



GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

Proposal from the GEOSTAT 3 project

Version 1.0

28 February 2019

Title: GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe - Proposal from the GEOSTAT 3 project

Project: Eurostat ESSnet grant project GEOSTAT 3

Grant agreement number: 08143.2016.002-2016.752

Author: Jerker Moström (editor), Karin Hedeklint, Marianne Dysterud, Erik Engelen, Vilni Verner Holst Bloch, Rina Tammisto, Pieter Bresters, Niek van Leeuwen, Ülke Valgma, Ingrid Kaminger, Anna Sławińska, Ana Santos, Pier-Giorgio Zaccheddu and Arvid Lillethun

It is permitted to copy and reproduce the content in this report. When quoting, please state the source.

© EFGS and Eurostat 2019

Foreword

This report presents the main results from the GEOSTAT 3 project, a Eurostat funded ESSnet grant project conducted from February 2017 until January 2019. The main objective of the project has been to make a proposal for a guide for the harmonised implementation of the Global Statistical Geospatial Framework (GSGF) in Europe. The implementation guide should cover the key aspects of statistical-geospatial integration as set out in the GSGF and its five principles, and adapt them for the European Statistical System and the wider European context. The focus should be on comparability of statistical outputs, harmonisation of geospatial data sources and methodologies, and on interoperability of various data sources and metadata. Furthermore, the project consortium should conduct the work in cooperation with national experts from NSIs and geospatial agencies and with UN-GGIM: Europe through its two working groups.

In order to assess the soundness of the requirements and recommendations proposed in the implementation guide, the project has undertaken a series of practical and technical tests, drawing on the requirement to geo-enable the indicators for the monitoring of Agenda 2030 Sustainable Development Goals (SDGs). A proof of concept has also been researched on automated linking between statistical tables encoded in SDMX and a web mapping service by means of a Table Joining Service (TJS). The results from these practical and technical tests are published in separate reports (*Automated Linking of SDMX and OGC Web Services* and *Testing the Global Statistical Geospatial Framework (GSGF Europe) by calculating a selection of SDG indicators*). The reports can be found at the website of the GEOSTAT 3 project (<https://www.efgs.info/geostat/geostat-3/>).

The project has been coordinated by Statistics Sweden. The project consortium comprised co-partners from seven other countries and three sub-contractors:

- Jerker Moström & Karin Hedeklint, Statistics Sweden
- Marianne Dysterud, Erik Engelen & Vilni Verner Holst Bloch, Statistics Norway
- Rina Tammisto & Marja Tammilehto-Luode, Statistics Finland
- Pieter Bresters & Niek van Leeuwen, Statistics Netherlands
- Ülle Valgma, Statistics Estonia
- Ingrid Kaminger, Statistics Austria
- Anna Sławińska, Statistics Poland
- Ana Santos, Statistics Portugal
- Pier-Giorgio Zaccheddu, German Federal Agency for Cartography and Geodesy (BKG)
- Arvid Lillethun, Norwegian Mapping Authority (Kartverket)
- Marina Backer Skaar, MD Mapping

The project consortium would like to thank all those who have contributed with comments and input through the course of the work, most significantly the members of UN-GGIM: Europe Working Group on Core Data and Data integration, for their reviews and productive critique. The consortium would also like to thank colleagues and peers from the geospatial and statistical communities around the globe that have contributed in a less formal way through questions, suggestions and discussions along the way.

Stockholm, February 28, 2019

List of terms and acronyms

API: Application programming interface
CC-BY: Creative Commons Attribution
CSDA: Common Statistical Data Architecture
CSPA: Common Statistical Production Architecture
DCAT: Data Catalog Vocabulary
DDI: Data Documentation Initiative
DGGs: Discrete Global Grid System (OGC standard)
DWG: Domain Working Groups
EBM: EuroGeographics EuroBoundaryMap
EFGS: European Forum for Geography and Statistics
EFTA: European Free Trade Association
EG-ISGI: UN Expert Group on the Integration of Statistics and Geospatial Information
EGM: EuroGeographics EuroGlobalMap
EIF: European Interoperability Framework
ELS: European Location Service
ERM: EuroGeographics EuroRegionalMap
ESDI: European Spatial Data Infrastructure
ESS: European Statistical System
ESSnet: A network of several ESS organisations aimed at providing results that will be beneficial to the whole ESS
ETRS89: European Terrestrial Reference System 1989
EuroDEM: EuroGeographics Euro(digital elevation model)
GCM: Generic Conceptual Model of INSPIRE
GDPR: General Data Protection Regulation
GFM: General Feature Model
GSGF: Global Statistical Geospatial Framework
GSBPM: Generic Statistical Business Process Model
GSIM: Generic Statistical Information Model
HLG-MOS: High-Level Group for the Modernisation of Official Statistics
IGIF: Integrated Geospatial Information Framework
INSPIRE: Infrastructure for Spatial Information in the European Community (INSPIRE Directive 14.03.2007)
ISO: International Organization for Standardization
IETF: Internet Engineering Task Force
ITRS: International Terrestrial Reference System
LAU: Local Administrative Unit
LOD: Linked Open Data
MoU: Memorandum of Understanding
NMCA: National Mapping and Cadastral Agency
NSDI: National Spatial Data Infrastructure
NSI: National Statistical Institute
NUTS: Nomenclature of territorial units for statistics
OGC: Open Geospatial Consortium
PSI: European legislation on the re-use of public sector information
RDF: Resource Description Framework
SDC: Statistical Disclosure Control
SDI: Spatial Data Infrastructure
SDMX: Statistical Data and Metadata eXchange
SDG: UN Sustainable Development Goals
SOA: Service Oriented Architecture
TERCET: TERritorial Classifications and Typologies
TJS: Table Joining Service
UNECE: The United Nations Economic Commission for Europe
UN GGIM: United Nations Committee of Experts on Global Geospatial Information Management

UN-GGIM: Europe: European (regional) committee of the United Nations Initiative on Global Geospatial Information Management.

W3C: World Wide Web Consortium

WFS: Web Feature Service

WMS: Web Map Service

Executive summary

In response to the growing need to add the “where” dimension in public information and statistics in general and in the monitoring of the SDGs in particular, the statistical and geospatial communities have a common task to build frameworks that support the production of relevant, accurate and timely information to inform evidence-based decision-making on all levels of society. The international statistical and geospatial communities recognised this challenge and responded by establishing the UN EG-ISGI to develop a Global Statistical Geospatial Framework (GSGF). Its five principles were adopted in August 2016 and it should act as a bridge between the statistics and geospatial world.

The European implementation guide (GSGF Europe) proposed by the GEOSTAT 3 project is intended as an enhancement to the global guidance, addressing the regional specifics of Europe. It aims to be more specific on the “how” to provide regional guidance on what elements should be available in countries to represent a meaningful GSGF. The implementation guide covers the key aspects of statistical-geospatial integration as set out in the principles of the GSGF and includes an adaption and enhancement to suit the European context and in particular the European Statistical System ESS. The global, generic descriptions of each principle of the GSGF are discussed and interpreted for the European context together with some requirements and recommendations set out for its implementation in Europe. The focus is on comparability of statistical outputs, harmonised geospatial data sources and methodologies, and on interoperability of various data sources and metadata including technical issues related to the implementation of the GSGF in Europe.

In order to utilise and materialise the potential of the GSGF, the GSGF Europe is targeting three major communities, or groups of stakeholders, and their different roles in terms of data provision and data integration: the geospatial community, the statistical and the administrative data communities. INSPIRE and the ESS constitute the institutional pillars of the recommendations as they provide rational infrastructures and mechanisms (legal, technical, collaborative and financial) to deploy a harmonised implementation of the GSGF Europe, both for ESS Member States and non-members.

Both in the GSGF and its European implementation guide, there is a specific focus on issues related to the enduring sustainability of the geospatial information management in a nation. This means that particular attention is given to longer-term financial sustainability, multi-stakeholder approaches, capacity and capability development, and innovation and communication.

Chapter 1 of this report presents the background, aim and scope of the project and introduces the Global Statistical Geospatial Framework.

Chapter 2 outlines the principles of the GSGF in relation to its proposed implementation in Europe. Each section starts with a short generic abstract derived from the Background Document on Proposal for a Global Statistical Geospatial Framework. The global, generic descriptions are accompanied by a short discussion on the interpretation of the principle for European ground. Under each of the five principles, follows a number of requirements that are considered crucial to start implementing the framework in Europe (the “What”). Each of these requirements connects to a set of more detailed recommendations to achieve the requirements (the “How”). In particular the recommendations are in their majority based on practical success stories from individual Member States. Consequently the

requirements and recommendations are linked to national good practices contained in Annex 2 to this report.

Chapter 3 identifies a number of additional generic and non-technical recommendations relating to governance and the process of implementing the GSGF and monitoring of its progress.

Chapter 4 contains concluding remarks and a few proposals on how countries could approach the recommendations in order to start the implementation process.

Content

1	Introduction.....	11
1.1	Vision for a better data integration.....	11
1.2	Background and aim of the project.....	12
1.3	The Global Statistical Geospatial Framework (GSGF).....	12
1.4	Why a European implementation guide?.....	14
1.5	Other stakeholders.....	16
1.6	An implementation guide for the ESS or for Europe?	17
1.7	Who is concerned by the implementation guide?	18
1.8	Technical and methodological guidance in focus.....	19
2	Requirements and recommendations for implementation of the GSGF in Europe.....	21
2.1	Principle 1: Use of fundamental geospatial infrastructure and geocoding of statistical information.....	23
	Global definition.....	23
	Relation with other principles.....	23
	Discussion.....	23
	Requirement 1.1 - Use data from National Spatial Data Infrastructures.....	25
	Requirement 1.2 - Use point-based location data for geocoding.....	27
	Requirement 1.3 - Build formal working relationships on institutional agreements	29
2.2	Principle 2: Geocoded unit record data in a data management environment	31
	Global definition.....	31
	Relation with other principles.....	31
	Discussion.....	31
	Requirement 2.1 - Build an effective and secure data management environment.....	32
	Requirement 2.2 - Store location only once.....	33
	Requirement 2.3 - Ensure consistency and quality of geocoding results	34
	Requirement 2.4 - Develop consistent approaches to manage non-matching data	36
	Requirement 2.5 - Use point-of-entry validation in collection of administrative or statistical data	36
2.3	Principle 3: Common geographies for production and dissemination of statistics	39
	Global definition.....	39
	Relation with other principles.....	39
	Discussion.....	39

Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies.....	40
Requirement 3.2 - Improve maintenance of the European framework of statistical geographies	42
Requirement 3.3 - Consolidate use of existing statistical grid systems and explore the potential of evolving global grid systems	45
2.4 Principle 4: Statistical and geospatial interoperability – Data, Standards and Processes	48
Global definition.....	48
Relation with other principles.....	48
Discussion.....	48
Requirement 4.1 - Improve geospatial workflows within statistical production.....	49
Requirement 4.2 - Enable data integration through consistent semantics and concepts across domains.....	51
Requirement 4.3 - Publish data once and leave it at its source to be reused many times.....	51
Requirement 4.4 - Increase use of services for merging geography and statistics.....	53
Requirement 4.5 - Explore the potential of Linked Open Data for increased interoperability.....	56
2.5 Principle 5: Accessible and usable geospatially enabled statistics.....	58
Global definition.....	58
Relation with other principles.....	58
Discussion.....	58
Requirement 5.1 - Implement clear and simple data licensing policies	59
Requirement 5.2 - Use service-oriented data portals supporting dynamic integration of data...	60
Requirement 5.3 - Define clear national and European rules to ensure protection of privacy....	61
Requirement 5.4 - Facilitate data search and use through cataloguing and improved guidance	62
3 Proposed procedure for the ESS on the implementation of the GSGF.....	65
4 Conclusions.....	66
5 References.....	68
Annex 1. List of requirements and recommendations.....	70

1 Introduction

1.1 Vision for a better data integration

European and global sustainable development programs increasingly require reliable and relevant information in terms of higher spatial and temporal resolution and increased abilities for spatial and thematic disaggregation. The UN Agenda 2030 and its Sustainable Development Goals (SDGs) is pushing for a closer integration of statistical and geospatial information. The work on achieving and monitoring the SDGs poses substantial challenges for the statistical and geospatial communities but it also offers a unique opportunity to demonstrate the power of statistical-geospatial data integration across a wide range of themes.

Another driver for statistical-geospatial data integration is the upcoming 2021 round of population censuses. The requirement to develop an integrated statistical and geospatial solution for the 2021 censuses has been repeatedly expressed in the UN context, e.g. by the UN Committee of Experts on Global Geospatial Information Management (UN-GGIM) in its report from 2013 and the UN Expert Group on the Integration of Statistics and Geospatial Information (EG-ISGI)¹.

In response to the growing need to add the “where” dimension in public information and statistics in general and in the monitoring of the SDGs in particular, the statistical and geospatial communities have a common task to build frameworks that can support the production of relevant, accurate and timely information to inform evidence-based decision-making on all levels of society.

The international statistical and geospatial communities recognised this challenge and responded by establishing the UN EG-ISGI to develop a Global Statistical Geospatial Framework (GSGF). At the Sixth Session of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), held in August 2016, the five principles of the GSGF were adopted². The GSGF should act as a bridge between statistics and geospatial information, between statistical institutes and geospatial agencies, and between statistical and geospatial standards, methods, workflows and tools.

One of the key areas of the European Statistical System (ESS³) Vision 2020⁴ is to harness new data sources comprising Big Data, administrative data and geospatial data. Using data from a range of sources, for multiple purposes, requires their integration into a common reference system of harmonised concepts, but also common location and temporal framework. Therefore, users have not only increased their demand for location information but they also require simpler integration of data across various data sources used in their analyses.⁵ Along with time, location and space are universal and well-defined concepts and, hence, can be used to integrate data from a wide range of topics.

¹ UNECE 2016b

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

³ <http://ec.europa.eu/eurostat/web/ess/about-us>

⁴ ESS Vision 2020.

<http://ec.europa.eu/eurostat/documents/7330775/7339647/ESS+vision+2020+brochure/4baffcaa-9469-4372-b1ea-40784ca1db62>

⁵ UNECE 2016a

On the European level there is a work ongoing towards integration of INSPIRE⁶ and SDMX⁷ data infrastructures for the 2021 population and housing census⁸. Also in the ESS context the 2021 population and housing census is the key driver for geo-enabling statistics and integrate geospatial information into statistical production.

Within the National and International Statistical Systems, there is a move towards an increased use of administrative data and registers for census purposes. In parallel, many countries have launched national geospatial strategies to geocode administrative records in order to support data integration.

1.2 Background and aim of the project

The ESS Committee in November 2013 requested to "... develop a strategy for a harmonised approach to geo-referenced statistics within the ESS ..." ⁹ As a response to this request, Eurostat launched the GEOSTAT 2 project in 2014, which proposed recommendations for a point-based foundation for statistics. The scope of the project also included an evaluation of the Generic Statistical Business Process Model (GSBPM) in terms of its fitness for purpose to describe the management of geospatial information in the production of statistics. ¹⁰

The main motivation behind the GEOSTAT 2 was the 2021 census but, with the arrival of the GSGF, the opportunity came up to fulfil the ESS ambition to broaden the scope of statistical-geospatial integration. Accordingly, in 2016, Eurostat launched an ESSnet grant to develop a consistent proposal for a harmonised implementation of the GSGF in the ESS countries.

The focus should be on comparability of statistical outputs, harmonised geospatial data sources and methodologies, and on interoperability of various data sources and metadata. The implementation guide should cover the key aspects of statistical-geospatial integration as set out in the principles of the GSGF and adapt them to the European and in particular the ESS context. Furthermore, the project consortium should conduct the work in cooperation with national experts from NSIs and geospatial agencies and with UN-GGIM: Europe through its two working groups.

1.3 The Global Statistical Geospatial Framework (GSGF)

The GSGF is a high-level framework consisting of five guiding principles that are considered essential for integrating geospatial and statistical information (Figure 2). This means that it is not intended to provide a detailed implementation instructions but rather guidance on what should be available in countries, leaving a lot of flexibility on the "how".

The GSGF provides the international statistical and geospatial community with a common framework to connect socio-economic and environmental data to appropriate locations, and improves the accessibility and usability of this geospatially-enabled data. Figure 1 below highlights the importance

⁶ Infrastructure for Spatial Information in Europe. <https://inspire.ec.europa.eu/>

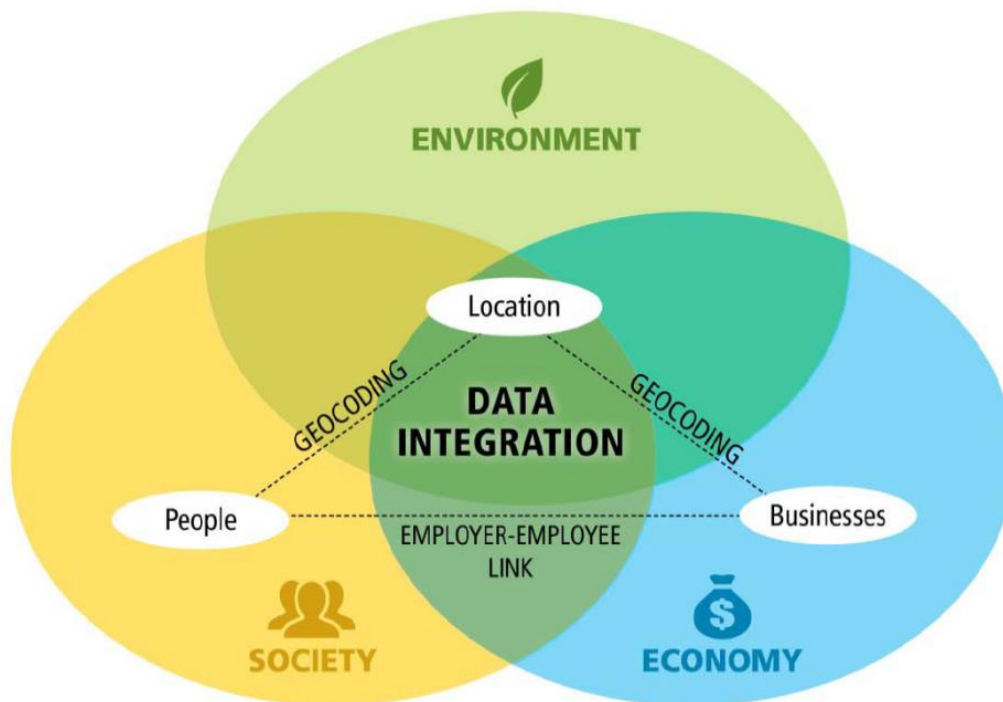
⁷ Statistical Data and Metadata eXchange. <https://sdmx.org/>

⁸ <https://statswiki.unece.org/display/Geo17/Workshop+on+Integrating+Geospatial+and+Statistical+Standards+2017>

⁹ WP DOC NR D/GIS/103. https://circabc.europa.eu/sd/a/8125fe32-6401-43b1-ab62-f380f4f84abb/D_GIS_103%20GISCO%20WP%20Mandate%20Revision_Feb%202015.doc

¹⁰ EFGS/GEOSTAT 2 2017. A Point-based Foundation for Statistics - Final report from the GEOSTAT 2 project: <https://www.efgs.info/geostat/geostat2/>

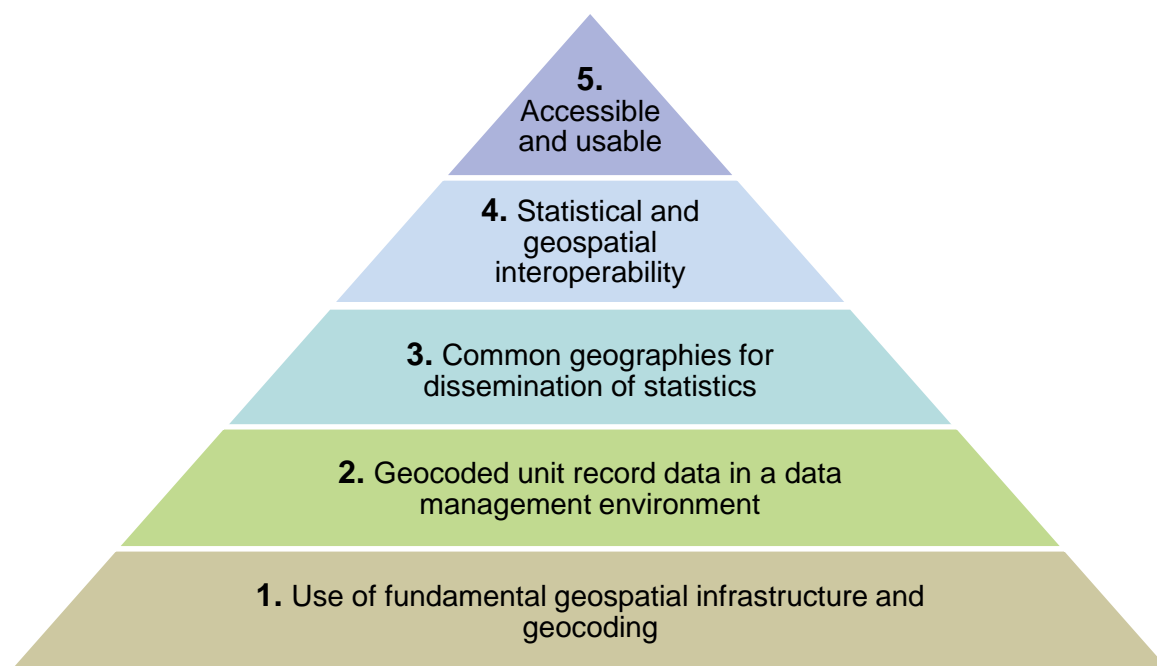
of location information as a tool, integrating the following three domains: the society, the economy and the environment.



SOURCE: Australian Bureau of Statistics (ABS)

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

In its first version, the GSGF focuses on the socio-economic and environmental statistical data traditionally produced by NSIs and their integration with location data (see Figure 1). The UN EG-ISGI will continue to develop the GSGF and monitor its implementation with a review point in the near future. The intention of the Expert Group is for the GSGF to be inclusive of all statistical and geospatial data and to possibly extend it to other relevant public data, and to enable and encourage NSIs and geospatial agencies to look beyond traditional data sources and methods. The UN EG-ISGI intends to submit an updated GSGF document to the UN-GGIM Committee of Experts in 2019 and later to the Statistical Commission.



SOURCE: Australian Bureau of Statistics (ABS) / UN-GGIM, illustration by Statistics Sweden

Figure 2: The Global Statistical Geospatial Framework (GSGF)

Each of the five guiding principles in the GSGF is defined by a set of objectives and is supported by international, regional and, where applicable, domestic standards and good practices. At the Sixth Session of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), held in August 2016, the principles of the Global Statistical Geospatial Framework were adopted. The background document presented at the GGIM6 explains this framework in further detail¹¹.

The standards and good practices that will provide a more detailed global guidance for countries implementing the GSGF are still under consideration by the UN EG-ISGI, and will be brought to the UN Statistical Commission and UN-GGIM for consideration in the near future. The GEOSTAT 3 project has followed the work of UN EG-ISGI closely to ensure consistency between the European implementation guide and the global guidance materials currently under preparation.

A description of each principle of the GSGF is presented briefly in chapter 2, together with a discussion on the suggested implication for the European context and the requirements and recommendations set out for its implementation.

1.4 Why a European implementation guide?

The adoption of the GSGF is believed to be an important step towards more coordinated geospatial information practices and better integration of statistical and geospatial information globally. However, as mentioned above, the GSGF in itself does not intend to provide a detailed plan or design for its implementation, rather leaving a lot of flexibility on the “how”. In contrast, the European implementation guide aims to be more specific on the “how” to provide guidance to the ESS and to

¹¹ <http://ggim.un.org/meetings/GGIM-committee/documents/GGIM6/Background-Paper-Proposal-for-a-global-statistical-geospatial-framework.pdf>

NSIs and geospatial agencies on what elements should be available in countries to represent a meaningful statistical geospatial framework. The implementation guide does not propose the introduction of a new independent framework with its own institutional structures, but rather to streamline the existing ones and pinpoint the synergies and links between existing initiatives for a common cause.

The GSGF Europe does not replace the global guidance currently under preparation by the UN EG-ISGI (to be published in the near future). It is rather intended as an enhancement to the global guidance, addressing the regional specifics of Europe. The work of EG-ISGI has been monitored closely by the project to assure compliance between global and European guidance and a few members of the GEOSTAT consortium are also members of the EG-ISGI. Furthermore, the GSGF Europe does not override the need for countries to define their own guidelines for implementation on national level.

The ultimate objectives of the GSGF Europe are to:

- Equip countries with a manual for implementing the GSGF that takes into account European specifics;
- Consolidate existing integration, standardisation and data sharing efforts for the purposes of statistics into a coherent framework;
- Harmonise and standardise methods for the integration of statistical and geospatial information within the European statistical system (ESS)/EU;
- Modernise the ESS and increase efficiency and flexibility in terms of statistical output;
- Ensure coherence of the GSGF Europe with the Modernisation of Official Statistics programme led by UNECE;
- Provide a better foundation for collaboration between the statistical and geospatial communities as well as between National Spatial Data Infrastructures (NSDIs) in providing society with more and better data for evidence based decision-making.

The relative regional homogeneity between countries within Europe, concerning National Spatial Data Infrastructures and National Statistical Systems, allows for a coordinated implementation of the GSGF in Europe. In order to do this, the GSGF Europe needs to expand the content of the GSGF and go beyond the generic global principles and guidance, by providing recommendations building on the specific European situation.

In particular, the implementation guide rests upon two major cornerstones and drivers for harmonisation and implementation, as foundation for the statistical-geospatial integration in Europe:

- The achievement undertaken for the availability and harmonisation of geospatial information through the implementation of the INSPIRE directive¹² and the National Spatial Data Infrastructures (NSDIs) set up by EU Member States;
- The existing and well-established structure for collaboration and harmonisation of European official statistical data provided through the ESS.

¹² Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE); <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002>

In essence, INSPIRE and the ESS constitute the institutional pillars of the recommendations as they provide rational infrastructures and mechanisms (legal, technical, collaborative and financial) to deploy a harmonised implementation of the GSGF in Europe.

1.5 Other stakeholders

Besides INSPIRE and the ESS, a number of other initiatives and bodies of European collaboration, not restricted to the EU/EFTA area, contribute to the European statistical-geospatial ecosystem.

An active regional UN-GGIM committee has also been set up in Europe, UN-GGIM: Europe¹³, conducting actions to demonstrate the benefits of authoritative, trusted geospatial information with focus on the achievement and monitoring of the Agenda 2030 SDGs. UN-GGIM: Europe has two dedicated work groups on Data Integration and Core Data that build on and have contributed to the development of the GSGF Europe.

Eurostat and the ESS, through their cooperation and development programmes carry out numerous activities and projects in non-EU countries to align their statistical system with the ESS, e.g. when preparing candidate countries for accession to the EU. An implementation guide for the GSGF that takes into account the specifics of the ESS and EU legislation will be a useful tool for Eurostat and the ESS to design and monitor development programmes in the use of geospatial information for statistics.

UNECE, which covers most of the European non-EU Member States works closely together with the ESS on modernising statistical methodology and statistical systems through the ModernStats programme which also looks into the integration of statistics and geospatial information.

EuroGeographics¹⁴ is an important association for voluntary collaboration between geospatial agencies for the development of the European Spatial Data Infrastructure. EuroGeographics and its members are currently building the European Location Services (ELS) and by that, aim to provide a single access point for international users of harmonised, pan-European, authoritative geospatial information and services.

Last but not least, the European Forum for Geography and Statistics¹⁵ (EFGS) provides a unique voluntary network of experts for knowledge exchange and harmonisation on geospatial statistics.

EuroGeographics, UN-GGIM: Europe and the EFGS have all been involved and consulted during the process of drafting the proposal.

All these initiatives work on the integration of statistical and geospatial information in Europe, fully or at least partially. It is therefore essential that they move in the same direction and towards the same objectives. This implementation guide together with the GSGF will help all stakeholders to achieve harmonisation and coordination.

¹³ <http://un-ggim-europe.org/>

¹⁴ <http://www.eurogeographics.org/>

¹⁵ <http://www.efgs.info/>

1.6 An implementation guide for the ESS or for Europe?

A question that has been raised several times during the project is how the implementation guide applies to non-EU/EFTA countries within geographic Europe. These countries are not bound by INSPIRE legislation and do not fully participate in the European Statistical System (ESS).

The promotion of EU methodologies and practices to non-EU member states is common in the European statistical community and should not be understood as ignoring their national sovereignty. Rather, starting from the ESS has been an effective way of achieving harmonisation which is essential for international comparisons and therefore in the interest of the global statistical community as a whole.

Also very few requirements and recommendations proposed in the implementation guide rely on EU/EFTA membership as such, though the legal and technical frameworks within the union can be effective drivers to enforce implementation. The INSPIRE legislation applies only within the European Union, yet its implementing rules, guidance documents, terminology and data models are recognised as good practice in candidate and potential candidate countries as well as in other EU neighbourhood countries. In fact this was the rationale for UN-GGIM: Europe to harvest European Core Data candidates from the INSPIRE annexes¹⁶.

Also the coordination and harmonisation efforts within ESS have impact beyond the ESS Member States. A good example is the production of the 2011 GEOSTAT population grid that was fuelled by technical and methodological development under the ESS but comprised contributions from several non-ESS countries.

¹⁶ UN-GGIM: Europe 2016.

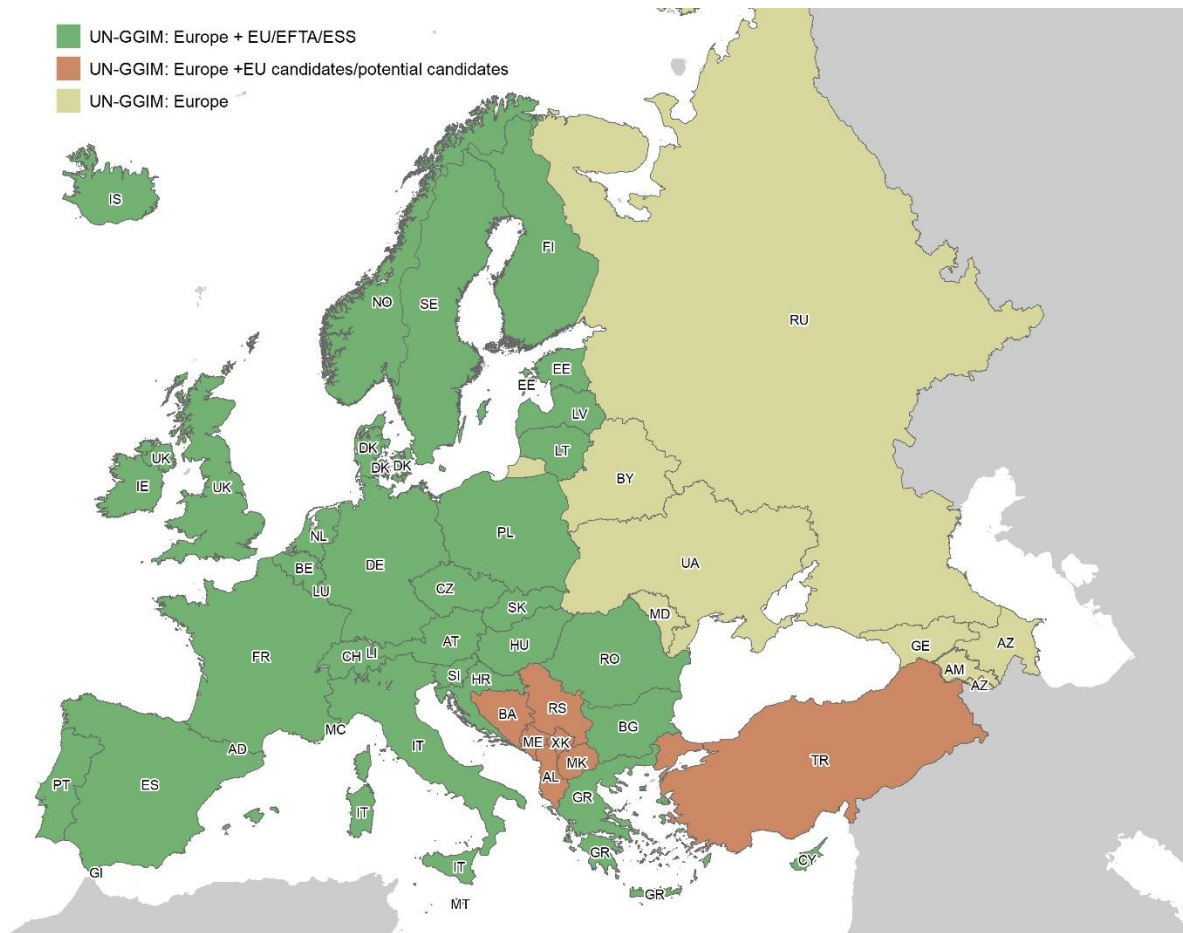


Figure 3: Geographic Europe and the status of countries in relation to EU/EFTA & ESS membership, EU candidate countries and UN-GGIM: Europe

In essence, both INSPIRE and the ESS can be considered as catalysts for technical and methodological development and for coordination needed for a harmonised implementation of the GSGF in Europe, both for EU Member States and non-EU members.

1.7 Who is concerned by the implementation guide?

The statistical and geospatial communities have a common task to build frameworks that supports the production of relevant, accurate and timely information to inform evidence-based decision-making on all levels of society.

In order to utilise and materialise the potential in statistical data, geospatial data and administrative data, a common understanding and working arrangements between communities is needed. The GSGF is targeting three major communities, or groups of stakeholders, and their different roles in terms of data provision and data integration:

- Statistical data community
 - NSIs or other public institutions responsible for the production of official statistics;
 - European institutions responsible for governance and coordination of the European Statistical System (Eurostat);
 - European bodies and initiatives for statistical collaboration and cooperation (UNECE).
- Geospatial data community

- National geospatial agencies or other public institutions responsible for the production and provision of authoritative geospatial data;
- National geospatial agencies or other public institutions responsible for governance and coordination of National Spatial Data Infrastructures;
- European institutions involved in INSPIRE regulation, its implementation and monitoring (European Commission, JRC, EEA etc.);
- European bodies and associations for geospatial data collaboration and cooperation (UN-GGIM: Europe, EuroGeographics).
- Administrative data community
 - National public institutions responsible for administrative data collection and maintenance of public administrative data repositories (population registries, land registries, tax authorities, business registries etc.);
 - European institutions responsible for legal frameworks for administrative data, e.g. the PSI directive¹⁷ (European Commission).

Due to the different roles and responsibilities of the three communities in relation to the GSGF Europe, they are not equally concerned by the different principles and the proposed requirements and recommendations.

Providing the fundamental geospatial infrastructure of Principle 1 is a task mainly for the geospatial community, but the statistical community has a role too, in defining requirements for data models and services.

Principle 2 is typically resting more on the statistical and administrative side of data communities, but requires participation from the geospatial community.

Principle 3 requires a shared burden between the geospatial and statistical communities but not so much the administrative data community.

Principle 4 is targeting all three communities but the statistical and geospatial communities have a particular role in defining standards and measures for increased interoperability in the data ecosystem.

Lastly, Principle 5 is also a shared concern for the statistical and geospatial community but may entail different tasks for respective community.

1.8 Technical and methodological guidance in focus

The GSGF Europe is primarily focusing on technical and methodological issues related to the implementation of the GSGF rather than issues related to governance and policy. That is not to say governance and policy are of less importance.

However, there are already other, more strategic initiatives developing guidance, in parallel to the GEOSTAT 3 project, with a main focus on these issues.

1. The Integrated Geospatial Information Framework (IGIF) is a United Nations endorsed Framework that was developed in collaboration between the United Nations and the World

¹⁷ <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>

Bank, originally to provide a basis and guidance for lower to middle income countries to reference when developing and strengthening their national and sub-national arrangements in geospatial information management and related infrastructures. However, as the IGIF has evolved, it has become apparent that many high income and developed countries will also significantly benefit, at least partially from its integrative and inclusive strategic nature.¹⁸

There is a specific focus in the IGIF on issues related to the enduring sustainability of geospatial information management in a nation. This means that particular attention is given to longer-term financial sustainability, multi-stakeholder approaches, capacity and capability development, and innovation and communication.

2. In a European context, the Executive Committee of UN-GGIM: Europe has identified a list of issues to be addressed nationally and internationally. Among these issues are strengthening of institutions and governance as well as improved communication with stakeholders considered top priority. UN-GGIM: Europe Work Group on Data Integration is currently preparing a draft paper called *The integration of statistical and geospatial information – a call for political action in Europe* presenting recommendations for coordinated action in all countries¹⁹. These recommendations offer complement to the strategic level of data integration.

¹⁸ UN-GGIM & World Bank 2018

¹⁹ UN-GGIM: Europe 2019

2 Requirements and recommendations for implementation of the GSGF in Europe

This main chapter outlines the elements of the GSGF in relation to its proposed implementation in Europe. Each section starts with a short generic abstract derived from the Background Document on Proposal for a Global Statistical Geospatial Framework²⁰. The global, generic descriptions are accompanied by a short discussion on the interpretation of the principle for European ground.

Under each of the five principles (P1-P5), follows a number of requirements that are considered crucial to start implementing the framework in Europe (the “*What*”). Each of these requirements connects to a set of more specific recommendations to achieve the requirements (the “*How*”). In particular the recommendations are in their majority based on practical success stories from individual Member States. Consequently, some recommendations are linked to national good practices contained in Annex 2 to this report. There are also a few references to cases which are not country specific.

Many good practice cases are mapped to more than one recommendation. In its current state, the collection of good practice cases does not aspire to be complete or even to systematically cover all the recommendations. However, this collection of cases could hopefully grow in the near future, to provide a solid foundation of good practices for implementation of the GSGF, not only for Europe but also globally.

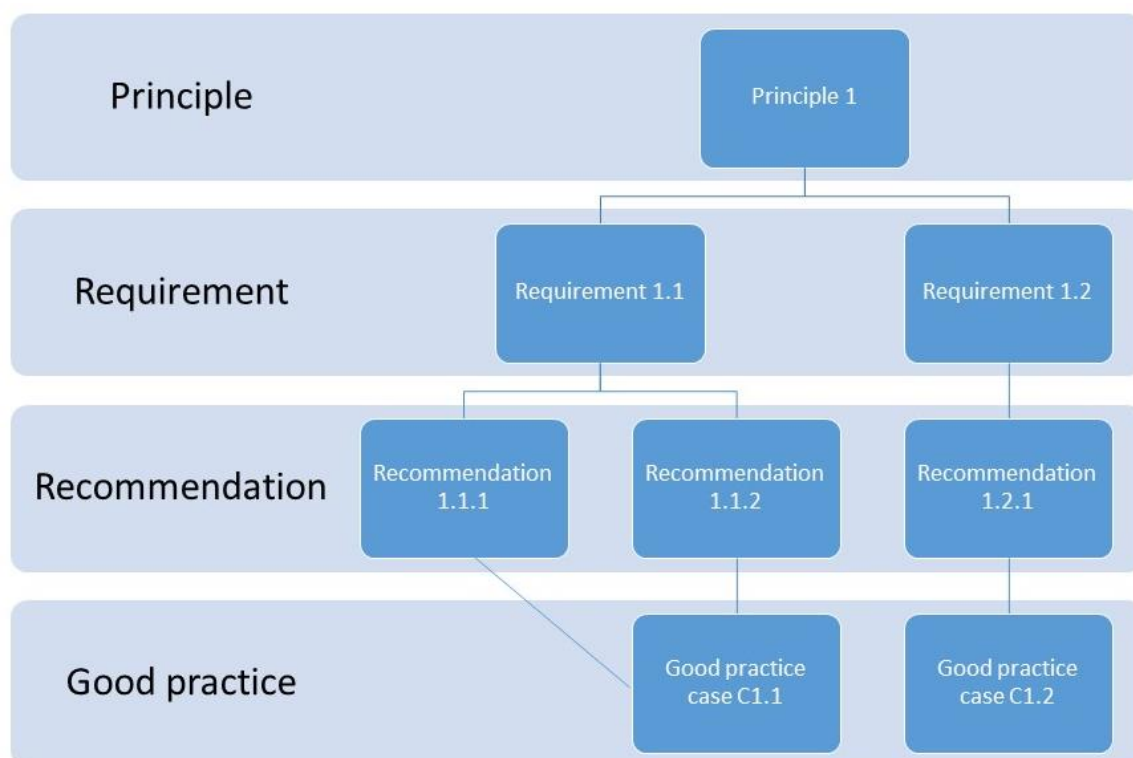


Figure 4: The structure of the implementation guide

²⁰ http://ggim.un.org/ggim_20171012/docs/meetings/GGIM6/Background-Paper-Proposal-for-a-global-statistical-geospatial-framework.pdf

The number of recommendations under each requirement and the level of concreteness vary depending on the complexity of the requirement to be achieved. In addition, the temporal scope of the recommendations may vary substantially. Some recommendations can, or should, be implemented as soon as possible whereas others should be seen in an intermediate or long-term perspective and may require further work to become operational. In essence, an immediate recommendation refers to actions to be carried out during the timespan from now until the upcoming round of census 2021. Intermediate recommendations refer to actions to be undertaken in a 1 to 5 year perspective, whereas long-term recommendations may need up to ten years to realise.

A complete list of all requirements, recommendations and the proposed time frame for their implementation is presented in Annex 1.

2.1 Principle 1: Use of fundamental geospatial infrastructure and geocoding of statistical information



Global definition

The Global Framework requires a common and consistent approach to establishing the location and a geocode for each unit in a dataset, such as a person, household, business, building or parcel/unit of land. A corresponding record of the relevant time or date for each instance of location information recorded should also be associated with each unit record.

The goal of this principle is to obtain a high quality, standardised physical address, property or building identifier, or other location description, in order to assign accurate coordinates and/or a small geographic area or standard grid reference to each statistical unit (i.e. at the microdata level). Time and date stamping these locations will place the unit both in time and in space. An alternative approach to geocoding for recording location is to use direct or indirect capture of coordinates (e.g. from GPS and maps respectively) from field work. 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 from National Spatial Data Infrastructures or other nationally agreed sources. These processes are generally referred to as geocoding.

Relation with other principles

Principle 1 is the foundational principle on which the rest of the principles of the GSGF, and in particular Principle 2, are built. In return, Principle 4 has a strong impact on Principle 1, as use of standards, data models and harmonisation of data content is a crucial in building the fundamental geospatial infrastructure.

Discussion

Principle 1 provides strong support for a point-based foundation for statistics in Europe, based on fundamental data from National Spatial Data Infrastructures (NSDIs).

At the global level the UN-GGIM has published a recommendation for fundamental data expected to be available in each nation²¹. Each category represents a theme consisting of several spatial data sets. The concept of Fundamental data is important in addressing data needs at a strategic level.

²¹ UN-GGIM 2018

In an EU context, the fundamental geospatial infrastructure requested by the GSGF is largely synonymous with the efforts made on the implementation of the INSPIRE directive. INSPIRE sets out the framework for sharing interoperable geospatial datasets and services and the deployment of a network of web services allowing to document, view and download the geospatial datasets. The main goal is to allow full accessibility and sharing of the national geospatial datasets owned by the public sector. Member States are obliged to establish NSDIs, building on international and national standards. The geospatial agencies shall guarantee well-established NSDI as the fundamental base on which the other four principles of the GSGF are built. This NSDI should implement INSPIRE in Europe and consider the requirements of the statistical community. The interoperability of geospatial data and services within a NSDI relies on the internationally agreed ISO 19100 family and the respective OGC standards and processes occasionally supported with feasible standards by W3C²², IETF²³ and other international standardisation bodies.

Though being the framework for the deployment of an interoperable geospatial infrastructure in Europe, the INSPIRE directive is limited to cover the discovery, accessibility and sharing of the datasets. This might implicate gaps and heterogeneity in the implementation between Member States, e.g. due to the many voidable attributes. To facilitate progress and harmonisation on the content-side and other data quality aspects such as scales, density, etc., UN-GGIM: Europe launched the Core data concept.²⁴ The Core data concept is an adaptation of the global Fundamental data concept, with slightly other themes and greater ambition to harmonise quality and other aspects of the delivery of geospatial data. The goal is to fulfil the main user requirements that are common to many countries and many use cases, above all for delivering data for analysing, achieving and monitoring SDGs.

UN-GGIM Fundamental data	UN-GGIM: Europe Core data	GEOSTAT point-based foundation
Geographical names	Geographical names	-
Addresses	Addresses	Addresses
Functional areas	Administrative Units, Statistical Units + Area management	-
Buildings and settlements	Buildings	Buildings
Land parcels	Cadastral parcels	Cadastral parcels
Transport networks	Transport networks	-
Elevation/Depths	Elevation	-
Population distribution	-	-
Land cover/land use	Land cover/land use	-
Geology and soils	-	-
Physical infrastructure	Basic services, utility and governmental services	-
Water	Hydrography	-
Orth imagery	Orth imagery	-

Table 1: Global Fundamental data, European Core data and reference data for a point-based foundation of statistics

The themes for this European initiative are using the thematic terms of INSPIRE and are closely linked to the INSPIRE guidance and implementation activities. Recommendations for content of Core data

²² World Wide Web Consortium (<http://www.w3.org/>)

²³ Internet Engineering Task Force (<https://www.ietf.org/>)

²⁴ <http://un-ggim-europe.org/content/wg-a-core-data>

will extend the INSPIRE data specifications by defining the priorities on the core content that is encouraged to be made available in Europe.

Table 1 above, illustrates the linkage between the global concepts of Fundamental data and the European Core data along with the reference data identified by the GEOSTAT projects (2 and 3) to constitute the basis for a point-based foundation for statistics.

The GEOSTAT 2 project identified three different tiers of information, reflecting that geospatial data can be used either as infrastructure data or as data to create statistical content or as a combination of both purposes²⁵.

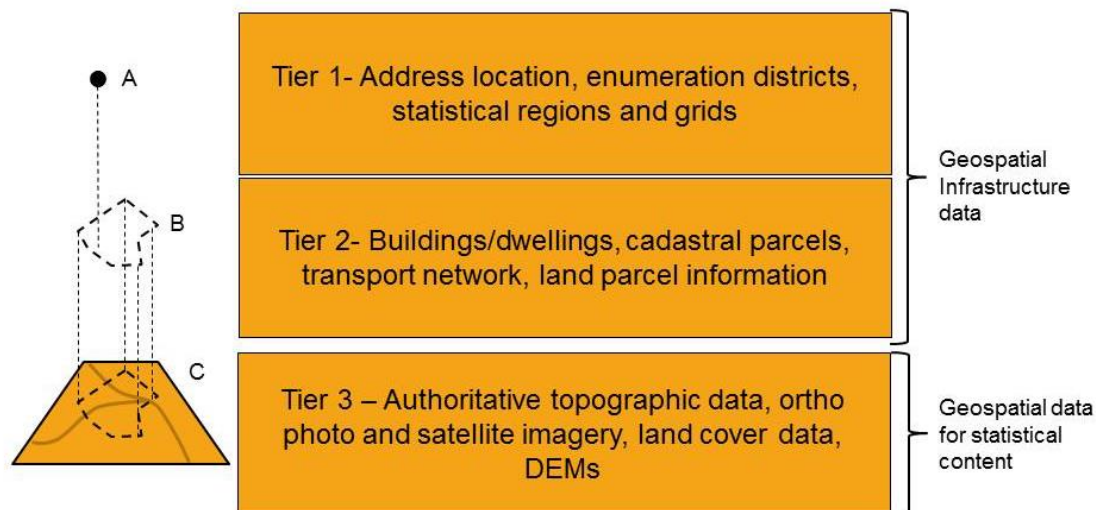


Figure 5: Different tiers of geospatial data for production and dissemination of statistics identified by the GEOSTAT 2 project. A workplace geocoded to an address location (A) can be linked to a cadastral parcel (B) in which land use can be computed by combining the parcel with a land use map (C).

The main focus of the GSGF, and Principle 1 in particular, is the geospatial infrastructure data of tier 1 and 2 illustrated in the graph above. Nonetheless, the role of tier 3 data is also recognised by Principle 1.

Requirement 1.1 - Use data from National Spatial Data Infrastructures

Thanks to INSPIRE, NSDIs have been put in place in EU Member States and it is rational to use data from these NSDIs as source for geocoding of statistical information. The Generic Conceptual Model of INSPIRE defines the elements necessary for interoperability of geospatial datasets and services, including cross-border aspects. It specifies requirements and recommendations with regard to data specification elements of common use, like the spatial and temporal schema, unique identifier management, object referencing, some common code lists, etc. Those requirements of the Generic Conceptual Model that are directly implementable are included in the Implementing Rule on Interoperability of Spatial Data Sets and Services²⁶. Relevant information can be taken from the

²⁵ EFGS/GEOSTAT 2 2017. A Point-based Foundation for Statistics - Final report from the GEOSTAT 2 project: <https://www.efgs.info/geostat/geostat2/>

²⁶ <https://ec.europa.eu/jrc/en/publication/reference-reports/conceptual-model-developing-interoperability-specifications-spatial-data-infrastructures>

following Technical Guidance documents²⁷ e.g. about addresses, buildings, cadastral parcels, geographical names, statistical units, population distribution and human health.

UN-GGIM: Europe has proposed address data, cadastral parcels and buildings, among other datasets, to be included in the Core Data concept. The Core data concept can be an important driver to enrich the NSDIs with data needed for geocoding purposes currently not available or not harmonised to the extent necessary. However, it has been recognized both nationally and in the European working groups, that the INSPIRE implementation is actually missing the data, integration methods and measures in order to realise data harmonisation aspects, which is an essential requirement of the GSGF for Europe.

Several European institutions like the European Commission, European Environment Agency (EEA), Eurostat and other associations, initiatives and services such as EuroGeographics, are working on pan-European data sets based on national data in the coming years, easing pan-European use. For example, the planned European Location Services (ELS)²⁸, envisaged by EuroGeographics and its members might cover the whole of geographic Europe and offer more than INSPIRE compliance. If successful and fully implemented, ELS may provide a single access point to the members' national reference data and services through a federated geo-information infrastructure in accordance with INSPIRE interoperability principles. Furthermore, ELS shall become a geospatial platform that will provide the framework for facilitating the provision of harmonised and cross-border reference datasets, targeting the master level in complement to the current pan-European products (ERM²⁹, EBM³⁰, EGM³¹, and EuroDEM³²). As for datasets, ELS shall provide INSPIRE compliant services giving access to those harmonised, cross-border datasets.

Recommendations

- Recommendation 1.1.1 - Any geospatial information used to obtain locations for statistical and administrative data (geocoding), or to produce statistical content, should build on relevant, authoritative and INSPIRE compliant geospatial data and services from NSDIs. INSPIRE compliance includes standardised and agreed formats, coordinate reference systems, metadata elements, data models and exchange services such as discovery, view and download services (see good practice cases C1.1, C1.2, C1.3, C1.4, C1.6, C1.7 and C3.3 in Annex 2).
- Recommendation 1.1.2 – Should reference data needed for geocoding of statistical information not exist within a country or should the quality not be adequate for geocoding purpose, it is of utmost importance that such information be collected, enhanced, improved and provided through the NSDI. Priority should be given to the reference data listed as Core data by UN-GGIM: Europe (Address, Building and Cadastral parcel), and meeting Core data or better quality (see good practice cases C1.1, C1.2, C1.3, C1.4, C1.6 and C1.7 in Annex 2).

²⁷ <http://inspire.ec.europa.eu/data-specifications/2892>

²⁸ <http://locationframework.eu> and <https://demo.locationframework.eu>

²⁹ EuroRegionalMap: <https://eurogeographics.org/products-and-services/euroregionalmap/>

³⁰ EuroBoundaryMap: <https://eurogeographics.org/products-and-services/ebm/>

³¹ EuroGlobalMap: <https://eurogeographics.org/products-and-services/open-data/>

³² EuroDEM: <https://eurogeographics.org/products-and-services/eurodem/>

- Recommendation 1.1.3 – Specifications of INSPIRE, on the implementation of systematic management of unique identifiers for features used for geocoding,³³ and corresponding (draft) recommendations by UN-GGIM: Europe Core data should be applied to facilitate integration of geospatial and statistical information through consistent identifier-key relationships, and to keep track of corrections in location changes as well as lifecycle of each statistical unit, i.e. at the micro-data level (see Figure 6 and good practice cases C1.1, C1.4 and C1.7 in Annex 2).
- Recommendation 1.1.4 - The roles and responsibilities of different agencies involved in production of geospatial information should be well defined, e.g. who maintains what information and how often data are updated. A custodianship model and stewardship model may need to be established, in order to identify the most relevant stakeholder for a geospatial data source (see good practice case C1.1).

Requirement 1.2 - Use point-based location data for geocoding

In the context of the GSGF, geocoding is generally defined as the process of geospatially enabling statistical unit records, or administrative information of any kind, so that they can be used in geospatial analysis³⁴. “Georeferencing” or “object referencing” are often used synonymously. As a result of geocoding the location of a phenomenon can be defined in relation to an existing spatial object, such as an address or building³⁵.

The concept of a point-based foundation for statistics was exhaustively explored and promoted in the GEOSTAT 2 project³⁶. The GEOSTAT 2 project concluded that a point-based foundation for geocoding of statistics allows for considerable adaptability to changes in geographic regions over time and for implementation of user-defined output geographies and, therefore, is the condition for implementing principle 3 on statistical output geographies.

It should be stressed that a point-based foundation for geocoding is applicable for information where a point location provides an accurate and feasible spatial representation of the statistical object, such as an individual, household or workplace, etc. There may be other statistical phenomena that should be more accurately represented using area or line-features³⁷. Unlike address locations, buildings and cadastral parcels cannot be truly represented by a point feature in the sense that they have an extent of space (a polygon). However, for the purpose of geocoding, point-representations of buildings and cadastral parcels (centroid, weighted coordinate, etc.) provide enough spatial accuracy and consistency with statistical objects to be considered as point-based data.

A point-based foundation must support the use of high quality standardised physical address locations, buildings and/or cadastral parcels. It must also comply with, and rely on, the INSPIRE

³³ National practice by Finland as an example: JHS 193 Unique identifiers of geographic data; Annex 5. Unique identifiers in INSPIRE data products, http://docs.jhs-suositukset.fi/jhs-suositukset/JHS193_annex5_en/JHS193_annex5_en.pdf (which includes in JHS 193, <http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/193;jsessionid=C46B0D681C8A6EDC8ED7BEDC62A54C70>)

³⁴ For further references to terminology, see EFGS website (<https://www.efgs.info/information-base/introduction/terminology/>)

³⁵ JRC 2012

³⁶ <http://www.efgs.info/geostat/geostat2/>

³⁷ E.g. traffic intensity reported for a road segment.

directive and its related implementation rules and regulations as well as to other international and national applicable standards, such as those published by the ISO, W3C, IETF and OGC.

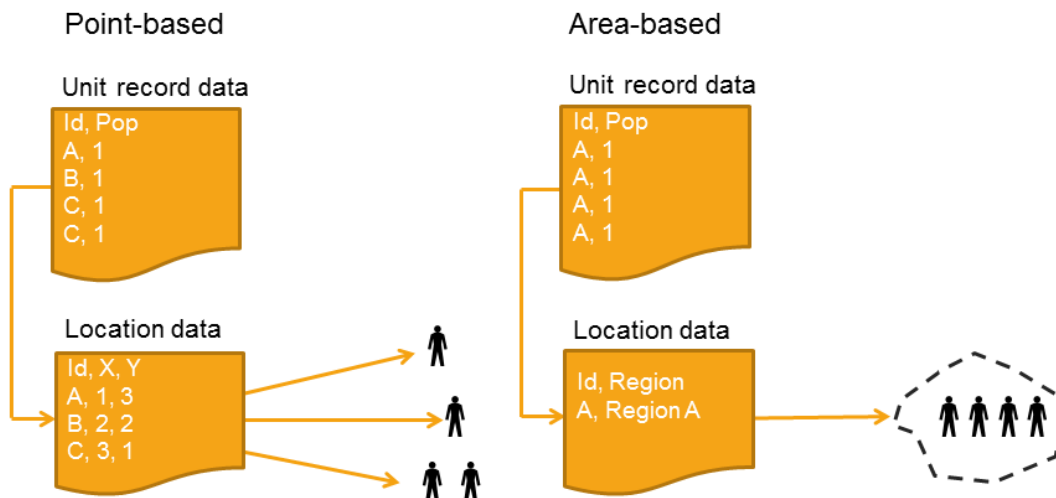


Figure 6: The conceptual difference between point-based and area-based geocoding infrastructures according to the GEOSTAT 2 project

The choice of national point-based location data for geocoding may vary between countries due to various national practises and history and maturity of the NSDIs. Physical address locations³⁸, buildings³⁹ and cadastral parcels⁴⁰ are recognised as equally preferred objects, as long as a minimum spatial accuracy requirement and completeness can be guaranteed⁴¹. Despite great diversity between countries in terms of standards for addressing, it can be assumed that address data is the strongest candidate for point-based location data in most Member States. UN-GGIM: Europe has published draft Core data recommendations for content for address data⁴². Though the recommendations for content are not legally binding themselves, they are based on INSPIRE. The Core data recommendations can hopefully trigger the creation of harmonised address registers in countries where such data currently does not exist, or has a poor coverage and quality. In addition to address, building and cadastral parcels, also other point-based data and standard, long-lasting point-locations may be relevant as basis for economic or environmental statistics (e.g. monitoring stations and water discharge points).

Recommendations

- Recommendation 1.2.1 - A point-based infrastructure (point locations) should be adopted as main and preferred approach for geocoding in the ESS, and is advised also for non-ESS European countries. Use of more general location descriptions, and/or larger geographies (such as enumeration areas or other statistical geographies), should be considered only as a *complementary* or secondary approach when point-based geocoding fails because of partially missing data. Countries should agree on one single uniform national infrastructure for

³⁸ Good practise for address data to be provided in annex.

³⁹ Good practise building/dwelling data to be provided in annex.

⁴⁰ Good practise cadastral parcel data to be provided in annex.

⁴¹ Core Data quality or better.

⁴² UN-GGIM: Europe 2017.

geocoding of all public and potentially private data (see good practice cases C1.1, C1.2, C1.3, C1.4, C1.6 and C1.7 in Annex 2).

- Recommendation 1.2.2 - In order to implement a point-based foundation for statistics, authoritative information on physical address locations, buildings and/or cadastral parcels should be made available within the NSDIs. In addition, information need to be accurate and consistent, have sufficient coverage and meet internationally and nationally agreed standards⁴³ (see good practice cases C1.1, C1.2, C1.3, C1.4, C1.6 and 1.7 in Annex 2).
- Recommendation 1.2.3 - All countries should have a single, national, authoritative, universal address register available for public institutions to include in their respective business processes (see good practice cases C1.1, C1.2, C1.4, C1.6 and C1.7 in Annex 2).
- Recommendation 1.2.4 - Geospatial agencies are encouraged to set up and provide national geocoding services, based on authoritative location data, within a service-oriented architecture referring to common reference data sets, common service configuration and common guidance for application. The ongoing initiative on a new OGC standard for geocoding APIs should be taken into consideration⁴⁴ (see good practice cases C1.1, C2.2, C2.3 and C3.3 in Annex 2).
- Recommendation 1.2.5 - National geocoding services should be open to authorities of other Member States in order to rely on similar methods and tools for the geocoding inside and outside countries and obtain consistent results. This would allow e.g. the recording of workplace addresses of citizens working abroad. Common services could provide a better basis for cross-border geocoding, hence improving calculation of statistics on cross-border commuting and migration (see good practice cases C1.1, C1.2, C1.5 and C3.3 in Annex 2).

Requirement 1.3 - Build formal working relationships on institutional agreements

Building formal working relationships between agencies responsible for production of geospatial information and NSIs, are crucial as to safeguard long-term provision and quality of geospatial data. Cooperation should ideally rely on (formal) agreements on roles and responsibilities of organisations or legislation and, when applicable, comply with the framework set by the INSPIRE directive, taking advantage of the national coordinating/steering committees and technical structures created for its implementation.

Examples are Legal Statutes, Memoranda of Understanding (MoU) or multilateral agreements between institutions. However, the agreements themselves are no guarantee for a good, flexible and solid cooperation. In a nutshell, the following conditions have been identified by UN-GGIM: Europe's Work Group on Data Integration for well-functioning working arrangements:

- a) Reliability of geospatial information are requested by all stakeholders;
- b) confidentiality of statistical data has to be observed;
- c) data of high quality for a wide variety of tasks is needed and
- d) quality standards have to be defined and met.

UN-GGIM: Europe Work Group on Data Integration recommends that, where organisations are not currently working on cooperative projects or in alliance for strategic development, a step-by-step

⁴³ E.g. comply with INSPIRE specifications and UN-GGIM: Europe recommendations for Core Data content.

⁴⁴ <http://www.opengeospatial.org/projects/groups/geocodeapiswg>

approach is considered and adopted⁴⁵. In addition to content and quality, cooperation needs to address terms of access, licensing and costs related to use of geospatial data for geocoding purposes.

Recommendations

- Recommendation 1.3.1 - NSIs and geospatial agencies should work actively to increase mutual exchange of knowledge between the geospatial and statistical community by initial or continuous training, information, communication or by working on concrete cooperative projects. Cooperation should be oversight and monitored by regular national and European steering groups composed of senior experts from both communities.
- Recommendation 1.3.2 - Agreements between NSIs and geospatial agencies should cover terms of access, licensing, governance and use of geospatial information. Agreements may also need to involve other stakeholders, such as municipalities or regional bodies responsible for data provision. Data from the NSDIs need to be easily accessible and usable, for NSIs or other public institutions conducting geocoding, at a low or affordable cost but preferably free of charge.
- Recommendation 1.3.3 - The GEOSTAT 3 project suggests that the fruitful collaboration between statistical and geospatial agencies on a European level established through the UN-GGIM: Europe Work group on Data Integration, should continue after the present task is completed. The group should act as the European steering group (mentioned in recommendation 1.3.1) to facilitate and monitor the continuous and mutual exchange of knowledge between the geospatial and statistical community on European level.

⁴⁵ <http://un-ggim-europe.org/content/wg-b-data-integration>

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



Global definition

The Global Framework recommends that the linkage of a geocode for each statistical unit record in a dataset (i.e. a person, household, business, building or parcel/unit of land) occur within a data management environment. Persistent storage of a high precision geocode enables any geographic context to be applied when preparing the data for release in the future (i.e. in aggregating data into a variety of larger geographic units or to adapt to changes in geographies over time). Moreover, geocodes can enable data linking processes that aim to integrate information of varying nature and sources.

This component of the Global Framework also recommends that established data management tools, techniques and standards be used to facilitate the integration and management of the geocode within the dataset, including address-to-geocode-linking mechanisms.

Relation with other principles

Whereas principle 1 describes the requirement for the geospatial infrastructure needed for a consistent geocoding, principle 2 is dealing with one specific use of the infrastructure: the actual process of geocoding of unit record data and the environment to support this process. Most recommendations concerning this principle follow on the assumption that the provisions of principle 2 are to be carried out mainly by NSIs or other institutions responsible for geocoding of statistical or administrative data. In turn, the successful implementation of principles 1 and 2 is the condition for the provision of statistics into flexible statistical geographies as set out in principle 3.

Discussion

General considerations for a good data management environment are expressed in the generic principles of the Common Statistical Data Architecture (CSDA) defined by the High-Level Group for the Modernisation of Official Statistics.⁴⁶ These principles are also compliant with the FAIR data principles⁴⁷. Though the CSDA is specifically targeting statistical institutions, its principles are relevant to any organisation managing large volumes of data. The CSDA bottom-line is to design an effective data management environment without compromising quality and security of the information it hosts. Despite the CSDA does not specifically address issues related to management of geospatial information, some of its principles are closely related to the principles of the GSGF⁴⁸.

⁴⁶ <https://statswiki.unece.org/display/DA/VI.+Key+principles>

⁴⁷ <https://www.fairdata.org.uk/who-its-for/>

⁴⁸ Most closely related is CSDA principle 5 - Use an authoritative source and principle 6 - Use agreed models and standards.

Requirement 2.1 - Build an effective and secure data management environment

An effective and safe data management environment is a priority to most organisations. However, in practice, effectiveness and safety can be conflicting targets. On the one hand, an “open” environment allowing a large number of users (e.g. internal staff) to directly access micro-data layers can be effective because it allows for processing and integration of a variety of data themes from the source, without time-consuming administration of permissions and security barriers. On the other hand, a compartmented data management environment with a restrictive access policy has a high level of security but perform less efficient in terms of data integration as micro-data layers may only be directly accessed by a small number of staff. Moreover, the different policies determining the access restriction of the datasets applied in the Member States will burden the effectiveness to elaborate a European data policy acting as a single access point to the national components of European datasets.

Hence, balancing between effectiveness and security in terms of data management can be a difficult task for many NSIs or organisations dealing with sensitive micro-data. Statistical production occupies a privileged position within the European General Data Protection Regulation (GDPR)⁴⁹.

Organisations that process personal data for statistical purposes may avoid restrictions on secondary processing and on processing sensitive categories of data as long as they implement appropriate safeguards. One of these safeguards encouraged by the GDPR is deployment of “pseudonymisation” of data. Pseudonymisation is the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable individual⁵⁰ (e.g. that data is key-coded).

For NSIs in Member States where personal IDs comprise part of the mechanisms for integration of data sources and for linking unit record data to location, the provisions of GDPR to restrict the use of personal data has posed as special challenge. However, the introduction of GDPR has also offered a timely trigger to improve the data management environment with smarter and more secure solutions for data integration to reduce un-necessary processing of personal data.

There are also other aspects related to the effectiveness and reliability of the data management environment. Such aspects entail consistent versioning with time stamps, and deployment of version ID where relevant, of unit record data and location data to ensure best possible temporal coherence between unit record data and location data.

Recommendations

- Recommendation 2.1.1 - Persistent storing of a high precision geocode for each statistical unit record in statistical and administrative datasets (i.e. a person, household, business etc.) should be the main and preferred approach for ESS Member States and is advised also for non-ESS European countries. A high precision geocode implies reference to an address location, building/dwelling or cadastral parcel (see corresponding recommendations for principle 1).

⁴⁹ <https://www.eugdpr.org/>

⁵⁰ GDPR, Article 89(1)

- Recommendation 2.1.2 - An efficient data management environment should allow linking of statistical and spatial objects at unit record level without compromising privacy of micro data. Implementation of data warehouse solutions could be an effective way to combine a widespread use of geocoded micro data with confidence that proper privacy measures has been applied in the data architecture.
- Recommendation 2.1.3 - Consistent synchronisation procedures should be assured for data sources involved, as to maintain the relationship between the geospatial infrastructure and the unit record data. The relation between microdata and statistical or administrative geographies should also be synchronised.

Requirement 2.2 - Store location only once

One single location object (e.g. a building or a physical address location) can serve geocoding of a multitude of statistical or administrative data. A good data management environment should be designed to avoid redundant storage, e.g. duplication, of coordinates or geometries. Ideally, location data with coordinates and geometries should be stored only once. A well-structured data management environment can support a division of the management of unit record data on one hand and management of location data on the other. If linking mechanisms has been set up using consistent, permanent and unambiguous IDs, there is no need to physically store coordinates or geometries alongside with unit record data. The only spatial reference needed in unit record data is the unique identifier to enable linkage to location. This is good practise to mitigate uncertainties regarding the origin of the coordinates or geometries. In fact, maintenance routines and clarity of governance can benefit from defining maintenance of location data and unit record data as separate roles.

To industrialise production of geospatial statistics and to simplify aggregation of data for common geographies (see principle 3), a good practise is to store references (codes) to any relevant statistical or administrative geography with each location data element. E.g. each point-location object should be clearly and unambiguously associated with the region, municipality, grid cell etc. in which it is located for a particular point in time. This enables simplified aggregations of data by regions and back-in-time by non-geospatial experts using standard relational database techniques.

Ideally, indexes of geographies for location data elements should be part of the NSDI and the institution responsible for collection and maintenance of location data should also be responsible for the synchronisation between location data elements and the administrative and statistical entity to which they belong.

NorthCoord	EastCoord	Parcel_ID	UUID	County	Municip	UrbanArea	ElectDistrict	Grid1000_SE	Start Date	End Date
6412239	321049	140811575	b0934777-e170-43c3-95fc-0b306f50cee5	14	1480	T4368	14804076	3210006412000	20150101	NULL
6409860	314444	140292849	f808fa66-bb31-4ca3-bea4-1d214178be35	14	1480	T0000	14804061	3140006409000	20141122	NULL
6409459	321683	140050428	660a6147-966e-44f7-94ab-5e2e6d73aaf8	14	1480	T4368	14804075	3210006409000	20120623	NULL
6404504	318210	140295195	42690817-3952-428c-adc3-12e200b9d564	14	1480	T4368	14804056	3180006404000	20111031	NULL
6406499	315689	140080726	9cde3b2d-de88-4e37-9bf4-25224be12f47	14	1480	T0000	14804061	3150006406000	20120804	NULL
6411848	314619	140014112	7eac82dc-13d7-46d0-aba4-cca0900bf77c	14	1480	T0000	14804081	3140006411000	20151118	NULL
6404836	316751	140081085	16b73e76-6d10-4ea1-bacf-09242f41ed43	14	1480	T4368	14804052	3160006404000	20133909	NULL
6410837	320648	140822904	be8e3a93-076c-43d1-8115-e85928fe0522	14	1480	T4368	14804071	3200006410000	20140213	NULL
6411415	320828	140812146	6597d6d7-ec69-484b-9afb-1df1c4a9cb57	14	1480	T4368	14804076	3200006411000	20160222	NULL
6407740	315983	140087473	ff520216-4847-45b8-9730-fac4181e59cb	14	1480	T0000	14804061	3150006407000	20110701	NULL

Figure 7: The table illustrates a location data table (in this case cadastral parcels) with references to administrative and statistical geographies. Besides references to county and municipality, each object is aware of its location in relation to national urban areas, electoral districts and 1 x 1km grid cells. By linking unit record data to this table, aggregates can be easily produced for the administrative and statistical geographies by use of standard database tools or tabulation software.

In case custodians of location data (e.g. geospatial agencies) have set up geocoding services based on authoritative address or building information accessible for NSIs, location data may not have to be stored at all within NSIs but can rather be reached via services and APIs on-demand.

Recommendations

- Recommendation 2.2.1 - Location data objects should be recognised and fully integrated in the general data architecture of NSIs in order to facilitate design of efficient workflows for data integration and geocoding (see good practice case C4.1 in Annex 2).
- Recommendation 2.2.2 - It is recommended to build repositories for location data (geocoding databases) holding references to a number of relevant and common administrative and statistical geographies also back-in-time, at each location data object, to allow simplified aggregations of data also by non-geospatial experts.
- Recommendation 2.2.3 - Address services or geocoding services provided by geospatial agencies need to fully support use of life-cycle attributes and versioning. Serving only the most up-to-date information is not sufficient to incorporate such services in statistical business processes. Hence, obsolete address objects should be retained in the data and their current status indicated using the INSPIRE mechanism of life-cycle attributes and versioning. Pre-allocated or provisional addresses, where available, should be managed in the same way (see good practices C1.1 and C2.2 in Annex 2).
- Recommendation 2.2.4 - Statistical microdata transmitted to Eurostat by ESS Member States should come with a reference to the ETRS89 1km² grid cell code as a minimum, if point references are not possible to provide.

Requirement 2.3 - Ensure consistency and quality of geocoding results

Geocoding is the process of assigning a geocode to a piece of information (e.g. a statistical unit record) using known location information, such as coordinates, or other, indirect location description, in order to assign direct location reference or link to the accurate coordinates to each

statistical unit. Geocoding can be conducted in different ways, using a variety of location data on various technical platforms and using different methods to deal with mismatch between datasets. To ensure a consistent result, countries should develop and apply standardised approaches to conduct geocoding by formulating and implementing geocoding guidelines. Such guidelines may state what type of location data to be used for which cases (if multiple options are available), what kind of rules to deploy if locations are missing and how to improve matching between records. Such guidelines help to make sure that the result will be the same regardless of the person or institution conducting the geocoding. In case several national institutions conduct geocoding activities, such standardised approaches may also need to be agreed between institutions to ensure a conform result⁵¹.

Considering the heterogeneity between countries in terms of frameworks for location data, geocoding guidelines may only apply effectively on national practices. However, as operability and data harmonisation efforts are moving forward on European level, commonly agreed and applied European geocoding guidelines should be a long-term goal in order to ensure coherence between countries.

In order to properly describe the quality of geocoding and geocoded data and to make the results from geocoding operations transparent, it is important that geocoding metadata be stored at object level. The aim of the quality declaration is to make sure that the method used for linking unit record data with location is traceable and repeatable. To obtain a fully geocoded record, typically different approaches may need to be iteratively applied with a gradually decreasing level of spatial accuracy. Preferably, linkage should be obtained by means of direct match between location data and unit record data, but, in absence of direct match, interpolation of location or references to larger geographies may be used as location proxies. Geocoding metadata may be divided in two parts; code lists describing the type of matching applied (direct match etc.), and code lists describing the type of location object used (building, address etc.).

Recommendations

- Recommendation 2.3.1 - Member States should develop and apply national guidelines for geocoding workflows in order to ensure a consistent and conform result within and between institutions. Such guidelines may include agreed decisions on what location data services to use to geo-enable which statistical information. They may also include agreed *ad hoc* methods to improve matching between location data and unit record data (see good practice C2.1 in Annex 2).
- Recommendation 2.3.2 – Defining geocoding guidelines on European level should be considered in order to ensure coherence and interoperability between countries. Drafting such guidelines requires a careful assessment of national conditions and practices.
- Recommendation 2.3.3 - Geocoding results should be as accurate and consistent as possible and documented according to agreed geocoding metadata. Geocoding metadata should be provided at object level so that the accuracy of the assigned location can be assessed for each observation (see good practice case C2.4 in Annex 2).

⁵¹ Good practice can be found in “SSF Guidance Material – Geocoding Unit Record Data Using Address and Location” prepared by Australian Bureau of Statistics (ABS 2018).

Requirement 2.4 - Develop consistent approaches to manage non-matching data

Complementary methods for geocoding are commonly deployed when mismatch occur due to partially missing data. Such methods may include area-based geocoding (administrative units, postal code areas, census enumeration areas etc.) or interpolation of locations from road network elements⁵².

Despite application of methods to improve matching between unit record data and location data, a certain number of non-allocated observations typically remains (e.g. homeless people or people without a permanent place of dwelling). These non-allocated observations need to be handled with a systematic approach in order to produce a consistent statistical output. For small output geographies (such as grid cells), this issue is particularly prominent.

Eurostat suggests that if the place of usual residence of a person is unknown within the territory of the reporting Member State, additional scientifically based, well-documented, and publicly available statistical estimation methods may be used to allocate this person to a specific grid cell⁵³. For the Census 2021 it has been agreed that there will be one additional synthetic grid cell without spatial representation that will contain all persons that cannot be geocoded to normal grid cells⁵⁴.

In case of other preferences for data on national level, countries need to define a consistent approach for non-matching observations. The problem of non-matching data was explored in the GEOSTAT 1 project and there are national cases describing disaggregation methods for geocoding of inhabitants with unknown place of residence⁵⁵.

Recommendations

- Recommendation 2.4.1 - For the provision of data to Eurostat for Census 2021, the ESS Member States should follow the agreed approach to use one additional synthetic grid cell without spatial representation to contain all persons that cannot be geocoded to normal grid cells.
- Recommendation 2.4.2 - Countries should define and describe a consistent approach for non-matching observations to be applied in production of small area or grid data on national level, for those cases where the approach set for European data is not applicable or suitable for national purposes.

Requirement 2.5 - Use point-of-entry validation in collection of administrative or statistical data

The challenge of high quality point-based geocoding comes with inconsistent or erroneous identifiers resulting in mismatch between location data and unit record data. A typical example is a piece of information regarding an individual or company stored in an administrative record where the

⁵² Good practise on workarounds to improve geocoding can be found in the GEOSTAT 2 report:

<https://www.efgs.info/geostat/geostat2/>

⁵³ Annex 1 to the Commission Implementing Regulations Deciding on a temporary direct statistical action in accordance with Regulation (EC) No 223/2009 to disseminate selected topics of the 2021 PHC geocoded to a 1 km² grid.

⁵⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018R1799&qid=1544707164955&from=EN>

⁵⁵ National case from Czech republic: <https://www.efgs.info/wp-content/uploads/geostat/1b/GEOSTAT1B-Appendix16-Case-study-CZ.pdf>

physical address (identifier) of the premise is misspelled or incomplete. Geocoding this information using an official record on address locations will fail because of mismatch between the official (true) address and the address provided for the individual or for the company.

There are numerous *ad hoc* techniques to improve the number of matching records including address repair tools, homogenisation of address information or interpolation of address location points, but the most efficient and sustainable way to increase quality is to make sure that the information is accurate from the source.

At its simplest, point-of-entry address validation checks the reported address (from a respondent, such as a person or a company) against an authoritative address index or gazetteer and, if there is a difference, asks the respondent to pick from one of the addresses in the index. By validating addresses at the point they are collected, against a national authoritative address record, many simple address reporting errors can be avoided, such as spelling mistakes or incorrect locality names⁵⁶.

One of the key features of INSPIRE is the employment of unique identifiers of spatial objects (UUIDs or PIDs). Due to their consistency and uniqueness, such identifiers are useful to identify information in computer systems. A unique identifier can be used to unambiguously identify an address location without having to bother about consistency of the actual address string. In practise, however, they can be difficult to use to set up linkages between location data and unit record data because data retrieved from administrative sources rarely comes with such consistent identifiers. For obvious usability reasons, references to addresses reported by respondents are expressed in natural language (e.g. Fisher Street no 4). However, with point-of-entry validation implemented in digital data collection platforms, the unique identifier of the address location retrieved from the gazetteer, can automatically accompany the information provided by the respondent (as a back-end procedure, not necessarily visible for the respondent). Having the unique identifier integrated with unit record data, can significantly improve the integration with location at a later stage. It also facilitates time stamping of data and life-cycle management of statistical records directly harmonised with the NSDI.

There are numerous examples on point-of-entry validation already being successfully implemented in collection of data for population or enterprise registries in many countries⁵⁷, all proving substantial improvement of quality. The reasons why point-of-entry validation is not widely used may be several:

- the institution responsible for collection of data may not have enough legal mandate to enforce the respondents to provide authoritative location references;
- the institution responsible for collection of data may not have access to authoritative location data such as address gazetteers or building registers, etc.;
- the institution responsible for collection of data may not fully understand the importance of correct and verified location references or may lack the capacity to technically implement procedures for point-of-entry validation;
- a uniform national address register is not available, only e.g. at municipality level.

⁵⁶ ABS 2018.

⁵⁷ Some examples are population registers in Finland, Estonia and Sweden and Business register in Slovenia.

As a growing amount of statistical information is retrieved from various administrative data sources collected by different public bodies, there is a need to actively involve a variety of government data custodians in this effort, not only geospatial agencies and statistical institutes. In order for custodians of administrative information (population registries, tax administrations, social security providers or any other governmental bodies) to be able to conduct validation of location references, correct and updated authoritative information must be easily accessible and affordable. Initiatives to release national authoritative address registers as open should be encouraged as to increase the prospects of improving location data verification broadly in society⁵⁸.

In addition to open data initiatives, legal measures may need to be considered as additional means to enforce quality improvement of collected data. In some countries, use of authoritative address data has become mandatory for collection of government data through regulations⁵⁹.

Recommendations

- Recommendation 2.5.1 – When creating and maintaining administrative or statistical records, point-of-entry validation mechanisms should be used to ensure the best possible quality of the location references (address, building ID etc.) stored in unit record data. All national public authorities in ESS Member States in charge of recording addresses into public files should be obliged to use the uniform geocoding infrastructure for entering addresses, to avoid inconsistencies. This policy is also strongly advised for non-ESS European countries (see good practise cases C1.4 and C2.5 in Annex 2).
- Recommendation 2.5.2 - The statistical and geospatial communities should collaborate to promote use of authoritative location data among public institutions collecting and managing administrative information.
- Recommendation 2.5.3 - Address services or geocoding services provided by geospatial agencies need to be available with defined APIs and accessible for public custodians of administrative data for easy integration in their data collection platforms (see good practice cases C2.2, C.2.3 and C3.3 in Annex 2).
- Recommendation 2.5.4 - To create strong incentives for the whole society (civil society and private sector) to use and implement authoritative national address registers in their business, release of address data under open data licenses should be considered (see good practice case C1.1, C1.2, and C1.5 in Annex 2).
- Recommendation 2.5.5 - Legal measures should be considered as means to enforce good quality of collected data. Such measures may concern both removal of legal barriers preventing point-of-entry validation to be conducted as well as legal instruments to enforce use of point-of-entry validation in public sector (see good practice case C1.1 in Annex 2).

⁵⁸ Address information is open data in Austria, Denmark, Switzerland and Netherlands, see use case C1.1, C1.2, C1.4 and C1.5 in Annex 2.

⁵⁹ In the Netherlands public authorities are obliged to use the official Base Register for Address information, see use case C1.1 in annex 2.

2.3 Principle 3: Common geographies for production and dissemination of statistics



Global definition

To enable comparisons across datasets from different sources, the Global Statistical Geospatial Framework recommends that a common set of geographies be used for the display, reporting and analysis of social, economic and environmental information.

The UN EG-ISGI recognises the importance of traditional statistical and administrative geographies. The Expert Group also recommends NSIs to consider the benefits of gridded data. Gridded data can be both a rich source of information and a consistent geography for disseminating and integrating information. Recent global efforts have culminated in the development of a Discrete Global Grid Systems (DGGs) standard which has been developed under the auspices of Open Geospatial Consortium (OGC). This System offers further options in the use of grids within the context of the principle of common geographies and in geospatially enabled statistics.

Use of a common set of geographies will ensure that all statistical data is consistently geospatially enabled and that users can discover, access, integrate, analyse and visualise statistical information seamlessly for geographies of interest.

Relation with other principles

The implementation of principles 1 and 2 is a condition for the full implementation of principle 3 in a statistical production system as their implementation allows for the flexible aggregation into any output geography. In turn, principle 3 is an important condition for principle 5, where common geographies form the basis for dissemination of geospatial statistics. Common geographies should also be included in the National Spatial Data infrastructure described in principle 1, for consistent use. If principle 2 is fully implemented, it is straightforward to aggregate geocoded statistics into any output geography.

Discussion

Geographies are spatial representations of the administrative, statistical or functional division of a country, also known from INSPIRE as administrative or statistical units. The following main groups of common geographies can be identified in a European context:

- national administrative, statistical and functional geographies;

- European statistical geographies within the TERCET⁶⁰ framework (LAU⁶¹ and NUTS⁶² areas);
- Global, European and national statistical grids.

In addition to the above-mentioned geographies, customised or user-defined geographies are becoming increasingly interesting for users. For obvious reasons such geographies cannot be included in the group of common and official geographies maintained by statistical or geospatial agencies. However, the capability among NSIs and/or geospatial agencies to deliver information for geographical areas defined by the user is increasingly important.

Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies

Each country has its own unique framework of administrative, statistical and functional geographies, ranging in scale from country parts, provinces and regions to local administrative units, census districts, mesh-blocks, parishes or neighbourhoods. Most frameworks comprise a hierarchical order of levels (nested geographies), where each level seamlessly covers the entire territory. However, there are also geographies covering only certain parts of a country, such as delimitation of localities or urban zoning, or functional areas grouping territories together based on geographic specificities or socio-economic criteria. The advantages of most national administrative levels are that they are publicly well-known and familiar for dissemination of national statistics.

Due to varying political and administrative structures and practises among countries, a full harmonisation of administrative, statistical and functional geography frameworks across Europe is neither possible nor desirable. However, statistical and administrative geographies⁶³ are covered by the INSPIRE regulation⁶⁴, and as such, a technical interoperability has been implemented mainly in terms of data models. Further data content and semantic harmonisation is taking place as part of the mandate of the UN-GGIM: Europe Work Group on Core data where administrative units and statistical units have been proposed candidates for Core data⁶⁵. This is also happening on a global level through the UN-GGIM Fundamental Global Geospatial Data Themes, where “Functional areas” has been agreed and approved as one of the fundamental themes needed to achieve and monitor the SDGs⁶⁶.

In spite of technical interoperability efforts through the existing INSPIRE specification, management of data on administrative and statistical geographies still suffer from inconsistencies in many ESS Member States. The following problems have been identified among Member States:

- unclear custodianship and poor collaboration between NSIs and geospatial agencies resulting in lack of coherence between coding systems, reference dates and geometries;
- inconsistencies in cases where the LAU level is not the lowest administrative level for building statistical geographies;

⁶⁰ TERCET = “TERRitorial Classifications and Typologies”: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016PC0788>

⁶¹ “Local Administrative Units”: <http://ec.europa.eu/eurostat/web/nuts/local-administrative-units>

⁶² “Nomenclature of Territorial Units for Statistics”: [http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Nomenclature_of_territorial_units_for_statistics_\(NUTS\)](http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Nomenclature_of_territorial_units_for_statistics_(NUTS))

⁶³ Administrative and statistical *units* according to INSPIRE nomenclature

⁶⁴ <http://inspire.ec.europa.eu/id/document/tg/su> and <http://inspire.ec.europa.eu/id/document/tg/au>

⁶⁵ UN GGIM: Europe 2018

⁶⁶ <http://un-ggim-europe.org/sites/default/files/7%20GGIM%204th%20Plenary%20FDTWG.pdf>

- lack of topological consistency between different hierarchical levels of geographies (resulting in gaps and overlaps) and geometries not sufficiently edge-matched between countries;
- vague extent and unclear role of cadastral areas for building administrative areas;
- lack of consistent rules to handle water areas (inland, estuaries, sea);
- lack of consistent rules for scale, accuracy and generalisation of geometries; and
- lack of versioning and data for “historical” geographies.

A uniform, complete, consistent, authoritative and widely accessible set of national administrative, statistical and functional geographies will facilitate data integration both on a national and European level. It will enable deployment of federated search technology such as Linked Open Data bringing together information from a variety of data sources. Good practise on the efforts to create one single, national access point for all relevant geographies can be found in the UK⁶⁷. The Open Geography portal from the Office for National Statistics (ONS) provides free and open access to the definitive source of geographic products, web applications, services and APIs. All content is compliant with INSPIRE and available under the Open Government Licence.

The GEOSTAT 3 project suggests that the specifications for administrative and statistical units provided through INSPIRE, the TERCET regulation and UN-GGIM: Europe Core data provide enough guidance to deal with the technical and semantic side of harmonisation of data if implemented duly. In addition to the technical challenges, countries need to provide the governance structure to effectively deliver consistent national statistical and administrative geographies.

The processes and elementary building blocks for the construction of the lowest national administrative layer should be clearly described including problems with under-coverage of data. Usually the cadastral parcels cover the full territory of a country and are the fundamental building blocks for creating all administrative areas (see recommendations on cadastral parcels by UN-GGIM: Europe)⁶⁸. However exceptions exist, such as coastal and transitional waters, beaches, river mouths, lakes, exclaves, etc. that are not part of the administrative or statistical geography of a country. This should be properly documented in the metadata.

Recommendations

- Recommendation 3.1.1 - Geospatial agencies and NSIs should define clear custodianship roles for coding systems and boundary data for statistical and administrative geographies respectively, on national level in order to enable a more efficient collaboration (see good practice case C3.1 in Annex 2).
- Recommendation 3.1.2 - All national administrative, statistical and functional geographies with relevance for production and dissemination of official statistics should be provided as authoritative geospatial data in compliance with the technical specifications of INSPIRE, whenever relevant, and the UN-GGIM: Europe Core data Recommendation for Content on Statistical Units and Administrative Units, including full topological and coding consistency (see good practice cases C3.2 and C3.3 in Annex 2).

⁶⁷ <https://www.ons.gov.uk/methodology/geography/geographicalproducts/opengeography>

⁶⁸ http://un-ggim-europe.org/sites/default/files/UN-GGIM-Europe_WGA_Recommandation_Content_CP-v1.1.pdf

- Recommendation 3.1.3 - When changes occur in coding systems or geometry of boundaries, data on national statistical and administrative geographies should be available to all users no later than six months after the reference date (see good practice case C3.1 in Annex 2).
- Recommendation 3.1.4 - Geospatial agencies and NSIs should work together to improve accessibility, speed of delivery and usability of national data on administrative and statistical geographies, with the aim to provide high quality data under Open Data licenses. Data should be machine-readable and provided through OGC/INSPIRE compliant view and download services (see good practice case C3.3 in Annex 2).
- Recommendation 3.1.5 - Geospatial agencies and NSIs should jointly start to explore the potential of dissemination of statistical and administrative geographies as Linked Open Data. In order to do so, the initiatives already being conducted in several ESS Member States should be evaluated and if needed, guidance and recommendations to be developed to facilitate harmonisation (see good practice cases C3.2, C4.4 and C4.5 in Annex 2).
- Recommendation 3.1.6 - Efforts to set up a consistent framework of national statistical and administrative geographies should give priority to current data, nevertheless inclusion of historical geographies is recommended, as these data are crucial for recasting current (point-based) data for past administrative or statistical divisions (see good practice case C3.1 in Annex 2).
- Recommendation 3.1.7 - Geospatial agencies and NSIs should agree on an approach on scale, reference dates and accuracy of administrative and statistical geographies building on UN-GGIM: Europe Core data recommendations. For analytical purpose, data capture and data processing, highest possible accuracy of boundaries is needed (Master Level 0 or 1⁶⁹). For dissemination and visualisation purpose, a generalised level (simplified geometries) is desirable.
- Recommendation 3.1.8 - The processes and elementary building blocks for the construction of the lowest national administrative layer should be clearly described and properly documented in the metadata, including problems with under-coverage of data.
- Recommendation 3.1.9 - Linked to the provision of statistical geographies, ESS Member States should also provide area statistics on the size of these geographies based on a harmonised methodology and national data⁷⁰. Land area statistics are important for density information (population density, etc.).

Requirement 3.2 - Improve maintenance of the European framework of statistical geographies

European regional statistics are widely used in the context of EU policies, in particular for regional, agriculture and transport policy and e.g. providing the reference geography for determining the eligibility of regions under the cohesion funds.

⁶⁹ Master levels are classification of levels of details in geospatial data defined by ELS. Master level 0 is equal to scales larger than 1: 5 000; typically, data at cadastral map level, for local level actions. Master level 1 is equal to scales between 1: 5 000 and 1: 25 000; data for local level actions (UN GGIM: Europe 2018)

⁷⁰ https://ec.europa.eu/eurostat/cache/metadata/Annexes/reg_area3_esms_an1.pdf

The geographical classification used for European statistics is laid down in the Regulation (EC) No. 1059/2003 on the establishment of a common classification of territorial units for statistics (NUTS) and provides the geographical framework.

The NUTS classification created and maintained by Eurostat is the approach to re-use the national administrative units, wherever possible, to create a more stable and comparable zoning across Europe. The classification consists of three NUTS levels and LAU (local administrative units) on lowest level.

Eurostat has expanded the range of statistics published to additional territorial typologies to address EU policy-makers' increasing need for information in their spatial context, mainly in the context of cohesion and territorial development policies.

Therefore in addition to the NUTS classification, the European Commission supported by OECD have over the past few years, developed several typologies of statistical geographies such as the DEGURBA (Degree of Urbanisation⁷¹):

LAU based typologies:

- Degree of urbanisation
- Coastal and non-coastal areas
- Functional Urban Areas (FUA)

NUTS level 3 based typologies:

- Urban-rural typology
- Metropolitan typology
- Coastal typology

These typologies are already in use in European Union legislation, for instance in Regulation (EU) No. 522/2014 which refers to the degree of urbanisation of local administrative units (LAUs) to define eligibility for European Regional Development Fund support to carry out innovative actions in cities or towns and suburbs.

In order to ensure a harmonised application of the typologies and allow for cross-referencing from other acts and programmes, Eurostat launched a legislative initiative called "TERCET"⁷², which is aiming at integrating the typologies into the NUTS Regulation.

⁷¹ http://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_DEGURBA

⁷² Regulation (EU) 2017/2391 of the European Parliament and of the Council of 12 December 2017 amending Regulation (EC) No 1059/2003 as regards the territorial typologies (TERCET): <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02003R1059-20180118&qid=1519136753473>

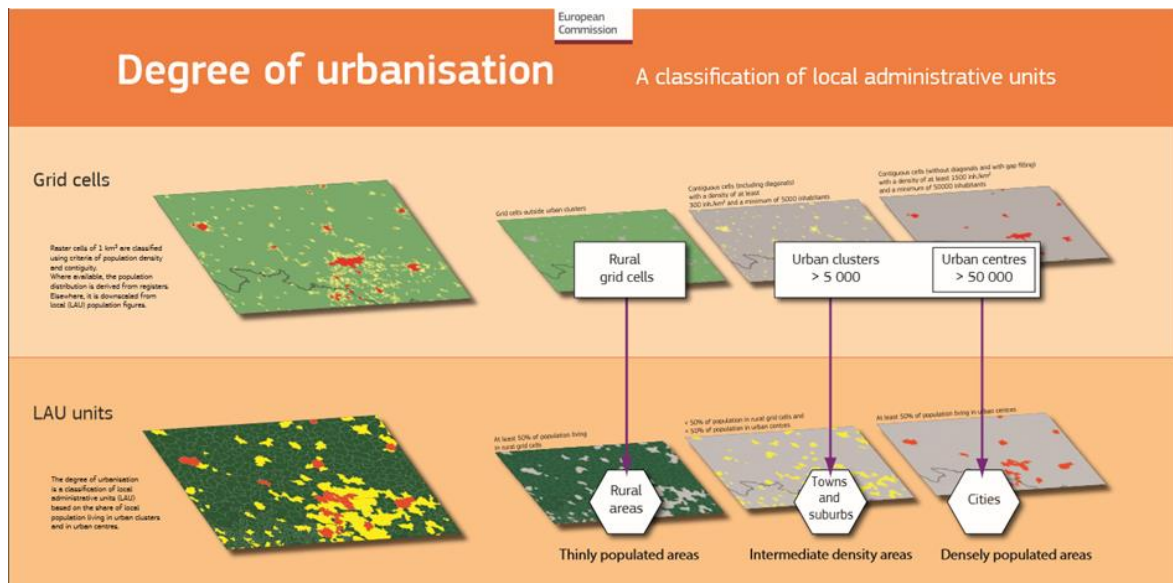


Figure 8: Illustration of the process to define the degree of urbanisation (source Directorate General Regional and Urban Policy of the European Commission)

These geographies rely on population grid data and other topographic data such as coastlines. However population grid data is currently not updated outside census years. There is an EU regulation⁷³ on the provision of 1km² population grids for the Census 2021. The ESS has launched an action on more frequent census information including population grids after 2021 and all NSIs are encouraged to support the regular and frequent production of population grids after 2021 to be able to update the territorial typologies on a regular basis.

To produce pan-European LAU and NUTS geographies, Eurostat needs to collect geospatial data from ESS Member States. In theory, compiling these data using national INSPIRE services would be a straightforward process. However, edge-matched geographies from all Member States, with harmonised scale and semantics, already aligned between NSIs and NMCA, are currently not available. As a result, Eurostat need a broker (EuroGeographics managing the product EuroBoundaryMap⁷⁴) to produce those data by furthering harmonisation and coordination of national data, which takes time and entails costs.

EuroGeographics plans to publish the content of EuroBoundaryMap (administrative units and NUTS/LAU units) via INSPIRE web services, likely as part of the European Location Services (ELS). The aim is to cover the whole of geographic Europe, not just the 28 current members of the European Union, to provide unique and powerful source of harmonised, authoritative geospatial information⁷⁵.

Currently most NSIs manage code lists of LAU used for statistics, while geospatial agencies manage the geographical boundaries. Until consistency of both code lists and geometry can be achieved, a lot of time goes by. As an example, LAU geometries with a reference date of 1 January 2016 only became available to Eurostat in February 2017. Similarly, the delay between the entry into force of a NUTS version and the availability of the geospatial data is far longer than 12 months. As a result,

⁷³ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018R1799&qid=1547812343941&from=EN>

⁷⁴ <http://www.eurogeographics.org/products-and-services/euroboundarymap>

⁷⁵ <https://eurogeographics.org/products-and-services/european-location-services/>

statistical areas may not yet be available upon arrival of the first statistics for these output geographies.

Recommendations

- Recommendation 3.2.1 - The coordination between NSIs and geospatial agencies in production of administrative units (LAU and NUTS) should improve to respond more efficiently to changes of administrative units. When changes occur in coding systems or boundaries, data on national statistical and administrative geographies should be available to all users no later than six months after the reference date.
- Recommendation 3.2.2 - All European geospatial agencies are encouraged to support the current work on developing Open European Location Services (Open ELS) coordinated by EuroGeographics, which would be the first operational step towards the implementation of the European Location Services (ELS), by targeting national open datasets and services first. If the Open ELS continues to an operational stage, it would provide a single access point for open national data on administrative geographies, cadastral parcels, addresses and buildings. In addition, it will give access to a pan-European regional geolocator service, and to a cadastral index map (see good practice C3.3 in Annex 2).
- Recommendation 3.2.3 - All NSIs within the ESS are encouraged to support the development of EU legislation currently under way on the regular and frequent production of population grids after 2021 to be able to update the territorial typologies on a regular basis. Such development is also advised for non-ESS European countries.

Requirement 3.3 - Consolidate use of existing statistical grid systems and explore the potential of evolving global grid systems

For purposes of European actions related to publishing statistical data on grids (e.g. the GEOSTAT project), the EFGS (European Forum for Geography and Statistics) created 1x1km grid datasets covering almost all EU and EFTA countries.

The TERCET regulation⁷⁶ as well as the regulation on population grids for the 2021 census define statistical grids as output geographies for European statistics. The grid coding system is compliant with INSPIRE principles (Statistical Units specification). The coordinate reference system is ETRS89 Lambert Azimuthal Equal-Area⁷⁷ or the International Terrestrial Reference System (ITRS) for areas beyond the ETRS89 geographical scope (over-sea territories).

Recent global efforts have culminated in the development of a Discrete Global Grid Systems (DGGS) standard, which has been developed under the auspices of Open Geospatial Consortium (OGC)⁷⁸. This System offers further options in the use of grids within the context of the principle of common geographies and in geospatially enabled statistics.

The goal of DGGS is to enable rapid assembly of spatial data globally without the difficulties of working with projected coordinate reference systems. The OGC DGGS Abstract Specification standard defines the conceptual model and a set of rules for building highly efficient architectures for spatial

⁷⁶ <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02003R1059-20180118&qid=1519136753473>

⁷⁷ EPSG: 3035, <http://spatialreference.org/ref/epsg/etrs89-etrs-laea/>

⁷⁸ <http://www.opengeospatial.org/projects/groups/dggswwg>

data storage, integration and analytics. In other words, the OGC DGGS describes a universe of valid grid systems. Valid grids could be used as “interlingua” between grids at different geographical levels.

For certain types of statistics with low counts in most grid cells (e.g. farm statistics) ESS Member States prefer grid cell sizes larger than 1km². The jump to the next power of 10 (grid size 10 km) is often too large for European statistics and unnecessarily limits the usability of grid statistics. Intermediate grid cell sizes might be an option but nesting and aggregations must be possible. For instance, a mix of 200 m and 250 m grid cells across countries should be avoided. The current revision of the INSPIRE data specification for Statistical Units will consider this requirement.

In addition to global and European grids, some countries use statistical grids based on national reference systems. Such national grids may be well established among domestic users, expecting continuous data releases based on the existing national grid systems. There is a potential conflict between the European and national grid systems. National grid systems in different projections are a problem if the European grid is not aggregated from re-projected microdata or if the recast from national to European grids is not done from very small grid cells. In addition, there are privacy issues due to the spatial differencing problem of intersecting grid systems.

Recommendations

- Recommendation 3.3.1 - For the time being, the current European ETRS89 grid system should remain the main output grid geography for pan-European statistics produced within the ESS. The grid system is already settled with the INSPIRE principles (Statistical Units specifications) and it is a well-established feature among users of pan-European geospatial statistics.
- Recommendation 3.3.2 - Introduction of additional grid sizes should be considered and agreed on for European level (e.g. 100 m, 125 m or 200 m or quad-tree) in line with the forthcoming revision of the INSPIRE implementing regulation on data interoperability for Statistical Units. The need for, and implications of, introducing additional grid sizes should be explored jointly by the statistical and geospatial community, taking into account a versatile and expanding use of spatio-statistical data and resources, even regarding earth observation data taking into account disclosure issues.
- Recommendation 3.3.3 - The geospatial and statistical communities should monitor the development of the DGGS and its application closely in order to prepare for a possible future implementation of the grid system for national and European data. Implementation of the DGGS will potentially be beneficial for the global data integration and for a seamless integration between the national and European level, however it is still at an early stage and more studies are needed on the benefits and implications of adopting the concept (see good practice case C3.4 in Annex 2).
- Recommendation 3.3.4 - The European ETRS89 grid system, or the International Terrestrial Reference System (ITRS) for over-sea territories, should be the main and preferred output grid geography on national level within ESS Member States. The European ETRS89 grid system is also recommended for non-ESS European countries. In case national grid systems are used as a complement to European grid systems, these grids (including their coding systems) should be made available as authoritative, open data in order to ensure a common

use on national level. Statistics disseminated on the EU grid system e.g. to meet EU legislation should have quality equal to statistics disseminated in national grid systems.

- Recommendation 3.3.5 - The parallel use of national, EU and global grid systems for statistics should be coordinated in the ESS to ensure that all grids have comparable quality.

2.4 Principle 4: Statistical and geospatial interoperability – Data, Standards and Processes



Global definition

Both the statistical and geospatial data communities operate their own general data models and metadata capabilities; however, these are often not universally applied. The statistical community uses the Generic Statistical Information Model (GSIM), the Statistical Data and Metadata Exchange (SDMX), and the Data Documentation Initiative (DDI) mechanisms. The geospatial community, on the other hand, makes use of the General Feature Model (GFM) and the ISO19115 metadata standard, plus a number of application specific standards.

Within the statistical community, there is a need to build geospatial processes and standards into statistical business processes in a more consistent manner. In consequence, the EG-ISGI has recognised that a top-down approach is required with a view to incorporating geospatial frameworks, standards and processes more explicitly into the Common Statistical Production Architecture (CSPA) and its components. In particular, the Generic Statistical Business Process Model (GSBPM) needs to refer, to a larger extent, to the use of geospatial data and methods in the statistical production process, and in particular the data, standards and methods that are incorporated into the GSGF. In addition, interoperability goes beyond technical and structural integration and should consider organisational and human aspects more intensively.

Relation with other principles

Interoperability issues, in most cases, cut across the other principles of the GSGF rather than belonging to one principle only. This is reflected in many of the recommendations provided for the previous principles dealing with data specifications and harmonisation of content. However, interoperability is so critical to the successful implementation of the GSGF, while also being a complex issue, that a separate principle is defined for it.

Discussion

In a nutshell, principle 4 is about the preconditions for statistical and geospatial data to occur as a data ecosystem, in which a number of actors interact with each other to exchange, produce and consume data. 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. According to INSPIRE, interoperability is defined as “the possibility for spatial datasets to

be combined, and for services to interact, without repetitive manual intervention, in such a way that the result is coherent and the added value of the datasets and services is enhanced”⁷⁹.

Based on the European Interoperability Framework (EIF)⁸⁰ interoperability should be divided into the following four layers:

- legal interoperability;
- organisational interoperability;
- semantic and conceptual interoperability;
- technical interoperability.

Elements of all four layers of interoperability can be identified as relevant in the GSGF:

- technical interoperability concerns mostly principles 1, 2, and 5. It covers the applications and infrastructures linking systems and services. Aspects include interface and services specifications;
- semantic and conceptual interoperability concerns mostly principle 2 and 5. It 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 as well describing the exact format of the information;
- legal and organisational interoperability concerns all principles. Legal interoperability is about ensuring that organisations operating under different legal frameworks, policies and strategies are able to work together and organisational interoperability refers to the way in which public administration align their business processes, responsibilities and expectations to achieve commonly agreed goals.

However, the full implementation of this principle is particularly important for principle 5, as failure in achieving interoperability in any of the other principles will often result in incomplete or less useful information for the end-user.

Requirement 4.1 - Improve geospatial workflows within statistical production

One of the key principles of the CSDA (Common Statistical Data Architecture) states that data should be described to enable reuse⁸¹. The rationale for this key principle is that data can be easily understood and used with confidence without requiring further information and that data, and its related metadata, can be easily reused by several business processes reducing the need to transform or recreate information. As geospatial objects do not fit well within traditional architecture of NSIs, these are currently handled externally to the rest of the statistical process through dedicated geospatial systems. In addition, there is a lack of metadata standards to describe geospatial information in such way that it can be easily and broadly incorporated in existing statistical business processes. There is also a general knowledge gap between the statistical and geospatial communities, which makes integrated data architecture more difficult⁸².

⁷⁹ <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC69484/lbna25280enn.pdf>

⁸⁰ https://ec.europa.eu/isa2/eif_en

⁸¹ <https://statswiki.unece.org/display/DA/VI.+Key+principles>

⁸² Coady 2018a.

Another logical step to build more consistent geospatial workflows in statistical business processes would be to incorporate geospatial frameworks, standards and processes more explicitly into the Common Statistical Production Architecture (CSPA)⁸³. In order to do so, the General Statistical Business Process Model (GSBPM)⁸⁴ needs to be developed and enhanced to make greater reference to the use of geospatial data and methods in the statistical production process. Such work is currently under way, the GSBPM is under revision by UNECE, and the need to incorporate geospatial data management is on the agenda of the revision committee. One idea presented recently by UNECE is to develop a geospatial view of the GSBPM⁸⁵.

As a response to the need to bridge the two communities over issues related to information interoperability, a Statistical Domain Working Group (DWG) has recently been proposed to be set up under the auspices of OGC. The Statistical DWG is chartered to identify requirements and use cases of how geospatial and statistical standards can support the integration of geospatial information into the statistical system and for the purposes of broad discovery, analysis and use⁸⁶.

Recommendations

- Recommendation 4.1.1 - The statistical community should engage more actively when new geospatial standards with relevance for data integration are developed and involve in evaluation of utility to statistical production (e.g. OGC TJS).
- Recommendation 4.1.2 - Forum for discussion and engagement regarding cross-domain issues are crucial to improve interoperability and should be provided through the UNECE⁸⁷, ModernStats, ESS, UN-GGIM, EFGS and initiatives like the OGC Statistical Domain Working Group.
- Recommendation 4.1.3 - Statistical production involving geospatial information should rely on both statistical models and standards as well as existing geospatial standards, e.g. in geospatial data collection and dissemination technologies (see good practice case C4.1 in Annex 2).
- Recommendation 4.1.4 - Interoperability requires consideration of geospatial data and the concept of location as a natural part of logical data warehouses and data architecture (see good practice case C4.1 in Annex 2).
- Recommendation 4.1.5 - The General Statistical Business Process Model (GSBPM) should be developed and enhanced to make greater reference to the use of geospatial data and methods in the statistical production process.
- Recommendation 4.1.6 - Geospatial services in a service-oriented architecture are recommended to standardise geospatial production components. NSIs should pursue to share common tools (see good practice cases C.1.1, C2.2, C2.3 and C4.1 in Annex 2).

⁸³ <https://statswiki.unece.org/display/CSPA/Common+Statistical+Production+Architecture>

⁸⁴ <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0>

⁸⁵ Vale 2018.

⁸⁶ Coady 2018a; Coady 2018b.

⁸⁷ The Joint UNECE/UN-GGIM Workshop on Integrating Geospatial and Statistical Standards in Stockholm, November 2017, was a good example on this:

<https://statswiki.unece.org/display/Geo17/Workshop+on+Integrating+Geospatial+and+Statistical+Standards+2017>

Requirement 4.2 - Enable data integration through consistent semantics and concepts across domains

A fundamental requirement for a successful integration of geospatial data and statistical information at unit record level is semantic and conceptual coherence between statistical and spatial objects. For decades, the statistical and geospatial communities have operated their own concepts and models. Simple, and seemingly obvious, concepts such as “building” may have slightly different meaning within the statistical and geospatial domains. This does not necessarily pose a problem in general production of statistics or geospatial data, but in terms of integration of geospatial objects and statistical objects on micro data level, semantic differences can be a significant problem for design of consistent and interoperable data architecture.

The general ambition in the ESS is to increase the use of administrative data sources and alternative data sources such as big data. In doing so, semantic interoperability between different domains poses even greater challenges. Whereas the statistical and geospatial communities share a common understanding on the value of consistent classification and modelling of information, other public or private data collecting communities may have other priorities and agendas.

Semantic models or Enterprise Data Models spanning across statistical and geospatial domains can help ensuring that coherence occur. One of the key recommendations from the joint UNECE and UN GGIM Workshop on Integrating Geospatial and Statistical Standards in Stockholm 2017, was an urge to the statistical and geospatial communities to look for opportunities to work on semantic interoperability issues, for example through ontologies for addresses and buildings⁸⁸.

Recommendations

- Recommendation 4.2.1 - Definition of common conceptual models for objects fundamental for both statistical and geospatial communities are needed. Statistical and geospatial agencies should look for opportunities to work on semantic interoperability issues, for example, ontologies for addresses and buildings (see good practice cases C1.1 and C4.5 in Annex 2).
- Recommendation 4.2.2 - NSIs and geospatial agencies should work together to inform administrative data custodians on the need for conceptual harmonisation as to obtain semantic interoperability between communities and data sources.

Requirement 4.3 - Publish data once and leave it at its source to be reused many times

One of the main objectives of INSPIRE is to reduce inefficiencies in the collection, handling, storing and distribution of geospatial information. Drawing on this objective, reducing the need to duplicate information is an important objective also for the GSGF. This guiding principle means that all data, both geospatial and statistical, should ideally be collected and published only once. In a federated ecosystem of data it can (and should) be published separate of each other but in a way that enables interaction between datasets with low or no technical or semantic barriers.

Managing the geometries of administrative and statistical geographies (which is defined by Principle 3) and compiling tables with statistical information follow different pathways throughout the

⁸⁸ <https://www.unece.org/index.php?id=45404>

producing organisations. Thus implying that, in the case of changing tabular or geometry data, the construction of a combined dataset is automatically following the most recent changes.

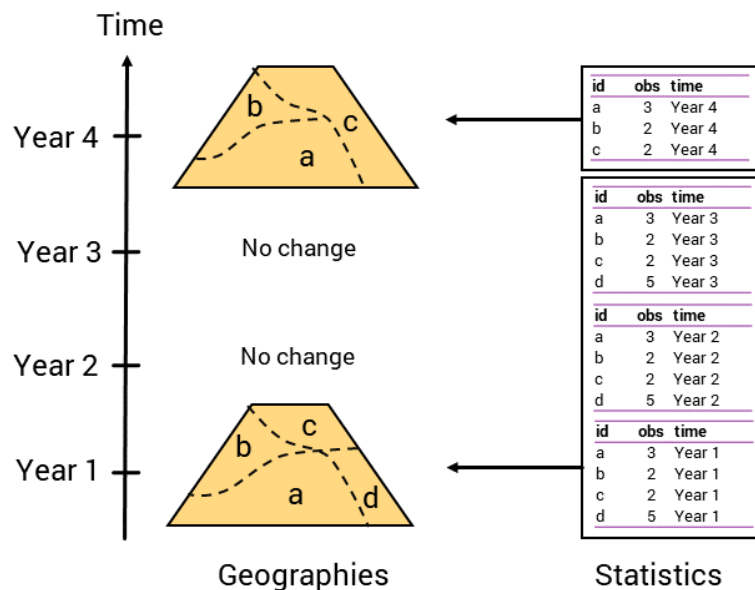


Figure 9: Illustration of the different pathways for geometry data (geographies) and tabular data (statistics). In this case the geography for year 1 remains unchanged until year 4. Accordingly, geometry data need only be published year 1 and 4 whereas the statistical data is published each year. Tabular data from year 1, 2 and 3 will be linked to the same geography.

Storing tabular data and geometry in open data format allows machine-to-machine based transformation and integration of data. This could reduce problems arising from duplication of data with unclear origin and actuality. It will also save time and resources through simplified publishing procedures. However, it will require mapping between statistical and geospatial standards and increased use of services that can automatically merge tabular statistical data with geospatial data online. Ideally, tailor made data fusion products can be created on-demand by the end user.

Many NSIs have since long implemented platforms to store and disseminate official statistics (e.g. statistical databases or databanks). The rapid development of web protocols for data exchange and APIs has provided new opportunities to search and harvest data from these platforms in new ways, with greater flexibility and possible integration of data in third party applications without a need to create physical copies of information. The use of these APIs also enables statistical services comprising machine-readable data for linkage with national and European geographies served through OGC Web Services (OWS).

Geospatial and statistical data and metadata are shared using different data formats, exchange methods and dissemination standards. Within EU, geospatial information is shared using the spatial data infrastructure INSPIRE whereas statistical information is exchanged following Standard for Data and Metadata eXchange (SDMX). Defining a mapping between these two standards is essential to support the combination of these two types of information and maximise the re-use of existing and accepted data infrastructures for statistics and geospatial data respectively. Such progress could also lead to a more efficient, service-oriented exchange of data between institutions on a national level and between the national and European level.

As an example, census information needs to be shared in an INSPIRE compliant manner as of 2021. This means that for the 2021 round of population and housing censuses in the EU, NSIs will have to share census data according to INSPIRE legislation in addition to the existing statistical dissemination infrastructure based on SDMX. Eurostat and NSIs have designed a technical solution that will enable the re-use of the Census Hub (see good practice) for population grids. The project has provided an important test bed for statistical and geospatial data integration using INSPIRE compliant services and SDMX. As a result, NSIs will implement automatically the requirements from INSPIRE without disruption of their established production systems and without double data sharing burden. Future work will have to investigate how this theme specific solution could be enhanced so that it works for all statistical areas and data providers that are concerned by INSPIRE.

Recommendations

- Recommendation 4.3.1 - Countries should intensify their efforts on service-oriented dissemination through APIs to provide machine-readable open data format for national geospatial statistics - Spatial Statistics as a service (see good practice cases C4.3, C3.3 and C5.1 in Annex 2).
- Recommendation 4.3.2 - The adaption of SDMX for INSPIRE implementation, successfully developed and tested by Eurostat in the context of population grids for the Census 2021, should be used as solution for INSPIRE harmonisation of Population distribution in the ESS. It is assumed to minimise the effects of double obligation on ESS Member States and to maximise the usability of the census information for the statistical and geospatial community (see good practice case C4.2 in Annex 2).
- Recommendation 4.3.3 - SDMX is evolving as data exchange format for a growing number of statistical domains. NSIs and Eurostat should work closely together to explore further applications of SDMX beyond the context of Census 2021 (see good practice cases C4.2 and C4.3 in Annex 2).
- Recommendation 4.3.4 – Open, centralised, INSPIRE compliant services with NUTS geographies and 1km and other sized grids should be set up by the ESS (or by other European body through agreement with ESS) for the whole of Europe, to avoid unnecessary duplication and efforts by the member states. Preferably this is done at different levels of detail. These services could be used for cross border mapping of Statistics when joined with the SDMX services (see good practice case C3.3 in Annex 2).

Requirement 4.4 - Increase use of services for merging geography and statistics

Implementing the guiding principle that all data, both geospatial and statistical, should be published only once in a standardised and machine-readable format, will set a good foundation for increased interoperability of geospatial and statistical information.

In theory, anyone, with access to geospatial data published as OGC Web Services and corresponding statistical data in SDMX format or statistical services provided through APIs, can set up their own statistical-geospatial data mashups. However, as machine-readable data is designed to overcome structural and syntactic barriers between machines and not for people, dealing with merging of data manually can be a very technical and complex exercise. Hence, it is strongly advisable to hide the complexity of the standards from as many people as possible through deployment of tools for automated linking of geographies and statistics.

Such concepts for machine-based integration of geography and statistics are generally referred to as Table Joining Services (TJS). A TJS is an online service that links statistical tables to web services with new OGC Web services as output. Tabular data can be services like the SDMX table service provided by Eurostat or it can be a table served through an API. The geometry can come from any existing geospatial information services or OGC Web Services (OWS). The TJS performs an automated online task generating an OWS like WMS and/or open web feature service like WFS, which can be used in different applications. This automated action replaces the work that typically has to be conducted manually by a GIS specialist.⁸⁹

The concept of TJS is particularly well designed for use where multiple themes of statistical data share the same reference geographies and/or when statistical tables from different years refers to stable geographies (like grid cells, national geographies or NUTS areas). Instead of publishing the same geometries over and over again each time a new statistical theme or variable is added to the statistical data, the geometry is published only once and multiple statistical contents can be added on demand.

A concern is the reference to the spatial features provided through INSPIRE services. The current allowed IDs for statistical areas in the SDMX files are listed in the code list for the GEO dimension. They contain mainly NUTS-codes consisting of the ISO country code and 1, 2 or 3 digits (e.g. NL, NL1, NL11 or NL111 to designate NUTS level 0, 1, 2, and 3). These codes should also be used as codes for the geospatial data provided as INSPIRE Statistical Units.

TJS is a concept rather than a specific tool, and as such it can be implemented in different ways and with different thematic content. There is an existing OGC standard for TJS that has also been implemented in open source platforms. One could also develop something similar, as long as the output is again OGC services with geometry data and corresponding tabular theme.

The concept of TJS can be adopted for data integration at European level, as has been demonstrated by the ELF project⁹⁰. However, it can also be adopted at national level as has been shown by Statistics Netherlands proof of concept for a TJS⁹¹. Successful tests have also been conducted within the GEOSTAT 3 project.⁹² The OGC standard for Table Joining Services is currently under revision to extend its capability to handle a variety of source data formats⁹³.

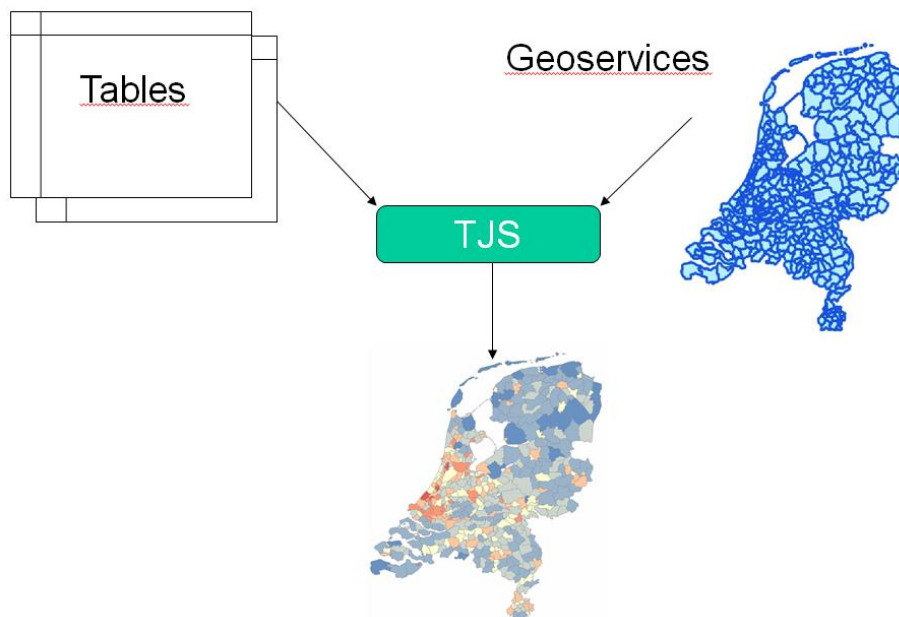
⁸⁹ An extensive description of Table Joining Services can be found at <http://geoprocessing.info/tjsdoc/index>

⁹⁰ The ELF project transformed into European Location Services: <https://eurogeographics.org/products-and-services/european-location-services/>

⁹¹ Statistics Netherlands 2006: https://www.efgs.info/wp-content/uploads/news/2016/NL_ImpactAnalysisTJS_v1.0.docx-1.pdf

⁹² See project report: Automated Linking of SDMX and OGC Web Services - A Feasibility Study by the GEOSTAT 3 project on the implementation of principles 4 and 5 of the Global Statistical Geospatial Framework <https://www.efgs.info/geostat/geostat-3/>

⁹³ https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2017/mtg3/S3_AARNIO_OGC_Table_Joining_Service_standard_revision_and_Oskari_1_.pdf



Source: Statistics Netherlands

Figure 10: Illustration of the concept of Table Joining Services (TJS)

Recommendations

- Recommendation 4.4.1 - European bodies, national geospatial agencies and NSIs should work closely together on developing and applying services using automated and dynamic linking of geographies and statistical services and on deciding on an underpinning governance on the provision of these services. This will enable more flexible and open statistical-geospatial integration providing more value for users (see good practice cases C4.2, C4.3 and C5.1 in Annex 2).
- Recommendation 4.4.2 - Though OGC Table Joining Services until now has not been widely implemented, the geospatial and statistical community should jointly involve in developing the standard and to consolidate and industrialise its implementation (see good practice cases C4.2, C4.3 and C5.1 in Annex 2).
- Recommendation 4.4.3 - For harmonised European data under the INSPIRE themes Population distribution and demography, a combined European action should be considered by means of a TJS operating on the Eurostat SDMX web service instead of setting up numerous different national services. This would also be a good business case and a large scale test, possibly paving way for application of the same technology also within other statistical domains (see good practice cases C3.3 and C4.2 in Annex 2).
- Recommendation 4.4.4 - Existing services using SDMX datasets should not be modified to additionally implement INSPIRE data models. Also no parallel INSPIRE download service should be implemented, as SDMX and the SDMX services already meet INSPIRE requirements as they are already machine readable and harmonised⁹⁴ (see good practice cases C4.2 and C4.3 in Annex 2).

⁹⁴ <https://ec.europa.eu/eurostat/web/sdmx-web-services/about-this-service>

Requirement 4.5 - Explore the potential of Linked Open Data for increased interoperability

Linked Open Data (LOD) is a set of design principles for sharing machine-readable interlinked data on the Web. It is one of the core concepts and pillars of the Semantic Web, also known as the Web of Data. The Semantic Web is all about making links between data understandable not only to humans but also to machines, and Linked Data provides the best practices for making those links.

Linking data has the power to greatly improve analysis and understanding of data, and this is increasingly recognised at national and international levels. This joining up of data has previously been done through methods such as data linkage but increasingly the diversity and complexity of administrative and big data sets compared to traditional surveys and census data requires new thinking to allow these datasets to be queried and exploited in a way that is not always possible with more traditional techniques. There is also a need to join statistical tables to allow users of statistical data to innovate and to build new platforms for analysis that move beyond the existing relationships between statistical tables. Linked data offers one methodology for exploiting these datasets by adding semantic structure to large datasets and by consolidating distributed datasets into a single queryable resource. The need to structure data to query and join it also makes statistics a useful case study as statistical tables and their associated geographic and non-geographic classifications means that - at least in theory – they could be mapped against a semantic vocabulary with relative ease⁹⁵.

NSIs and geospatial agencies in countries like UK, the Netherlands, Poland, Finland, Norway and the European Commission already provide, or are about to provide, geometry data as Linked Open Data and yet others are under way to do it more or less. In most cases the data provided is reference data (administrative and statistical geographies or other classification systems) rather than statistical content.

Linked data solutions demand a clear definition of objects and their life cycle management. A system of unique identifiers is needed. This approach has been taken into implementation in many NSIs and it may turn out as a quite straightforward way to integrate statistics and geospatial data.

Recommendations

- Recommendation 4.5.1 - Though Open Linked Data is still an emerging concept, geospatial agencies and NSIs should start exploring its potential. LOD and semantic web technology may be one fruitful way to improve the integration of geospatial and statistical datasets (see good practice cases C1.1, C4.4 and C4.5 in Annex 2).
- Recommendation 4.5.2 - The ESS should increase its efforts to facilitate common tools, concepts and methods in the field of LOD. Current progress made in individual ESS Member States on both statistical and geospatial data needs to be consolidated and exploited in a more systematic way to stimulate harnessing of LOD and semantic web technology in the whole of Europe (see use case 4.5 in Annex 2).
- Recommendation 4.5.3 - Building on experiences from ESS Member States, a good practice is to approach the work on LOD with one consistent reference dataset, e.g. statistical and administrative geographies (see good practice cases C4.4 and C4.5 in Annex 2).

⁹⁵ ONS 2016

- Recommendation 4.5.4 - As published statistical data and corresponding geographies may be maintained in different locations and by different organisations, a standard for coding of common geographies is needed for seamless integration of table information and related geographies by machine reading (see good practice cases C4.4 and C4.5 in Annex 2).

2.5 Principle 5: Accessible and usable geospatially enabled statistics



Global definition

The goal of principle 5 is to make sure that geospatial statistics is accessible and usable in the best possible way. This principle of the GSGF emphasises the need to identify or, where required, develop policies, standards and guidelines, which support the release, access, analysis and visualisation of geospatially-enabled information.

There is a wide range of legislative and operational issues that organisations need to be aware of when releasing and analysing information about people and businesses in a spatial context. One important aspect of this principle is to ensure that data can be accessed using safe mechanisms that not only protect privacy and confidentiality but also enable access to data in order to undertake various analyses that foster decision-making.

Relation with other principles

Principle 5 has a strong relation to principle 3 as common geographies are the basis for dissemination of geospatial statistics. But it also has a strong relation to principle 4, where the interoperability of services for data dissemination is defined.

Discussion

In this first version of this implementation guide, the principles are generic without a specific geospatial statistics product in mind. At a later stage, the definition of the intended use of geospatial statistics products should also be considered to improve the performance of the framework.

One of the key benefits of data integration is to have more flexible production systems in NSIs that can respond to new and emerging user demands without the need to change data collections or production systems (Principles 1, 2, and 3 define the requirements for this flexibility).

From a geospatial perspective, the main benefits are finer spatial granularity of statistics, a richer thematic content and more flexible output geographies. The ultimate scenario would be 'Statistics as a service', which means that users could request services from NSIs or Eurostat and e.g. demand spatial aggregations according to their spatial and temporal requirements. The increasing interest for statistics with higher spatial resolution and user defined data requests should be embraced by NSIs but needs to be paired with careful considerations regarding data protection and privacy issues.

There is a great variety of solutions throughout Europe for dissemination of statistics and geospatial data using on-line mapping tools and services. Besides dissemination platforms for traditional tabular data, most NSIs offer data access through some kind of mapping platform, ranging from simple atlases for viewing data to more advanced portals with built-in capabilities for spatial analysis and statistical visualisation. In addition, there are national geospatial platforms or NSDI services and data platforms for SDG monitoring.

Typically, many of the existing mapping tools do not fully support the use of machine-readable services for geospatial and statistical integration provided through APIs and OGC web services. Data has to be migrated from one technical environment to another and integration of statistical and geospatial data have to be conducted by staff responsible for maintaining the mapping tools.

Requirement 5.1 - Implement clear and simple data licensing policies

A good and transparent licensing policy is critical in order to release the full potential of geospatial statistics. Unclear terms of use of information will hamper innovation and development of new services. Even though data are openly accessible, conditions for use may be complicated and restrictions partially apply to some use (e.g. commercial use).

Most NSIs, as well as Eurostat, deploy business models whereby data, at least official statistics, can be used without restrictions. However, the approach for geospatial or small area statistics can be slightly different due to two main reasons:

- The production of geospatial statistics involves use of third party geospatial information where re-use of data is associated with royalties;
- Business models for production of geospatial statistics in some NSIs are depending on commissions. As a result, production costs may not be covered by adequate appropriations and NSIs often have to recover, at least partly, the cost of their production by offering them to the market.

Lately however, there has been a strong momentum for open data, even in countries where production of geospatial statistics is commission-based. The push for open data has encouraged NSIs to offer at least some of their geospatial statistics products under open data licenses. The 2021 round of Census is going further in this direction by incorporating the European grid, as a dissemination level, in the census hyper-cube. The forthcoming revision of the PSI directive will also increase the re-use of public data under open data license terms⁹⁶.

There is clearly a need to find a balance between the demand for more open data and the necessity of NSIs to cover the costs of their production. The GEOSTAT 3 project suggests a combined approach of open mid-resolution, standardised products, including 1km² statistical grids and a core set of statistical variables, and provision of chargeable services for high-resolution data or tailor-made services, as to allocate resources to improve the production setup and to increase the thematic content.

Recommendations

- Recommendation 5.1.1 - ESS Member States should aim to release at least a core set of statistical variables (such as total population) for mid-resolution grids (1 km²) or other small-area statistics under open data licenses. This is advised also for non-ESS European countries.
- Recommendation 5.1.2 - Geospatial statistics should be released as open data. The license of geospatial statistics created by combining statistical and geospatial information shall be as little restrictive as possible considering the respective open data licenses of the source data combined.

⁹⁶ <https://ec.europa.eu/digital-single-market/en/proposal-revision-public-sector-information-psi-directive>

- Recommendation 5.1.3 - The data provider should promote and implement the inclusion of the license information (fees, access constraints, etc.) in the INSPIRE metadata for the geospatial data set and/or service offered.

Requirement 5.2 - Use service-oriented data portals supporting dynamic integration of data

INSPIRE services and development of new and improved OGC standards for geospatial data, in combination with increased access to statistical services via APIs, have provided new opportunities for solutions building on machine-to-machine access, as well as dynamic linkage of information. Such solutions may provide a greater flexibility and save resources in the long-term perspective as data can be used from the source in numerous ways and through a variety of interfaces.

A good example on successful service-based integration of geospatial and statistical data is the Oskari platform, jointly configured for dissemination by the Finnish National Land Survey and Statistics Finland. The platform reads data live from Statistics Finland's dissemination platform via an API and perform an automatic merge with corresponding geographies retrieved from OGC web services⁹⁷. Through harmonisation and common configurations of APIs for statistical information between countries, such solutions are scalable and could be used to retrieve information from several countries simultaneously into one single federated search.

Also on a European level, services enabling machine-to-machine access and dynamic linking should be set up based on SDMX and OGC web services. This way the speed of interaction could be enhanced, since dissemination of maps on European level would not depend on aggregation of national web services. To be able to make cross boarder maps on pan-European data, the data needs to be harmonised technically and semantically (as discussed already under principle 4). This is already the case with the SDMX data as hosted by Eurostat.

At present, no operational service exists for providing authoritative and harmonised pan-European geospatial data by a single access point. However, European Location Services (ELS)⁹⁸, from EuroGeographics, aims to provide this based on services from European geospatial agencies. As a practical example of INSPIRE implementation, the forerunner of ELS, the ELF Project⁹⁹, has supported the delivery of several national web feature services and provided valuable feedback on the data specifications as they are implemented in different countries. It has also delivered the technical infrastructure to incorporate data content into an application environment, as well as tools for harmonisation and edge-matching and other tools for identifying areas of interest and products.

The realisation of Open ELS service offering harmonised pan-European geospatial data as open data would represent a next, important step towards new, open cross-border products and services for pan-European applications.

⁹⁷ See good practice case C5.1 in Annex 2

⁹⁸ <http://openels.eu>

⁹⁹ <http://www.elfproject.eu>

Recommendations

- Recommendation 5.2.1 - Countries should explore use of service oriented dissemination platforms, providing a greater flexibility in terms of usability and supporting data access through a variety of interfaces via APIs. OGC compliant services and non-proprietary formats should be used (e.g. OGC Geopackage for file deliveries¹⁰⁰) for dissemination in order to ensure flexibility also from an end-user perspective (see good practice case C5.1 in Annex 2).
- Recommendation 5.2.2 - The ESS should increase its efforts to facilitate common concepts and shared solutions on service-oriented and dynamic linking of data. Good practise and current progress made in individual Member States or elsewhere need to be consolidated and better exploited to stimulate development in the whole of Europe (see good practice case C4.2 in Annex 2).
- Recommendation 5.2.3 - EuroGeographics' work on establishing the Open ELS service might be an important step towards new, open cross-border products and services for pan-European applications. An open and service-based provision of the high-resolution EuroBoundaryMap would provide a strong business case for statistical-geospatial integration on pan-European level (see good practice case C3.3 in Annex 2).
- Recommendation 5.2.4 - NSIs and Eurostat should work together to develop concepts for 'Statistics as a service' including collection of good practice and proof of concepts for solutions.
- Recommendation 5.2.5 - Geospatial agencies and NSIs should consider to publish simple features as defined by OGC and in ISO 19125 next to the complex features as defined by the INSPIRE data models. This will improve usability in GIS systems. Also consider following modern encodings like LOD, WFS 3.0 and GeoJSON to improve accessible services also in the future.

Requirement 5.3 - Define clear national and European rules to ensure protection of privacy

Principle 5 seeks to create a balance between the need to protect the data and confidentiality of individuals and enterprises with the need to produce high quality, informative, spatio-temporal statistics. Data needs to be edited in pre-processing before release (data disclosure and confidentiality) to ensure that the data is not causing disclosure issues. Data custodians should be able to release data with confidence, with privacy and confidentiality protected.

The evolvement of geospatial statistics portfolios in many countries has been followed by discussions and concerns regarding the risks of violating confidentiality. The interconnectedness of different non-nested small area geographies (e.g. grid cells vs census tracts) presents a challenge, for the existing statistical disclosure control methods. There is a built-in conflict between the risk of disclosure on one hand and data utility on the other. While combining different small area geographies for dissemination will increase data utility for spatial planning, understanding how disclosure risk is increased due to geographic differencing is crucial.¹⁰¹

¹⁰⁰ <http://www.geopackage.org/>

¹⁰¹ Lukan & Smukavec 2017

There are a variety of confidentiality methods at hand which can be briefly categorised as perturbative vs non-perturbative and pre-tabular vs post-tabular.¹⁰² Member states are deploying different approaches due to various traditions, varying strictness in confidentiality policies and technical solutions.

Harmonised protection of census data in the ESS has become a topical issue due to the upcoming Census 2021 where dissemination of key variables per 1 km² grid cells is planned. Eurostat's working assumption is that the main application of grid statistics are accessibility studies for planning and analysis. For this purpose, users need to know where people live and if few or many. Exact count (e.g. 4 or 5) may not be essential. With regard to this, Eurostat has defined the following specific requirements for Statistical Disclosure Control (SDC) on grid statistics:¹⁰³

- SDC should minimise information loss, in particular as regards inhabited and non-inhabited grid cells;
- Recommended method should be applicable in all NSIs possibly with different parameters.

Recommendations

- Recommendation 5.3.1 - For dissemination of Census data following EU regulation for 2021 population grids¹⁰⁴, ESS Member States should follow the recommendations on methods for Statistical Disclosure Controls provided by the ESS centre of excellence on SDC¹⁰⁵. The method proposed should be applicable in all NSIs, possibly with different parameters.
- Recommendation 5.3.2 - For dissemination of national grid data or small area statistics, countries should define, describe and publish their own principles for the preservation of privacy with respect to existing national legislation and policy.
- Recommendation 5.3.3 - ESS Member states need to be aware of potential confidentiality risks, due to geographical differencing¹⁰⁶, that come with an increasing number of national and European data being disseminated on small areas. The quality of European statistics should be given priority. Therefore national geospatial statistics should not be released/published if stricter disclosure control (to avoid geographical differencing) lead to loss of quality of the European statistics.

Requirement 5.4 - Facilitate data search and use through cataloguing and improved guidance

The concept of Open Data Portals has significantly gained grounds and a corresponding metadata standard, building on Dublin Core¹⁰⁷ and describing data for these portals has been developed by the W3C: DCAT. Relatively simple and fast, the potential of DCAT to make

¹⁰² Antal, L et al 2017; Petri, E 2017

¹⁰³ Petri, E 2017

¹⁰⁴ https://eur-lex.europa.eu/eli/reg_impl/2018/1799/oj

¹⁰⁵ https://ec.europa.eu/eurostat/cros/content/recommendations-protection-census-data_en

¹⁰⁶ Geographic differencing is the process where the same data is obtained for two different but overlapping regions and the data from the smaller of these regions is subtracted from the data for the larger region. By utilising this method it is possible to obtain data for the area that is not common to both regions. Obtaining data for small areas using this method may result in a risk to privacy or confidentiality.

¹⁰⁷ <http://dublincore.org/>

geospatial and statistical metadata interoperable across portals was discussed in both the statistical and geospatial communities in Europe; related projects using e-government and interoperability programs as ISA and ISA² have been launched.

DCAT¹⁰⁸ is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. Its basic use case is to enable cross-data portal search for data sets and to make public sector data better searchable across borders and sectors. This can be achieved by the exchange of descriptions of datasets among data portals. By using DCAT to describe datasets in data catalogues, publishers increase discoverability and enable applications to easily consume metadata from multiple catalogues. DCAT further enables decentralised publishing of catalogues. It is now widely used as a metadata specification in Open Data portals. DCAT is a W3 recommended specification.

The DCAT Application profile for data portals in Europe (DCAT-AP)¹⁰⁹ is a specification based on the Data Catalogue vocabulary (DCAT) for describing public sector datasets in Europe. The application profile is a specification for metadata records to meet the specific application needs of data portals in Europe while providing semantic interoperability with other applications on the basis of reuse of established controlled vocabularies (e.g. EuroVoc) and mappings to existing metadata vocabularies (e.g. Dublin Core, SDMX, INSPIRE metadata, etc.).

Specific application profiles of DCAT to geospatial information (GeoDCAT) and statistics (StatDCAT) have been developed, assuring interoperability with ISO19115 and SDMX respectively.

It is suggested to propose DCAT-AP as the overarching vocabulary to achieve interoperability between statistical and geospatial metadata. This would have the advantage that geospatial and statistical resources following this approach would automatically be ready for publication in Open Data Portals implementing DCAT.

Besides facilitating discoverability of data, more efforts need to be invested in understanding the users' community in order to provide relevant data and appropriate guidance on the use of it. Usability was the core theme of the "What if...?" sessions at the 2017 INSPIRE Conference. The conclusions converged around the need to make INSPIRE more user-centric, to focus on data content, improve communication and improve its success stories¹¹⁰. The GEOSTAT 2 project argued that a good understanding of users' needs is required in order to be aware not only of *what* to deliver, but also *when*, *how*, and, perhaps most importantly, *why*¹¹¹. However, approaching the stakeholders of geospatial statistics may be particularly challenging, not only because the users' community is heterogeneous and sometimes difficult to identify, but also because the typical end-user of geospatial statistics consumes statistical information in a way that may be quite different from using traditional, non-geospatial statistics, involving GIS software and integrated spatial analyses. A population grid or census small statistical areas provide limited values unless they are analysed together with other geospatial information. The statistical and geospatial communities have a shared responsibility to respond to the users and their demands.

¹⁰⁸ <http://www.w3.org/TR/vocab-dcat/>

¹⁰⁹ <https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe>

¹¹⁰ JRC 2018.

¹¹¹ EFGS/GEOSTAT 2 2017. A Point-based Foundation for Statistics - Final report from the GEOSTAT 2 project

Recommendations

- Recommendation 5.4.1 - DCAT-AP should be used as the overarching vocabulary to achieve interoperability between statistical and geospatial metadata and support discovery in open data portals. DCAT extensions to statistical and geospatial data (GeoDCAT-AP¹¹² and StatDCAT-AP¹¹³ respectively) should be considered, allowing a richer data description and maintaining easy interoperability with the generic DCAT-AP profile (see use case 4.5 in Annex 2).
- Recommendation 5.4.2 - It is advisable to establish procedures for systematic consultations with the geospatial statistics users' community. Consultations with the users can be conducted in numerous and more-or-less formalised ways. They can take the form of user councils, focus groups or information seminars, etc.
- Recommendation 5.4.3 - The ESS should investigate the need for an EU official geospatial statistics portfolio based on user needs analysis. User centred product design might be a method to better meet user requirements of geospatial statistics.
- Recommendation 5.4.4 - Statistical and geospatial communities should reach out to users about the possibilities in use of the infrastructure information elements, tools and geospatial services/APIs, encourage online consumption into end user applications.

¹¹² <https://joinup.ec.europa.eu/release/geodcat-ap/v101>

¹¹³ <https://joinup.ec.europa.eu/solution/statdcat-application-profile-data-portals-europe>

3 Proposed procedure for the ESS on the implementation of the GSGF

Besides the technical and methodological recommendations relating to the principles of the GSGF, the GEOSTAT 3 project has also identified a number of generic recommendations relating to governance and the process of implementing the GSGF in Europe:

- The ESS should adopt the methodological aspects of the GSGF Europe as an official ESS methodology, and reach out broadly to seek support for endorsement also from other bodies working on statistical-geospatial integration in a European context, most notably UN-GGIM: Europe, UNECE and EuroGeographics. A broad consensus is beneficial for the implementation as well as a clear mandate to enforce the provisions of the framework.
- The ESS should work together with UNECE and UN-GGIM: Europe to support implementation of the GSGF Europe also in non-ESS European countries.
- Consistency between this implementation guide and other relevant methodological frameworks such as the GSBPM and CSPA will be vital for achieving the common goal of better integration of statistical and geospatial information.
- The ESS should work on an implementation plan for the GSGF Europe. Building on the results from the GEOSTAT 3 project, an officially agreed road map, including milestones and priorities, should be drafted. This road map should consolidate the recommendations, from this project into an implementation plan. Due to the cross-cutting nature of the implementation guide, such a road map needs to be supported jointly by the geospatial and statistical community, nationally and internationally.
- A number of recommendations provided in the proposal will require further elaboration through operational projects or collaboration initiatives involving both the statistical and geospatial community before implementation in a harmonised way. Some of these tasks should be addressed within the planned GEOSTAT 4 project, including:
 - Test requirements in actual production;
 - Develop show cases
 - Carry out benchmarking between countries
 - Work on operational workflows and tools;
 - Training.
- The implementation of the GSGF in Europe needs to be carefully and continuously monitored. A quality assessment framework for measuring the current situation and future performance of NSIs and the ESS towards achieving the recommendations of the GSGF needs to be developed. This task is one of the proposed goals of the forthcoming GEOSTAT 4 project.

4 Conclusions

The implementation guide presented in this report has the status of a proposal from the GEOSTAT 3 project. Hence, GEOSTAT 3 has been tasked to draft the requirements and recommendations, whereas the mandate to enforce them is out of scope for the project. As suggested in chapter 3, the GEOSTAT 3 project recommends the ESS to endorse the proposal and to reach out to seek support for endorsement also from other stakeholders working on statistical-geospatial integration in a European context, most notably UN-GGIM: Europe, UNECE and EuroGeographics.

The GSGF is still under development and the management of geospatial information in is still developing in many countries due to emerging technologies and new data sources. The implementation guide should therefore be regarded as a first attempt (1.0) to outline a consistent road map for implementation of the GSGF in Europe, addressing the most fundamental conditions and high priority issues identified through the work of the GEOSTAT 3 project. The proposal cannot, and does not, intend to cover the full scope of relevant aspects related to integration of geospatial and statistical information. Accordingly, the implementation guide should evolve as a living document, to be gradually extended and revised, to stay relevant in an ever changing technical and institutional landscape in the years ahead.

As stressed in Chapter 1, the GSGF Europe does not intend to replace the global guidance currently under preparation by the UN EG-ISGI. It should rather be considered as an extension to the global guidance, addressing the regional specifics of Europe and proposing common solutions to common, regional challenges. However, the implementation guide does not override the need for countries to define their own guidelines for implementation on national level.

The goal of GEOSTAT 3 has been to discuss the principles of the GSGF in the broadest possible way and be as inclusive as possible to all aspects of geospatial statistics and the integration of statistical and geospatial information. This report therefore represents the most comprehensive overview of the situation of statistical-geospatial data integration in Europe. Still the situation between countries varies greatly, and some countries may need to implement additional measures not yet identified here before having the full framework in place. As a result, implementing all recommendations may not necessarily be sufficient for a complete and full integration of statistical and geospatial information management in statistical production and for an operational GSGF in all countries.

The implementation guide identifies some 20 requirements and 80 recommendations. The requirements and recommendations address a wide range of issues and goals, some of which are fairly concrete while others are more strategic. Some concern statistical production, others the way how organisations should work together. It may seem like an overwhelming number of tasks, but the implementation of the GSGF in Europe does not require a big-bang approach and a complete redesign of enterprise architectures, production processes and legislation.

So what should NSIs and geospatial agencies do next?

Small and stepwise improvements are possible, or even recommended, and countries are advised to start with simple “traffic light assessments” of the implementation guide to find out to which degree they have already succeeded in implementing the proposed recommendations, or to find out where they have the biggest challenges in a national context. Such exercise could preferably be conducted by NSIs and geospatial agencies together.

The results and experiences from these smaller implementation steps should also feed-back into the future revisions of the GSGF and GSGF Europe, e.g. into the planned GEOSTAT 4 project.

Experts and managers in NSIs and geospatial agencies will find many good examples in Annex 2 that contains a collection of national good practices for all principles of the GSGF. They should look at these individual use cases as a source for inspiration for future improvement projects that suit best their national context. Already by implementing a few recommendations set out in this implementation guide, countries can already improve greatly the condition for the production of high-quality geospatial statistics and for the integration of geospatial information management into statistical production. The table in Annex 1 highlights those recommendations with priority 'immediate action' that yield the best return on investment. This system should help countries in defining a roadmap and action plan that fits best their specific situation and will result in the biggest improvements.

It is also worth to revisit the results from the GEOSTAT 2 project, providing rich guidance on how to set up and use a point-based foundation for statistics. Both projects share the same goal to provide rational solutions for a better integration of statistical and geospatial information. Also the GEOSTAT 2 project compiled a collection of good practice cases, of which most are still relevant as illustrations to the implementation of the GSGF.

The GEOSTAT 3 project consortium wishes all colleagues a lot of success on the long but rewarding journey to implement the provisions of the Global Statistical Geospatial Framework.

5 References

- ABS 2018. SSF Guidance Material – Geocoding Unit Record Data Using Address and Location. Third release.
([http://www.abs.gov.au/websitedbs/d3310114.nsf/home/Statistical+Spatial+Framework+Guidance+Material/\\$File/Geocoding+Unit+Record+Data.pdf](http://www.abs.gov.au/websitedbs/d3310114.nsf/home/Statistical+Spatial+Framework+Guidance+Material/$File/Geocoding+Unit+Record+Data.pdf))
- Antal, L et al (2017). Statistical disclosure control methods for harmonised protection of census data. Harmonised protection of census data in the ESS.
- Coady, I (2018a). Statistical Domain Working Group Charter. OGC Project Document 18-064.
- Coady I, (2018b). OGC Statistical Domain Working Group. Presentation at the GISCO Working Party 2018 (<http://un-ggim-europe.org/sites/default/files/Item12-UN-GGIM%20-%20OGC%20Standards%20-%20Coady.pdf>)
- EFGS/GEOSTAT 2 (2017). A Point-based Foundation for Statistics - Final report from the GEOSTAT 2 project.
- JRC (2012). A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures.
- JRC (2018). INSPIRE – what if? Summary report from the What if...? sessions at the 2017 INSPIRE conference.
- Lukan, J & Smukavec, A (2017). Statistical disclosure control considerations of publishing data on grid squares and territorial units in the application STAGE. Work Session on Statistical Data Confidentiality 20-22 September 2017.
- ONS (2016). ONS Linked Statistical Data Pilot – Final Report.
- Petri, E (2017). Statistical Disclosure Control on 2021 EU Census grid data – the fast track. Presentation at the 2017 EFGS conference, Dublin.
- Statistics Netherlands (2006). Impact analysis Table Joining Service (https://www.efgs.info/wp-content/uploads/news/2016/NL_ImpactAnalysisTJS_v1.0.docx-1.pdf).
- UNECE (2016a). Statistical and Geospatial Information – an Australian perspective on challenges and opportunities. Note by the Australian Bureau of Statistics.
- UNECE (2016b). In-depth review of developing geospatial information services based on official statistics. Note by the UK Office for National Statistics.
- UNECE (2018). <https://statswiki.unece.org/>.
- UN-GGIM (2018). Global Fundamental Geospatial Data Themes (v1.0).
- UN-GGIM & World Bank (2018). Integrated Geospatial Information Framework – A Strategic Guide to Develop and Strengthen National Geospatial Information Management. Part 1: Overarching Strategic Framework.

UN-GGIM: Europe (2016). Core Data Scope. Working Group A - First Deliverable of Task 1.a. Version 1.2 - 2016-07-27.

UN-GGIM: Europe (2017). Core Spatial Data Theme Address. Recommendation for Content. Version 1.0 2017-11-10.

UN-GGIM: Europe (2018). Core Spatial Data Theme Statistical Units (SU) Recommendation for Content. Version 0.6 - 2018-08-17 (draft)

UN-GGIM: Europe (2019). The integration of statistical and geospatial information – a call for political action in Europe. Draft.

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

Vale, S (2018). The review of the GSBPM and other statistical standards. Presentation at EFGS 2018 conference.

Annex 1. List of requirements and recommendations

This annex contains a full list of all requirements and recommendations and the proposed time-frame for their implementation. For each recommendation is also indicated the institution(s) mainly responsible for its implementation.

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
P 1	Requirement 1.1 - Use data from National Spatial Data Infrastructures	Recommendation 1.1.1 - Any geospatial information used to obtain locations for statistical and administrative data (geocoding), or to produce statistical content, should build on relevant, authoritative and INSPIRE compliant geospatial data and services from NSDI. INSPIRE compliance includes standardised and agreed formats, coordinate reference systems, metadata elements, data models and exchange services such as discovery, view and download services.	Immediate action	No	Geospatial agencies or other institutions responsible for production of geospatial data for the NSDI
P 1	Requirement 1.1 - Use data from National Spatial Data Infrastructures	Recommendation 1.1.2 – Should reference data needed for geocoding of statistical information not exist within a country or should the quality not be adequate for geocoding purpose, it is of utmost importance that such information be collected, enhanced, improved and provided through the NSDI. Priority should be given to the reference data listed as Core data by UN-GGIM: Europe (Address, Building and Cadastral parcel), and meeting Core data or better quality.	Immediate action	No	Geospatial agencies or other institutions responsible for production of geospatial data for the NSDI
P 1	Requirement 1.1 - Use data from National Spatial Data Infrastructures	Recommendation 1.1.3 – Specifications of INSPIRE, on the implementation of systematic management of unique identifiers for features used for geocoding, and corresponding (draft) recommendations by UN-GGIM: Europe Core data should be applied to facilitate integration of geospatial and statistical information through consistent identifier-key relationships, and to keep track of corrections in location changes as well as lifecycle of each statistical unit, i.e. at the micro-data level.	Immediate action	No	Geospatial agencies or other institutions responsible for production of geospatial data for the NSDI
P 1	Requirement 1.1 - Use data from National Spatial Data Infrastructures	Recommendation 1.1.4 - The roles and responsibilities of different agencies involved in production of geospatial information should be well defined, e.g. who maintains what information and how often data are updated. A custodianship model and stewardship model may need to be	Immediate action	No	NSIs, geospatial agencies or other institutions responsible for production of geospatial data for the NSDI

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
		established, in order to identify the most relevant stakeholder for a geospatial data source.			
P 1	Requirement 1.2 - Use point-based location data for geocoding	Recommendation 1.2.1 - A point-based infrastructure (point locations) should be adopted as main and preferred approach for geocoding in the ESS, and is advised also for non-ESS European countries. Use of more general location descriptions, and/or larger geographies (such as enumeration areas or other statistical geographies), should be considered only as a complementary or secondary approach when point-based geocoding fails because of partially missing data. Countries should agree on one single uniform national infrastructure for geocoding of all public and potentially private data.	Immediate to intermediate action	Yes, some Member states still lack the necessary infrastructure to fully implement a point-based foundation	NSIs, geospatial agencies or other institutions conducting geocoding
P 1	Requirement 1.2 - Use point-based location data for geocoding	Recommendation 1.2.2 - In order to implement a point-based foundation for statistics, authoritative information on physical address locations, buildings and/or cadastral parcels should be made available within the NSDIs. In addition, information need to be accurate and consistent, have sufficient coverage and meet internationally and nationally agreed standards.	Immediate action	No	NSIs, geospatial agencies or other institutions responsible for production of geospatial data for the NSDI
P 1	Requirement 1.2 - Use point-based location data for geocoding	Recommendation 1.2.3 - All countries should have a single, national, authoritative, universal address register available for public institutions to include in their respective business processes (see good practice cases).	Immediate action	No	Geospatial agencies or other institutions responsible for production of geospatial data for the NSDI
P 1	Requirement 1.2 - Use point-based location data for geocoding	Recommendation 1.2.4 - Geospatial agencies are encouraged to set up and provide national geocoding services, based on authoritative location data, within a service-oriented architecture referring to common reference data sets, common service configuration and common guidance for application. The ongoing initiative on a new OGC standard for geocoding APIs should be taken into consideration.	Immediate action	Yes, new OGC standard for geocoding APIs	Geospatial agencies or other institutions responsible for production of geospatial data for the NSDI
P 1	Requirement 1.2 - Use point-based location data for geocoding	Recommendation 1.2.5 - National geocoding services should be open to authorities of other Member States in order to rely on similar methods and tools for the geocoding inside and outside countries and obtain consistent results. This would allow e.g. the recording of workplace addresses of citizens working abroad. Common services could provide a	Immediate to intermediate action	No	Geospatial agencies or other institutions responsible for production of geospatial data for the NSDI

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
		better basis for cross-border geocoding, hence improving calculation of statistics on cross-border commuting and migration.			
P 1	Requirement 1.3 - Build formal working relationships on institutional agreements	Recommendation 1.3.1 - NSIs and geospatial agencies should work actively to increase mutual exchange of knowledge between the geospatial and statistical community by initial or continuous training, information, communication or by working on concrete cooperative projects. Cooperation should be oversight and monitored by regular national and European steering groups composed of senior experts from both communities.	Immediate	No	NSIs, geospatial agencies or other institutions responsible for production of geospatial data for the NSDI
P 1	Requirement 1.3 - Build formal working relationships on institutional agreements	Recommendation 1.3.2 - Agreements between NSIs and geospatial agencies should cover terms of access, licensing, governance and use of geospatial information. Agreements may also need to involve other stakeholders, such as municipalities or regional bodies responsible for data provision. Data from the NSDIs need to be easily accessible and usable, for NSIs or other public institutions conducting geocoding, at a low or affordable cost but preferably free of charge.	Immediate action	No	NSIs, geospatial agencies or other institutions responsible for production of geospatial data for the NSDI
P 1	Requirement 1.3 - Build formal working relationships on institutional agreements	Recommendation 1.3.3 - The GEOSTAT 3 project suggests that the fruitful collaboration between statistical and geospatial agencies on a European level established through the UN-GGIM: Europe Work Group on Data Integration, should continue after the present task is completed. The group should act as the European steering group (mentioned in recommendation 1.3.1) to facilitate and monitor the continuous and mutual exchange of knowledge between the geospatial and statistical community on European level.	Immediate action	No	UN-GGIM: Europe
P 2	Requirement 2.1 - Build an effective and secure data management environment	Recommendation 2.1.1 - Persistent storing of a high precision geocode for each statistical unit record in statistical and administrative datasets (i.e. a person, household, business etc.) should be the main and preferred approach for ESS Member States and is advised also for non-ESS European countries. A high precision geocode implies reference to an address location, building/dwelling or cadastral parcel.	Immediate to intermediate action	Yes, some Member states still lack the necessary infrastructure to fully implement a point-based foundation	NSIs or other institutions responsible for geocoding and storing of unit record data.

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
P 2	Requirement 2.1 - Build an effective and secure data management environment	Recommendation 2.1.2 - An efficient data management environment should allow linking of statistical and spatial objects at unit record level without compromising privacy of micro data. Implementation of data warehouse solutions could be an effective way to combine a widespread use of geocoded micro data with confidence that proper privacy measures has been applied in the data architecture.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs or other institutions responsible for geocoding and storing of unit record data.
P 2	Requirement 2.1 - Build an effective and secure data management environment	Recommendation 2.1.3 - Consistent synchronisation procedures should be assured for data sources involved, as to maintain the relationship between the geospatial infrastructure and the unit record data. The relation between microdata and statistical or administrative geographies should also be synchronised.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs or other institutions responsible for geocoding and storing of unit record data.
P 2	Requirement 2.2 - Store location only once	Recommendation 2.2.1 - Location data objects should be recognised and fully integrated in the general data architecture of NSIs in order to facilitate design of efficient workflows for data integration and geocoding.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs or other institutions responsible for geocoding and storing of unit record data.
P 2	Requirement 2.2 - Store location only once	Recommendation 2.2.2 - It is recommended to build repositories for location data (geocoding databases) holding references to a number of relevant and common administrative and statistical geographies also back-in-time, at each location data object, to allow simplified aggregations of data also by non-geospatial experts.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs or other institutions responsible for geocoding and storing of unit record data.
P 2	Requirement 2.2 - Store location only once	Recommendation 2.2.3 - Address services or geocoding services provided by geospatial agencies need to fully support use of life-cycle attributes and versioning. Serving only the most up-to-date information is not sufficient to incorporate such services in statistical business processes. Hence, obsolete address objects should be retained in the data and their current status indicated using the INSPIRE mechanism of life-cycle attributes and versioning. Pre-allocated or provisional addresses, where available, should be managed in the same way.	Immediate to intermediate action	Yes, some Member states still lack the necessary infrastructure to fully implement a point-based foundation.	Geospatial agencies
P 2	Requirement 2.2 - Store location only once	Recommendation 2.2.4 - Statistical microdata transmitted to Eurostat by ESS Member States should come with a reference to the ETRS89 1km ²	Immediate action	No	NSIs

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
		grid cell code as a minimum, if point references are not possible to provide.			
P 2	Requirement 2.3 - Ensure consistency and quality of geocoding results	Recommendation 2.3.1 - Member States should develop and apply national guidelines for geocoding workflows in order to ensure a consistent and conform result within and between institutions. Such guidelines may include agreed decisions on what location data services to use to geo-enable which statistical information. They may also include agreed ad hoc methods to improve matching between location data and unit record data.	Immediate to intermediate action	Yes, some Member states still lack the necessary infrastructure to fully implement a point-based foundation.	NSIs, geospatial agencies or other institutions conducting geocoding operations.
P 2	Requirement 2.3 - Ensure consistency and quality of geocoding results	Recommendation 2.3.2 – Defining geocoding guidelines on European level should be considered in order to ensure coherence and interoperability between countries. Drafting such guidelines requires a careful assessment of national conditions and practices.	Immediate to intermediate action	Yes, collection and assessment of national practices will be needed.	Eurostat and UN-GGIM: Europe
P 2	Requirement 2.3 - Ensure consistency and quality of geocoding results	Recommendation 2.3.3 - Geocoding results should be as accurate and consistent as possible and documented according to agreed geocoding metadata. Geocoding metadata should be provided at object level so that the accuracy of the assigned location can be assessed for each observation.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs or other institutions responsible for geocoding and storing of unit record data.
P 2	Requirement 2.4 - Develop consistent approaches to manage non-matching data	Recommendation 2.4.1 - For the provision of data to Eurostat for Census 2021, the ESS Member States should follow the agreed approach to use one additional synthetic grid cell without spatial representation to contain all persons that cannot be geocoded to normal grid cells.	Immediate action	No	NSIs
P 2	Requirement 2.4 - Develop consistent approaches to manage non-matching data	Recommendation 2.4.2 - Countries should define and describe a consistent approach for non-matching observations to be applied in production of small area or grid data on national level, for those cases where the approach set for European data is not applicable or suitable for national purposes.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs or other institutions responsible for geocoding and storing of unit record data.
P 2	Requirement 2.5 - Use point-of-entry validation in	Recommendation 2.5.1 – When creating and maintaining administrative or statistical records, point-of-entry validation mechanisms should be used to ensure the best possible quality of the location references (address, building ID etc.) stored in unit record data. All national public	Immediate to intermediate action	No, but there is a need for good practise on solutions	NSIs, geospatial agencies and other institutions responsible for collection and management of administrative data

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
	collection of administrative or statistical data	authorities in ESS Member States in charge of recording addresses into public files should be obliged to use the uniform geocoding infrastructure for entering addresses, to avoid inconsistencies. This policy is also strongly advised for non-ESS European countries.		to share between institutions and countries	
P 2	Requirement 2.5 - Use point-of-entry validation in collection of administrative or statistical data	Recommendation 2.5.2 - The statistical and geospatial communities should collaborate to promote use of authoritative location data among public institutions collecting and managing administrative information.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs and geospatial agencies
P 2	Requirement 2.5 - Use point-of-entry validation in collection of administrative or statistical data	Recommendation 2.5.3 - Address services or geocoding services provided by geospatial agencies need to be available with defined APIs and accessible for public custodians of administrative data for easy integration in their data collection platforms.	Immediate action	Yes, some countries still lack the necessary authoritative data to fully implement geocoding services. Also, the current work on an OGC geocoding standard should be taken into account.	Geospatial agencies
P 2	Requirement 2.5 - Use point-of-entry validation in collection of administrative or statistical data	Recommendation 2.5.4 - To create strong incentives for the whole society (civil society and private sector) to use and implement authoritative national address registers in their business, release of address data under open data licenses should be considered.	Immediate to intermediate action	Yes, some countries does not have the business models to provide address data under open licenses.	Geospatial agencies
P 2	Requirement 2.5 - Use point-of-entry validation in collection of administrative or statistical data	Recommendation 2.5.5 - Legal measures should be considered as means to enforce good quality of collected data. Such measures may concern both removal of legal barriers preventing point-of-entry validation to be conducted as well as legal instruments to enforce use of point-of-entry validation in public sector.	Immediate to intermediate action	No, but legal structures are complex and may take time to change.	National government policy institutions
P 3	Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies	Recommendation 3.1.1 - Geospatial agencies and NSIs should define clear custodianship roles for coding systems and boundary data for statistical and administrative geographies respectively, on national level in order to enable a more efficient collaboration.	Immediate action	No	NSIs and geospatial agencies

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
P 3	Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies	Recommendation 3.1.2 - All national administrative, statistical and functional geographies with relevance for production and dissemination of official statistics should be provided as authoritative geospatial data in compliance with the technical specifications of INSPIRE, whenever relevant, and the UN-GGIM: Europe Core data Recommendation for Content on Statistical Units and Administrative Units, including full topological and coding consistency.	Immediate action	No	NSIs and geospatial agencies
P 3	Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies	Recommendation 3.1.3 - When changes occur in coding systems or geometry of boundaries, data on national statistical and administrative geographies should be available to all users no later than six months after the reference date.	Immediate action	No	NSIs and geospatial agencies
P 3	Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies	Recommendation 3.1.4 - Geospatial agencies and NSIs should work together to improve accessibility, speed of delivery and usability of national data on administrative and statistical geographies, with the aim to provide high quality data under Open Data licenses. Data should be machine-readable and provided through OGC/INSPIRE compliant view and download services.	Immediate action	No	NSIs and geospatial agencies
P 3	Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies	Recommendation 3.1.5 - Geospatial agencies and NSIs should jointly start to explore the potential of dissemination of statistical and administrative geographies as Linked Open Data. In order to do so, the initiatives already being conducted in several ESS Member States should be evaluated and if needed, guidance and recommendations to be developed to facilitate harmonisation.	Intermediate to long-term action	Yes, more efforts needed to harmonise and standardise work on LOD. Also, consolidation and better exploitation of existing knowledge.	NSIs and geospatial agencies
P 3	Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies	Recommendation 3.1.6 - Efforts to set up a consistent framework of national statistical and administrative geographies should give priority to current data, nevertheless inclusion of historical geographies is recommended, as these data are crucial for recasting current (point-based) data for past administrative or statistical divisions.	Intermediate action	No	NSIs and geospatial agencies
P 3	Requirement 3.1- Set up and maintain a consistent framework of national	Recommendation 3.1.7 - Geospatial agencies and NSIs should agree on an approach on scale, reference dates and accuracy of administrative and statistical geographies building on UN-GGIM: Europe Core data	Immediate to Intermediate action	Yes, work needed to define scale, accuracy etc.	NSIs, geospatial agencies and European stakeholders.

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
	statistical and administrative geographies	recommendations. For analytical purpose, data capture and data processing, highest possible accuracy of boundaries is needed (Master Level 0 or 1). For dissemination and visualisation purpose, a generalised level (simplified geometries) is desirable.			
P 3	Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies	Recommendation 3.1.8 - The processes and elementary building blocks for the construction of the lowest national administrative layer should be clearly described and properly documented in the metadata, including problems with under-coverage of data.	Immediate action	No	Geospatial agencies
P 3	Requirement 3.1- Set up and maintain a consistent framework of national statistical and administrative geographies	Recommendation 3.1.9 - Linked to the provision of statistical geographies, ESS Member States should also provide area statistics on the size of these geographies based on a harmonised methodology and national data. Land area statistics are important for density information (population density, etc.).	Immediate action	No	NSIs and geospatial agencies
P 3	Requirement 3.2 - Improve maintenance of the European framework of statistical geographies	Recommendation 3.2.1 - The coordination between NSIs and geospatial agencies in production of administrative units (LAU and NUTS) should improve to respond more efficiently to changes of administrative units. When changes occur in coding systems or boundaries, data on national statistical and administrative geographies should be available to all users no later than six months after the reference date.	Immediate action	No	NSIs and geospatial agencies
P 3	Requirement 3.2 - Improve maintenance of the European framework of statistical geographies	Recommendation 3.2.2 - All European geospatial agencies are encouraged to support the current work on developing Open European Location Services (Open ELS) coordinated by EuroGeographics, which would be the first operational step towards the implementation of the European Location Services (ELS), by targeting national open datasets and services first. If the Open ELS continues to an operational stage, it would provide a single access point for open national data on administrative geographies, cadastral parcels, addresses and buildings. In addition, it will give access to a pan-European regional geolocator service, and to a cadastral index map.	Immediate action	No	EuroGeographics and geospatial agencies

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
P 3	Requirement 3.2 - Improve maintenance of the European framework of statistical geographies	Recommendation 3.2.3 - All NSIs within the ESS are encouraged to support the development of EU legislation currently under way on the regular and frequent production of population grids after 2021 to be able to update the territorial typologies on a regular basis. Such development is also advised for non-ESS European countries.	Intermediate to long-term action	Yes, regulations for regular updates on population grids needed for ESS	Eurostat and NSIs
P 3	Requirement 3.3 - Consolidate use of existing statistical grid systems and explore the potential of evolving global grid systems	Recommendation 3.3.1 - For the time being, the current European ETRS89 grid system should remain the main output grid geography for pan-European statistics produced within the ESS. The grid system is already settled with the INSPIRE principles (Statistical Units specifications) and it is a well-established feature among users of pan-European geospatial statistics.	Immediate action	No	NSIs and Eurostat
P 3	Requirement 3.3 - Consolidate use of existing statistical grid systems and explore the potential of evolving global grid systems	Recommendation 3.3.2 - Introduction of additional grid sizes should be considered and agreed on for European level (e.g. 100 m, 125 m or 200 m or quad-tree) in line with the forthcoming revision of the INSPIRE implementing regulation on data interoperability for Statistical Units. The need for, and implications of, introducing additional grid sizes should be explored jointly by the statistical and geospatial community, taking into account a versatile and expanding use of spatio-statistical data and resources, even regarding earth observation data taking into account disclosure issues.	Intermediate action	Yes, work needs to be undertaken to define common concepts for additional grid sizes.	Eurostat, UN-GGIM: Europe, NSIs and geospatial agencies
P 3	Requirement 3.3 - Consolidate use of existing statistical grid systems and explore the potential of evolving global grid systems	Recommendation 3.3.3 - The geospatial and statistical communities should monitor the development of the DGGS and its application closely in order to prepare for a possible future implementation of the grid system for national and European data. Implementation of the DGGS will potentially be beneficial for the global data integration and for a seamless integration between the national and European level, however it is still at an early stage and more studies are needed on the benefits and implications of adopting the concept.	Intermediate to long-term action	Yes, work needs to be undertaken to prepare for possible future implementation of DGGS.	Eurostat, NSIs, geospatial agencies, EFGS, UN-GGIM: Europe
P 3	Requirement 3.3 - Consolidate use of existing statistical grid systems and	Recommendation 3.3.4 - The European ETRS89 grid system, or the International Terrestrial Reference System (ITRS) for over-sea territories, should be the main and preferred output grid geography on national level within ESS Member States. The European ETRS89 grid system is also recommended for non-ESS European countries. In case national grid	Immediate action	No	NSIs and geospatial agencies

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
	explore the potential of evolving global grid systems	systems are used as a complement to European grid systems, these grids (including their coding systems) should be made available as authoritative, open data in order to ensure a common use on national level. Statistics disseminated on the EU grid system e.g. to meet EU legislation should have quality equal to statistics disseminated in national grid systems.			
P 3	Requirement 3.3 - Consolidate use of existing statistical grid systems and explore the potential of evolving global grid systems	Recommendation 3.3.5 - The parallel use of national, EU and global grid systems for statistics should be coordinated in the ESS to ensure that all grids have comparable quality.	Immediate action	No	NSIs
P 4	Requirement 4.1 - Improve geospatial workflows within statistical production	Recommendation 4.1.1 - The statistical community should engage more actively when new geospatial standards with relevance for data integration are developed and involve in evaluation of utility to statistical production (e.g. OGC TJS).	Immediate action	No	Eurostat, UNECE, NSIs
P 4	Requirement 4.1 - Improve geospatial workflows within statistical production	Recommendation 4.1.2 - Forum for discussion and engagement regarding cross-domain issues are crucial to improve interoperability and should be provided through the UNECE, ModernStats, ESS, UN-GGIM, EFGS and initiatives like the OGC Statistical Domain Working Group.	Immediate action	No	UNECE , ModernStats, ESS, UN-GGIM
P 4	Requirement 4.1 - Improve geospatial workflows within statistical production	Recommendation 4.1.3 - Statistical production involving geospatial information should rely on both statistical models and standards as well as existing geospatial standards, e.g. in geospatial data collection and dissemination technologies.	Immediate action	No	Eurostat, UNECE, NSIs
P 4	Requirement 4.1 - Improve geospatial workflows within statistical production	Recommendation 4.1.4 - Interoperability requires consideration of geospatial data and the concept of location as a natural part of logical data warehouses and data architecture.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
P 4	Requirement 4.1 - Improve geospatial workflows within statistical production	Recommendation 4.1.5 - The General Statistical Business Process Model (GSBPM) should be developed and enhanced to make greater reference to the use of geospatial data and methods in the statistical production process.	Immediate action	Yes, review is currently in progress	UNECE
P 4	Requirement 4.1 - Improve geospatial workflows within statistical production	Recommendation 4.1.6 - Geospatial services in a service-oriented architecture are recommended to standardise geospatial production components. NSIs should pursue to share common tools.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs and geospatial agencies
P 4	Requirement 4.2 - Enable data integration through consistent semantics and concepts across domains	Recommendation 4.2.1 - Definition of common conceptual models for objects fundamental for both statistical and geospatial communities are needed. Statistical and geospatial agencies should look for opportunities to work on semantic interoperability issues, for example, ontologies for addresses and buildings.	Immediate to Intermediate action	Yes, work needs to be done on different levels and by different institutions	NSIs, geospatial agencies, Eurostat/EC, UNECE, OGC, UN-GGIM
P 4	Requirement 4.2 - Enable data integration through consistent semantics and concepts across domains	Recommendation 4.2.2 - NSIs and geospatial agencies should work together to inform administrative data custodians on the need for conceptual harmonisation as to obtain semantic interoperability between communities and data sources.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs and geospatial agencies
P 4	Requirement 4.3 - Publish data once and leave it at its source to be reused many times	Recommendation 4.3.1 - Countries should intensify their efforts on service-oriented dissemination through APIs to provide machine-readable open data format for national geospatial statistics - Spatial Statistics as a service.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs and geospatial agencies
P 4	Requirement 4.3 - Publish data once and leave it at its source to be reused many times	Recommendation 4.3.2 - The adaption of SDMX for INSPIRE implementation, successfully developed and tested by Eurostat in the context of population grids for the Census 2021, should be used as solution for INSPIRE harmonisation of Population distribution in the ESS. It is assumed to minimise the effects of double obligation on ESS Member States and to maximise the usability of the census information for the statistical and geospatial community.	Intermediate action	Yes, will be part of the Census 2021 programme	Eurostat and NSIs
P 4	Requirement 4.3 - Publish data once and leave it at its	Recommendation 4.3.3 - SDMX is evolving as data exchange format for a growing number of statistical domains. NSIs and Eurostat should work	Intermediate to long-term action	Yes, studies and projects needs to be undertaken	Eurostat and NSIs

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
	source to be reused many times	closely together to explore further applications of SDMX beyond the context of Census 2021.			
P 4	Requirement 4.3 - Publish data once and leave it at its source to be reused many times	Recommendation 4.3.4 – Open, centralised, INSPIRE compliant services with NUTS geographies and 1km and other sized grids should be set up by the ESS (or by other European body through agreement with ESS) for the whole of Europe, to avoid unnecessary duplication and efforts by the member states. Preferably this is done at different levels of detail. These services could be used for cross border mapping of Statistics when joined with the SDMX services.	Immediate action	No	Eurostat, EuroGeographics and UN-GGIM Europe
P 4	Requirement 4.4 - Increase use services for merging geography and statistics	Recommendation 4.4.1 - European bodies, national geospatial agencies and NSIs should work closely together on developing and applying services using automated and dynamic linking of geographies and statistical services and on deciding on an underpinning governance on the provision of these services. This will enable more flexible and open statistical-geospatial integration providing more value for users.	Immediate to Intermediate action	No	Eurostat, EuroGeographics, UN-GGIM Europe, NSIs and geospatial agencies
P 4	Requirement 4.4 - Increase use services for merging geography and statistics	Recommendation 4.4.2 - Though OGC Table Joining Services until now has not been widely implemented, the geospatial and statistical community should jointly involve in developing the standard and to consolidate and industrialise its implementation.	Immediate to Intermediate action	Yes, use of TJS may be part of the INSPIRE compliance of Population distribution – demography and Human health and safety themes by 2020.	OGC, Eurostat, EuroGeographics, NSIs and geospatial agencies
P 4	Requirement 4.4 - Increase use services for merging geography and statistics	Recommendation 4.4.3 - For harmonised European data under the INSPIRE themes Population distribution and demography, a combined European action should be considered by means of a TJS operating on the Eurostat SDMX web service instead of setting up numerous different national services. This would also be a good business case and a large scale test, possibly paving way for application of the same technology also within other statistical domains.	Immediate to Intermediate action	Yes, standards may need to be developed	Eurostat, EuroGeographics, NSIs and geospatial agencies
P 4	Requirement 4.4 - Increase use services for merging geography and statistics	Recommendation 4.4.4 - Existing services using SDMX datasets should not be modified to additionally implement INSPIRE data models. Also no parallel INSPIRE download service should be implemented, as SDMX and	Immediate action	Yes, data specifications may need to be altered	

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
		the SDMX services already meet INSPIRE requirements as they are already machine readable and harmonised.			
P 4	Requirement 4.5 - Explore the potential of Linked Open Data for increased interoperability	Recommendation 4.5.1 - Though Open Linked Data is still an emerging concept, geospatial agencies and NSIs should start exploring its potential. LOD and semantic web technology may be one fruitful way to improve the integration of geospatial and statistical datasets.	Intermediate to long-term action	Yes, more efforts needed to harmonise and standardise work on LOD. Also, consolidation and better exploitation of existing knowledge.	Eurostat, NSIs and geospatial agencies
P 4	Requirement 4.5 - Explore the potential of Linked Open Data for increased interoperability	Recommendation 4.5.2 - The ESS should increase its efforts to facilitate common tools, concepts and methods in the field of LOD. Current progress made in individual ESS Member States on both statistical and geospatial data needs to be consolidated and exploited in a more systematic way to stimulate harnessing of LOD and semantic web technology in the whole of Europe.	Intermediate to long-term action	Yes, more efforts needed to harmonise and standardise work on LOD. Also, consolidation and better exploitation of existing knowledge.	Eurostat, NSIs and geospatial agencies
P 4	Requirement 4.5 - Explore the potential of Linked Open Data for increased interoperability	Recommendation 4.5.3 - Building on experiences from ESS Member States, a good practice is to approach the work on LOD with one consistent reference dataset, e.g. statistical and administrative geographies.	Intermediate to long-term action	Yes, more efforts needed to harmonise and standardise work on LOD. Also, consolidation and better exploitation of existing knowledge.	Eurostat, NSIs and geospatial agencies
P 4	Requirement 4.5 - Explore the potential of Linked Open Data for increased interoperability	Recommendation 4.5.4 - As published statistical data and corresponding geographies may be maintained in different locations and by different organisations, a standard for coding of common geographies is needed for seamless integration of table information and related geographies by machine reading.	Intermediate to long-term action	Yes, more efforts needed to harmonise and standardise work on LOD. Also, consolidation and better exploitation of existing knowledge.	Eurostat, NSIs and geospatial agencies
P 5	Requirement 5.1 - Implement clear and simple data licensing policies	Recommendation 5.1.1 - ESS Member States should aim to release at least a core set of statistical variables (such as total population) for mid-resolution grids (1 km ²) or other small-area statistics under open data licenses. This is advised also for non-ESS European countries.	Immediate action	No	NSIs

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
P 5	Requirement 5.1 - Implement clear and simple data licensing policies	Recommendation 5.1.2 - Geospatial statistics should be released as open data. The license of geospatial statistics created by combining statistical and geospatial information shall be as little restrictive as possible considering the respective open data licenses of the source data combined.	Immediate action	No	NSIs
P 5	Requirement 5.1 - Implement clear and simple data licensing policies	Recommendation 5.1.3 - The data provider should promote and implement the inclusion of the license information (fees, access constraints, etc.) in the INSPIRE metadata for the geospatial data set and/or service offered.	Immediate action	No	NSIs, geospatial agencies and other producers of geospatial statistics products
P 5	Requirement 5.2 - Use service oriented data portals supporting dynamic integration of data	Recommendation 5.2.1 - Countries should explore use of service oriented dissemination platforms, providing a greater flexibility in terms of usability and supporting data access through a variety of interfaces via APIs. OGC compliant services and non-proprietary formats should be used (e.g. OGC Geopackage for file deliveries) for dissemination in order to ensure flexibility also from an end-user perspective.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs and geospatial agencies
P 5	Requirement 5.2 - Use service oriented data portals supporting dynamic integration of data	Recommendation 5.2.2 - The ESS should increase its efforts to facilitate common concepts and shared solutions on service-oriented and dynamic linking of data. Good practise and current progress made in individual Member States or elsewhere need to be consolidated and better exploited to stimulate development in the whole of Europe.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	Eurostat, UN GGIM, NSIs and geospatial agencies
P 5	Requirement 5.2 - Use service oriented data portals supporting dynamic integration of data	Recommendation 5.2.3 - EuroGeographics' work on establishing the Open ELS service might be an important step towards new, open cross-border products and services for pan-European applications. An open and service-based provision of the high-resolution EuroBoundaryMap would provide a strong business case for statistical-geospatial integration on pan-European level.	Immediate to Intermediate action	Yes, Open ELS is not operational yet and access and terms of data use is not clarified	EuroGeographics and geospatial agencies
P 5	Requirement 5.2 - Use service oriented data portals supporting dynamic integration of data	Recommendation 5.2.4 - NSIs and Eurostat should work together to develop concepts for 'Statistics as a service' including collection of good practice and proof of concepts for solutions.	Intermediate to long-term action	Yes, both collaboration projects is needed to develop such a concept as well as good practise on solutions to share	Eurostat and NSIs

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
				between institutions and countries	
P 5	Requirement 5.2 - Use service oriented data portals supporting dynamic integration of data	Recommendation 5.2.5 - Geospatial agencies and NSIs should consider to publish simple features as defined by OGC and in ISO 19125 next to the complex features as defined by the INSPIRE data models. This will improve usability in GIS systems. Also consider following modern encodings like LOD, WFS 3.0 and GeoJSON to improve accessible services also in the future.	Immediate action	No	NSIs and geospatial agencies
P 5	Requirement 5.3 - Define clear national and European rules to ensure protection of privacy	Recommendation 5.3.1 - For dissemination of Census data following EU regulation for 2021 population grids, ESS Member States should follow the recommendations on methods for Statistical Disclosure Controls provided by the ESS centre of excellence on SDC. The method proposed should be applicable in all NSIs, possibly with different parameters.	Intermediate action	Yes, will be part of the Census 2021 programme	Eurostat and NSIs
P 5	Requirement 5.3 - Define clear national and European rules to ensure protection of privacy	Recommendation 5.3.2 - For dissemination of national grid data or small area statistics, countries should define, describe and publish their own principles for the preservation of privacy with respect to existing national legislation and policy.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs
P 5	Requirement 5.3 - Define clear national and European rules to ensure protection of privacy	Recommendation 5.3.3 - ESS Member states need to be aware of potential confidentiality risks, due to geographical differencing, that come with an increasing number of national and European data being disseminated on small areas. The quality of European statistics should be given priority. Therefore national geospatial statistics should not be released/published if stricter disclosure control (to avoid geographical differencing) lead to loss of quality of the European statistics.	Immediate action	No	NSIs
P 5	Requirement 5.4 - Facilitate data search and use through cataloguing and improved guidance	Recommendation 5.4.1 - DCAT-AP should be used as the overarching vocabulary to achieve interoperability between statistical and geospatial metadata and support discovery in open data portals. DCAT extensions to statistical and geospatial data (GeoDCAT-AP and StatDCAT-AP respectively) should be considered, allowing a richer data description and maintaining easy interoperability with the generic DCAT-AP profile.	Immediate to Intermediate action	No	Eurostat/EC, NSIs and geospatial agencies responsible for publishing open data

GSGF Europe - Implementation guide for the Global Statistical Geospatial Framework in Europe

GSGF Principle	Requirement	Recommendation	Temporal scope for implementation	Further action before implementation	Responsible institution(s)
P 5	Requirement 5.4 - Facilitate data search and use through cataloguing and improved guidance	Recommendation 5.4.2 - It is advisable to establish procedures for systematic consultations with the geospatial statistics users' community. Consultations with the users can be conducted in numerous and more-or-less formalised ways. They can take the form of user councils, focus groups or information seminars, etc.	Immediate action	No, but there is a need for good practise on solutions to share between institutions and countries	NSIs and geospatial agencies
P 5	Requirement 5.4 - Facilitate data search and use through cataloguing and improved guidance	Recommendation 5.4.3 - The ESS should investigate the need for an EU official geospatial statistics portfolio based on user needs analysis. User centred product design might be a method to better meet user requirements of geospatial statistics.	Immediate to Intermediate action	No	Eurostat and NSIs
P 5	Requirement 5.4 - Facilitate data search and use through cataloguing and improved guidance	Recommendation 5.4.4 - Statistical and geospatial communities should reach out to users about the possibilities in use of the infrastructure information elements, tools and geospatial services/APIs, encourage online consumption into end user applications.	Immediate action	No	Eurostat, UN GGIM, EuroGeographics, NSIs and geospatial agencies