information to inform and facilitate data-driven and evidence-based decision making to support local, sub-national, national, regional, and global development priorities and agendas, such as the 2020 Round of Population and Housing Censuses and the 2030 Agenda for Sustainable Development.

> *"There is an urgent need for a mechanism, such as a global statistical-spatial framework, to facilitate consistent production and integration approaches for geo-statistical information."*
>
> *The Global Forum on the Integration of Statistical and Geospatial Information, New York 2014*

This paper presents information that will support countries and users to understand the GSGF — its value, application, infrastructure and implementation requirements. This is achieved through an overview of the GSGF, along with an elaboration of its five Principles and other key elements. Further information to support implementation is being compiled in the wiki of the Expert Group on the Integration of Statistical and Geospatial Information. The GSGF has been developed through a collaborative process, engaging statistical and geospatial information agencies globally.

> *"…. develop the Global Statistical Geospatial Framework as a common method for geospatially enabling statistical and administrative data to ensure that data from a range of sources can be integrated based on location and can be integrated with other geospatial information."*
>
> *United Nations Committee of Experts on Global Geospatial Information Management, New York 2015*

Fundamentally, the GSGF enables:

- Integration of data to support the measuring and monitoring of the targets and global indicator framework for the Sustainable Development Goals of the 2030 Agenda for Sustainable Development and the 2020 Round of Population and Housing Censuses;
- Comparisons at local, sub-national, national, regional, and global levels for decision-making processes within and between countries and thematic domains;
- Data sharing between institutions, through interoperability of geospatial and statistical information and the development of common tools and applications;
- Unlocking of new insights and data relationships that would not have been possible by analysing socio-economic, environmental, or geospatial data in isolation;
- Increased information on smaller geographical areas;
- Increased awareness of methods and tools to assess and manage disclosure risks and to enhance privacy in collection, storage, and dissemination of information;
- Conditions for investment and capability building in geospatial and statistical information;
- Integration of new sources of data to inform the production of high-quality geospatial information, for example Earth observations and other complementary data sources; and
- Strengthening of institutional collaboration between the geospatial and statistical communities.

List of Abbreviations

API – Application Programmable Interface
CSDA – Common Statistical Data Architecture
CSPA – Common Statistical Production Architecture
DDI – Data Documentation Initiative
DGGS – Discrete Global Grid System
EG-ISGI – (United Nations) Expert Group on the Integration of Statistical and Geospatial Information
ESS – European Statistical System
GAMSO – Generic Activity Model for Statistical Organizations
GFM – General Feature Model
GGRF – Global Geodetic Reference Frame
GSBPM – Generic Statistical Business Process Model
GSIM – Generic Statistical Information Model
GSGF – Global Statistical Geospatial Framework
HLG-MOS – High–Level Group for the Modernisation of Official Statistics
IHO – International Hydrographic Organisation
ISO – International Standards Organisation
MAUP – Modifiable Areal Unit Problem
MOU – Memorandum of Understanding
NGIA – National Geospatial Information Agency[1]
NSDI – National Spatial Data Infrastructure
NSO – National Statistical Organisation
NSS – National Statistical System
OGC – Open Geospatial Consortium
RDF – Resource Description Framework
SDGs – Sustainable Development Goals
SDO – Standards Developing Organisation
SEEA – System of Economic and Environmental Accounts
SDMX – Statistical Data and Metadata eXchange
UN-GGIM – The Committee of Experts on Global Geospatial Information Management
UNSC – United Nations Statistical Commission
UNSD – United Nations Statistics Division
UNECE – United Nations Economic Commission for Europe
W3C – World Wide Web Consortium
WMS – Web Mapping Services
WFS – Web Feature Services

[1] NGIA is used as an encompassing term to cover National Mapping, National Cartographic, National Geospatial Information Agencies and Authorities.

Contents

Executive Summary .... **vii**

List of Abbreviations .... **viii**

Introduction **1**

Part 1: The Global Statistical Geospatial Framework **6**

Principle 1: Use of fundamental geospatial infrastructure and geocoding 9

Principle 2: Geocoded unit record data in a data management environment 10

Principle 3: Common geographies for dissemination of statistics 11

Principle 4: Statistical and geospatial interoperability 12

Principle 5: Accessible and usable geospatially enabled statistics 13

Complementary Initiatives 14

Call to Action and the Way Forward 15

Part 2: Elaborating in detail the Global Statistical Geospatial Framework **16**

Principle 1: Use of fundamental geospatial infrastructure and geocoding 17

Principle 2: Geocoded unit record data in a data management environment 20

Principle 3: Common geographies for dissemination of statistics 24

Principle 4: Statistical and geospatial interoperability 27

Principle 5: Accessible and usable geospatially enabled statistics 31

Part 3: Annexes **34**

Annex A: Definition of Terms 35

Annex B: Standards, Quality and Enabling Frameworks 37

Annex C: Further Reading 46

List of Figures

Figure 1 Location as a link between society, the economy and the environment 1

Figure 2 The Global Statistical Geospatial Framework – inputs through to outputs 5

Figure 3 The five Principles of the GSGF 7

Figure 4 Administrative and Gridded Geographies 25

Figure 5 The European Interoperability Framework 28

Figure 6 The 14 Global Fundamental Geospatial Data Themes 42

Introduction

The 2030 Agenda for Sustainable Development, with its 17 Sustainable Development Goals (SDGs), 169 targets and global indicator framework, conveys a compelling need to understand, sustain, and develop a country's three pillars for development: Economy, Society, and Environment[2]. Positive outcomes in these three areas will be maximised by enabling data-driven, evidence-based decision-making within both the public and private sectors. The integration of information from both statistical (which includes a wide range of socioeconomic and enterprise data) and geospatial (which includes increasingly important environmental and earth observation data) domains has been recognised as important for understanding relationships between and within these three pillars. The integration of this information is vital to improve the quality of the evidence we use to make present and future decisions.

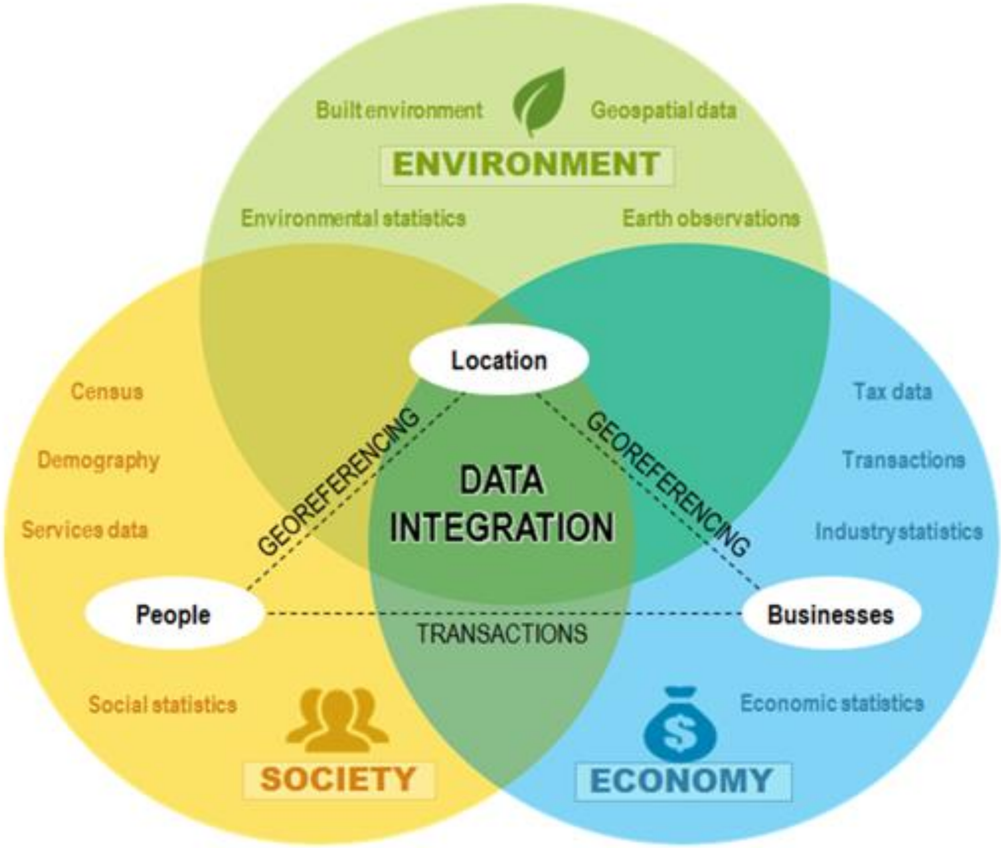


Figure 1 Location as a link between society, the economy and the environment

Put simply, **linking data about people and businesses to a place or geographic location, and its integration with geospatial information through the medium of location, can result in an improved understanding of social, economic, and environmental issues; much greater than is possible if viewing statistical or geospatial information in isolation** (Figure 1 above highlights this role).
Thus, the GSGF provides the underlying mechanism to achieve this integration, with the resulting output being geospatially enabled statistics which provide a more complete picture of our world. Considerable effort must be made to fully reap the benefits of integrated statistical and geospatial

[2] https://www.un.org/ecosoc/en/sustainable-development

information. Foremost, as a global community we must work together to increase awareness of the benefits to local, sub-national, national, regional, and global development that can be achieved through integrating statistical and geospatial information. When the management and use of geospatial information is a routine operation and an integral part of the production of statistics, a wider range of information and enabling technologies will become accessible and available for use. Efforts to improve the integration of statistical and geospatial information occur in an environment where NSOs are also seeking to modernise their statistical production systems and processes, to transform their operations, reduce complexities, and to derive new relevant metrics and indicators for statistical purposes. Critically, this includes the introduction of interoperable, standards-based, metadata-driven infrastructure and processes.

Geospatially enabled statistics will serve to strengthen the analysis of data to support informed, data-driven, evidence-based decision-making. This has several significant benefits, including the ability to create and analyse local geographies at a finer scale and to take advantage of new data sources. These 'new' data sources can include the use of Earth observations to generate statistics, which have great potential to provide data where traditional statistical and/or geospatial information methods are proving inefficient, not timely enough or do not exist due to the current national statistical and/or geospatial information capacity. Further pressure to examine these alternate data sources comes from the demand placed on both the statistical and geospatial communities to deliver information and insights for national development priorities and the 17 SDGs.

For NGIAs, this work occurs at a time when there is a collective effort that will enhance the management and use of geospatial information. This includes efforts to specify the fundamental or core geospatial datasets that are required to support geospatial activities within national and international efforts; for example, to support the implementation of the SDGs. These fundamental datasets are part of a country's geospatial infrastructure, which supports the geocoding of statistics and includes administrative and statistical geographies that facilitate the integration of statistical information with geospatial information.

The GSGF is situated in a rapidly developing ecosystem that is strengthening the global statistical and geospatial communities. Key elements of this ecosystem include:

- The Integrated Geospatial Information Framework, which is a basis and guide for developing, integrating, and strengthening geospatial information management;
- The High–Level Group for the Modernisation of Official Statistics, which is facilitating the development of statistical models and architectures; and,
- The Inter-agency and Expert Group on the Sustainable Development Goal Indicators Working Group on Geospatial Information, which is providing expertise and guidance on the critical role of geospatial information in supporting the attainment of the SDGs.

The usefulness and potential of the GSGF has already been demonstrated through its adoption by several countries, each at different starting points in their statistical and geospatial infrastructure development (e.g. Australia, Egypt, Mexico, New Zealand, South Africa, and Sweden among others)[3]. Furthermore, its value has been recognised and adoption of the GSGF is already being promoted at

[3] http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Global-Statistical-Geospatial-Framework-July-2018.pdf

the regional level (e.g. Africa, Europe, and Latin America). Further information on the implementation process can be obtained from documents and information published on the EG-ISGI wiki[4] on good practice approaches and implementation methodologies, as well as country case studies.

How to use this Paper

This paper specifies the current, recommended high-level implementation of the GSGF. This is achieved through three parts:

Part 1 provides the GSGF, including its data inputs, five Principles, key elements, and outputs. This is a high-level summary and intended to provide distilled information on the importance of geospatially enabled statistical information and how to achieve its integration;

Part 2 provides a detailed elaboration on the five Principles, providing further information and background to each of the Principles; and,

Part 3 consists of a series of Annexes:

- Annex A details definitions of common terminology used throughout this paper;
- Annex B provides background information regarding standards and data quality; and,
- Annex C lists further reading.

The EG-ISGI will progressively release material to further support implementation. This material will support countries in implementing the GSGF and will share details on lessons learned from country and regional level implementations, along with good practices, standards developments and pathways of implementation. This supporting material will be made available primarily via the EG-ISGI wiki, which offers a collaboration space for the EG-ISGI to publish information and update the GSGF.

[4] https://unstats.un.org/wiki/display/ISGI/United+Nations+Expert+Group+on+the+Integration+of+Statistical+and+Geospatial+Information

[This page intentionally left blank]

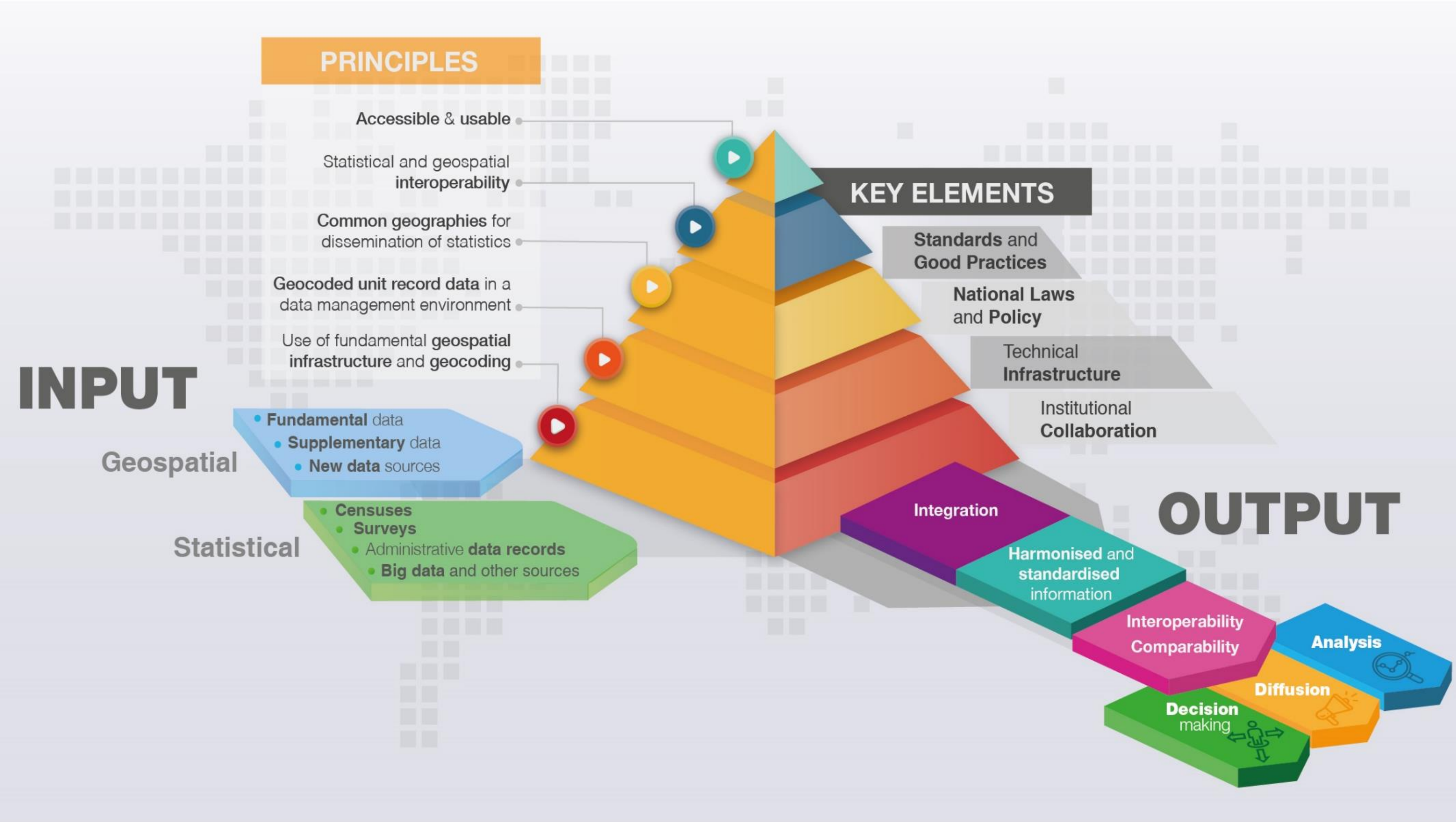


Figure 2 The Global Statistical Geospatial Framework – inputs through to outputs

Part 1: The Global Statistical Geospatial Framework

A holistic view of the GSGF is illustrated in Figure 2. It features Inputs, Principles, Key Elements, and Outputs. Through these components the GSGF acts as a bridge between statistical and geospatial professional domains, between NSOs and NGIAs, and between statistical and geospatial standards, methods, workflows, and tools.

GSGF Input Data

The GSGF begins with critical inputs of Fundamental Geospatial Data, and supplements these with other geospatial data sources as necessary, which can be provided by NGIAs and the broader geospatial community[5]. It primarily serves to geospatially enable traditional and authoritative statistical data, and increasingly data from administrative and other sources, much of which comes from NSOs and administrative data custodians within the broader NSS. Each of these inputs could be augmented by complementary datasets, such as those derived from crowdsourcing, new technology or sources (e.g. Big Data).

The five Principles of the Global Statistical Geospatial Framework

The five Principles of the GSGF outline the broad processes by which a range of geospatial and statistical infrastructures and processes are applied to input data to enable integration. Firstly, the statistical data are geospatially-enabled to the finest level possible. Then, geospatial tools and methods, such as common geographies and common standards of good practice, are used to ensure the data are interoperable, accessible, and usable. The five Principles are:

1. Use of fundamental geospatial infrastructure and geocoding;
2. Geocoded unit record data in a data management environment;
3. Common geographies for the dissemination of statistics;
4. Statistical and geospatial interoperability; and,
5. Accessible and usable geospatially enabled statistics.

Figure 3 visually depicts the hierarchical structure of each of the five Principles.

[5] Fundamental Geospatial Data has been defined by UN-GGIM. Annex B discusses the Global Fundamental Geospatial Data Themes this in more detail.

Figure 3 The five Principles of the GSGF

These Principles guide implementing countries in establishing and strengthening their processes for geospatially enabling statistical and administrative data and allow for the identification of capacity gaps within their respective national statistical and geospatial infrastructure.

The GSGF – Four Key Elements

Four key elements, which cut across the five GSGF Principles, play a critical enabling role that allow data to be obtained from the various sources and the GSGF Principles to be applied. These elements comprise:

- Standards and Good Practice
  - Apply across the geospatial and statistical communities and extend into information technology and other domains; and,
  - Include formal statistical standards (such as classifications, concepts, definitions), other standards (such as ISO standards), good practices/guidelines (such as OGC, ISO, and W3C Best Practices), and other *de facto* standards widely used within the geospatial community (such as GeoJSON).
- National Laws and Policies

    - Are key pieces of legislative, professional and social infrastructure that enable and, in some cases, constrain activities; and,
    - Can include international and national data protection, privacy, and confidentiality legislation, ethics and social licence requirements, open data policies and data access agreements.
- Technical Infrastructure
    - Includes a broad range of country- and regional- technical capabilities – spanning people skills, established and agreed methodologies and processes, and systems infrastructure; and,
    - Can include data management systems, computing and communications infrastructure.
- Institutional collaboration
    - Requires a commitment to institutional collaboration across key stakeholders, particularly across statistical, geospatial and administrative agencies of government; and,
    - Can be supported by formal agreements and/or positive institutional collaborative relationships and should include outreach and education initiatives.

GSGF Outputs

As a result of the GSGF processes, output data has a higher degree of structural harmonisation and standardisation, as well as geospatial flexibility, as compared with outputs previously produced by ad-hoc approaches that have largely characterised geospatial enablement of data. The GSGF outputs will support access to interoperable geospatial and socio-economic data that can be efficiently used across disciplines in support of country- and global-level priorities.

These enhanced outputs have an inherently greater capacity for integration based on location and have a substantially greater capacity to be further used in more complex statistical data integration processes. A significant feature of outputs supported by the GSGF is that they allow for higher confidence in creating reproducible results across people and time. Ultimately, these outputs support more efficient and powerful analysis and applications that support informed decision-making processes.

It is also important to stress that the statistical and geospatial data used to create geospatially enabled statistics should only be publicly released in a manner which complies with privacy and confidentiality legalisation, and prevailing community expectations. The ability to identify private and confidential information may be possible within data when collected and securely stored but should be confidentialised to an appropriate degree when disseminated.

Principle 1: Use of fundamental geospatial infrastructure and geocoding

Principle 1 specifies the adoption of a common and consistent approach to place each statistical unit of a dataset in time and space, using fundamental geospatial infrastructure.

The goal of Principle 1 is to obtain high quality, standardised location references (such as physical addresses, property or building identifiers, or other location descriptions), in order to assign accurate coordinates, and/or a small geographic area or standard grid reference, to each statistical unit at the microdata/unit record level. In addition, time and date-stamping these locations clearly places the unit both in time and in space, due to the strong statistical requirement for establishing data in a time series. Preferably, location is recorded through direct or indirect capture of x- and y-coordinates. Where this level of precision is not possible using current geospatial and statistical infrastructure within a country, adaptations using more general location descriptions and/or larger geographies will be necessary.

The process of obtaining locations and geocodes should use relevant, fundamental geospatial data[6], and support system capabilities from a NSDI or other nationally agreed sources. Ideally, a country should have a geo-referenced address, building register, land parcel, and/or place name as part of its statistical-geospatial infrastructure. If such a register is not available, countries are urged to test the implementation of alternative point-based referencing for unit record data[7].

Implementation of Principle 1 achieves the following objectives:

- Address, property, building, and location information are accurate and consistent, meeting country-level agreed standards and good practices;
- Geocoding results are as accurate and consistent as possible using common approaches or systems; and,
- Any geocoding issues are consistently managed through application of standardised approaches.

1

Principle 1 focuses on creating infrastructure that enables the implementation and socialisation of the GSGF.

This infrastructure supports the creation of high-quality, standardised location references such as a physical address, property or building identifiers, or other location description, and ensures the accurate assignment of coordinates and standard grid references. The addition of a time and date stamp links the time component to the unit.

[6] As described by the Global Fundamental Geospatial Data Themes detailed within Annex B
[7] See Geocoding definition within the annex

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

Principle 2 supports the process of linking or storing high-precision geographic references (i.e. geocodes – coordinates, small geographic area codes, or linked-data identifiers) to each microdata/statistical unit record. This is often referred to as geospatially enabling data, and must occur within a secure, standards-based data management environment. This process applies the address coding infrastructure and fundamental data from Principle 1.

The goal of Principle 2 is to allow all statistical unit records to be linked to a location, wherever it is possible to do so. This will then enable integration of data from a wide variety of sources, such as other socio-economic statistical data, administrative data, and geospatial information about the built and natural environment. The incorporation of these data, using geospatial processing, can then deliver new, geospatially enabled statistical variables for analysis. Principle 2 also enables flexible application of any geographic content when preparing data for release and analysis. This includes supporting future aggregation of statistical data into new geographical units or adapting to changes to existing geographies over time.

Principle 2 includes the use of data management tools, techniques, standards and good practices to facilitate the linking and management of geocodes within statistical datasets. This also serves to ensure that privacy and confidentiality requirements are correctly managed for the released data.

Implementation of Principle 2 achieves the following objectives:

- All statistical microdata is geospatially enabled for flexible use in analysis, visualisation, dissemination and statistical data integration processes;
- Aggregation of data for larger geographies is simplified through storage of a unique identifier or code for a small area geography or standard grid cell for each unit record;
- Adaptation to changes to existing geographies or to allow compilation of data for new geographies is enabled;
- Data can be effectively managed, including the protection of privacy and confidentiality;
- Clear data maintenance and custodianship roles are defined; and,
- Geocoded information and metadata are consistent, interpretable and systematically maintained.

2

Principle 2 supports the linking of each statistical unit record to a geographic reference (e.g. a coordinate or small geographic area) that will allow for statistics to be applied to any geographic context.

This will support integration or linkage of data from other data sources and mitigate challenges that arise with new geographies or changes in existing geographies.

Principle 3: Common geographies for dissemination of statistics

Principle 3 applies geography as a tool for integrating data. It uses a common and agreed set of geographies for the display, storage, reporting, and analysis of social, economic and environmental comparisons across statistical datasets from different sources. Principle 3 establishes the fundamental importance of balancing existing statistical and administrative geographies with other geographic referencing systems, such as grids, as a basis for establishing common geographies across datasets.

The goal of Principle 3 is to support the provision of a common set of geographies that ensure the consistent geospatial aggregation and dissemination of statistical data, irrespective of whether they are in gridded or administrative boundaries. Data is uniformly allocated to smaller administrative segments or statistical units (such as mesh blocks) that are divided according to political, property or topological subdivisions, or consistently assigned to differently sized grid units (i.e. squares or pixels). Furthermore, Principle 3 also allows for the translation and mapping of statistical information between gridded and administrative boundaries.

Implementation of Principle 3 achieves the following objectives:

- Data from different sources can be integrated using a common geography;
- The visualisation, analysis and interpretation of statistical and geographic information is simplified;
- Metadata supports data aggregation, integration and use;
- Identification and application of aggregation and disaggregation methods will enhance data quality and the assessment, consistency and increased use of data; and,
- The conversion of data between geographies is supported, through standard conversion mechanisms (e.g. through correspondences[8]).

3

Principle 3 determines definitions of geographic regions and aggregation/ disaggregation of data to regions and enables the consistency and comparability of integrated statistical and geospatial data.

A common set of geographies ensures that statistical data is geospatially enabled in a consistent manner and is capable of being integrated at the aggregate level; and also ensures that users can discover, access, integrate, analyse, and visualise statistical information seamlessly into geographies of interest.

[8] For more information on correspondence methods see: http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Correspondences

Principle 4: Statistical and geospatial interoperability

Data, Standards, Processes, and Organisations

Principle 4 defines the preconditions for statistical and geospatial data to work as a data ecosystem, in which those involved interact with each other to exchange, produce and consume data. Interoperability between statistical and geospatial data and metadata standards, is needed to overcome structural, semantic and syntactic barriers between data and metadata from different communities and providers.

Furthermore, it is necessary to enhance the efficiency of discovery, access, and use of geospatially enabled data. Often, full interoperability of data first requires the removal of obstacles in country-level laws, policies, and organisations that hamper cooperation between stakeholders and create barriers between producers and end users.

Enhancing interoperability allows both the statistical and geospatial communities to continue to operate their own general data models, metadata capabilities and architectures, while efficiently and seamlessly accessing, integrating, and linking datasets across different systems and applications. Therefore, Principle 4 urges the use of internationally adopted standards and good practices from both communities to enable greater interoperability of statistical and geospatial data, standards, processes and organisations.

Implementation of Principle 4 achieves the following objectives:

- Greater efficiency and simplification in the creation, discovery, integration and use of geospatially enabled statistics and geospatial data;
- Ensures service-based or machine-readable access mechanisms (e.g. through APIs) are implemented to provide greater efficiency of access and use, and to allow adaptation and evolution of uses through time; and,
- Increases the potential application of a larger range of data and technologies.

4

Principle 4 enables greater standardisation and interoperability of data which will lead to improved efficiency and simplification in the creation, discovery, integration, and use of geospatially enabled statistics and geospatial data.

This will increase the potential application of a larger range of data and technologies, and thereby enable a wider range of information to be available and accessible for use in decision making. It will also facilitate better cooperation between all stakeholders producing and using statistical and geospatial information.

Principle 5: Accessible and usable geospatially enabled statistics

Principle 5 highlights the need for data custodians to make geospatially enabled statistics accessible and usable according to agreed standards and good practices, so that data users can discover, access, integrate, analyse and visualise this information seamlessly for geographies of interest. It addresses the need to identify or, where required develop, policies, standards, good practices, and technologies that support these uses.

NSOs and NGIAs need to be aware of a wide range of legislative and operational issues when releasing and analysing information about people and businesses. As such, one important aspect of this Principle is to ensure that data can be accessed using safe mechanisms, which protect privacy and confidentiality, while also enabling the analysis of data to support data-driven, evidence-based decision-making. Other issues of relevance include: data quality in its different dimensions (particularly regarding reliability, timeliness, and relevance) and access to analysis, dissemination and visualisation capabilities.

The goal of Principle 5 is to support the release of geospatially enabled statistical information in a usable and accessible form. It specifically promotes the use of standard web services and linked data methods to provide dynamic, machine-readable access to these data with the necessary assurances regarding the integrity of the data.

Implementation of Principle 5 achieves the following objectives:

- Data custodians can release data, with data privacy and confidentiality protected;
- Data users can discover and access geospatially enabled statistics;
- Data users can undertake analysis and visualisation;
- Web services and linked data methods enable machine-to-machine access, as well as dynamic linkage of information; and,
- Data users can know the status of modifications / changes in the data provided by data custodians through data integrity checks.

5

Principle 5 supports data custodians to release data with confidence, improve the discovery and access of geospatially enabled statistics (particularly through promoting web services to provide machine readable and dynamic linkage to data), and to support analysis and evaluation of data in decision making.

Complementary Initiatives

The GSGF is situated in a rapidly developing information ecosystem that is strengthening the global statistical and geospatial communities. A range of frameworks and associated information resources that complement and interact with the GSGF include:

- The Integrated Geospatial Information Framework[9], adopted by the UN Committee of Experts on Global Geospatial Information Management (UN-GGIM) by resolution 8/113 in 2018, promotes, supports, and provides the leadership, coordination and standards necessary to deliver integrated geospatial information that can be leveraged to find sustainable solutions for social, economic and environmental development. Composed of 7 underpinning principles, 8 goals, and 9 strategic pathways, the Integrated Geospatial Information Framework provides a basis, a reference and a mechanism for Member States to take positive steps at the local, sub-national, national, regional, and global levels to address challenges faced in the collection, processing, analysis and dissemination of reliable, timely, accessible and consistent geospatial data and associated information.
- The High–Level Group for the Modernisation of Official Statistics is facilitating the development of modernised statistical models and architectures. This group supports the development of the Common Statistical Production Architecture[10], which offers a practical link between the more specific conceptual models of GSIM, GSBPM, and with practical production of statistical data; and,
- The Inter-agency and Expert Group on the Sustainable Development Goal Indicators Working Group on Geospatial Information is providing expertise and guidance on the critical role of geospatial information in supporting the attainment of the SDGs, including on topics such as modelled disaggregation of data and the potential use of earth observations.

These frameworks and information resources will serve to strengthen and compliment the implementation of the GSGF. This ultimately enables countries to help bridge the statistical and geospatial communities, foster and support more efficient production of geospatially enabled statistical information, improve the comparison of data across geographies and countries, and ultimately enable more integrated information to be available for analysis and evidence-based decision-making. Additionally, the GSGF can also act as an enabler for these frameworks within the broader information ecosystem of a country.

[9] http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Part%201-IGIF-Overarching-Strategic-Framework-24July2018.pdf
[10] https://statswiki.unece.org/display/CSPA/Common+Statistical+Production+Architecture

Call to Action and the Way Forward

The GSGF has been developed to support the underpinning data needs of the 2020 Round of Population and Housing Censuses and the 2030 Sustainable Development Agenda. The integration of statistical and geospatial information is crucial to unlocking the potential of data to make informed and evidence-based decisions at all levels within a country, as well as informing regional and global action. The resulting geospatially enabled statistics and integrated data will support development plans and provide valuable information on the underpinning pillars of a country's society, economy, and environment.

As an important first step for countries to consider in making progress towards producing geospatially enabled statistics, the GSGF requires the availability of data that is geocoded as a fundamental component. As such, countries are urged to consider how to take advantage of novel and emerging technologies and methods to ensure the production of geocoded data, and to embrace the GSGF, along with new technologies and approaches, within their statistical and geospatial production architecture. This will include acquisition of geospatial and statistical analysis and data management skills, which will ultimately strengthen a country's broader information ecosystem, not just within the statistical and geospatial communities.

The GSGF is complemented by overarching frameworks, including the Integrated Geospatial Information Framework, and upcoming initiatives including the OGC's Statistical Domain Working Group[11]. As such, it is important to consider that implementing the GSGF will be a journey for all countries, regardless of the starting point. Many will need assistance from their neighbours, from within their region and from those that have already made progress in implementation.

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 NSOs and NGIAs. The information in this paper provides essential guidance on the key elements to be considered in this process and in the cooperative work that will need to be accomplished by all parties. The United Nations Regional Commissions will also provide a useful resource for capacity building and promotion activities, including information sharing within regions.

Against the background of the 2020 Round of Population and Housing Censuses and the 2030 Sustainable Development Agenda, the urgent need for the GSGF is clear. Implementing the GSGF may be a challenging process, but the rewards in improved information, enhanced interoperability, and better decisions will follow this effort.

[11] http://www.opengeospatial.org/projects/groups/statisticaldwg

Part 2: Elaborating in detail the Global Statistical Geospatial Framework

Principle 1: Use of fundamental geospatial infrastructure and geocoding

Principle 1: *Use of fundamental geospatial infrastructure and geocoding* focuses on creating infrastructure that enables the implementation and socialisation of the GSGF. This infrastructure supports the creation of high-quality, standardised location references such as a physical address, property or building identifiers, or other location description, and ensures the accurate assignment of coordinates and standard grid references. The addition of a time and date stamp can add a temporal component to the unit.

Why do we need this Principle?

The collection of accurate information on location provides the input to obtain a high-quality, persistent, standardised location, such as a physical address, property or building identifier, or other location description. This allows accurate coordinates and/or a small geographic area, or standard grid reference to be assigned to each statistical unit (i.e. at the microdata/unit record level in statistical datasets). The addition of a time and date stamp can add a temporal component to the statistical unit record. Where the current geospatial, statistical, or legal and policy infrastructure within a country does not permit the records to be assigned coordinate precision geocodes, then geocoding using more general location descriptions and/or larger geographies will be necessary and is a suitable compromise. Furthermore, using persistent unique identifiers can provide links to non-spatial data and provide linkages to historical information and the mechanisms for the version control of features.

What does this Principle cover?

Principle 1 empowers countries and implementers by providing information on the key data and infrastructure that will enable a common and consistent approach to establishing the location and a geocode for each statistical unit in a dataset, such as a person, household, business, building or parcel/unit of land, with a corresponding record of a time and date for each unit record.

The data inputs to Principle 1 should be from standardised data sources to assure the quality, accuracy, currency, and consistency of data from data custodians within the NSS and NSDI. This helps to meet country-level priorities, international agendas, and internationally agreed and recognised standards and good practices. A key driver of the GSGF is to support the needs of statistical aggregation and disaggregation, and data integration, to inform data-driven decision-making at different geographic levels. To facilitate this, it is recommended by the Statistical Commission[12] that all statistical unit record data should be collected or associated with a location reference, preferably a coordinate. If this is not possible, then the association of the statistical units with a geography (i.e. a polygon) or grid cell is necessitated. Also, wherever the infrastructure[13] is existent, fundamental

[12] E/CN.3/2018/33

[13] When implementation of x- and y- coordinate georeferencing is not possible, it is advised that location descriptions and more general or larger geographies are used as the geospatial references within datasets. These can be found in authoritative datasets such as address or building registers.

geospatial data[14] from country-level agencies should be used to support geospatial referencing and other activities within the statistical and administrative data communities.

Objectives

The main objective of Principle 1 is the creation of a resilient, fundamental geospatial infrastructure that handles statistical and geospatial data. This in turn enables interoperability to combine data from various sources, such as a high-quality, standardised location, for example a physical address, a property or building identifier, or any other accurate, current, consistent, and standardised location element, which subsequently allows for the assignment of precise coordinates.

Requirements and benefits

The following key benefits are directly related to a harmonised fundamental geospatial infrastructure, as well as being relevant for all five Principles of the GSGF:

- Allows the sharing of data and information for general analysis;
- Promotes the use of geocoded information;
- Promotes the use of standards and good practices;
- Makes data transparently available with easy access to information;
- Unlocks new insights and relationships between data;
- Promotes the visualisation of data;
- Supports strategic and decision-making processes (local, sub-national, national, regional, and global levels);
- Supports time consistent linking of statistics and geographic location information by maintaining lifecycle information of geospatial features;
- Allows reproducibility of results to produce transparent and high-quality results;
- Gives impact and legibility to statistics;
- Offers new geospatial information sources to measure and monitor targets and indicators for the SDGs; and,
- Allows new methods, evaluation processes and visualisation techniques to be implemented in Population Censuses and other statistical collection efforts, thus allowing comparison between similar data through time.

Relationship to other Principles

Principle 1 is the substantive basis on which the subsequent Principles, in particular Principle 2, are built. In return, Principle 4 has a strong impact on Principle 1 as the use of standards, data models, and harmonisation of data content is a crucial element in building a fundamental geospatial infrastructure. Principle 3 provides statistical geographies for the fundamental geospatial data, which form a key component of this Principle.

[14] Such as the Global Fundamental Geospatial Data Themes (see Annex B)

Inputs

Statistical Data

Territorial/Administrative geographical units or geographies defined for statistical purposes (including census enumeration areas and blocks). Social statistics, economic statistics, demographic statistics, censuses data, agricultural statistics, environmental statistics and other statistical and administrative datasets.

Fundamental Geospatial Data

The Global Fundamental Geospatial Data Themes are a set of 14 themes, endorsed by UN-GGIM[15], to support the capacity development of NGIAs and NSOs. Implementing the themes will require the integration of statistical data and information from NGIAs, NSOs, and other institutions to produce standardised, fundamental geospatially-enabled data for use within member states and to support initiatives such as the SDGs. The management of these fundamental geospatial data themes usually occurs within the area of a NSDI. Implementing these themes provides a fundamental component of the geospatial infrastructure for countries.

Key Stakeholders

Stakeholders that can contribute to, adopt and implement Principle 1 include NSOs (encompassing the NSS), NGIAs (encompassing a country's geospatial information system), NGOs, civil society, private enterprise, data suppliers, and citizens. As many stakeholders are affected, an open data environment is recommended due to the efficiencies and potential opportunities afforded by such a backdrop. This is counterbalanced by the necessity to consider how best to ensure privacy of personal data of citizens in line with evolving country- and regional-level privacy legislation and compliance regulations, while considering international norms and recommendations.

To enhance informed decision-making, implementers (primarily NSOs and NGIAs) are urged to have strong links of communication and institutional collaboration. This can include, but is not limited to, country-level laws and policies, MoUs, and/or communities of practice.

[15] As annexed in the report: http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/E-C20-2018-7-Add_1-Global-fundamental-geospatial-data-themes.pdf

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

Principle 2: *Geocoded unit record data in a data management environment.* Using the infrastructure from Principle 1, this Principle supports the linking of each statistical unit record to a geographic reference (e.g. a coordinate or small geographic area) that will allow for statistics to be applied to any geographic context. This will support integration or linkage of data from other data sources and mitigate challenges that arise with new geographies or changes in existing geographies.

Why do we need this Principle?

All statistical unit records need to be geospatially enabled - geocoded to a specific location - in a secure, standards-based data management environment. This in turn supports the integration or linkage of data from a wide variety of data sources and allows statistical outputs to be applied to any geographic context, including new or changing geographies.

What does this Principle cover?

Principle 2 covers the process of linking and storing a high-precision geographic reference (i.e. geocodes – coordinates or small geographic areas) for each statistical unit record within a secure, standards-based data management environment. This allows for the statistics generated from these datasets to be produced for a wide range of geographic contexts; for example, various administrative and statistical geographies, grid systems, and supports future aggregation of statistical data into new geographical units or adaptation to changes to existing geographies over time.

This enables integration or linking of data from a wide variety of sources, such as other administrative statistics, or geospatial data about the built and natural environment, like earth observation data, the incorporation of these data into geospatial analyses, such as measures of proximity of households to public open space, health, or other services and activities. It also enables data linkage processes that are dependent on, or enhanced by, having precise location information, for example, in matching census and administrative records based on location and other demographic characteristics. Ultimately, this ensures that detailed data can be managed securely to ensure safe and appropriate access to potentially private or sensitive information.

Objectives

In implementing Principle 2 the following objectives should be targeted:

- Enable the production of geospatially enabled statistics;
- All statistical unit records should include or be linked to a precise geographic reference (ideally an x- and y- coordinate), and if not, the smallest geographic area possible. This fosters the greatest opportunity for reuse and aggregation of geospatially enabled statistical information. Where a foundational geospatial infrastructure cannot support this activity then other methods should be pursued, such as direct coordinate capture or use of small area statistical/census enumeration areas;

- Ensure the effective implementation of fundamental geospatial and geocoding infrastructure and demonstrate its broader value. This means working in partnership with NSOs, NGIAs, and other providers of fundamental data to ensure requirements are understood and the data and infrastructure are used appropriately;
- Implement effective data management of statistical and geospatial data. This requires good technical data and metadata management practices, in accordance with country-level and international standards and good practices;
- Ensure appropriate protection of privacy and confidentiality of unit record or microdata level datasets. This includes securing sensitive and confidential data while ensuring appropriate access to meet user needs for analysis and decision making;
- Storage of consistent and interpretable geographic references preferably linked from a "point-of–truth" (e.g. linked to a centrally managed address register). This requires the establishment and implementation of data and metadata standards to ensure that geographic references are well documented and consistent across datasets, allowing them to be used effectively over time in different applications; and,
- Ensure that data is stored in a way that will facilitate flexible application of any geographic content when preparing data for release and analysis. This includes supporting future aggregation of statistical data into new geographical units or in adapting to changes to existing geographies over time and ensuring they are kept secure to prevent misuse.

Relationship to other Principles

There is a strong relationship to Principle 1 as activities in Principle 2 are dependent on, and should use, the fundamental data provided through a country's geospatial data infrastructure and geocoding capabilities.

Principle 2 is also a key enabler for Principle 3 as these geographies are dependent on, and use, the definition of common geographic regions for the dissemination of data defined in Principle 3 - these may also be identified within the fundamental or country-level geospatial data identified in Principle 1. Associated metadata and data for the common geographic regions defined in Principle 3 should also be applied in the activities associated with Principle 2. Additionally, the geographies, and associated data, standards, and good practices used from Principle 3 may be drawn from the geospatial and/or statistical community's data, tools, standards and good practices.

Inputs

Statistical Data

From the statistical perspective, the GSBPM and GSIM[16] managed by the UNECE HLG-MOS are models intended to enhance, integrate, promote, support, and facilitate statistical modernisation[17]. The GEOSTAT[18] project is an ESS network initiative that fosters improved integration of statistical and geospatial information for the statistical community to provide more qualified descriptions and

[16] The current versions of these models incorporate geospatial tools, methods and processes into their definitions and resources. As of the start of 2019, there are limited examples of current country applications that incorporate geospatial processes, but this is expected to increase over time.
[17] https://statswiki.unece.org/display/hlgbas
[18] The GEOSTAT projects provide countries and regions with an example of GSGF implementation, along with further, more detailed resources and practical implementation guidance. https://www.efgs.info/geostat/

analyses of society and environment. GEOSTAT3 has developed recommendations for a harmonised implementation of the GSGF within the overarching European statistical community, considering existing conditions, initiatives and European and country-level frameworks.

Additional resources include:

- UNECE's HLG-MOS CSDA[19];
- Data management principles[20];
- Privacy laws, policies, and/or agreed country-level and international privacy protocols. Some current resources include:
    - United Nations Fundamental Principles of Official Statistics[21]
    - Principles and Guidelines for Managing Statistical Confidentiality and Microdata Access[22];
    - Results of the UNSD/UNECE Survey on organizational context and individual projects of Big Data[23];
    - Relevant country- and regional-level laws and policy.

Geospatial Data

From the Geospatial perspective, OGC/ISO/IHO standards, including those listed in "A Guide to the Role of Standards in Geospatial Information Management" are the international geospatial standards best practice guide for geospatial data infrastructure. These standards will evolve and change through OGC/ISO/IHO processes and will include input from a recently established Statistical Domain Working Group[24] under the auspices of the OGC.

Additional resources include:

- The GEOSS Data Management Principles[25]
- Addressing and/or location reporting standards and infrastructure (see GSGF Principle 1);
- Geocoding infrastructure tools, metadata standards and good practices, including batch and point of contact address validation and geocoding (see GSGF Principle 1);
- Promotion of point-of-entry address validation and geocoding (see GSGF Principle 1); and,
- Privacy laws, policies, and/or agreed country-level and global privacy protocols. Some current resources include:
    - Agreed geographic regions, etc. and associated data and metadata infrastructure (see Principle 3);
    - Global or national/regional Geodetic Reference Frames, such as the Global Geodetic Reference Frame[26].

---

[19] https://statswiki.unece.org/display/hlgbas/Modernisation+Groups
[20] https://statswiki.unece.org/display/DA/VI.+Key+principles
[21] https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx
[22] https://unstats.un.org/unsd/statcom/doc07/BG-Microdata-E.pdf
[23] https://unstats.un.org/unsd/statcom/doc15/BG-BigData.pdf (page 10)
[24] http://www.opengeospatial.org/projects/groups/statisticaldwg
[25] https://www.earthobservations.org/documents/dswg/201504_data_management_principles_long_final.pdf
[26] http://www.unggrf.org

Key Stakeholders

Geospatial community

- Provision of fundamental geospatial data and infrastructure, and geocoding capabilities (from GSGF Principle 1);
- Country, regional, and global-level Geodetic Reference Frames and implementations;
- Geospatial data management frameworks;
- Geospatial data standards and good practices, particularly geocoding metadata specifications; and,
- Supporting common geographic concepts and areas.

Statistical Community

- Country- and global-level privacy protocols (e.g. UN Fundamental Principles of Official Statistics);
- Statistical data management frameworks, including base registers (e.g. address, building, and business registers;
- Supporting common statistical geographic concepts and areas;
- Non-survey data, including scanner and sensor data sources, and crowdsourced data; and,
- Implementation of Principles to statistical and administrative unit record data and their storage and management.

Administrative Data Community

Implementation of Principles to administrative unit record data and their storage and management[27].

[27] Across these stakeholder roles differences between community roles may occur at the country-level.

Principle 3: Common geographies for dissemination of statistics

Principle 3: *Common geographies for the dissemination of statistics*. This Principle determines definitions of geographic regions and aggregation/disaggregation of data to regions and enables consistency and comparability of integrated statistical and geospatial data. A common set of geographies ensures that statistical data is geospatially enabled in a consistent manner and is capable of being integrated at the aggregate level; and also ensures that users can discover, access, integrate, analyse, and visualise statistical information seamlessly into geographies of interest.

Why do we need this Principle?

Informed, evidence-based decision-making necessitates the comparison of statistics across geographies. This is achieved through common geographies. Common geographies can include census blocks, grids, administrative and cadastral boundaries. Common geographies enable basic statistical reporting, geostatistical analyses and visualisation at different scales (such as at local, sub-national, national, regional[28], and global, as well as institutional[29]), where the resulting outputs can be compared and assessed on a consistent basis. These geographies also provide a mechanism to enable the management of privacy and confidential statistical and geospatial data outputs.

What does this Principle cover?

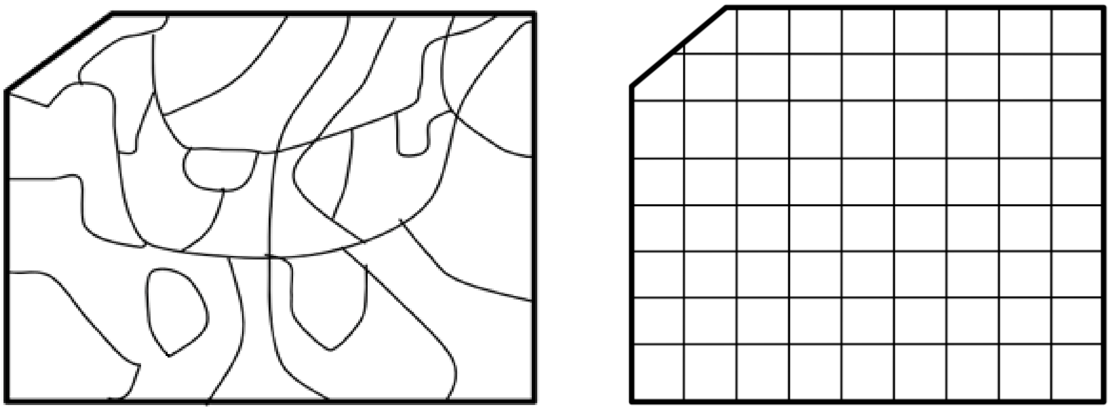
A common geography enables the production and dissemination of integrated statistics and geospatial information within a country to support informed decision-making. This can be achieved by promoting stakeholder participation in the creation of geographic classifications (a classification constituted of one or more common dissemination geographies) and geographic grids. This ultimately promotes increased data consistency and efficiency among participating NSOs, NGIAs, international and regional organisations, and other relevant institutions, through the generation and maintenance of the geographic reference layer.

Accordingly, the goal of Principle 3 is to support the provision of a common set of geographies that ensure that statistical data is geospatially enabled in a consistent manner, whether in gridded form or using administrative or statistical boundaries. These are either uniformly assigned boundaries by size when a grid, or by administrative boundaries or statistical units (such as mesh blocks or functional areas) that are divided by political, property, or topological subdivisions.

A key element of this Principle is to enable the aggregation of data accurately from the smallest common dissemination geography to higher level dissemination geographies in order to meet the needs of the widest possible array of decision makers irrespective of geographic boundaries and/or scale. This includes aggregations at both higher level administrative and statistical geographies (e.g. electoral areas, census tracts, and other functional areas)[30] and the smallest possible cell size when using a gridded geography. It is also important to note that there are two options for the dissemination

[28] Such as the European Union, North America, and Africa
[29] Such as la Francophonie, G20
[30] There is nothing inherently incorrect with a unit record-to-small area geography approach. However, this approach will become less optimal over time as the demand from stakeholders for statistical data for new and additional common higher-level geographic areas are made of institutions responsible for integrating statistical data and geospatial infrastructure.

of geospatially enabled statistics: through Administrative/Statistical Geographies, or Gridded Geographies (see Figure 4), these are discussed in further detail in Annex B.



Administrative Geographies

Gridded Geographies

Figure 4 Administrative and Gridded Geographies

NSOs are urged to consider the benefits of gridded data alongside established administrative geographies created by NGIAs and other geographic institutions. Gridded data can be both a rich source of information and a consistent geography for disseminating and integrating information[31].

Objectives

In implementing Principle 3, the following objectives should be targeted:

- Ensuring that the broadest array of potential common dissemination geographies, legal-administrative[32], statistical[33] and more recently initiated integrative[34] geography types are considered and eventually represented;
- Common dissemination geographies should be collaboratively assessed and acknowledged by interested stakeholders prior to their adoption. They should be integrable within existing and emergent statistical geospatial infrastructures. Information on changes should be documented by using time stamping and lifecycle information on births, splits, mergers, and border shifts within these geographies (such as when new cadastral or political boundaries are created). They should also be able to include the production of comparable and integrated social, economic and environmental data, indicators and other information. This will enable an enhanced capacity to produce data and indicators for domestic purposes and supports the attainment of the monitoring and reporting needs of global initiatives (such as the 2020 Round of Population Censuses, and the SDGs);
- Viewing proposed and adopted common dissemination geographies as complementary to existing administrative and statistical geographies, and enabling the integration between them is critical;

[31] Recent global efforts have culminated in the development of a DGGS standard which has been developed under the auspices of OGC. This offers further options in the use of grids within the context of the principle of common geographies and in geospatially enabled statistics. More information on the DGGS: http://www.opengeospatial.org/projects/groups/dggsswg

[32] Geographies defined in law, regulations or constitution (e.g. the legal definition of major sub-national geographic areas – states and provinces)

[33] Geographies defined by a set of rules or a methodology meant to represent a geographic concept (e.g. the metropolitan regions or core-based functional areas, labour market areas outside of metropolitan regions).

[34] Geographies designed to efficiently integrate social, economic and environmental data. (e.g. the grid-based approach). In Europe the GEOSTAT population-grid dataset is promoted as a first example of a European Union (EU) population grid.

- Providing metadata and other related documentation for all geographies, for example, information on delineation methodologies for any new common dissemination geography and the lifecycle and changes to these geographies through time. This is essential to ensure that the potentially complex concepts for understanding delineation methodologies used to maintain common dissemination geographies are understood and used by stakeholders. As such, the continuing adoption of existing international statistical and geospatial metadata standards are a means to achieving this goal (see Principle 2);
- Establish tools and methods to enable simplified geographic aggregation of data. This will include implementing standard code lists, or allocation tables, that enable statistical tools and applications to consistently aggregate and display or map geographic aggregations of data;
- Ensure that the evolving country-level and international data privacy and data quality Principles, frameworks and practices are considered and respected in the design of common dissemination geographic areas, their adoption and subsequent implementation (for example with respect to data and indicator reporting, distribution and visualisation); and,
- Undertake regular and cooperative reviews of any adopted common dissemination geography to ensure continuing relevance to key country-, regional- and global-level programs.

Relationship to other Principles

Implementing Principles 1 and 2 is a pre-condition for the full implementation of Principle 3 in a statistical production system as their implementation allows for flexible aggregation of data into any output geography. In turn, Principle 3 is an important condition for Principle 5, as common geographies form the basis for disseminating geospatially enabled statistics. If Principle 2 is fully implemented, it is straightforward to aggregate geocoded statistics into any output geography.

Inputs

Statistical Geographies

- Statistical units/grid cells;
- Statistical regions; and,
- Census Enumeration Areas.

Geospatial Data

- Administrative boundaries;
- Functional areas;
- Address locations;
- Topography; and,
- Buildings/dwellings and/or cadastral parcels.

Key Stakeholders

Key stakeholders in this Principle are analogous to Principle 2.

Principle 4: Statistical and geospatial interoperability

Data, Standards, Processes, and Organisations

Principle 4: *Statistical and geospatial interoperability (data, standards, processes and organisations)* enables greater standardisation and use of data which will lead to improved efficiency and simplification in the creation, discovery, integration, and use of geospatially enabled statistics. It also increases the potential application of a larger range of data and technologies, and thereby enables a wider range of information to be available and accessible for use in decision-making, and addresses aspects of better cooperation between all stakeholders producing and using statistical and geospatial information.

Why do we need this Principle?

Greater interoperability between statistical and geospatial data and metadata standards is required to overcome structural, semantic, and syntactic barriers between data and metadata from different communities and providers. This also improves the discovery, access, and use of geospatially enabled statistical data. Enhancing interoperability improves the fitness-for-use of geospatial and statistical data for their use in a range of applications and data management systems, including data modelling and production planning. Clear agreement on standards and commitment to their implementation are therefore critical to realise the benefits of interoperability.

What does this Principle cover?

Principle 4 covers the interoperability of all data, metadata, standards, and good practices that facilitate the integration and output of geospatially enabled statistical data. This includes tools and methods which are used in all stages of the statistical production process. It also addresses supporting processes, including reproducibility, quality management and the mechanisms by which stakeholders and users interact. Principle 4 recognises that both the statistical and geospatial communities operate their own general data models, metadata capabilities, architectures and data infrastructure. For example, the statistical community use the GSIM, SDMX, and DDI mechanisms. In parallel, the geospatial community commonly use the GFM and developed the ISO:19115 metadata standard, plus several application specific standards[35] and good practices to support interoperability of data.

Within the statistical community there is a need to consistently build geospatial processes, standards, and good practices into statistical business processes and data management systems as a fundamental element, not just to disseminate statistical data. To ensure this occurs, countries are urged to consider how to incorporate existing geospatial frameworks, standards, good practices, and processes more explicitly into the CSPA and its components. This would in turn provide greater efficiency and

[35] For a discussion on these statistical and geospatial models and metadata standards, see: http://ggim.un.org/meetings/2015-2nd_Mtg_EG-ISGI-Portugal/documents/Connecting%20Geographic%20and%20Statistical%20Information%20Standards%20EG-ISGI%202015.pdf and http://ggim.un.org/meetings/2015-2nd_Mtg_EG-ISGI-Portugal/documents/Metadata%20interoperability%20cover%20paper%20EG-ISGI%202015.pdf

simplification of the creation, release, discovery, and use of geospatially enabled statistics, while also supporting the integration of geospatial, statistical, and administrative data.

Principle 4 covers not only technical aspects, but also addresses all aspects of how organisations produce and share information with each other and data users. The integration of statistics and geospatial information is a genuinely cross-cutting effort that specifically concerns various departments within NSOs and NGIAs, but also requires input and institutional collaboration from other government agencies and organisations, as well as producers and users of information. As a result, in addition to technical matters, cooperation models and agreements need to be adjusted and aligned so that they do not hamper the integration and use of data. The European Interoperability Framework, shown in Figure 5 below[36], has been used to provide a mechanism for describing the required dimensions.

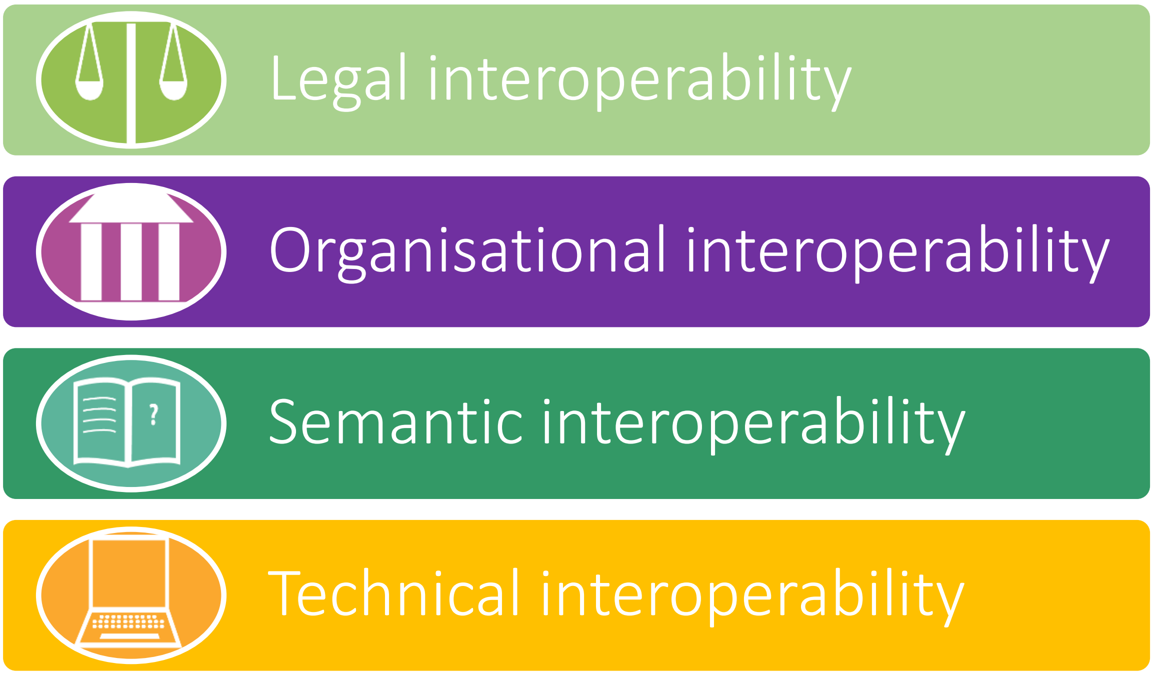


Figure 5 The European Interoperability Framework

1. **Legal interoperability** enables organisations operating under different national legal frameworks, policies and strategies to work together. National laws and policies should not block cooperation and there should be clear agreements about how to deal with differences in legislation across borders. As an example, national laws and policies on statistics should include the right of NSOs to have access to essential geospatial information with defined quality and ideally without charging.
2. **Organisational interoperability** refers to the way in which public administrations (i.e. government agencies and organisations) align their business processes, responsibilities and expectations to achieve commonly agreed goals. In practice organisational interoperability means documenting and integrating or aligning business processes and relevant information exchanged. This also covers meeting the requirements from the user community and the NSS.

[36] https://ec.europa.eu/isa2/eif_en

3. **Semantic interoperability** ensures that the precise format and meaning of exchanged data and information is preserved and understood: "What is sent is understood". This includes syntactic aspects, such as the terminology used to describe concepts, as well describing the exact format of the information.
4. **Technical interoperability** covers the linking systems and services of applications and infrastructures. Aspects include interface and services specifications, and data and metadata standards and formats.

Each of these items are crucial towards the integration and output of geospatially enabled statistical data and share a close interlinkage with each other.

Objectives of this Principle

In implementing Principle 4, the following key elements should be targeted:

Standards and good practices

- Enabling experts from the statistical and geospatial community to fully understand the nature, potential and limitations of other data domains;
- Facilitating smooth communication between experts by using consistent and understandable terminology;
- Provision should be made such that data, tools, processes and methodologies are documented in the official languages of the country;
- Guaranteeing preservation and persistence of data and tools;
- Ensuring that only open and international standards and good practices are implemented, ideally by using or further developing existing standards or connecting between existing standards; and only creating new standards and practices when necessary and doing so collectively;
- Ensuring transparency and visibility of data and metadata; and,
- Safeguarding common quality Principles.

National Laws and Policies

- Supporting cooperation of stakeholders through arrangements and legislation.

Infrastructure

- Ensuring that geospatially enabled/integrated data can flow freely between statistical and geospatial data producers, and from data producers to data users without having to worry about technical, national laws and policies, organisational, economic, language and conceptual barriers or national borders;
- Implementing service based or machine-readable access mechanisms (e.g. through APIs) that provide greater efficiency of access and use and allow adaptation and evolution of uses through time;
- Developing common solutions, so data and tools can be re-used, avoiding duplication of effort, through a single national fundamental geospatial infrastructure;

- Ensuring that users, no matter if they come from both a geospatial or statistical background, can discover and access the required information via defined technical and user interfaces that do not require cross domain knowledge; and,
- Ensuring that data and tools are open and free, wherever possible, so that users have access to the full range of information with no information loss due to technical or other interoperability issues;

Relationship to other Principles

Interoperability concerns how data travels from the source to the end-user; for example, across the full statistical production process, including dissemination to intermediate and end-users. As a result, interoperability issues in most cases cut across the other Principles of the GSGF rather than belonging to one Principle only. Interoperability is critical to the successful implementation of the GSGF.

The full implementation of interoperability described in this Principle is particularly important for Principle 5, as failure to achieve interoperability in any of the other Principles will often result in incomplete or less useful information for the end-user.

Key Stakeholders

Often, NSOs and NGIAs are augmented by administrative data custodians, which also act as providers of statistical data, but which are often not interoperable with statistics and geospatial information (for example administrative boundaries – see Principle 3). Other stakeholders are the main global standard setting bodies such as ISO, OGC and IHO and the organisations driving the Modernisation of Official Statistics, such as UNECE[37].

The **European Commission** is the custodian of INSPIRE as the most important standard setting framework for geospatial information in Europe with **Eurostat** maintaining the European Statistical System and contributing with respect to standard setting activities, such as SDMX and ModernStats. The regional overview is further supported by the UN Regional Commission for Europe, **UNECE**, in supporting the “Modernstats” initiative for the modernisation of official statistics. The European Forum for Geography and Statistics focuses on the development of best practices in the production of geospatially enabled statistics in Europe and acts as the professional network and organises the European annual conference on the integration of statistics and geospatial information, further enhancing knowledge exchange and communication. Regionally, Europe have led with various initiatives and bodies supporting interoperability and the UNSC acts as the global custodian for statistics and geospatial information and their integration, while supporting capacity building.

European Efforts towards Interoperability

[37] This is not limited to those countries within the geographic bounds of Europe and includes several non-European Member States: https://www.unece.org/oes/nutshell/member_states_representatives.html

Principle 5: Accessible and usable geospatially enabled statistics

Principle 5: *Accessible and usable geospatially enabled statistics*. This Principle supports data custodians to release data with confidence, improve the discovery and access of geospatially enabled statistics (particularly through promoting web services to provide machine readable and dynamic linkage to data), and to support analysis and evaluation of data in decision making.

Why do we need this Principle?

The data derived from geospatially enabled statistics needs to be accessible and usable by policy-makers to make informed, data-driven decisions. As such, data custodians need to be able to release data in confidence, in accordance with national and regional laws and policies, while ensuring privacy protection and other good practices are followed.

What does this Principle cover?

Principle 5 supports the access, analysis, visualisation, and dissemination of geospatially enabled statistical information to support informed decision-making. This is achieved through utilising, identifying, and developing applicable policies, standards, and good practices that increase the accessibility and use of geospatially enabled statistics.

One important aspect of Principle 5 is to ensure data is maintained at regular intervals and can be accessed in a timely manner using safe mechanisms that protect privacy and confidentiality and enable the analysis of data to support informed, data-driven decision-making. This includes the provision of metadata that denotes when the information was last updated, published, and when it is expected to be updated. Other issues of relevance include the different factors of data quality (particularly regarding reliability, timeliness, and relevance), and the data's analysis, dissemination and visualisation.

Accessibility is about providing the policies, standards, good practices, and technologies to make geospatial data readily available. These policies, standards, good practices, and guidelines highlight the wide range of legislative and operational issues that organisations need to be aware of when releasing and analysing information about people and businesses. Usability is about ensuring that those outputs are easily understood by a wide range of technical and non-technical users and can be used to support policy and decision-making. Through protecting the privacy and confidentiality of data, data custodians can release data with confidence and data users can easily discover and access geospatially enabled statistics which are reusable and can inform visualisation, analysis, and decision-making.

Accessibility and usability can be mutually achieved through the proper tagging and cataloguing of data by appropriate standardised metadata, and subsequently disseminated through standardised and open web services or file formats. This provides machine-readable data access and the potential to support linked data, which in turn creates opportunities for the provision and development of integrated knowledge systems across sectors and across local, regional and global implementations.

Objectives

In implementing Principle 5, the following objectives should be targeted:

- Geospatially enabled statistical data that can be rapidly and easily accessed and can be effectively utilised to address policy analysis and rapid evidence-based decision-making;
- Countries should strive to suitably protect the privacy and confidentiality that may be required of input data, and the people and businesses providing this information, while also producing high-quality, informative, spatio-temporal statistics;
- Dissemination platforms that support automated data discovery to ease the exchange, aggregation, and presentation of the data;
- Ideally "change" is tracked and reported to differentiate purely statistical change (population increase/decrease) from geospatial change (large annexation/de-annexation of territory) and representational change including datum shifts and adjustments to the cadastre;
- Ensure that geospatial data, statistics, administrative data, and other data sources are integrated, available, and accessible. Ideally, one access method should include web services, which enable machine-readable access, where data is accessible, discoverable, and provided in a standardised, ingestible, platform-agnostic format;
- Redundancy should be built into the system, in the form of multiple platforms, access points, and distribution methodologies. Mirrored websites, alternative data formats, and other methods should be implemented; and,
- Access to a broad audience should be enabled through web-based GIS platforms with the ability to generate custom visualisations, aggregations, and eventually data analysis.

Relationship to other Principles

Principle 5 influences the implementation of all other Principles, as it stresses the need for high-quality geospatially enabled and integrated data for evidence-based decision-making and is structured in a way that enables access and use. The geospatially enabled outputs should be designed to meet specific reporting requirements such as the SDGs, the SEEA framework, or country-level development priorities. Therefore, the inputs from the previous Principles must be designed with these requirements as a central consideration.

Principles 1 and 2 are important because of their influence on the timeliness and quality (relevance, completeness, definitions, classifications, topology, etc.) of the data available to users. In turn, the data collection and maintenance decisions when applying Principle 1 should be informed by the usability and applications identified within Principle 5.

Principle 3 is related through the clear need for common output geographies, which form the basis for analysis and comparisons, and Principle 4 is closely related through technical standards that drive interoperability of data and technical infrastructure. The importance for interoperability within technical infrastructure, extends to front end user interfaces and back end databases, software and hardware components.

It is important to note that broad objectives of Principle 5 can and should be pursued to the greatest extent possible even if the whole GSGF is incapable of being implemented. This will ensure that geospatially enabled data is made available in the most accessible ways possible.

Inputs

The inputs for this Principle are ultimately the data, elements and processes defined in the other Principles. As such, the inputs to this Principle are:

- Timely and accurate geospatially enabled data;
- Standardised metadata and proper tagging of data;
- Involvement, knowledge, and skills from both the geospatial and statistical sides are important;
- Technical infrastructure capable of providing access to published datasets, including in machine readable forms; and,
- Current technical standards for web-based services and accessibility.

Key Stakeholders

The statistical and geospatial communities will play a large role in the successful implementation of Principle 5. The exact nature of this role, together with detailed use cases for implementation, should be identified through mutually agreed institutional collaboration arrangements. Principle 5 represents the interface between public use, and the data collected and curated through the GSGF processes. External stakeholders and NSOs should play an important part in developing and advising on best practices to prevent the disclosure of private data and to prevent the use of data which might personally identify respondents or lead to sensitive information being accessed.

Part 3: Annexes

Annex A: Definition of Terms

Common Geography

A common geography is an agreed set of geographies for the display, storage, reporting, and analysis of social, economic and environmental comparisons across statistical datasets from different sources. Common geographies enable the production and dissemination of integrated statistics and geospatial information within a country or region to support informed decision-making.

Data Management Environment

A data management environment holistically encompasses the tools, storage, and environment for acquiring, validating, storing, protecting, and processing required data to ensure the accessibility, reliability, and timeliness of the data for its users.

Discrete Global Grid System (DGGS)

A DGGS represents the Earth as a hierarchy of equal area cells with progressively finer geospatial resolution. Individual observations can be assigned to a cell corresponding to both the position and size (or uncertainty) of the phenomenon being observed. DGGSs provide significant benefits when integrating and analysing geospatial data.

Fundamental Geospatial Infrastructure

A Fundamental Geospatial Infrastructure encapsulates the NSDI (see below), standards, technologies, policies, and best practices, amongst other key elements to enable the provision of geospatial information within a country. The UN-GGIM endorsed Integrated Geospatial Information Framework[38] can provide this infrastructure, if such country-level infrastructure is not yet established.

Geocoding

For the purposes of the GSGF, geocoding is generally defined as the process of geospatially enabling statistical unit records so that they can be used in geospatial analysis. More specifically, geocoding is the process of linking unreferenced location information (e.g. an address), that is associated with a statistical unit, to a geocode (i.e. a geospatially referenced object). Alternatively, the geocode can be directly incorporated into the statistical unit record. Geocodes are, preferably, fine scale geospatially referenced objects that are stored as a geometry data type, such as location coordinates (i.e. x, y, z coordinates), and/or small area geographies (e.g. mesh blocks, block faces or similar small building block geographies). Larger geographic units, such as enumeration geographies, can be used as geocodes where finer scale geospatial units are not available. The linkage of a geocode to a statistical unit record can occur through use of standard geographic coding systems, a Uniform Resource Identifier (URI) or through other computer-based linkage mechanisms.

Geographic Feature

A geometric representation of a feature. This can be a physical feature such as a unit record, a dwelling, or property or a functional area such as an administrative boundary or an economic area.

[38] http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Part%201-IGIF-Overarching-Strategic-Framework-24July2018.pdf

Georeferencing

Georeferencing is a set of broad processes that includes geocoding. It is the process of referencing data against a known geospatial coordinate system, by matching to known points of reference in the coordinate system (e.g. image rectification to survey points or addresses linked to parcel centroids), so that the data can be viewed, processed, queried and analysed with other geographic data.

Geospatially Enabled Statistics

Location or geospatial-extent are the main characteristics of geospatially enabled statistics. Furthermore, it is recommended that all statistical unit record data should be collected or associated with a location reference, and that ideally it should allow for geospatial coordinates with x- and y-values to be produced for each record.

Interoperability

Interoperability is the ability of a system to exchange and make use of information, enabled through the application of open standards.

Location information

Location information can include addresses, property or building identifiers, as well as other location descriptions, such as enumeration geographies and other standardised and non-standardised (e.g. village names) textual descriptions of a location.

National Spatial Data Infrastructure

National Spatial Data Infrastructure is the technology, policies, standards, good practices, and human resources necessary to acquire, process, store, distribute, and improve utilisation of geospatial data. A successful NSDI implementation addresses the following considerations:

- Maintenance of data and systems;
- Redundancies should be built into the dissemination solution to prevent a single point of failure;
- Final review and pre-processing before release (data disclosure and confidentiality) to prevent disclosure issues; and,
- Generalisation and thinning of geospatial data should be implemented to ensure that the data meets the minimum level of quality and is useable at defined scales supporting both large and small-scale needs. This can impact both cartographic and data storage issues.

Statistical Unit Records

Statistical Unit Records can include persons, households and living quarters, businesses, buildings or parcels/units of land.

Annex B: Standards, Quality and Enabling Frameworks

Just as the need to work together is fundamental to implement the GSGF, so is the need to adopt and implement the respective statistical and geospatial standards, and good practices, as well as the need to assess the quality of inputs and outputs of the data used across the GSGF.

It is important to recognise that there are many standards in both the statistical and geospatial communities and while the following information will support and guide implementers of the GSGF, it is by no means an exhaustive list. NSOs and NGIAs are urged to be aware of the quality Principles and standardisation efforts in both the statistical and geospatial communities, as well as other related communities, especially considering the high-quality work already occurring in the development of standards. Countries are urged to contribute to these efforts and to consider the GSGF as part of a set of wider efforts to modernise the production of statistics and geospatial information, as well as the general advancement of technology and information.

The following sections briefly introduce the key relevant standardisation efforts related to statistical and geospatial information, as well as the key quality Principles of both communities. It is not the goal of this section to introduce a comprehensive list of standards that are relevant but to raise awareness that data integration requires awareness of the use of standards and an understanding of the wider context, particularly that standards development can be very dynamic.

Standards

Standards are crucial to facilitating the development, sharing, and use of statistical and geospatial data and services, but it is important to note that statistical and geospatial data and standards have different criteria and development pathways. The GSGF urges the use of internationally adopted standards from both communities to enable greater interoperability of statistical and geospatial data, standards and processes.

Within the statistical realm:

- The HLG-MOS has developed and supports the GAMSO[39], GSBPM[40], GSIM[41], and CSPA[42];
- The SDMX ISO/TS 17369 [43] guideline, SDMX 2.1[44];
- The “Fundamental Principles of National Official Statistics” endorsed in 1994 by the UN General Assembly and accompanied by a resolution (renewed in 2014) summarise the fundamental Principles to enable the production of high-quality official statistical information for analysis and informed policy decision-making; and,
- The UNSD developed a Data Interoperability Guide[45] that supports interoperability efforts within the development sector.

---

[39] https://statswiki.unece.org/display/GAMSO
[40] https://statswiki.unece.org/display/GSBPM
[41] https://statswiki.unece.org/display/gsim
[42] https://statswiki.unece.org/display/CSPA
[43] https://www.iso.org/standard/52500.html
[44] https://sdmx.org/?page_id=5008
[45] https://unstats.un.org/wiki/display/InteropGuide/Introduction

Within the geospatial realm:

- international standards organisations such as the OGC, IHO and ISO work on the definition of geospatially relevant standards and the transformation of standards and have released the Guide to the Role of Standards in Geospatial Information Management[46];
- The ISO 19000 series (ISO/TC 211) is used to describe geographic information and services. ISO 19111 'geospatial referencing by coordinates' or ISO 19112 'spatial referencing by geographic identifiers', and, ISO 19115 'geographic information – metadata', are commonly useful standards. ISO 19115-2 is an extension of this standard for imagery and gridded data. ISO 19157 also covers quality Principles for geographic information. Furthermore, ISO 19160 defines a conceptual model for address information; and,
- OGC offers standards on standardised services to publish and perform geospatial computing processes over the web, such as WFS and WMS. When focussing on geospatial data, the GGRF (e.g. ETRS89) also needs to be considered. The DGGS also provides a standard for defining grids at local through to global levels, this is especially useful if comparing across countries.

Although there is already a rich body of statistical and geospatial standards, few standards address interoperability of statistical and geospatial information. Hence, working on an approach to overcome interoperability challenges resulting from different data models, interfaces or even file formats is highly desirable. For example, Table Joining Services defined by OGC can overcome gaps to exchange tabular encoded in SDMX data for a range of geographies.

The following priority areas taken from the Principles of the GSGF require further standardisation efforts:

- Common geographic classification, including use of administrative and statistical geographies that are complemented by use of grid type geographies; (Principle 3);
- Standards or guidance on the use of geographies for dissemination of data (Principle 3);
- International statistical and geospatial metadata standards (Principle 4);
- Systems and methods to correspond data between geographies[47]; (Principles 2 and 3); and,
- Country-level privacy laws, policies, and/or agreed privacy Principles, in accordance with the United Nations Fundamental Principles for Official Statistics. (Principles 3 and 5).

However, the GSGF explicitly does not call for the development of new standards unless they are absolutely needed. Rather, the communities should work on developing bridges between standards, or by extending existing standards. As an example, Eurostat has extended SDMX to cover the INSPIRE compliant dissemination of the 2021 Census data.

Administrative and Gridded Geographies

Historically, the boundaries of administrative areas have been seen for some as a more reliable system for surveying populations, housing units, and geographical areas because they conform to the physical features on the ground. These areas align with physical boundaries (i.e., roads, railroads, and other visible features), landscape features (i.e., rivers, lakes, and other drainage), as well as urban

[46] http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Standards_Guide_2018.pdf
[47] http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Correspondences

development patterns. Geospatial ground-truthing is perhaps the strongest argument to favour administrative areas over grids. However, there is a strong counter argument that a reliance on physical boundaries that have potential for change is precisely the reason not to use administrative areas in favour of the statistical grid. Table 1 provides an overview and comparison of the positives and negatives of administrative and gridded geographies[48].

Irrespective, when using administrative or gridded geographies, it is important to consider the potential bias introduced by statistical units having different shapes and sizes. This issue is known as the Modifiable Areal Unit Problem[49], is where the aggregation of units can be influenced by how boundaries are drawn to segment and subsequently aggregated.

Administrative Geographies

Advantages of using established administrative or statistical boundaries as a common geography is that they are usually based upon historical patterns of human settlement. For the most part, humans relate and understand administrative boundaries well. The geospatial accuracy of these areas, as well as the data itself can be verified on the ground through field verification, satellite imagery, and geocoding/address canvassing. Authoritative sources of this data often come from NSOs, NGIAs and other country-level bodies. Statistical geographies are designed for the release of statistics and can overcome many of the other issues associated with administrative boundaries. They are usually created to capture areas containing the same or similar population sizes. However, in many cases this is not possible therefore data must be normalised before comparison between areas can be performed. In order to capture consistent population sizes, statistical areas in rural land may cover large areas and contain a disproportionate mix of land uses and population areas.

Disadvantages of using administrative areas as a common geography stem from change and variability resulting from population growth, expanding urban settlement patterns, and inconsistencies between legal authorities and governments, that can affect comparability within a single country, and between countries. This challenge is especially acute in developing countries with high levels of population growth, urban and rural change, among many other demographic factors. For this reason, the comparability of census geographies that are based upon population densities, and settlement patterns that can change between census years is difficult. Non-visible boundaries are subjective, and can also change, depending upon the local land management systems and legislation. For example, if a road is constructed, or a river redirected, consequently, a non-visible boundary changes to align to a visible feature.

The challenge of comparability is amplified as there are even greater differences between countries on how administrative areas are defined as well as the number and type of different geographic areas used to disseminate data. Census blocks might be the smallest area in one country, with county or provinces in another. Cartographic generalisation also differs from country to country. Differences in

[48] This is drawn from Tim Trainor's presentation at the International Workshop on Integrating Geospatial and Statistical Information, Beijing, China 9-12 June 2014 http://ggim.un.org/meetings/2014-IGSI_Beijing/documents/04_USA_UN_Grid_Admin_Trainor_6_5_14.pdf

[49] Gehlke, Charles E., and Katherine Biehl. "Certain effects of grouping upon the size of the correlation coefficient in census tract material." Journal of the American Statistical Association 29.185A (1934): 169-170.

classification systems, units of measure, sizes, and shapes of administrative boundaries makes it difficult to create a consistent statistical model or measure of population density and land use.

Regional definitions based upon economic, cultural, or land cover/land use, as well as the physical topography of administration areas add another level of variability within countries and between countries and thematic domains. Within countries, there may be various participation or response rates based upon regional biases, or the physical isolation of respondents within specific administrative areas. On a global scale the overriding concern may be the accessibility of administrative areas in remote areas, especially given the overarching mission of the 2030 Agenda of "leaving no-one behind".

Gridded Geographies

The main advantage of using grids as a common geography is comparability. Grids are stable over time and unaffected by frequent administrative boundary changes (for example, when redrawing administrative boundaries to better reflect the population within a certain geographical area)[50]. As most statistical units at the country-level use point-based location frameworks, i.e., addresses and building registers, and the use of an x- and y coordinate is recommended for each statistical unit record, data aggregation and disaggregation algorithms can be applied to any type of area, including grid cells. Grid cells can then be combined to form areas for a specific purpose or study area, such as urban, rural, mountain, and coastal regions etc. Size is the important factor within a grid, and the size of each grid cell must be the same size. A hierarchical system of grids can be created to be proportionate to the size of the study area, from global to local. The smallest building block is point-based georeferenced microdata. Grids can be sized so that you can avoid a "dilution effect" of too large a building block.

Gridded geographies are also becoming a global standard that will allow global comparisons of country-level data to be made. Grids, particularly global grids like a DGGS, can serve as a piece of infrastructure that provides both a location referencing system and a container for data.

The disadvantages of using grid-based statistical areas pertain primarily to the size of the grid cell, projection, and coding systems. The arbitrary nature of defining grid borders and their capacity to cut homogenous features that cannot be further subdivided (such as large buildings) can pose practical problems of data allocation and use. As gridded data can represent a small statistical area, there is also a need for implementers to seriously consider the protection of privacy and the prevention of disclosing identifiable information. This risk is highest if a grid contains a small population. Data suppression in these areas can eliminate essential data that is necessary for government allocations and assistance. The quality of data included in grids is very dependent on the accuracy of the geocoded input data, where errors are difficult to find in grid-based statistics.

[50] Over 20 countries and institutions have begun the transition to a grid-based statistical system for statistical reporting including, Austria, Belgium, Cyprus, Denmark, Estonia, Finland, France, Ireland, Netherlands, Norway, Poland, Republic of Kosovo, Serbia, Slovak Republic, Slovenia, and Sweden and elsewhere.

	Positives	Negatives
Administrative	• Spatial accuracy of data • Field verification • Imagery verification • Geocoding / address verification • Authoritative sources • Local government involvement • Local knowledge • Nesting relationship with other geographic areas • Cadastral boundaries • Data thresholds • Separate land & water area • Response rates • Response options • Response quality • Sample frame • Controls on disclosure • Ready to use within a GIS analysis • Defines places and communities • Can be normalised by population	• Comparability • Boundary changes • Traditional census data collection is becoming more infrequent for countries • Use of non-visible boundaries • Number of different geographic areas • Cartographic considerations / generalisation of boundaries • Varying participation • Irregular sizes • Irregular shapes • Variable density measures • High costs to maintain the data • Legal variation • Regional variation • Topographic variation • Insufficient understanding of micro characteristics inside macro-scale units • Data integration is difficult • There can be differences within a country's administrative geography • Often contains a mix of land uses/geographic units • Not suitable for time-series analyses • Not suitable for many geospatial analysis techniques • Not aesthetically pleasing for cartography
Grids	• Global and local scope—fully scalable • Uniform scale conducive to cross-border studies • Comparability; better suited for NSDI • More attention to problem-oriented science • Can locate people in space with more precision • Good territorial framework for sampling • Can aggregate to different kinds of territorial units • Ready to use with GIS analysis • Easily generated from point-based georeferenced data • Able to see clusters • Easy and cost-efficient to generate • Micro-scale analysis using flexible size grids • Data integration is possible with newer data sources, (i.e. ground-based, imagery, internet) • Stable over time; time-series not affected by admin unit changes • Independent from traditional data collection procedures • Widely used in science and practice • Can include higher resolutions in rural areas • Can become a global standard	• Arbitrary • Can cut through significant features, including features and building that are not able to be split • Disclosure control /cell size • Grid cell sizes in rural areas • Dependant on highly accurate geocoding of units • When merging datasets, there is a need to change from one coordinate system to another before the data is compiled into grids • European terrestrial reference system (ETRS80) is based on Lambert Azimuthal Equal Area coordinate reference system with fixed projection centre; different projections may be needed in other parts of the world • Coding systems [scale intervals vs quadtree solutions] • Due to high data volume, errors are difficult to find and correct • Various grids may be adopted within regions or countries • Areas with dynamic or transient population fluctuations pose numerous complications for regional analysis • Spatial and temporal cross-validation models using multiple sources of geographic, physiographic, and socio-economic data in conjunction with imagery analysis is necessary • The size of the cell differs. Consider the requirements between the size of the cell when data is collected to the size of the cell when data is disseminated. • Can be lower resolution in urban areas

Table 1 Comparison of the Advantages and Disadvantages of Administrative and Gridded Geographies

The Global Fundamental Geospatial Data Themes

Theme	Description
Global Geodetic Reference Frame	The **Global Geodetic Reference Frame** is the framework which allows users to precisely determine and express locations on the Earth, as well as to quantify changes of the Earth in space and time. It is a crucial prerequisite for the accurate collection, integration and use of all other geospatial data.
Addresses	**Addresses** are a structured label, usually containing a property number, a street name and a locality name. Addresses are often used as a proxy for other data themes such as Land Parcels and usually are linkable to geographic coordinates.
Buildings and Settlements	A **Building** refers to any roofed structure permanently constructed or erected on its site, for the protection of humans, animals, things, or the production of economic goods. **Settlements** are collections of buildings and associated features where a community carries out socio-economic activities
Elevation and Depth	The **Elevation and Depth** theme describes the surface of the earth both on land and under a body of water, relative to a vertical datum.
Functional Areas	**Functional Areas** are the geographical extent of administrative, legislative, regulatory, electoral, statistical, governance, service delivery and activity management areas.
Geographical Names	**Geographical Names** provide orientation and identity to places. They are location identifiers for cultural and physical features of the real world, such as regions, settlements, or any feature of public or historical interest.
Geology and Soils	**Geology** is the composition and properties of geologic materials (rocks and sediments) underground and outcropping on the Earth's surface. **Soil** is the upper part of the earth's crust, formed by mineral particles, organic matter, water, air and living organisms.
Land Cover and Land Use	**Land Cover** represents the physical and biological cover of the earth's surface. **Land Use** is the current and future planned management, and modification of the natural environment for different human purposes or economic activities.
Land Parcels	**Land Parcels** are areas of land, cadastral parcels, or areas of the Earth's surface (land and/or water) under common rights (such as ownership or easements), claims (such as minerals or indigenous land) or use.
Orthoimagery	**Orthoimagery** is geo-referenced rectified image data of the Earth's surface, from satellite or airborne sensors.
Physical Infrastructure	The **Physical Infrastructure** theme includes industrial & utility facilities, and the service delivery facilities associated with administrative & social governmental services such as public administrations, utilities, schools and hospitals etc.
Population Distribution	The **Population Distribution** theme covers data for the geospatial distribution of population and its characteristics, as well as how population impacts urbanisation, regional development or sustainability.
Water	The **Water** theme covers the geographic extent and conditions of all water features including rivers, lakes and marine features.
Transport Networks	The **Transport Networks** theme are the suite of road, rail, air and water transport routes and their connectivity.

Figure 6 The 14 Global Fundamental Geospatial Data Themes

Figure 6 lists and describes the 14 Global Fundamental Geospatial Data Themes[51]. Implementing the themes will necessitate the integration of information from NGIAs, NSOs and other data custodians.

Statistical and geospatial quality frameworks

NSOs believe that the high quality of official statistics represents a "… *competitive advantage in a world experiencing a growing trend of instant information which often lacks the necessary proof of quality.*"[52]

For this reason, official statistics have developed a comprehensive quality management framework[53]. As an example, Europe has developed a comprehensive, internationally agreed quality framework, with the European Statistics Code of Practice at its apex[54]. In addition to the institutional environment of official statistics, and the quality of statistical processes, it covers the quality of statistical products. The quality of a statistical product can be defined as fitness-for-purpose and be further broken down into the following dimensions:

- Relevance;
- Accuracy and reliability;
- Timeliness and punctuality;
- Accessibility and clarity; and,
- Coherence and comparability.

The integration of statistical and geospatial information can contribute to all these dimensions, but in particular to the relevance of statistical outputs for the user, and must consider all phases within the statistical process such as the requirement detection and analysis, design and testing, execution, analysis and dissemination[55]. Quality systems in official statistics are typically structured in two levels, one with the Principles and goals, and one with the indicators measuring the achievement of the goals.

As an example, the implementation of the Principle on **accuracy and reliability** with the achievement of the goal: Statistics accurately and reliably portray the reality, can be measured using the following indicators: *Source data, intermediate results and statistical outputs are regularly assessed and validated.* As an example, quality measures of data in a register can include:

Criterion	Indicator	
	Name	Value
Accuracy	Over-coverage – units outside the population	The percentage of units that do not belong to the population
	Under-coverage – missing population units	The percentage of missing population units
	Unjustified repetition of records regarding the same population unit	The percentage of repeated records
	Missing data for variables	The percentage of information characteristics for which values are missing

[51] https://undesa.maps.arcgis.com/apps/Cascade/index.html?appid=4741ad51ff7a463d833d18cbcec29fff
[52] Quality Declaration of the European Statistical System https://ec.europa.eu/eurostat/documents/4031688/8188985/KS-02-17-428-EN-N.pdf/116f7c85-cd3e-4bff-b695-4a8e71385fd4
[53] The UN has developed the fundamental principles of official statistics https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf and UN guidelines for statistical quality https://unstats.un.org/unsd/dnss/qualitynqaf/nqaf.aspx
[54] https://ec.europa.eu/eurostat/web/quality/overview
[55] Colombia's implementation (in Spanish): https://www.dane.gov.co/files/sen/normatividad/NTC_Proceso_Estadistico.pdf

		The percentage of units for which values of specific information characteristics are missing
	Adjustment, imputation	The percentage of adjusted units
		The percentage of adjusted values
		The percentage of supplemented values
	Integration of data from various sources	The accurate matching – the percentage of matched units
		Integration errors – the percentage of inaccurately matched units
		No match – the percentage of non-matched units

Table 2 Quality Principle accuracy applied to data in register

In this regard, they are similar in structure to this statistical geospatial framework, which should facilitate communication among statisticians and decision makers generally. On the other hand, statisticians are usually less familiar with the quality concepts of geospatial information that are relevant to produce geospatially enabled statistics and are also addressed in this framework - in particular, Principles 1 and 3.

For this reason, the communication and exchange on quality should be bi-directional:

- NSOs by implementing this framework should also pay attention to the quality concept set out therein and understand the impact of geospatial quality on their products and processes; and,
- NGIAs who are concerned by this framework and are asked to contribute to its implementation should become familiar with the way NSOs use different data sources, produce official statistics, assess the quality of data sources, production processes and final products and document quality for the end users.

Nevertheless, the overall definition of quality of any product as fit-for-purpose, also holds for geospatial information and the results of its integration with statistics. Similarly, several of the quality Principles of official statistics are also applicable to NGIAs and their products. For statisticians and users, understanding the quality of geospatial information as set out in the GSGF and throughout the NSS is therefore vital. This applies both for its temporary use and integration during statistical production that might be invisible for the end user and for final geospatial statistics products that are designed for a range of users. As for statistics, good quality geospatial data underpins notions of credibility and authority that are also important for the resulting integrated products.

In addition to the more traditional management of statistical and geospatial data quality there are five areas of specific concern regarding the quality of geospatial data, not only for statistics that emerge from digitalisation and data revolution:

1. There is an increasing availability, exchange and use of geospatial data;
2. There is a growing group of users less aware of geospatial data quality;
3. A Geospatial Information System enables the use of geospatial data in a range of applications, without considering the appropriateness of the quality of data;

4. Current Geospatial Information Systems offer few tools for handling geospatial quality; and,
5. There is an increasing distance between those who use the geospatial data (intermediate end users) and those who are best informed about the quality of the geospatial data (producers).

The targets and indicators should be integrated into existing statistical quality frameworks to ensure uptake and use by statisticians and to facilitate their communication. Ultimately, these targets and indicators should be an important part in the forthcoming implementation guide to the GSGF.

Annex C: Further Reading

National and International Frameworks

- INSPIRE[56]
- The Global Fundamental Geospatial Data Themes[57]
- Integrated Geospatial Information Framework[58]
- Australian Statistical Spatial Framework[59]
- GEOSTAT3[60]
- Geographic Information Framework Data Content Standard - FGDC[61]

Accessibility

- World Wide Web Consortium Web Accessibility Initiative[62]
- OGC Spatial Data on the Web Working Group[63]
- U.S. Federal Government Section 508 Accessibility Program[64]
- ISO/IEC Guide for addressing accessibility in standards [65]

Metadata and data format standards

- JavaScript Object Notation "JSON" [66]
- ISO/TC 211 Geographic information /Geomatics (standards catalogue)[67]
- SDMX[68]
- OGC Web Feature Service[69]
- Geopackage [70]
- ISO 19115 Metadata Standard[71]
- OGC Table Joining Service[72]

Privacy and Data Disclosure Prevention

- U.S. Census Bureau – Statistical Disclosure Control[73]
- Australian Bureau of Statistics Confidentiality Series[74]

Resources of the EG-ISGI – available through the EG-ISGI Wiki

- Common Statistical and Geospatial Definitions[75]
- Process and history of the GSGF[76]

---

56 https://inspire.ec.europa.eu/about-inspire/563
57 https://undesa.maps.arcgis.com/apps/Cascade/index.html?appid=4741ad51ff7a463d833d18cbcec29fff
58 http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Part%201-IGIF-Overarching-Strategic-Framework-24July2018.pdf
59 https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Statistical+Spatial+Framework
60 https://www.efgs.info/geostat/geostat-3/
61 http://www.fgdc.gov/standards/projects/framework-data-standard/GI_FrameworkDataStandard_Part0_Base.pdf
62 https://www.w3.org/WAI/fundamentals/accessibility-principles/
63 http://www.opengeospatial.org/projects/groups/sdwwg
64 https://www.section508.gov/
65 https://www.iso.org/standard/57385.html
66 http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf
67 https://www.iso.org/committee/54904/x/catalogue/p/1/u/0/w/0/d/0
68 https://sdmx.org/
69 http://www.opengeospatial.org/standards/wfs
70 https://www.geopackage.org/
71 https://www.iso.org/standard/53798.html
72 http://www.opengeospatial.org/standards/tjs
73 https://www.census.gov/srd/sdc/
74 http://www.abs.gov.au/ausstats/abs@.nsf/mf/1160.0
75 https://unstats.un.org/wiki/display/ISGI/Common+Statistical+and+Geospatial+Definitions
76 https://unstats.un.org/wiki/display/ISGI/The+Process+of+Developing+the+GSGF