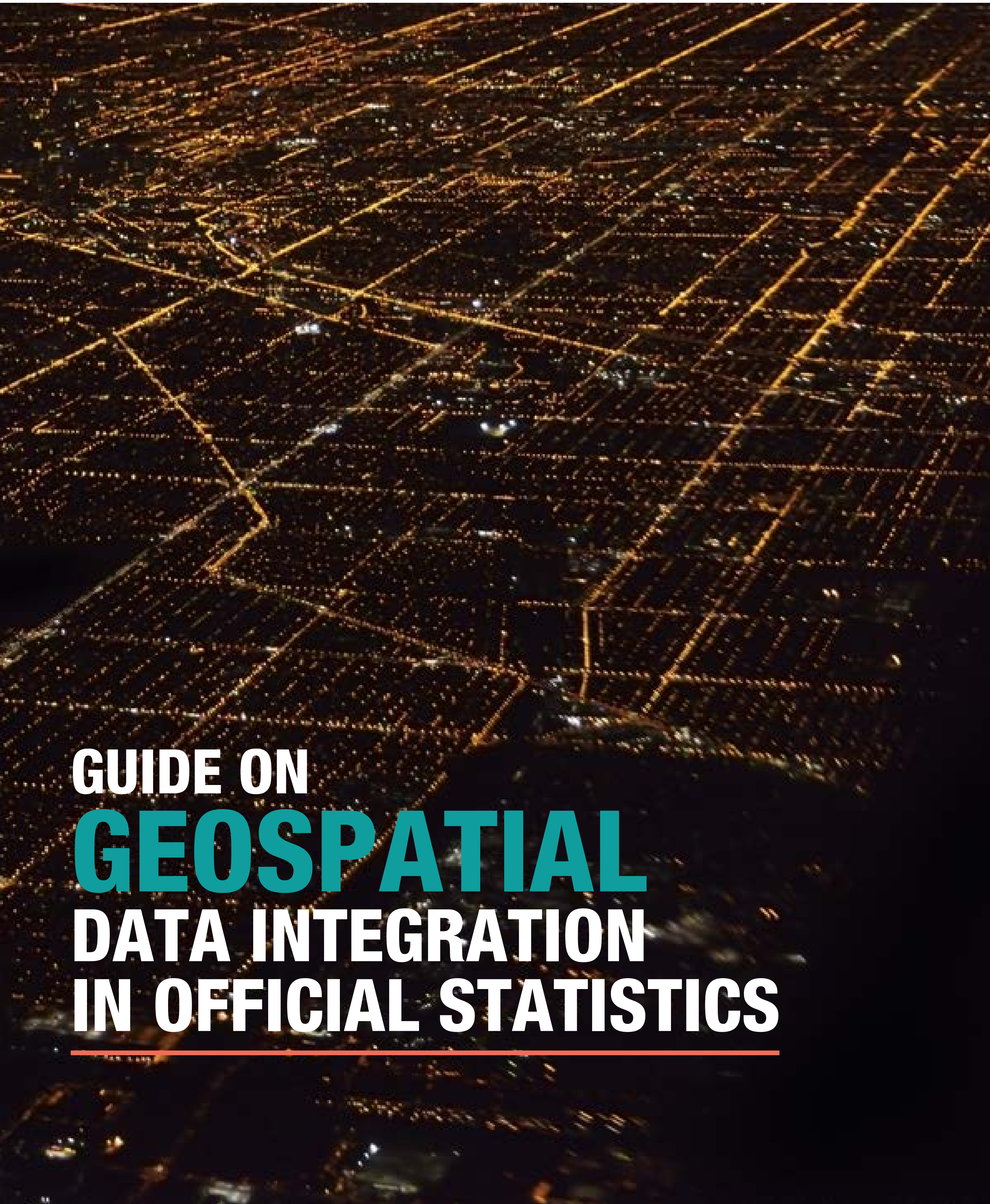
SCB

PARiS 21
Partnership in statistics for development in the 21st century



GUIDE ON GEOSPATIAL DATA INTEGRATION IN OFFICIAL STATISTICS

Acknowledgements

This guide is a joint product of the Partnership in Statistics for Development in the 21st Century (PARIS21) and Statistics Sweden. It was developed under the supervision of François Fonteneau, Deputy Head of PARIS21, and Cathy Kruger, International Cooperation Office, Statistics Sweden. The guide was prepared and drafted by a team led by Philippe Gafishi, Inter-Regional Advisor at PARIS21.

The lead author of the guide is Javier Carranza-Torres, PARIS21 consultant. Jerker Moström from Statistics Sweden and Jørn Kristian Undelstvedt from Statistics Norway provided the methodological design and the logical structure for the step-by-step guidance and both were closely involved in the final drafting process. They also suggested most of the literature review, pre-existing related guidelines and other documentation to turn this into a standalone guide. The UN Expert Group on Integration of Geospatial and Statistical Information provided feedback on the final outcome.

Alex Mudabeti, Director for Spatial Data at the Namibia Statistics Agency, provided valuable contents for Step 3. Ivan Ojeda-Aguilera, Director of INE in Paraguay, suggested insightful ideas for future developments. Useful writing style comments were provided by Olga Henker from Fundación GeoCensos. Sasha Ramirez-Hughes provided support with editing as well all communication and dissemination activities.

The report also benefitted from significant inputs from the PARIS21 networking with the NSOs of Finland, Honduras, Mauritius, Namibia, Paraguay and Sweden related to the implementation of steps.

The paper is designed by Karamelaki Creatives.

Available at: https://paris21.org/geospatial

Contact the PARIS21 Secretariat at contact@paris21.org
Visit www.PARIS21.org
Follow PARIS21
@contactparis21
/company/paris21/

Visit www.scb.se
Follow SCB
@StatsSweden
/company/scb/

FOREWORD

National statistical offices (NSOs) and national geospatial integration agencies (NGIAs) are central elements of the national statistical system (NSS) and play a pivotal role in improving the understanding of social, economic and environmental issues. Indeed, knowing when and where the phenomena that traditional statistics represent happen can greatly aid public policies by seamlessly combining multiple sources.

However, many NSOs from low- and middle-income countries (LMICs) lack the needed degree of collaboration and skills, rely on outdated technologies, deal with disparate geographies and unarticulated systems, or work in an absence of consistent interoperability or dissemination methodologies. Given such circumstances, many NSOs may be struggling in the short term to keep up the requirements of modern census operations, including long-term commitments such as national development plans and the 2030 Agenda for Sustainable Development.

The COVID-19 pandemic has, moreover, challenged many traditional face-to-face statistical methodologies. In this context, NSOs from LMICs will need more than ever before to adapt and develop novel responses that could eventually facilitate a data revolution within their administrations. To tackle this hurdle, they must optimise the use of all the qualitative data that can be made available. This will require, in many cases, a step-change in their practices.

This guide provides comprehensive, step-by-step guidance for NSOs from LMICs. It is focused on those that are conscious of their need to integrate their statistics and geospatial data seamlessly across all geographic scales – not only in terms of administrative structure disaggregation but mindful of the everyday local spaces where the population lives, socialises and works.

This guide will help NSOs tactically approach the implementation of geo-enabled statistics, including the use of unevenly distributed data sources already available across their statistical ecosystem. Recommended steps in this guide are aligned with the National Strategy for the Development of Statistics (NSDS) lifecycle to help NSSs strategically plan for geospatial data integration in official statistics during the NSDS design process.

This document is designed to be self-contained and also offers case studies from national statistical offices. It includes references to specialised material that should help to refine the integration of official statistics and geospatial data. The outlined experiences highlight the potential of applying advanced technologies and learning from good practices for the next round of censuses already started in 2020 and for the fulfilment of the 2030 Agenda. They also demonstrate how the extension of outreach and integration to NSS data producers, and especially with NGIAs, has the potential to lead to better coverage of disaggregated data and to expand the meaningfulness of statistics.

TABLE OF CONTENTS

1. INTRODUCTION . . . 7

2. STATISTICAL-GEOSPATIAL DATA INTEGRATION . . . 11

3. THE GLOBAL STATISTICAL GEOSPATIAL FRAMEWORK . . . 15

3.1 The principles perspective and key elements of the GSGF . . . 17

3.2 Highlights of the relationship between the five principles and the four key elements . . . 19

4. THE STRUCTURE OF THIS GUIDE . . . 23

4.1 NSDS in a nutshell . . . 24

5. STEP-BY-STEP GUIDANCE: THE “WHEN”, THE “HOW TO” AND THE “WHAT” OF IMPLEMENTATION . . . 27

5.1 Step 1 - Get the right people together and put them on a map . . . 28

5.2 Step 2 - Assess and secure necessary human resources . . . 32

5.3 Step 3 - Assess and secure technical infrastructure . . . 36

5.4. Step 4 - Inventory fundamental geospatial data and address geocoded infrastructures, safeguarding access to all scales . . . 40

5.5 Step 5 - Set up a basic framework of geographies for analysis and dissemination of data . . . 45

5.6 Step 6 - Design a data management environment fit for multi-purpose use . . . 48

5.7 Step 7 - Define goals for dissemination and procure the necessary tools for it . . . 51

5.8 Step 8 - Use good practices and standards to obtain interoperability . . . 54

6. DEFINITIONS AND ADDITIONAL RELEVANT AND RESOURCES . . . 61

6.1 Definitions and terminology . . . 61

6.2 Resources and further reading . . . 62

REFERENCES BIBLIOGRAPHY . . . 64

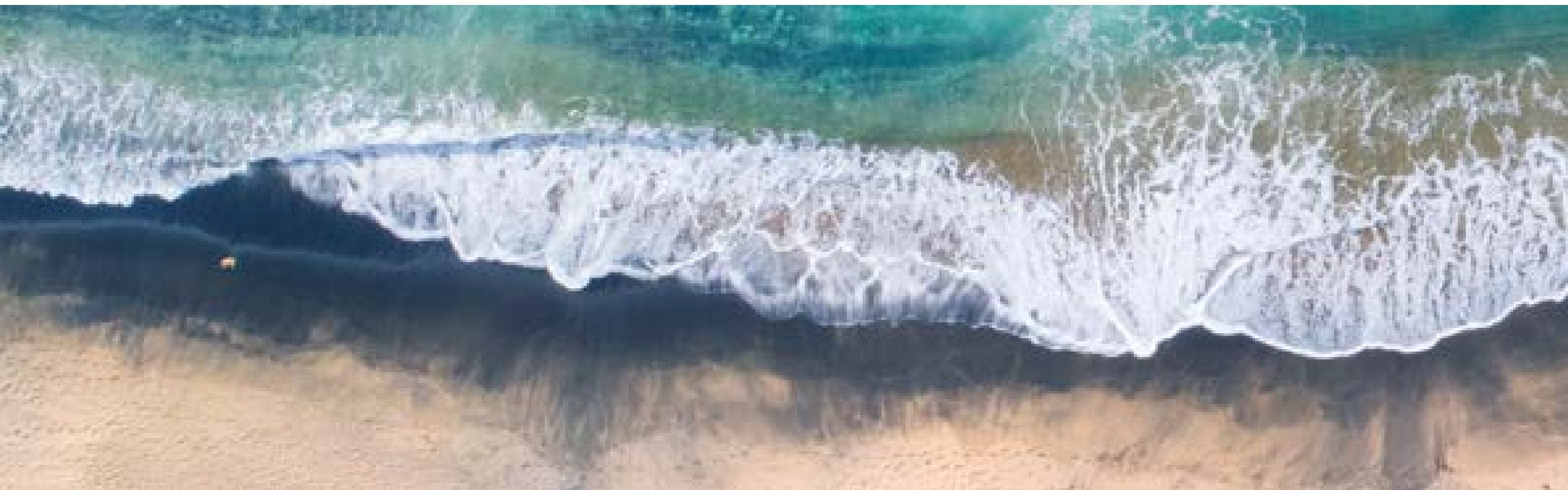


TABLE OF FIGURES

Figure 1. Location as a link between society, the economy and the environment 7

Figure 2. A conceptual illustration of statistical-geospatial integration 12

Figure 3. The Global Statistical Geospatial Framework – inputs through to outputs 16

Figure 4. The five Principles of the GSGF. 17

Figure 5. Illustration of the relations between principles, requirements, recommendations and good practice 20

Figure 6. The life cycle perspective of the NSDS 25

Figure 7. A general overview of the sequential order of the steps to be taken 28

Figure 8. GEOSTAT 4: Key actors of the geospatial and statistical ecosystem 30

Figure 9. Example of a power mapping exercise undertaken in Honduras 32

Figure 10. Example of the organisation chart of the cartography unit in a LMIC. 33

Figure 11. Example of aggregated group score for 11 areas of the TASC tool 34

Figure 12. Levels 1 and 2 of the Generic Statistical Business Process Model. 38

Figure 13. Building points in an enumeration area polygon 44

Figure 14. User interface of REGINA Web Map 47

Figure 15. Example of tabulated statistics projected on maps using common administrative units 53

Figure 16. Geospatial data conceptual model scheme 56

Figure 17. Statistics Finland's logical data warehouses. 57

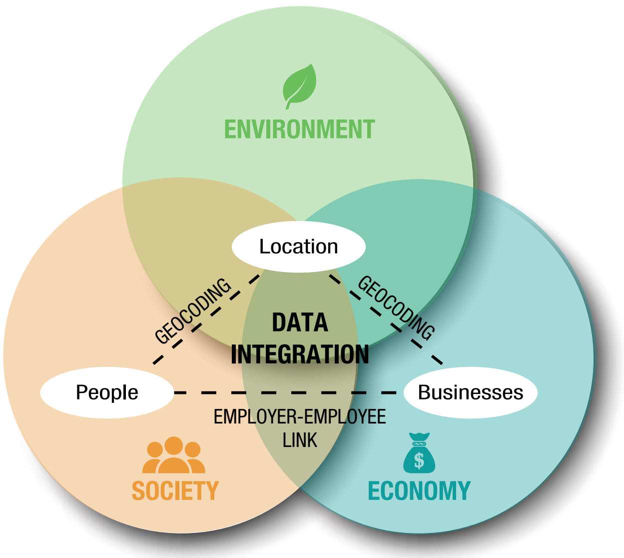
1

1. INTRODUCTION

Achieving the mission of sustainable development, especially in low- and middle-income countries (LMICs), is unquestionably linked to the capacity to provide reliable and relevant information for effective and efficient decision making. This means it is essential not only to produce reliable statistics based on fundamental principles but also to involve the spatial and temporal dimensions by developing the necessary capacity to, among other purposes, achieve disaggregation at the highest possible level of detail (or greater scale) to assure no one is left behind.

Linking data about people, businesses, buildings, infrastructures, agriculture, natural resources and anthropogenic impacts on the biosphere to a geographic location, including the temporal dimension, can result in an improved understanding of social, economic and environmental issues. This outcome will always be far more revealing, as shown in Figure 1, than visualising statistical or geospatial information separately.

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



Source: (UN DESA, 2019[1]), The Global Statistical Geospatial Framework, http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf.

In many low-income countries, among them Bangladesh (Government of Bangladesh, 2014[2]), governments are implementing a multi-stakeholder approach to effect the integration of statistics and geospatial information. The National Survey together with the country's national statistical office (NSO) have been collaborating in projects where data integration is provided to governmental and non-governmental stakeholders. This valuable partnership has been making available layers of data for educational and health as well as industrial purposes.

Compelling examples of the endurance of effective geospatial infrastructures emerged during the COVID-19 crisis. In several developing countries (Carranza, 2020[3]), NSOs demonstrated how statistical-geospatial data integration can support emergency decision making amid the unprecedented impact of a pandemic. For instance, the NSO of Colombia developed an interactive application to identify, at the municipal level, distribution of populations most vulnerable to COVID-19. The integration of geospatial data and statistics was based on estimates obtained by linking statistical morbidity data from the Ministry of Health and Social Protection with demographic characteristics of territorial units pulled from the country's most recent census and other administrative records.

The 2030 Agenda promoted by the United Nations (UN) and its Sustainable Development Goals (SDGs) recommend the integration of statistical and geospatial information, which makes geo-enabled statistics particularly relevant to the development aspirations of countries. Such integration poses significant challenges to all national and international data communities, including official as well as unofficial sources. But it also presents a unique opportunity to demonstrate how integrating statistical and geospatial data can improve monitoring of SDG progress and provide much-needed information to achieve sustainable development.

The Partnership in Statistics for Development in the 21st Century (PARIS21) is a global partnership of national, regional and international statistics experts and policy makers seeking to improve evidence-based decision making in developing countries. As demand for data increases, the Partnership supports the strengthening of statistical capacity at both country and regional level to improve the provision, availability and use of high-quality data for the achievement of national and international development goals.

Given the growing need to geo-enable the indicators for the monitoring of SDGs, LMICs have increasingly called for the PARIS21 Secretariat to provide a guide to national statistical systems under the co-ordination of NSOs on the implementation of statistical-geospatial data integration in official statistics.

In this context, the Global Statistical Geospatial Framework (GSGF) aims to structure the integration of a variety of data from both the statistical and geospatial realms (UN DESA, 2019[1]). The framework recommends five principles and supporting key elements for the production of harmonised and standardised geospatially enabled statistical data.

This thematic guide to implementation of the GSGF acknowledges, refers to and builds on other existing resources, including the GEOSTAT 3 project proposal, GSGF Europe – Implementation Guide for the Global Statistical Geospatial Framework in Europe (Hedeklint et al., 2019[4]). Thanks to the GEOSTAT 3 project, the implementation guide for Europe covers the key aspects of statistical-geospatial integration as established in the GSGF and adapts them for the European Statistical System (ESS) and the wider European context. This thematic guide was produced in close collaboration with the UN Expert Group on the Integration of Statistical and Geospatial Information Task Team working on the implementation guidance for the GSGF to provide specific support for LMICs.

The statistical community related to LMICs has also manifested a need for a thematic guide on the integration of geospatial data into official statistics. A key priority of PARIS 21 is to provide the necessary support and make available unprecedented reference resources to strengthen the National Strategy for the Development of Statistics (NSDS) guidelines.

This thematic guide introduces and explains, section by section and step by step, the principles, frameworks, guidelines, key elements and good practices to achieve a better integration of statistical and geospatial data for official statistics based on diverse sources. It aims to inform and guide NSOs and the heads of national geospatial information agencies (NGIAs) at the strategic level and is written to be readily understood by those who are not experts in the subject matter.

The intended readership comprises strategic officers from NSOs and NGIAs, including strategic advisors and high-level technical staff. Although this is not a technical handbook, its implementation is expected to deepen the implementation of global frameworks and general guidelines by addressing the characteristics, limitations and prevailing conditions at statistical and geospatial administrations of LMICs.

The rationale and structure of the guide are to provide specific guidance regarding, among others, the "when" and "how" of implementing data integration and to point out diverse answers to the "what", including references to the key elements that should be strengthened in the statistical offices of LMICs to deploy the GSGF.

The implementation guide covers all relevant aspects of statistical-geospatial integration — starting from the very basic conditions of availability of data and technologies in the data ecosystem on to the most advanced levels of development of geo-enabled statistics. It specifically considers and includes adaptations for the targeted countries where these are considered necessary. Nevertheless, the rationale of the guide is closely linked to the GSGF principles, including the framework's key elements, input and outputs.

Finally, one of the most popular and advanced outcomes of statistical and geospatial integration proposed in this guide is the dissemination of solutions that feature visually impactful results for decision makers and many other types of users. Nonetheless, readers are warned that this is not the ultimate goal of data integration. Rather, it is to make available more, better and reliable information by bringing together different kinds of sources for evidence-based decision making for sustainable development.

2

2. STATISTICAL-GEOSPATIAL DATA INTEGRATION

As noted in Section 1, using sources of different kinds for evidence-based decision making supports the SDGs and the 2030 Agenda. This implies that data integration must be accomplished using the best resources and available knowledge. Broadly defined, data integration "is the practice of consolidating data from disparate sources into a single dataset with the ultimate goal of providing users with consistent access and delivery of data across [a wider] spectrum of subjects" (OmniSci, 2020[5]). In particular, at the 60th International Statistical Institute World Statistics Congress in 2015, the global statistical community recognised the importance of the integration of statistical and geospatial information due, among other reasons, to its contribution to support social, economic and environmental policy decision making (Hedeklint et al., 2019[4]).

In this regard, the concept of integration of geospatial data and statistics refers to the practice of incorporating and consolidating both kinds of sources into a single dataset, with the ultimate goal of providing users with consistent access to and delivery of information across the geographical, social, economic and environmental spectrums.

Integration of these traditional, although dissimilar, datasets has long been performed throughout the history of statistics. In fact, many tools were originally developed (Esri, n.d.[6]) as a practical means to describe spatial patterns and interpolate values for locations where samples were not available. Similarly, until fairly recently and before the disruptive arrival of digital technologies, NSOs from LMICs in Central America regularly used compass and counting footsteps to draw paper maps in census rounds. In those days, the resulting paper-based "census cartographies" were integrated with statistics after the census was completed to disseminate combined information in different aggregated ways.

Only in the last decade has the practice of incorporating digitalised data using geographic information systems (GIS) acquired special relevance for NSOs around the world, given the growing enabling impact of information technologies for official organisations. Advances in geocoding methods and standards have supported the practice. Other significant drivers of this change include the sophistication in data management processing techniques, the generalised digitalisation of statistics and other data, the massive use of global positioning systems in handheld devices, and the decreasing costs of satellite images.

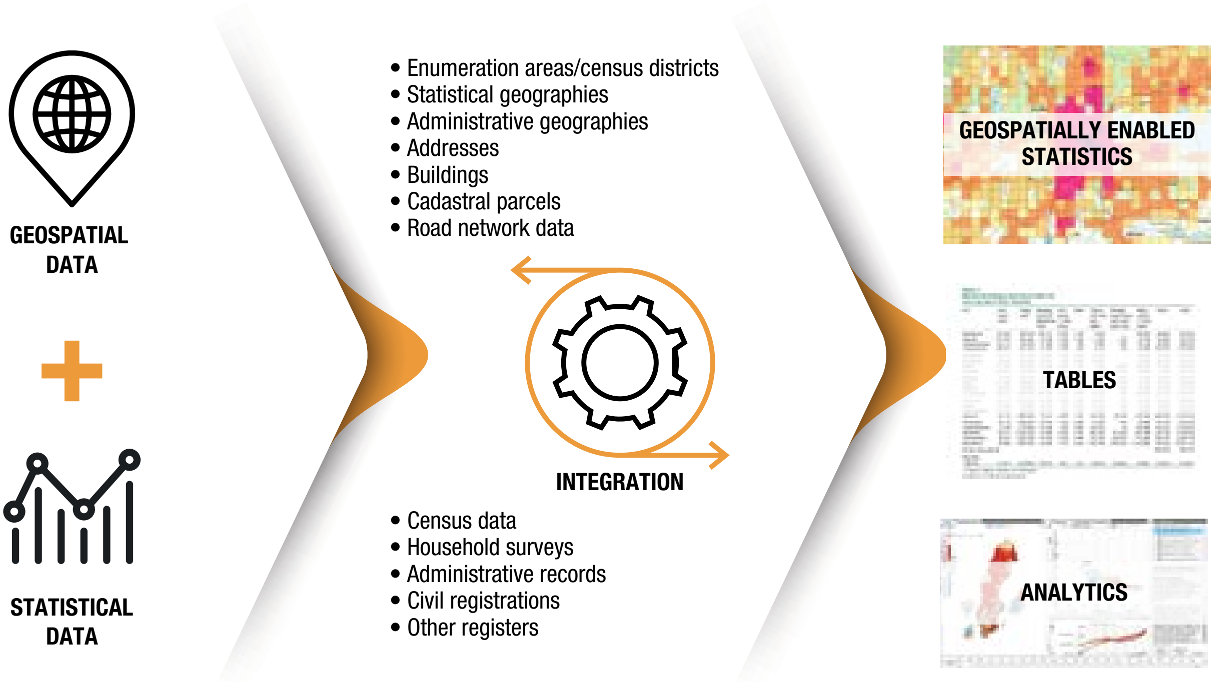
The outcomes of the integration of digitalised statistics and geospatial data are used for several traditional as well as innovative purposes in the production of statistics — for example, providing more accurate census cartographies, allowing sophisticated spatial-temporal statistical analysis and enabling the visualisation of non-geographical data in a spatial environment. These advances, in a broad sense, empower all official data stakeholders by supplying rich information for data analysis and making it available for decision making.

At the NSO and NGIA level, most of these advances have benefited from the availability of powerful geospatial tools that enhance the value and usability of official statistics by leveraging the application of the spatial context. Additionally, the integration of geospatial data with statistics has greatly contributed to full-fledged disaggregation of data that enables comparisons through various administrative levels. Driven by a growing demand over the last two decades for better data for more informed decision making, a growing body of knowledge has become available about the technical aspects and benefits of bringing together statistical and geospatial information.

The Ghana Statistical Service (2014[7]) developed a valuable exercise of this kind that covered both urban and rural areas with a unique code for each enumeration area (EA) in order to analyse poverty from a territorial perspective throughout the entire country. Ghana took an important step towards statistical-geospatial integration by empowering the statistical service's Geo Information Systems Unit to utilise different sources including surveys, censuses and administrative datasets. This allowed the NSO to integrate the EAs with data from the Population and Housing Census and the Ghana Living Standard Survey and produce a series of maps to inform decision making on poverty on a local level.

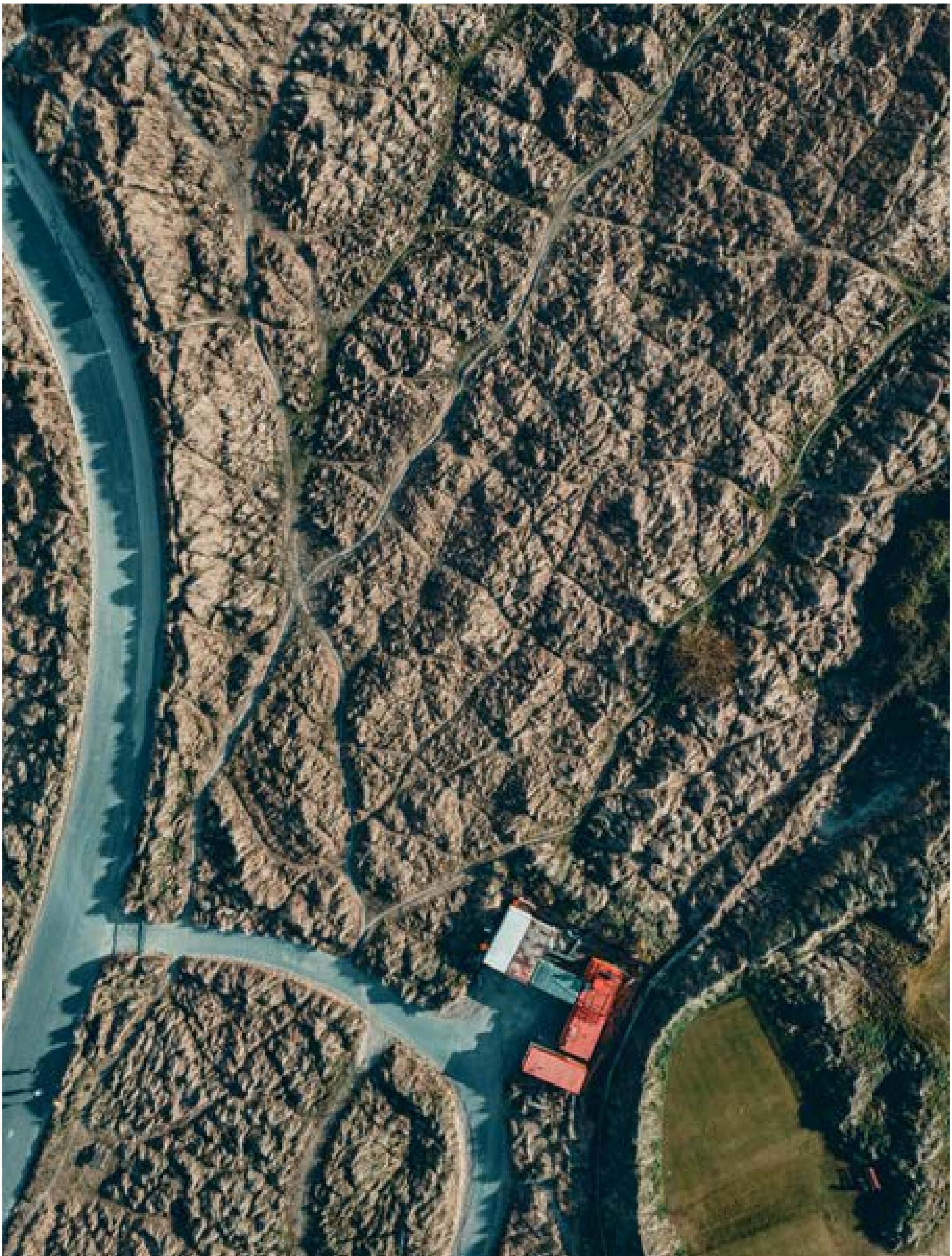
Figure 2 shows a conceptual illustration of statistical-geospatial integration. NSOs that have implemented a full-fledged geospatial infrastructure including appropriate arrangements, technologies, data sources and technical skills can transform data into a diversity of outputs such as geospatially enabled statistics, regular statistics, indicators, analytics and visualisations. Data integration can occur in many iterations, for example the geocoding of census records using address locations to produce geospatially enabled microdata, which in turn can be spatially aggregated and released using statistical geographies. Additional iteration can be added, for example by relating the geospatially enabled census microdata to a road network to measure accessibility before aggregation and release of the result using statistical geographies. The more layers of information available and the higher degree of interoperability between data sources, the richer the possible outputs.

Figure 2. A conceptual illustration of statistical-geospatial integration

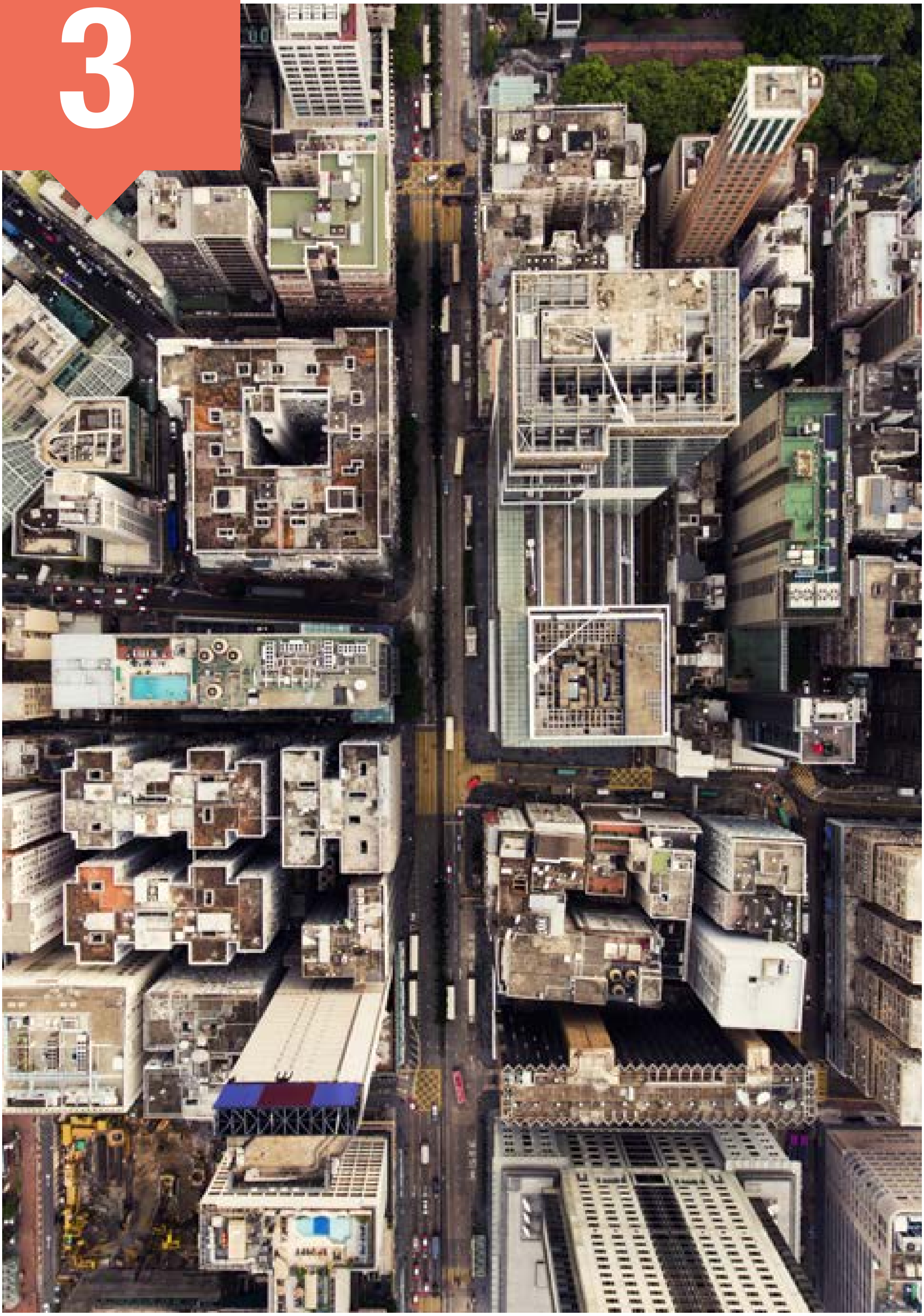


Source: Authors' adaptation based on information from Statistics Sweden.

These kinds of integrations enable NSOs of all sizes and capacities to access new data production methods with a clear focus on the output. They can also enable the dissemination of a continuous flow of visually impactful outcomes that can be produced in the form of a news scoop to enrich story-telling narratives in newspapers and the national media (Statistics Sweden, 2014[8]), among other outreach impacts.



3



3. THE GLOBAL STATISTICAL GEOSPATIAL FRAMEWORK

Producing a varied portfolio of geospatially enabled statistical products and visually impactful outcomes requires NSOs and the entire statistical community to develop the necessary skills, capabilities and good practices within proven integration frameworks.

First, stakeholders need to work together to increase awareness that local, subnational, national, regional and global development is possible and that more focused decision making can be achieved through integrating statistical and geospatial information. Second, the management and use of geospatial information should be mainstreamed within existing processes and made an integral part of the production of statistics. With these inputs, a wider range of information and enabling technologies will become accessible and available for use and dissemination.

The above arrangements are not meant to be carried out as an overnight task. They need to be patiently and carefully prepared, designed, implemented and monitored. For instance, NSOs need to create the ability to develop and analyse local geographies at a finer scale. They also need to be aware that the process of integrating geospatial data requires an institutional openness to incorporate new knowledge and technology and take advantage of new data sources as a complement.

One benefit of carrying out these initiatives is that these "new" data sources include the use of accessible Earth Observations, or EO, such as satellite imagery as a basis to generate statistics. Indeed, satellite imagery has great potential to provide data where traditional statistical and/or geospatial information methods prove inadequate or not timely enough or where they do not exist due to structural capacity gaps. Furthermore, considering alternate data sources will help in the quest to deliver information and insights for national development priorities and the corresponding 17 SDGs.

The GSGF is a high-level framework that facilitates consistent production and integration approaches for geospatial and statistical information. However, fully implementing the GSGF requires decision making at all levels in a national statistical office. Although the framework is generic, it allows the application of principles to the specific circumstances and possibilities of each country seeking an impactful transformation of all processes in the production of official statistics.

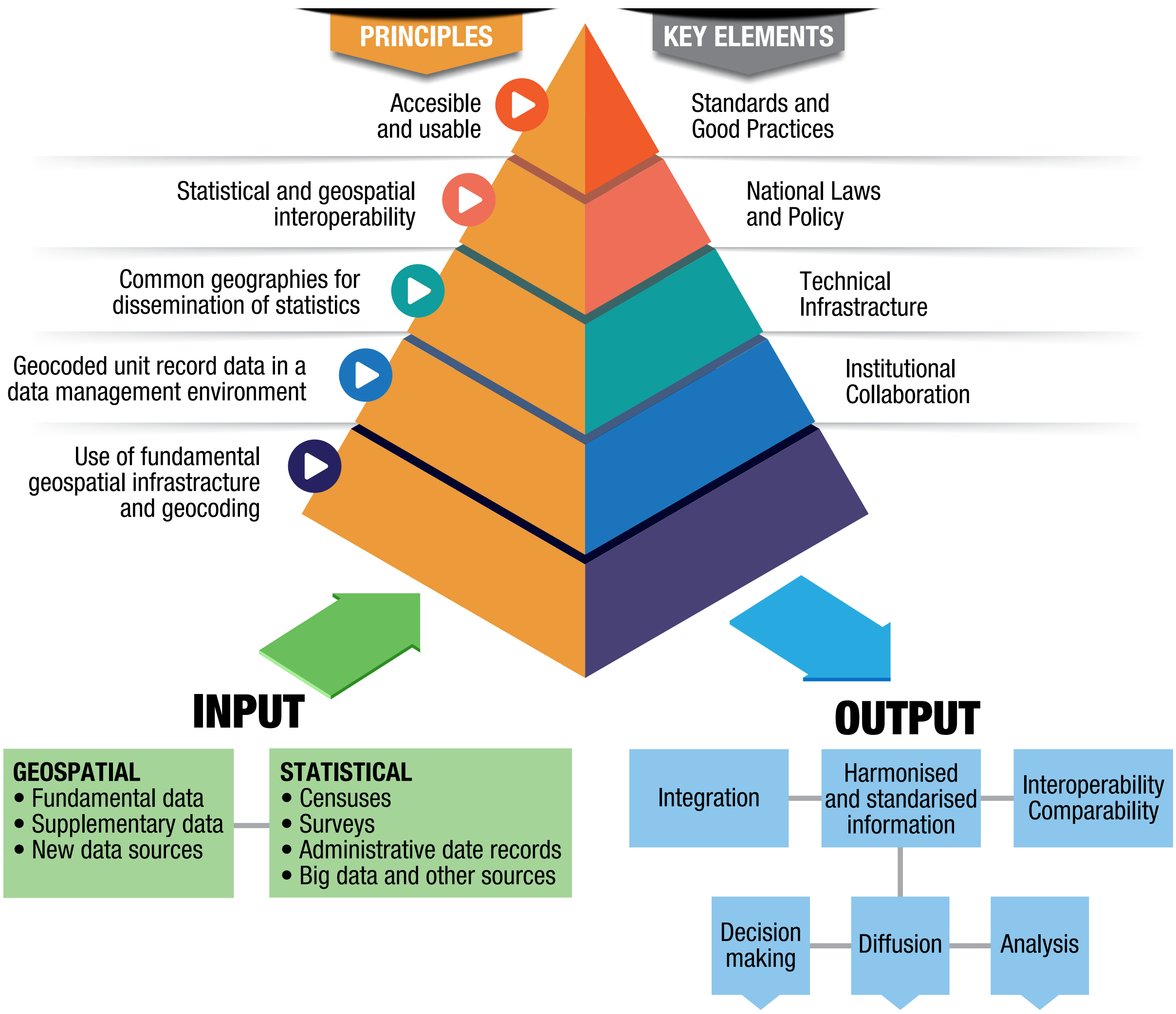
As an aid to governments through their NSOs and NGIAs, the GSGF (UN DESA, 2019[1]) was tested and implemented in several pilot countries (Australia, Egypt, Mexico, New Zealand, South Africa and Sweden, among others) that started at different points in their statistical and geospatial infrastructure development. It has already demonstrated its usefulness and potential.

As a global framework, the GSGF stands on the shoulders of other general frameworks that aim to connect the road of sustainable development and the 2030 Agenda. One of those overarching references is the Integrated Geospatial Information Framework (IGIF), which provides a basis and overall guidance for developing, integrating, strengthening and maximising geospatial information management and related resources in all countries (UN DESA, 2021[9]). This framework especially assists in bridging the geospatial digital divide to secure socio-economic prosperity and to leave no one behind. In this regard, it is a starting point to address the gaps in geospatial digital data.

The GSGF is based on the IGIF and also reflects the convergence of ideas that emerged in discussions around the drafting of documents and agreements and involved experts from bodies like the High-Level Group for the Modernisation of Official Statistics and national representatives from the Working Group on Geospatial Information within the Inter-agency and Expert Group on the Sustainable Development Goal Indicators.

The GSGF is conceived as a “conceptual workflow”, as illustrated in Figure 3, composed of inputs, principles, key elements and outputs. Through these components, the GSGF matches the statistical and geospatial professional domains in an input-output logical sequence and identifies key solutions to combine the efforts of NSOs and NGIAs involving statistical and geospatial standards, methods, data processes and tools.

Figure 3. The Global Statistical Geospatial Framework



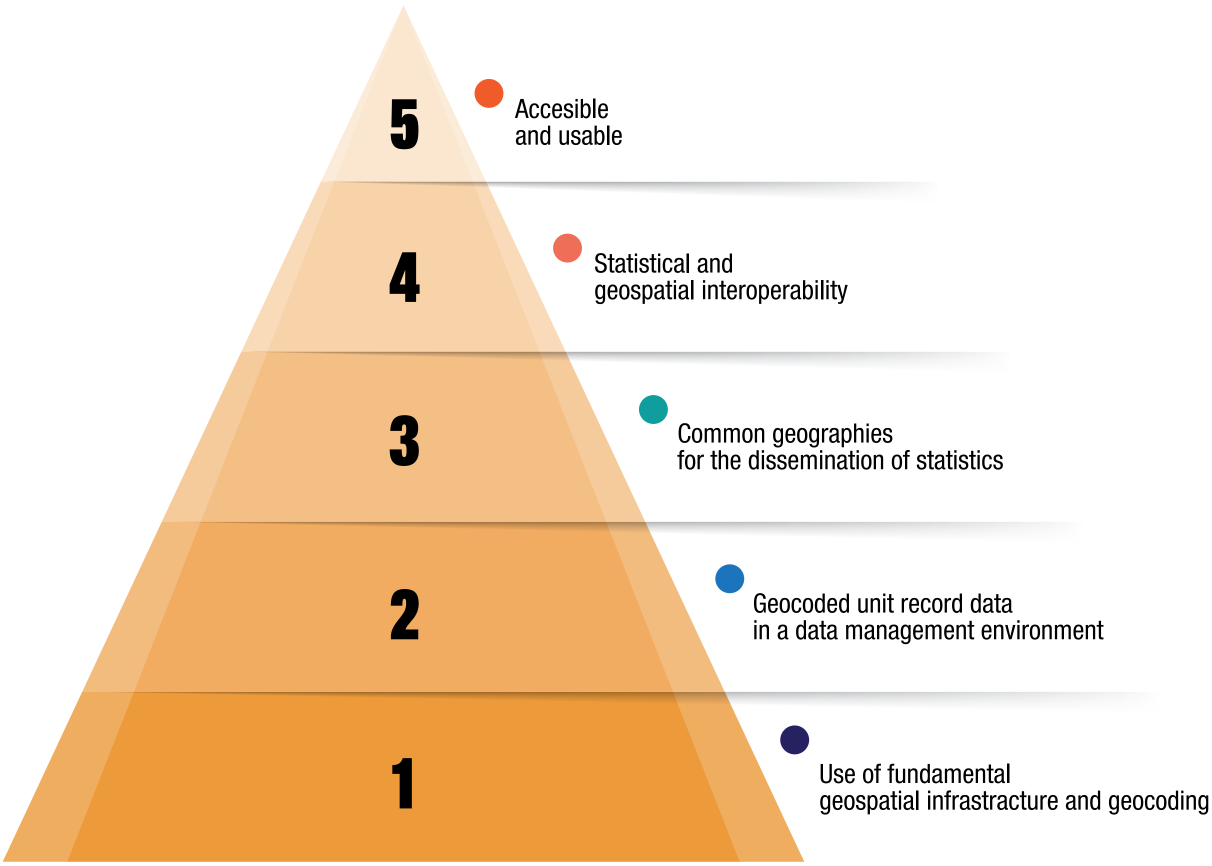
Source: (UN DESA, 2019[1]), The Global Statistical Geospatial Framework, http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf.

3.1 THE PRINCIPLES PERSPECTIVE AND KEY ELEMENTS OF THE GSGF

This subsection briefly introduces the Global Statistical Geospatial Framework's five principles and key elements. The classic differentiation between these principles is presented in Figure 4.

The representation of the GSGF as a pyramid communicates the idea of having solid bases for statistical-geospatial integration. Its basic message aims to identify and describe five thematic "layers" that need to be in place to ensure the integration of statistics and geospatial data.

Figure 4. The five principles of the GSGF



Source: (UN DESA, 2019[1]), The Global Statistical Geospatial Framework, http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf.

The subsection starts with an intuitive interpretation of the principles when this is needed. It then focuses on how the four key elements and the five principles interact to enable the processes that provide the interoperable outputs of the conceptual workflow of the framework (Figure 3). The following five principles are one of the core components of the GSGF:

1. Use of fundamental geospatial infrastructure and geocoding. This principle specifies the need for a common and consistent approach to place each statistical unit of a dataset in time and space, using fundamental geospatial infrastructure. This means the requirement for a geospatial infrastructure in the country, i.e. that it has a georeferenced address, building register, land parcel and/or place name as part of its statistical-geospatial infrastructure.

2. Geocoded unit record data in a data management environment. This principle provides support to the process of linking or storing high-precision geographic references (i.e. geocodes such as coordinates, small geographic area codes or linked-data identifiers) to each microdata and/or

statistical unit record. It refers to the actual process of geocoding of unit record data and the required environment to support it. This involves the best possible use of data management tools, techniques, standards and good practices to facilitate the linking and management of geocodes within statistical datasets.

3. Common geographies for the dissemination of statistics. This principle refers to geography as a tool for integrating data. It uses a common and agreed set of geographies for the display, storage, reporting and analysis of social, economic and environmental comparisons across statistical datasets from different sources. It pursues the goals of integrating data from different sources through a common geography; simplifying the visualisation, analysis and interpretation of statistical and geographic information; making available metadata for aggregation, integration and use of the data by implementing aggregation and disaggregation methods to enhance data quality; and supporting of conversion of data between geographies using methods as, for example, correspondences.

4. Statistical and geospatial interoperability. This principle defines the preconditions for statistical and geospatial data to work as a data ecosystem in which those involved interact with each other to exchange, produce and consume data. To attain interoperability in this context means that the data themselves and the metadata standards must transcend existing structural, semantic and syntactic barriers so that statistical and geospatial data from different communities and providers are swiftly integrated.

5. Accessible and usable geospatially enabled statistics. This principle highlights the need for data custodians to make geospatially enabled statistics accessible and usable according to agreed standards and good practices so that data users can discover, access, integrate, analyse and visualise this information seamlessly for geographies of interest. It addresses the need to identify or, where required, develop policies, standards, good practices and technologies that support third party uses. The rationale of the principle is that while data custodians release data, third party users are able to discover, access, analyse, visualise and check the integrity of geo-enabled statistics. The logic of the principle also seeks to grant machine-to-machine access through web services and linked data methods.

The following four key elements — the other core component of the GSGF — are present across all five principles and play a critical enabling role that allows data to be obtained from the various sources together with the application of the principles:

Standards and good practices. This key element refers both to repeatable, harmonised, agreed and documented methods and to practices that have been shown to work well and produce good results for integrating geospatial data and statistics.

National laws and policies. This key element refers to all pre-existing arrangements and legislation that affect the access, processing, use and dissemination of data.

Technical infrastructure. This key element comprises all interfaces and databases from the front-end user side to the back end, especially including software and hardware components.

Institutional collaboration. This key element expresses the need for key stakeholders to commit to working together — particularly (but not only) statistical, geospatial and administrative agencies of government — through positive agreements and relationships.

3.2 HIGHLIGHTS OF THE RELATIONSHIP BETWEEN THE FIVE PRINCIPLES AND THE FOUR KEY ELEMENTS

An understanding of the five principles helps strategic-level decision makers and implementers to comprehensively apply the GSGF. For instance, the first three principles are chiefly preconditions for the other two principles, and some depend more on one specific principle than another. While this allows a certain flexibility in the order they can be implemented, it also demands a deep understanding of their interrelationships.

For instance, Principles 2 and 3 are strongly linked because producing geocoded units of statistical records allows statistics to be provided into flexible statistical geographies. Additionally, Principle 5 relies heavily on Principle 3 as the flexible aggregation of statistics into any output geography is needed as a basis for dissemination of geospatial statistics to ensure that such statistics are accessible and usable in the best possible way.

Evaluating the relationships between the five principles and the key elements provides useful insights:

- Principle 1 needs a solid technical infrastructure as a key element to be able to provide physical and digital storage space including map libraries, servers, cloud hosting services and corresponding data-handling considerations.
- Principle 2 requires standards and good practices to apply formal statistical classifications, concepts and definitions. In general, standards can be translated into recommendations regarding data specifications and harmonisation of content, a very important issue when building a homogeneous geocoding.
- Principle 3 needs the key element of national laws and policies to organise, recognise and, if possible, update the legislation and mission statements of authoritative and referential mapping organisations. This principle also requires technical infrastructure, another of the key elements that includes trained human resources, appropriate software, enabling applications and techniques that allow boundary harmonising in general.
- Principle 4 requires that national laws and policies comply with international and national data protection, privacy and confidentiality legislation.
- Principle 5 requires that accessible and usable geospatial statistics are released in the best possible way, meaning technical infrastructure for the necessary communication technologies is required in the processing of the front end-facing users. This infrastructure includes software, applications and other developments, especially those related to geospatial computing processes over the web.

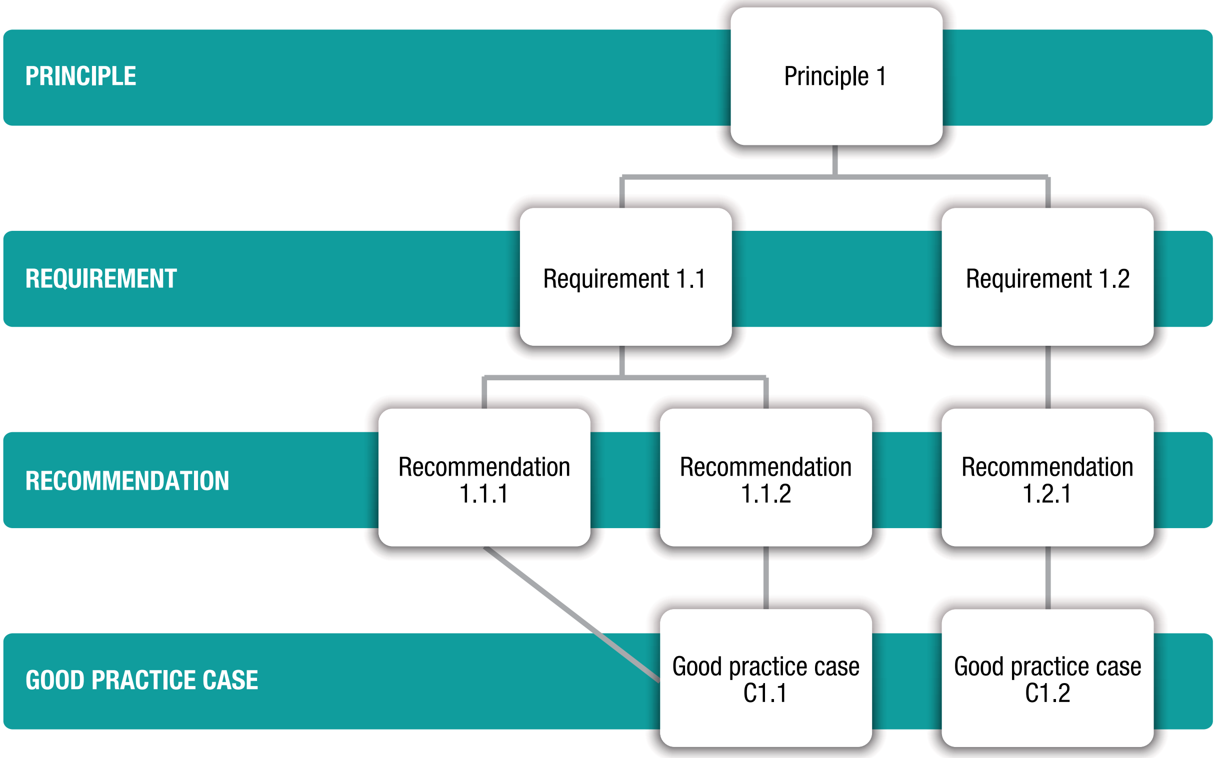
The tables in Annex 1 outlines additional relationships between principles and key elements and their specifications. A general conclusion from cross-analysis and combinations of the principles and the relationships between principles and key elements is that the GSGF is a tight conceptual construct wherein the components interact in such a way that the situations of most countries are considered and anticipated inclusively.

By way of illustration, the process needed to implement a framework like the GSGF is analogous to the incremental process of software development. The software development practice builds partial functionalities and delivers efficient results in a piecemeal fashion, at different moments in time and not necessarily sequentially, one after the other. Each "piece", or increment, of the software represents a complete subset of functionalities. The increment of one subset to the other may be either small or large, perhaps ranging from just a system to provide a back-end functionality to a highly flexible set of data management tools.

Learning from requirements and recommendations for implementation of the GSGF coming from the European experience

As noted, this thematic guide recognises the executive proposal from the European Commission's GEOSTAT 3 project as a valuable resource for the guidance of successful implementation of the GSGF (Hedeklint et al., 2019[4]). It recommends a detailed architecture of requirements, recommendations and good practices (Figure 5).

Figure 5. Illustration of the relations between principles, requirements, recommendations and good practice

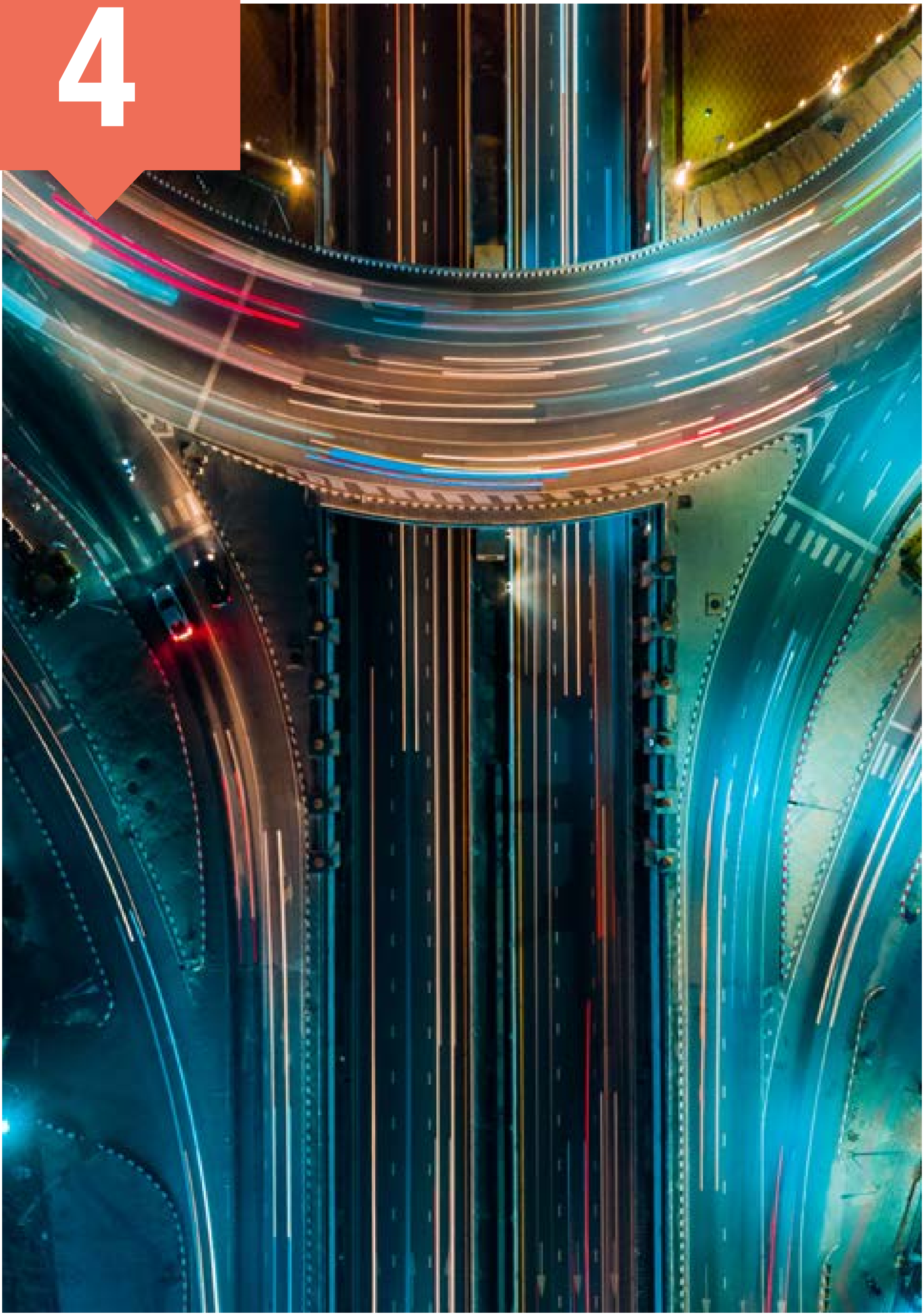


Source: (Hedeklint et al., 2019[4])Hedeklint et al. (2019[4]), GSGF Europe – Implementation Guide for the Global Statistical Geospatial Framework in Europe: Proposal from the GEOSTAT 3 Project, https://www.efgs.info/wp-content/uploads/geostat/3/GEOSTAT3_GSGF_EuropeanImplementationGuide_v1.0.pdf.

This proposal also recommends that to implement the five principles and the framework, European countries should fulfil several crucial requirements (also called the "what"). Each of these requirements connects to a set of more specific recommendations to achieve them (also called the "how").

Although the structure illustrated in Figure 5 is a well-proven methodology to implement the GSGF in developing countries, this guide will follow an alternative structure, learning from the European experience and at the same time adapting it to the needs, conditions and realities of the data ecosystems in LMICs.

4



4. THE STRUCTURE OF THIS GUIDE

This guide has six sections. As seen, the first three sections discuss the general context to be considered in any statistical-geospatial data integration and outline the reasoning and basis for integrating geospatial data and statistics. Section 3 also provides a concise overview of the Global Statistical Geospatial Framework and the interrelationships between its components.

This fourth section introduces the role of the NSDS framework to support implementation of geospatially enabled statistics within strategies of statistics development. It examines commonalities and differences between the GSGF and the NSDS guidelines in line with the aim of this guide to build on the principles and key elements to integrate them with the more dynamic, process-oriented perspective embodied in the NSDS. The process-oriented perspective represented by the presented frameworks helps interpret and apply the subsequent step-by-step process that will guide both the implementation of the GSGF and the mainstreaming of geospatial data and statistics into the NSDS.

Statistical-geospatial integration should rest on a solid base — as conveyed in the pyramid representation of the GSGF in Figure 4 — and the five thematic "layers", or principles, need to be in place to ensure the integration of statistics and geospatial data. However, the GSGF itself, as noted, does not set out a strict order or timeline for implementation of the principles. Nor does it specify what the first step should be. In some cases, the framework suggests how one principle relates to another, but this does not imply that they should be implemented in a particular order, for instance starting with Principle 1 and proceeding consecutively to Principle 5.

By mainstreaming and comparing the GSGF and NSDS frameworks, Section 4 also lays the groundwork for the step-by-step guidance presented in Section 5. Consideration of a life cycle framework entails a temporal dimension and the processing approach that can help answer the "when", the "how to" and the "what" questions at each stage of a country's NSDS deployment while partially using the European approach to apply the GSGF as noted in Section 3.

The core of this guide is Section 5. It proposes eight general, consecutive steps to be undertaken in order to achieve a successful statistical-geospatial integration.

The guidance provided in Section 5 integrates examples, mostly from LMICs, of notable NSDS progress to raise awareness and interest among the target audience and demonstrate the support that the GSGF principles can provide countries. The section also highlights success stories from other developing and European countries. These examples and case studies especially focus on answering the "how to" and "what" questions by walking the talk for future implementers through specific steps. Strategical implementations are linked in some cases to GSGF principles, but also describe key elements and innovations that were used in other countries.

Although it is not meant to be a handbook per se, this guide includes some technical details on certain issues for reference. It also includes carefully selected resources, links and articles on available frameworks, principles, procedures, requirements, standards, case studies, good practices, knowledge bases and other useful material. These are included to enrich the guide without making it unnecessarily long.

In sum, the contents of this guide are meant to foster and enhance the global and strategic frameworks, taking into account the specificities of LMICs and implementing the process that will lead to geospatially enabled statistics in a sequential order. Once the logic of processes is linked to the basic principles and key elements, each step will incorporate the LMIC reality and perspective.

4.1 NSDS IN A NUTSHELL

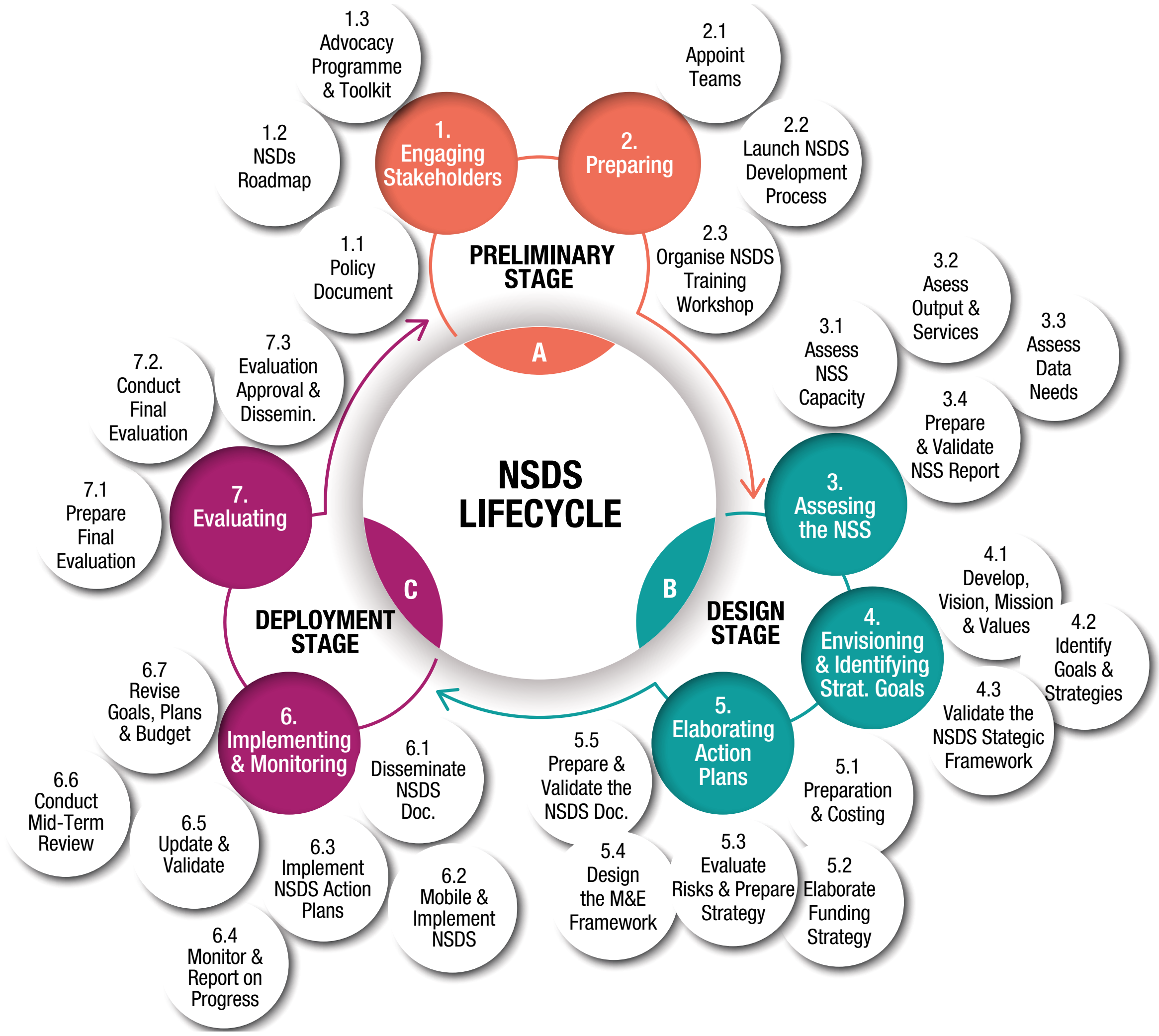
Simply put, a National Strategy for the Development of Statistics is the national framework, process and product for statistics development aimed at mainstreaming statistics into national policy and planning process (PARIS21, 2018[10]). It also seeks to produce information that meets the needs of the various users by integrating and co-ordinating the ecosystem's diverse sectors and actors into the national statistical system (NSS).

By sharing good practices, adapting, making more flexible arrangements and responding to demanding data challenges, the NSDS should also be able to deliver a country-led data revolution and build statistical capacity across the statistical value chain. The process to elaborate and implement an NSDS should be consultative and inclusive in nature and involve all the major actors from the NSS: producers and users of statistics, decision makers, technical and financial partners, civil society, the private sector, universities, and others.

The NSDS has three generic design and implementation activities: the preliminary stage, the design stage and the stage of implementation activities for deployment (Figure 6). These are subdivided into sub-activities involved in the process. The NSDS design activity includes two types of processes: design processes carried out at a specific moment in time (sequenced processes) and essential processes that are to be considered as running throughout the overall design process.

In terms of implementing an NSDS, the first proposed step is to define a technical framework to prepare the new strategy, using a participative approach and with the benefit of strong political support. This should lead, in turn, to a participatory process of creating a roadmap. The main steps from thereon in include preparation of a diagnosis of the NSS, defining a medium- and long-term vision for the NSS, defining a medium-term strategy, and preparing action plans that take available resources into account.

Figure 6. The lifecycle perspective of the NSDS



Source: PARIS21 (2018[10]), NSDS Guidelines: Guidelines for the Elaboration of a National Strategy for the Development of Statistics, https://nsdsguidelines.paris21.org/NSDS-GUIDELINES-full-lang-en.pdf.

To successfully prepare and implement its NSDS, a country needs strong political support and an explicit recognition of the role of statistics in development; constructive dialogue with the main categories of data users (government, private sector, civil society); integration of all components of the NSS in a common strategy responding to the needs of the users and to the requirements of the monitoring and evaluation of the SDGs; and mobilisation of government funds and support from co-ordinated partners.

All these features and processes of an NSDS have commonalities with those of the GSGF. Both instruments, for instance, refer to mainstream statistics by integrating them with available information and administrative resources to unleash their power. One difference between the two instruments is the nature of their goals. While the GSGF is conceived as a generic global framework for approaches to the consistent production and integration of geo-statistical information, the NSDS is a national framework to integrate statistics into national policy and planning processes. Both seek to develop statistics, but they envision different kinds of integration — on data in the case of the GSGF and into the planning processes of the country in the case of an NSDS. The two frameworks also differ in how the temporal dimension is conceived. The GSGF is a non-temporal, multi-layered enumeration of basic conditions to be fulfilled in the integration of data; the NSDS is a lifecycle strategy that aims to iteratively develop national statistics over time.

5

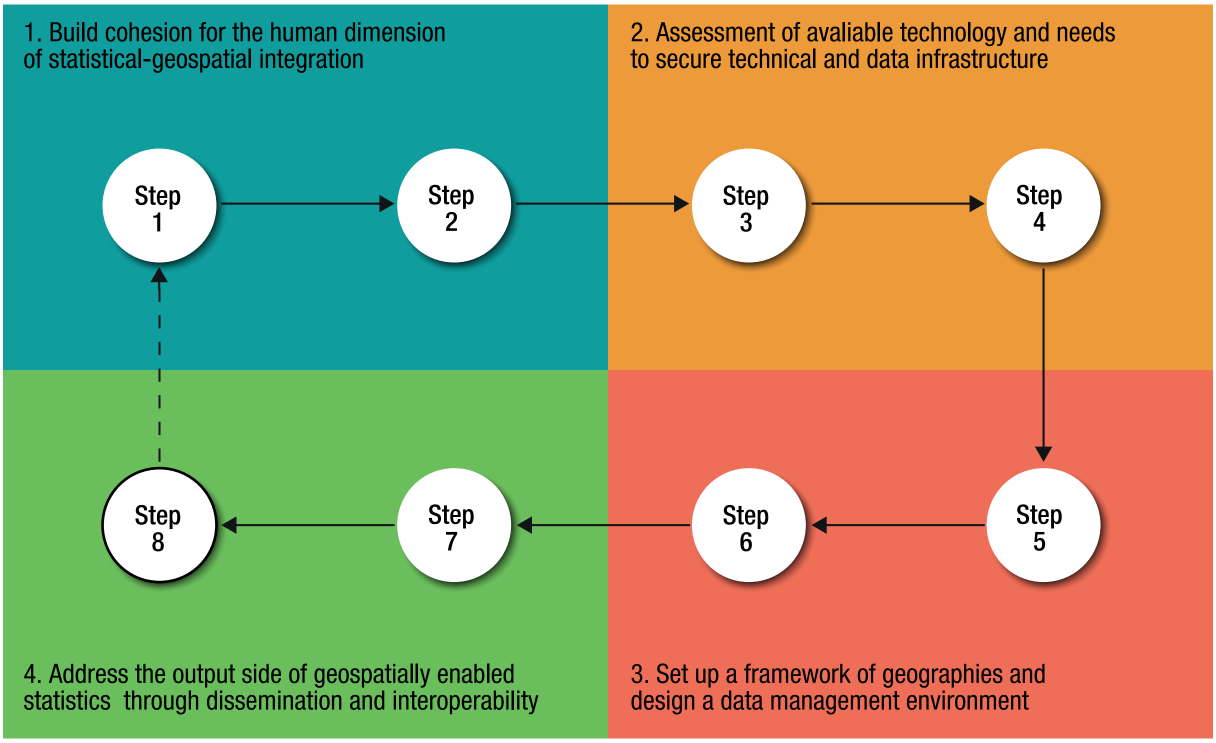
5. STEP-BY-STEP GUIDANCE: THE "WHEN", THE "HOW TO" AND THE "WHAT" OF IMPLEMENTATION

This section, the core of this guide, describes, develops and illustrates the sequence of steps to bring the integration of geospatial data and statistics to a state of solid interoperability — the highest level of development of an effective integration of geospatial data and statistics. The step-by-step guidance identifies eight main steps (labelled as Step 1 through Step 8" and presented in the corresponding subsections 5.1 to 5.8) that are best conducted in consecutive order. Steps are mostly independent transformations within the geospatial and statistical integration enablement. Steps focus on the transformation of specific external resources, such as the teams of experts, data, technologies or data interoperability. Each step can be implemented through a coherent set of tasks. Tasks comprise all the needed inputs to develop specific steps, although tasks might be related to other consecutive and non-consecutive steps. In this regard, tasks contribute to the fulfilment of one step or many steps. However, in some steps, tasks should be implemented in the suggested ordinal sequence, starting with task "1" up to "n". However, this rationale does not necessarily apply to all steps. Few steps are developed without the assistance of tasks.

Inspired by the structure of the European GEOSTAT 3 project, each of the steps also may involve specific recommendations especially formulated for each resource or input. Where necessary and feasible, good practices and cases studies are also portrayed within each step to help derive recommendations and also link the nature of the principles of the GSGF to real-life actions that leading countries have undertaken to implement it. The progressive, step-by-step strategical implementation is based on the aforementioned NSDS lifecycle perspective, which is in turn interwoven with the five principles and key elements of the GSGF.

Figure 7 presents a general overview of the steps and the order in which they should be taken within a process-oriented perspective. The starting point refers to the people who will be involved in the geospatial infrastructure and to the consolidation of expert teams, whether by their data stakeholders or by labour resources (Steps 1 and 2); a comprehensive assessment of needs to secure the technical and data infrastructure follows (Steps 3 and 4). Once the required human resources and technical necessities are known and secured, the next steps outlined in this guide focus on the consolidation of the previously assessed and diagnosed resources, including the recommendation to first set up a framework of geographies and to later design a data management environment that includes the resulting geographies by enabling the storage of data and their coordinates (Steps 5 and 6). Approaches for the next two, and final, steps focus on the output side of the workflow of geospatially enabled statistics, emphasising the need to first define goals and procure tools for dissemination purposes (Step 7) and then to learn from good practices and standards. This last step leads to reaching the highest ambition of integration: stepping up interoperability goals to enable data ecosystems in order to evolve and get involved (Step 8). In line with the lifecycle perspective of the NSDS, after reaching Step 8, a new iteration should be initiated, gradually improving and enhancing the steps.

Figure 7. A general overview of the sequential order of the steps to be taken



Source: Authors' own elaboration.

The case studies discussed in the step-by-step strategic recommendations include, as available, a wide range of examples from different locations and involving different kinds of actors. LMICs are the main focus, though some useful examples from the experiences of developing countries are provided. Examples and successful data integration stories from Europe that are particularly pertinent are also reviewed. Some of these cases are based on a data collaborative philosophy that may come from data partnership experiences that enhance the incentives of users and data providers such as governmental agencies within the NSS ecosystem, international organisations, academia, the private sector, philanthropic bodies and civil society organisations.

5.1 STEP 1 - GET THE RIGHT PEOPLE TOGETHER AND PUT THEM ON A MAP

With countries under great pressure to meet increasing demand for better data to support more informed decision making, experts and scholars have produced a growing body of specialised material addressing the technical aspects and benefits of bringing together statistical and geospatial information. Nevertheless, one of the key features of geospatial and statistical integration relies on the engagement of these two diverse professional communities — the statistical and the geospatial — and their awareness of the need to work together.

Although not related to any of the previously described five principles of the GSGF, the initial step of assembling and mapping the right people should be considered a core foundation of this process-oriented guide. This step is tightly linked to the key element of institutional collaboration. Indeed, as most countries have already established a working National Spatial Data Infrastructure (NSDI) or system to share data and knowledge from geospatial data, GSGF implementation requires these arrangements to be leveraged and further developed in the national statistical system. Given that this step is a cornerstone of GSGF implementation, a set of interrelated tasks are recommended to effectively reach realistic goals.

Task 1. Identifying key actors

An important initial task for this step is to identify key actors of the geospatial and statistical ecosystem so they can later be mapped according to their relative power and other engagement characteristics. This is also called stakeholder analysis. Given the range of historical, geopolitical, legal, institutional, governmental and political situations, the actors and their different roles in the data ecosystem will be unique to each country.

A few generic roles presumably exist in all countries. On the supply side, all producers in particular of geospatial data, statistical data and administrative data are identifiable, especially if their countries take part in higher-level bodies of the global statistical and data systems (Van Halderen et al., 2016[11]) such as the United Nations Statistical Commission, which includes NSOs and administrative agencies, and the United Nations Committee of Experts on Global Geospatial Information Management, comprising NSDIs, NGIAs and other geospatial data stakeholders from most of the countries in the world. It is also important to identify producers of subnational and local data such as cadastres and local governments as well as relevant programmes and projects by international development co-operation, non-governmental organisations (NGOs) and civil society, as these also provide disaggregation of data.

On the demand side, users can also be identified by their interests, including other governments from the national and local levels, international organisations, academia, private companies, civil society and the general public.

Task 2. Mapping the actors

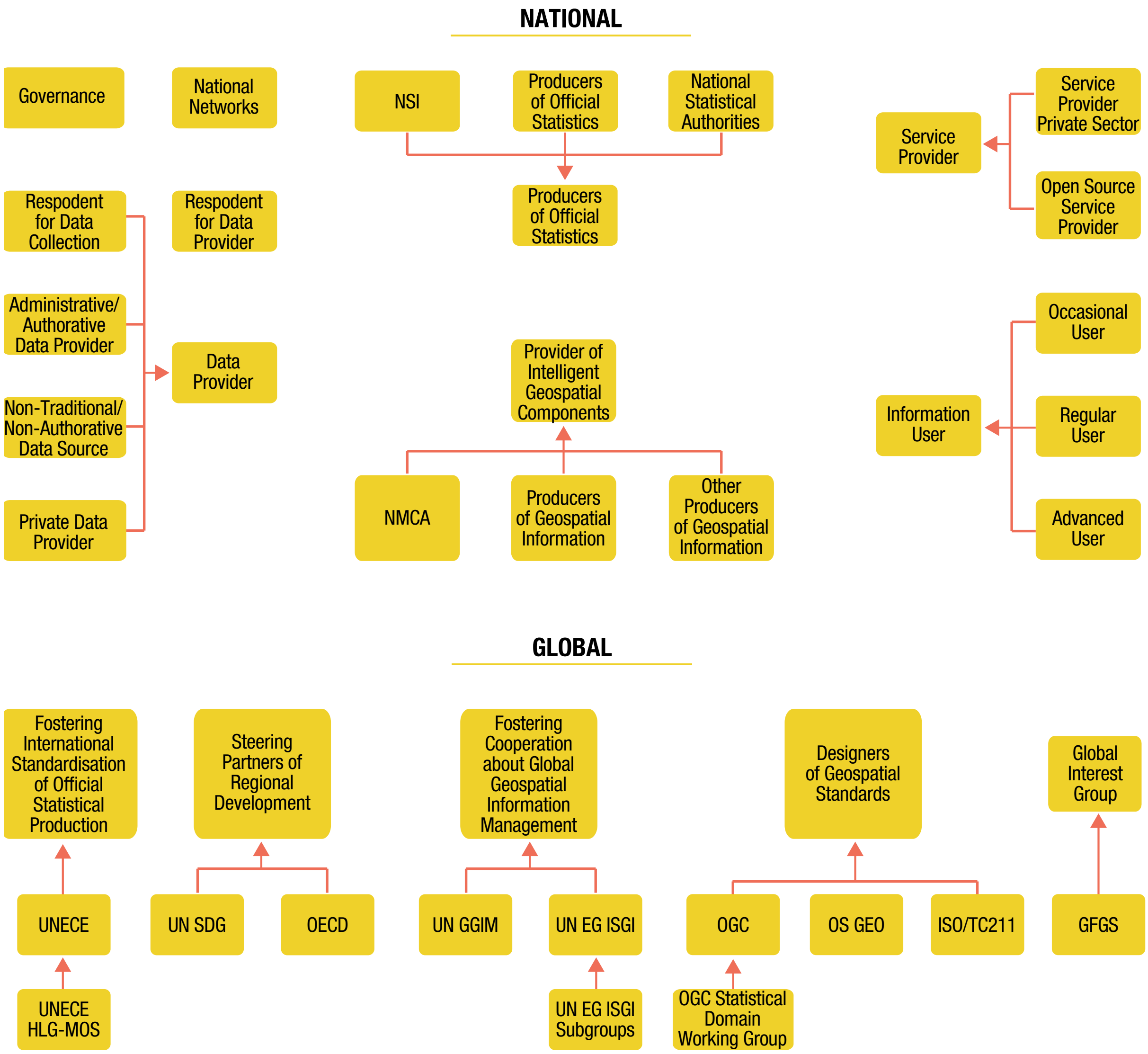
Once parties are carefully identified and summoned for the data integration quest, their strengths, skills and relationships must be mapped (Figure 8). The mapping of actors can be a useful tool to set clear goals and inform strategic decisions. A map with just a few actors, or one where few roles have been assigned to actors, would probably imply that the co-ordinating NSO has a more demanding task ahead to build a solid foundation for statistical-geospatial integration.

A map showing a richer network of institutions with clearly assigned duties will balance the burden of duties and decisions across more contributors. On the other hand, a larger number of institutions involved will require a higher degree of co-ordination and checks and balances for the co-ordinating NSO to manage. Finding the right balance should be guided by a formal political decision to support the co-ordinators of the network and to commit all stakeholders at the same time.

The engagement and mapping tasks account for potential geospatial and statistical and/or administrative data producers as well as users and other stakeholders. One recommendation is to begin the process of building trust and positive relationships among stakeholders with general arrangements and formal agreements among organisations with similar interests. This could be especially helpful in preparing to converge the ecosystem to address Step 4, which relates to inventorying and securing fundamental geospatial data to be shared.

While the final map may be more or less advanced for a variety of reasons, it should at a minimum deliver a common and accepted understanding on what is expected from whom and why.

Figure 8. Examples of key actors of the geospatial and statistical ecosystem



Source : Eurostat - GEOSTAT 4 Project, draft documents.

Task 3. Raising awareness

An important task in the sequential implementation of Step 1 is to raise awareness on the need for integration of geospatial data and statistics. This requires, first, creating the necessary commitment for a top-down transfer of institutional knowledge and related policies regarding data integration frameworks and arrangements. Second, it calls for training spaces to adapt the traditional processes and incorporate new capabilities for each of the participating organisations and their personnel. Third, the co-creation of awareness solutions is recommended.

Issues to be considered in the planning of this task include the negative perceptions about data integration and the misinterpretation on the part of both of individuals and administrations that integrating data is just another policy for budget cuts or a disguised strategy to replace workers with technologies. These kinds of perceptions should be understood, contained and avoided at all costs. The narrative and related messaging should create awareness of the benefits and costs of data integration and of its implementation and aim to improve information knowledge and help people and institutions embrace the technological leap.

Raising awareness is crucial for a successful and effective implementation, and communications should be crafted in such a way to enable new scenarios to be envisioned in the context of a data revolution that will open opportunities to all and not shut down available resources.

The promotion of awareness of the usefulness of geospatially enabled statistics is a matter of knowledge and perceptions and should permeate all involved organisations and their staff. Implementing a consistent plan to face the challenge of enhancing vitally needed information for decision making demands a holistic communication approach that not only informs but targets and addresses uncertainties.

Activities for this task should include:

- producing an awareness plan at strategic and higher technical levels that responds to political commitments and deeply embeds top-level mandates adopted either in the implementation of the NSDS and/or the GSGF
- designing, planning and implementing training programmes to effectively integrate knowledge and guidelines, including this guide, in internal processes and organising kick-off meetings with selected middle-level officers and operating personnel
- formulating co-creation programmes to consolidate the needed awareness to lay the groundwork for the next step of evaluation and human resources. Co-creation, in this context, refers to a multi-stakeholder design of solutions to address the empowerment of the geo-enabled statistics concept.

Task 4. Facilitate a general agreement environment

This task refers to the need to prepare, negotiate and sign formal agreements under the auspices of positive relationships. Once all organisations and their personnel are made aware of the need for integration of geospatial data and statistics, a nurturing environment of common understanding can be facilitated. This requires disclosing geospatial inventories and transparently assessing available data resources.

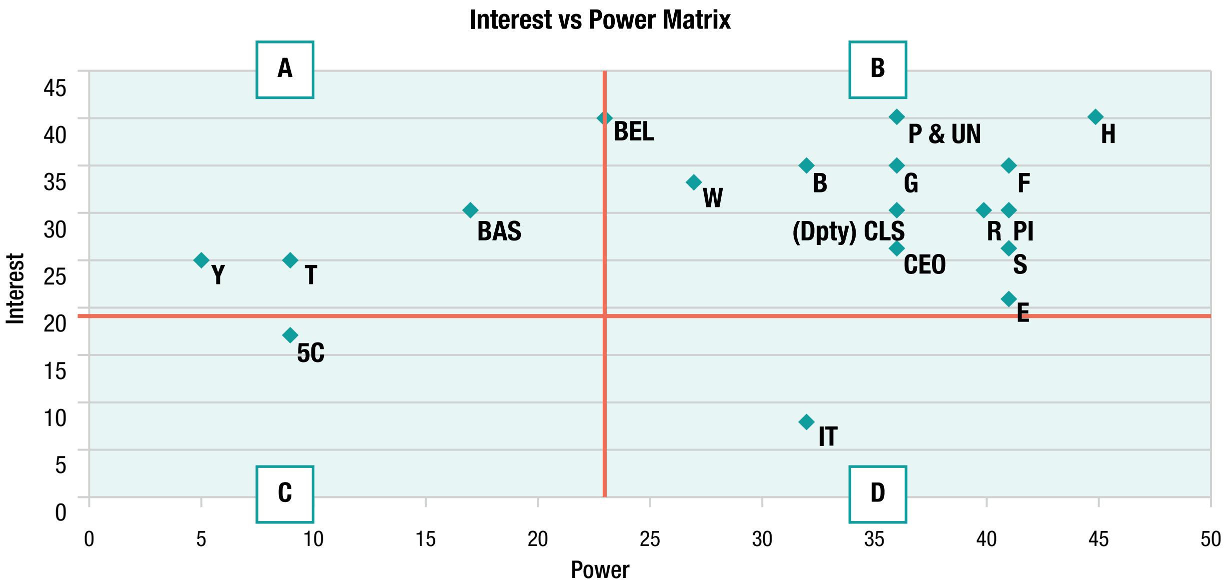
A good way to start to build alliances between organisations, especially in countries where there is a low level of institutional trust, is to sign and exchange letters of intent. Letters of intent may not offer binding agreements, but they may be important as declarations of a willingness to contribute to a common goal. As collaboration matures, letters of intent can be gradually replaced with signed formal agreements to clarify mandates and obligations among organisations.

Good practice use case: Power mapping in Honduras

In a power mapping exercise undertaken in 2009 in Honduras under the auspices of Inter-American Development Bank, a situational diagnosis for the prospective launch of an NSDI in the country was implemented. A workshop that brought together the participants to discuss the proposed infrastructure developed this intervention (Carranza Tresoldi, 2009[12]). Once the basic concepts and cases of data infrastructures were explained to participants, a questionnaire was distributed to elicit their perceptions and impressions.

The survey results were used to map respondents' characteristics in terms of interest, perceived power, capacities and dependence to assess their plans and capacities for a potential NSDI project in Honduras. Among the official data agencies that participated were the NSO (National Statistics Institute), the NGIA (National Geographic Institute) and the national cadastre (Instituto de la Propiedad). Results combined the answers for interest and self-perceived power by adapting the classical logical framework approach, as illustrated in the power mapping matrix in Figure 9.

Figure 9. Example of a power mapping exercise undertaken in Honduras



Source: Carranza Tresoldi (2009[12]), *Metodología de Taller de Presentación y Difusión de Infraestructuras de Datos Espaciales para Honduras [Methodology of a Workshop for Dissemination of Spatial Data Infrastructures for Honduras]*, https://www.academia.edu/218876/Metodolog%C3%ADa_de_Taller_de_Presentaci%C3%B3n_y_Difusi%C3%B3n_de_Infraestructuras_de_Datos_Espaciales_para_Honduras

The results demonstrated a greater grouping of the summoned data agencies and institutions in the fourth quadrant, which suggests they perceive themselves as mostly prepared to embark on an NSDI project. This also suggests a promising engagement for the future NSDI.

Specific recommendations

Although this step does not relate to any specific principle of the GSGF, it should be regarded as an important component of the solid architecture needed to pursue the quest of geospatial data and statistics integration. Under the GSGF, this step and its tasks would be considered part of the key element of institutional collaboration that cements the needed relationships to leverage the powers of the principles within the conceptual workflow of the framework. The important message of this step is that people’s commitment and the creative force they can bring to the implementation process are foundational and that trusted arrangements can ensure smooth implementation of subsequent steps.

5.2 STEP 2 - ASSESS AND SECURE NECESSARY HUMAN RESOURCES

Once the leading statistical agency within the NSS is mandated to co-ordinate the map of actors developed for Step 1 — i.e. getting the right people from the relevant organisations and putting them on a map — it is time to assess the availability of and needs for human resources for the next step. Although this step does not relate to any specific GSGF principle, the need to assess and secure well-trained and technically sound skilled officers and competent personnel is a basis of a solid strategy to undertake the planned tasks ahead.

Nevertheless, this step requires a combination of key elements, especially those related to institutional collaboration and to national laws and policies, to enable a fluid exchange of knowledge and key technical human resources across the organisations depicted in the power map, particularly those that belong to the official realm.

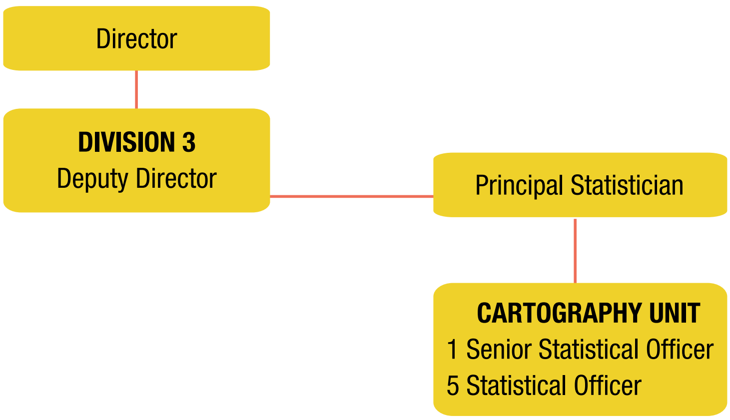
Task 1. Focus on management issues with the assistance of the cartography department and leveraged by enhanced awareness

Consideration of the specificities of human resources management by the heads of NSOs and support staff should start with a general assessment of the organisational chart and job descriptions. Developing the integration of statistics and geospatial data requires, among other things, a genuine spirit of co-operation and relevant technical skills. Planning to meet the new demand for geospatially enabled statistical data, including new expertise to handle different kinds of data from different sources, involves working backward from the delivery of new products to the new requirements, recommendations and good practices needed to complete this task.

In this context, census cartography departments at NSOs, regardless of whether they primarily use digital or analogue technologies, should have a crucial role in planning, deploying and following up the strategy for integrating geospatial data and statistics. Cartography departments traditionally organise their work in units of data capture or processing of third party information such as satellite imagery, thematic mapping production, analysis of resulting data and dissemination. Directors and managers should ask themselves how they should adapt this structure, if needed, to the demand of integrating statistics in their processes.

Usually, cartographic units at NSOs are more or less complex according to what the organisation must produce, either official or referential mapping products, according to the NSO's national mandate. In Mauritius, for instance, the mandate of the Central Statistics Office is to provide timely, relevant and reliable statistics but not official cartographies. Its cartography unit is thus as unidimensional as pictured in Figure 10.

Figure 10. Example of the organisation chart of the cartography unit in a LMIC



Source: Central Statistics Office of Mauritius (2007[13]), "PowerPoint presentation", https://unstats.un.org/unsd/demographic/meetings/wshops/Zambia_8Oct07/docs/Countries_presentations/Mauritius_presentation.ppt.

Whatever the concerned cartography unit's organisation chart, each of the process-oriented functions should be the basis for a needs assessment to anticipate future needs.

TASK 2. RECOGNISE NEEDS, IDENTIFY GAPS, AND PLAN POSSIBLE WAYS TO SECURE SKILLS AND CAPACITIES

Census mapping units also should plan for and be aware of how they will assess the level of complexity of the skills of the required personnel and whether they will give priority to the initial or to more advanced-level profiles after the assessment. Finally, the appraisal should include a consideration of good practices from other countries' processes that could provide valuable lessons.

Needs assessment tools are often helpful instruments to understand potential gaps in service processes. As a designer with Industry Expansion Solutions at North Carolina State University put it, "A needs assessment is the 'what' (what the organization needs) that precedes the gap analysis, which is the 'how' (how to close the gap between where the organization is currently and where they want or need to be)" (Weisberg, 2017[14]).

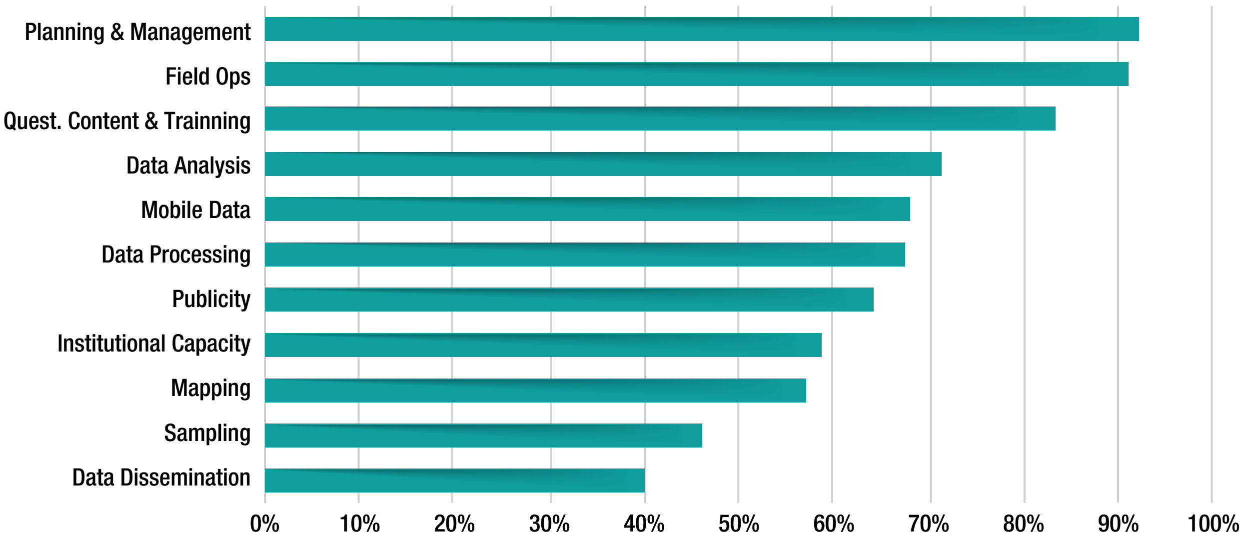
In the context of statistical-geospatial data integration, the "what" refers to all human resources needed to geospatially enable statistical processes in compliance with the GSGF. The "how" refers to the following, among other questions:

- How will the work ahead be organised regarding data capture, thematic mapping, integrated analysis and dissemination?
- How will the level of complexity be tackled? Through entry-level or advanced positions?
- How are the produced choices comparable to processes in other countries?

A recommended tool to undertake an effective needs assessment exercise, although it needs adjustment, is the Statistical Capacity Assessment Tool (TASC), developed by the United States Census Bureau. Among other goals, the tool aims to aid NSOs in identifying areas where improvement is needed; help target beneficiaries and donors justify the need for funding for training in specific areas; and, as it is administered at two points in time, before and after, provide a measure of the impact of capacity-building activities (United States Census Bureau, 2017[15]).

The administrator's manual notes that deploying the TASC assessment in an NSO could help stakeholders make evidence-based decisions. Figure 11 identifies the relative relevance of the needs by contrasting strong areas with other issues where the focus should be set.

Figure 11. Example of aggregated group score for 11 areas of the TASC tool



Source: United States Census Bureau, (2017[15]), Tool for Assessing Statistical Capacity (TASC), https://www.census.gov/data/software/tasc.html.

The example shown in this figure depicts an aggregated group score for the 11 areas that the tool evaluates. It is clear that the areas of planning and management and field operations are strengths in the capacity of the selected NSO. Areas that require additional solutions and resources are mapping, sampling and dissemination.

This example also suggests that it could be useful to strengthen the technical skills of officials working in the cartography unit for mapping and disseminating data, as scores show poor results in those areas. Prioritising areas according to the scores could save and reallocate valuable resources for statistical processing, always judging by the leads that the relative scoring of this assessment suggests.

A final question for decision makers regarding the human resources that are potentially needed is whether competences could be pooled from other external organisations if they are not available in house and the needs assessment exercise shows gaps. Of course, the first place to look for such competencies would be in other organisations within the NSS, particularly in terms of the key element of national laws. In general, resources could be procured from the larger data ecosystem including from the private and civil society sectors as well as academia and international organisations.

TASK 3. DEFINE ROLES AND THEIR NEEDED SKILLS AND CAPACITIES

Once skills and needed capacities are assessed and identified, it is time for the implementers to assign roles in the new geospatial statistical framework. When implementing the GSGF, an NSO must choose whether to rely on the technical capacity and local knowledge of its staff or, alternatively, to seek advice and contributions from impartial experts in relation to technical deployment. It is advisable to make an informed decision based on the availability of, and access to, proven, well-documented sources of expertise so that the most appropriate personnel are selected to manage the transformed infrastructure according to its actual needs.

Although roles and capacities might emerge naturally from the needs assessment, a working structure should usually consider the following roles:

- An NSO director or working group with knowledge of the presented frameworks and strategies for implementation should direct the strategy and be able to develop positive relationships to assure continuous and constructive dialogue with actors of the NSDI, NSS and other data-sharing spaces.
- The head of census cartography or survey cartography should lead the planning, deployment and monitoring of the chosen strategy within the NSO. This role demands a professional with strong technical skills, mastery of geographic information systems and a-state-of-the art background in data interoperability management. Capabilities required for this role should also enable technical translation for the operational implementation of the strategy.
- A liaison person with the NGIA is needed who has strong GIS knowledge capabilities, is knowledgeable about the integration strategy and has strong communication skills to ensure linkages to other members of the NSS.
- Supervisors and operation officers are needed who are trained in specific functionalities for the functioning of the geospatial infrastructure, managing geocoding processes, generating database harmonisation and operating dissemination tools.
- Consideration should be given to roles related to technical interoperability and legal and-technical advice, as circumstances warrant.

It is possible that some or most of the profiles to fulfil these roles may not be available in certain cases. The reality for a number of LMICs is that they may need to figure out how to deal with a situation where the expertise to be hired in the market is extremely limited or the fees are prohibitively high relative to the NSO's budget.

Where NSOs are limited to working only with existing staff, it is advisable to take advantage of co-operation agreements with other organisations within the NSS or NSDI ecosystem. In this regard, there are multiple options to consider in planning comprehensive programmes to train human resources for the geospatial realm:

- In the short run, establish collaborative spaces like communities of practice where expertise could be freely shared and reinforced, including with the active contribution of the more acknowledged officers from other data agencies such as those dealing with territorial issues in the NGIA, the environmental agency and the land management administration, among others.
- In the short run, the NSO also could consider including specific clauses in the agreements initiated in Step 1 to help mobilise resources by offering a temporary status and to use transferred human resources to support training programmes through the previously planned strategies.
- In a medium-term scope, the cartography department or division could open up possibilities for internships for students who are interested in geospatial-related careers and are in the last stages of graduate studies, with interns recruited domestically or from other countries. These pre-graduate information systems and engineering students, especially those with geospatial programming skills, could be a valuable source for innovation and modernisation of the data integration processes and could benefit from the experience.
- Take advantage of the international co-operation support for the 2020 census round from UN agencies like the UN Population Fund and development banks, especially those that could offer non-reimbursable funds to train and provide assistance to update the technical expertise of officers in the cartographic unit and census department.

If the above or similar strategies prove unsuccessful, the organisation should consider a comprehensive retraining programme for existing staff working in statistics or other related areas. To develop specialised skills and abilities in the geospatial field, it is important to identify and train individuals whose profile demonstrates analytical skills and mid-level mathematical and/or statistical knowledge with computer literacy, especially officers oriented to data processing and preferably with database management experience.

Specific recommendations

Assessing and securing competent geospatial personnel for different positions and needs to implement the GSGF requires a balance of efficiency and pragmatism, including cost-benefit analysis and consideration of trade-offs.

- An effective strategy is to determine the roles needed first and identify the human resources with appropriate geospatial technical knowledge and skills to fit the roles identified.
- Needed software and hardware technologies should be seen as a strategic investment.

5.3 STEP 3 - ASSESS AND SECURE TECHNICAL INFRASTRUCTURE

Once human resources are secured, such as support from stakeholders and manpower, the next step is to focus on the assessment of needs to secure efficient technical infrastructure management. This step thus logically follows Step 2, which addressed the need for staff with technical capabilities to move the geospatial statistics enabling processes forward on a reliable technical basis.

Step 3, while not directly linked to any specific GSGF principle, is nonetheless related to the key element of technical infrastructure and it partially contributes to fulfilling Principle 1. This is because a successful implementation of Principle 1 requires that addresses and property and building data be processed and that all location information, including geocoding, is accurate and sufficiently consistent as a result of using common systems. Accomplishing these requires the availability of powerful and specialised software and other processing or disseminating tools that, in turn, require significant storage capacity.

This step covers, from a strategic point of view, all angles of technological infrastructure including software, database selection and storage solutions to enable the required capacities to handle geospatial data processing and their dissemination. In this respect, the need for an organisational transformation under the leadership of the NSO's cartographic department, as recommended in task 2, incorporates the recommendation in this step to transform the technological infrastructure to bring about the changes needed for implementing the integration of statistics and geospatial information.

A particular issue to consider when planning adjustments to the technological infrastructure is that NSOs need more powerful technologies to produce geospatial-enabled statistics. This is because geospatial technology is generally more complex and hence more demanding, particularly in terms of computing capacities and storage, than the typical technology used to process statistics at NSOs.

The Generic Statistical Business Process Model

An instructive way to understand the role of geospatial technologies for the transformation of the technological infrastructure is to identify geospatial processes in the context of the Generic Statistical Business Process Model (GSBPM).

This model is a means to describe statistics production in a general and process-oriented way. Because of the granularity of the model, the UN Economic Commission for Europe (UNECE) is reviewing whether it is fit for the purpose of describing the management of geospatial information in the production of statistics (Vale, 2018[16]).

The Joint UNECE/Eurostat/OECD Group on Statistical Metadata originally developed the GSBPM in 2008, and its structure is based on the business process model used by the NSO of New Zealand. A review of version 5.0 of the GSBPM produced version 5.1, with a clear mandate from the High-Level Group for the Modernisation of Official Statistics to only introduce changes that had a strong business case.

One of the introduced changes recognises the growing importance of integrating statistical data with geospatial data. Version 5.1 is presently an input for both within and between statistical offices' management plans and serves as a common basis for work with statistics production from multiple points of views.

The model is designed to give answers to all processes needed to produce all types of surveys. The word "business" in its name refers to the business undertaken by the statistical office, in a broader-than-usual sense. Moreover, the value of this frameset is that its details enable key inputs to be identified at different moments of pre-established processes for statistical production.

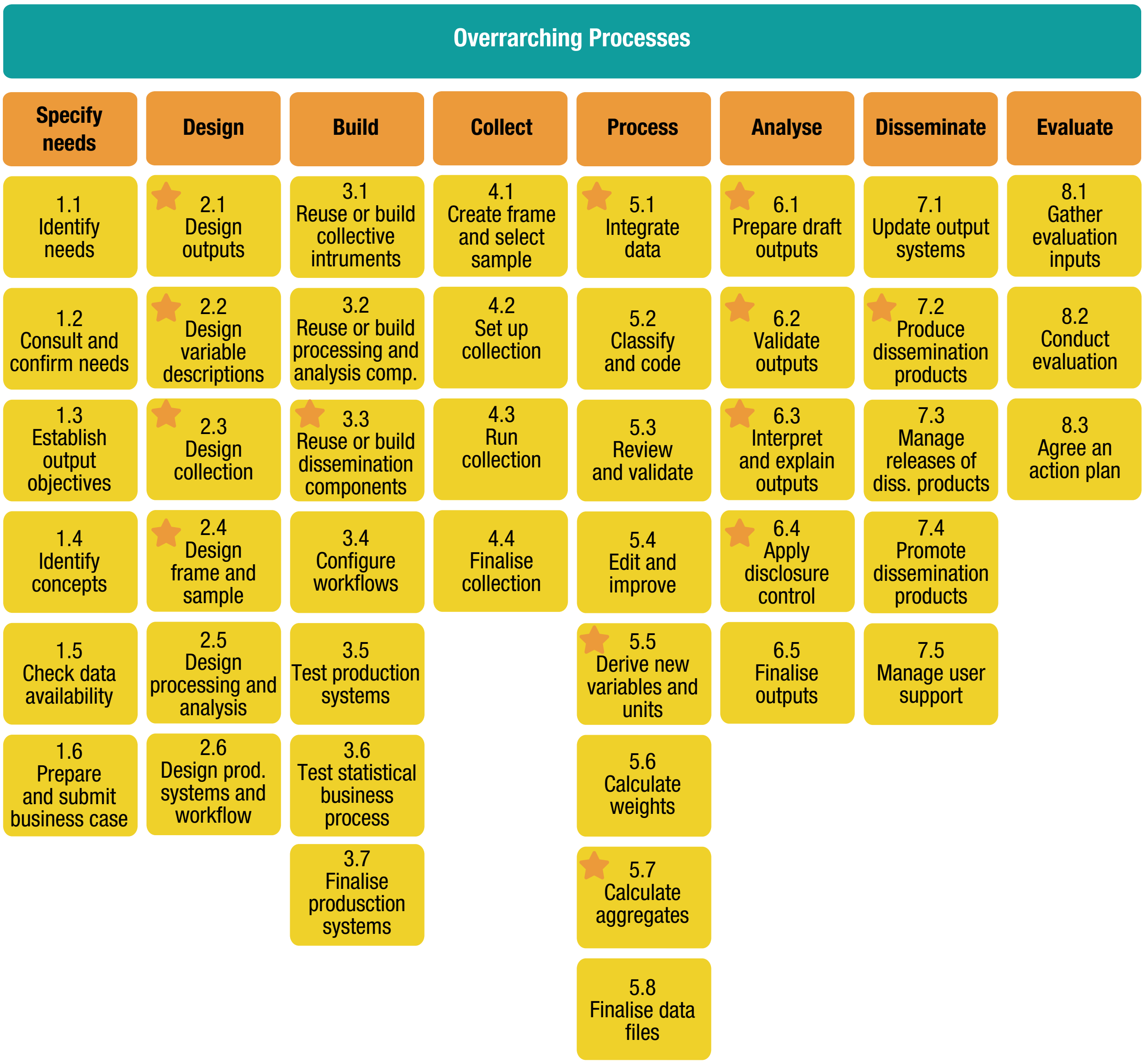
The level of detail described in the model comprises not only the traditional processes but also other processes that are usually assumed or taken for granted, such as specifying needs, building instruments and components, or evaluating the outputs.

A mapping of geospatial aspects is presented in Figure 12, which shows all cells or pieces of sequences of sub-processes through the processes detailed in each of the columns marked with a red star.

Among other sub-processes, the GSBPM identifies with geospatial inputs the design collection of data, the design of a frame and sample, the running of the collection of data, and the integration of data based on geographies or the interpret outputs along several stages including the design, collection, analysis and dissemination processes.

Focusing on the traditional processes considered in the statistics production realm — for instance, collection, processing, analysis and dissemination — an informed decision maker for the integration of geospatial technologies should consider which technological inputs are needed for the transformation of the infrastructure to geo-enable statistics.

Figure 12. Levels 1 and 2 of the Generic Statistical Business Process Model



Source: United Nations Economic Commission for Europe (2019[17]), Generic Statistical Business Process Model GSBPM, Version 5.1, https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1.

Departing from the statistical business process model, and once all geospatially related issues have been diagnosed to be needed for the technological transformation of the NSO, it is the right moment to consider some basic questions such as the following.

What technology will be integrated for the data collection process and how? What are the best cost-effective tools and applications for field data collection?

There are many possible answers to these questions, depending on the available degree of integration of the GSBPM implicitly used by the NSO and its hardware and software inventories. The market, including groups of developers and data producers advocating for open technologies, offers a broad portfolio of technologies ranging from sophisticated devices with precise built-in global positioning systems and software development kits based on open data and open source software to hardware that uses regular smartphone technologies.

What technology will be used, and how, to undertake processing and analysis of data? Which are the best software and hardware efficient enough to do the work in the NSO?

The answers to these questions depend on the scale of the trade-off between proprietary source, but assisted, technologies such as commercial GIS software and free licensed technologies that require more skilled programming and community working capacities from the users. A broad diversity of solutions is offered for geospatial integration analysis and processing, which are featured in the more popular commercial solutions including Esri ArcGIS, ArcGIS Online, Google Earth Pro, Google Maps API and Maptitude. Among the interesting open source solutions are open source and non-licensed software such as QGIS, GVSIG and their complementary developments as well as applications from the OpenStreetMap community.

What level of processing speed, performance specifications and storage space will the NSO need for data integration?

The answer to this question is generally linked to the minimum requirements recommended by the selected software. However, some examples of geospatial infrastructures could be found where minimum and maximum capacities are determined (to be developed further).

What kind of processing spatial databases needs to be in place?

This question is linked to the two previous questions, as all GIS need to rely on spatial databases to query the data from and to apply their tools and developments. A spatial database is an enhanced version of statistical classic databases that can store and access spatial data or data that define a geometry element, as in the case of boundaries of administrative units.

The data that spatial databases store are often associated with geographic locations or, in the case of cities or administrative locations, with spatial entities. Data in spatial databases are stored with geometry elements such as points, lines, polygons and topology. Given that these kinds of tools integrate multiple dimensions, often including the temporal one, they require more storage space capacities.

The kinds of spatial databases to be considered for the implementation of the GSGF are either structured query language (SQL) databases or non-SQL (NoSQL) databases. SQL is an international standard that defines a language for querying relational databases. NoSQL databases are instead prepared to work with changing data structures or semi-structured and unstructured data sources.

One of the most popular relational databases is PostGIS, a free and open source extension to the PostgreSQL database. SQL Server, from Microsoft, is also one of the company's top-level databases. Spatial and Graph are the spatial extensions to the well-known Oracle database. Among the most used NoSQL spatial databases and extensions are the Google App Engine with GeoModel and GeoDataStore extensions and the MongoDB with MongoDB Geospatial Queries extension.

What technology will be implemented, and how, in the front end? Are service-oriented solutions like map servers or storage space over cloud services going to be used?

On the front end of the GIS platform, the GIS portals are currently one of the most used technologies in NSOs for the purpose of dissemination of geo-enabled statistics and metadata. These are web-based, interactive platforms that use databases that are modelled for different data types including the domains of microdata, macro data and geospatial data. Often, statistics are also combined with maps through impacting visualisations leveraged by publicly available platforms, either those that use proprietary copyrights of software and data, such as ArcGIS and Google Maps, or those that advocate for an open data philosophy, like OpenStreetMap.

A general recommendation to implementers at this point is to consider that dissemination should acknowledge the needs of internal and final users at the same time. This is especially vital when balancing the autonomy of having internally administered systems versus reliance on externally provided assistance through software services. In many cases, having on-premises servers for GIS platforms might be preferred to cloud-based GIS platforms. However, as desktop solutions need to be installed locally on the organisation's servers (or provided on lease from a third party), this will impact in the costs and the capacity of managing locally stored maps. The previously described option also depends on the capacity of internet provision. A second option is to host on a vendor's server, usually in the cloud, a software as a service platform and allow the public to access tailored geospatial products through a third party browser. The decision of whether to choose a desktop GIS or software as a service solution should depend not only on consideration of laws and regulations but also should be based on the corresponding policies, including adherence to the 2030 Agenda, related to the kind of access (open or more restricted) to internal and end users. The final decision should consider several aspects such as the affordability of software licenses and copyright limitations, internet service capabilities, privacy and security regulations, and other related matters.

Specific recommendations

As discussed, a general rule of thumb for making correct decisions is to focus on balancing the pros and cons of choosing a specific technological infrastructure. Whether its components are commercial, open source or a hybrid combination of solutions, any final outcome for transforming the technological infrastructure should result from a careful study of all possible angles.

Among the angles to be considered are sustainable financial decisions, legal and statutory issues (legislation that requires the use of open source software, for example), fit-to-purpose solutions, and technical knowledge capabilities available at the NSO or within the NSS. No matter the final outcome for the chosen solution, the aforementioned mapping module of the TASC tool can also help as a guide to making the right technological infrastructure decisions, especially if an adapted version of the tool is considered.

Finally, given that this step responds to a partial implementation of Principle 1 of the GSGF, it should be seen as a foundational matter once support is secured from the data ecosystem and specialised staff. Some could argue that the cost of acquiring the needed infrastructure to simply publish geospatially enabled statistics is too high, and this is an understandable argument. However, many experts emphasise countries should approach the reuse of mainstream data like census cartographies and other related sources as a continuing, project cycle profiled process (UN DESA, 2009[18]; United Nations Economic Commission for Europe, 2019[17]; Luaces et al., 2004[19]; Massachusetts Institute of Technology, 2012[20]).

Implementation of the GSGF is more than a sequence of mapping processing activities designed for dissemination of geospatially enabled statistics. It aims to achieve a technological leap to enable data ecosystems to grow inside and between organisations.

5.4. STEP 4 - INVENTORY FUNDAMENTAL GEOSPATIAL DATA AND ADDRESS GEOCODED INFRASTRUCTURES, SAFEGUARDING ACCESS TO ALL SCALES

As the GSGF implementation advances, a clearer picture will emerge of the actors and their roles, capabilities and technologies. Once the human resources and technical capabilities are secured, it is time to shift focus towards the data to be fed into the framework and the output information that the framework will produce.

The next level of transformation, which the leading NSO must build around geographic information, involves establishing the relationship between a country's geospatial data assets and the various sets of statistics that the NSO uses and produces. In many LMICs, NSOs will find themselves facing a situation in which official geospatial data are not immediately available or are unfit for their purposes.

In many LMICs, multiple government agencies produce analogue and digital geospatial data. NGIAs in most countries have mainstreamed digital techniques through the whole geospatial data production process. Furthermore, other government departments and organisations such as environmental or water resources administrations, departments responsible for utility services, public and private transportation services, and health care planning and implementation offices use and produce prodigious geospatial technology to manage the information they collect and analyse for their decision making. Facilitation of data-sharing practices through detailed documentation and metadata across user networks can help formalise agreements when trusted data are identified.

From a statistical production perspective, it is relevant to distinguish geospatial information needed as infrastructure for georeferencing from geospatial information needed to create, or support the creation of, statistical content.

The role of "infrastructure data" is to provide a basis for georeferencing statistical information. Infrastructure data typically comprise addresses or buildings and play a critical role when assigning a location to a unit record in census operations, etc. Within the European Statistical System, there is a widespread use of address registers or building registers as uniform infrastructure for georeferencing by means of geocoding. In many LMICs, however, there may not be such data infrastructures in place, leaving NSOs the sole option of building one from scratch by means of field collection of location data for buildings, dwellings or settlements.

The role of "complementary data" is to support statistical-geospatial production. A typical example is the satellite imagery and road network data needed to construct good enumeration areas in the preparation of a census operation.

Step 4 assumes that in order to complete the implementation of Principle 1 of the GSGF, two kinds of interrelated tasks will be needed. The first task refers to the inventory of fundamental data; the second refers to the georeferencing infrastructure as a vital component to enable statistics to be georeferenced and allowing their flexible aggregation from the most precise location of single points all the way up to the national scale.

Task 1. Inventory of fundamental data

A good starting point for any inventories of fundamental data is offered by the UN initiative on Global Geospatial Information Management list of Global Fundamental Geospatial Data Themes (UN DESA, 2019[21]). The list presents (and discusses) 14 themes considered fundamental to strengthening a country's geospatial information infrastructure and implementation of the 2030 Agenda for Sustainable Development and the 17 SDGs in particular. The list does not aspire to be exhaustive or complete, but rather is a baseline goal.

The active search and harmonisation of geospatial data sources should consider, among other needs and possibilities, a number of specific thematic layers. However, before considering these, the Global Geodetic Reference Frame is needed to allow users to precisely determine and express locations on the Earth, as well as to quantify Earth changes in space and time. A Global Geodetic Reference Frame is a prerequisite for the precise collection, integration and use of all other geospatial data. The thematic layers include:

- Addresses: Labels that usually contain a property number, a street name and a locality name are used to identify buildings, other constructions and land plots and are related with coordinates indicating their geographic position.
- Buildings and settlements: A building refers to any roofed structure permanently constructed or erected on a site. Settlements are collections of buildings and associated features where a community carries out socio-economic activities.
- Elevation and depth: These describe the surface of the Earth relative to a vertical datum.
- Functional areas: These are the geographical extent of administrative, legislative, regulatory, electoral, statistical, governance, service delivery and activity management areas.
- Geographical names: These provide orientation and identity to places such as regions, settlements or historical places.
- Geology and soils: Geology refer to the composition and properties of geologic materials like rocks and sediments; soils are the upper part of the Earth's crust.
- Land cover and land use: Land cover represents the physical and biological cover of the Earth's surface, and land use is the current and future planned management of the natural environment for human purposes.
- Land parcels: These are areas of land or more generally of the Earth's surface (land and/or water) under common rights. This theme can include individual fields and cadastral parcels.
- Ortho-imagery: These are georeferenced, rectified image data of the Earth's surface that are a widely used data source for many other data themes.
- Physical infrastructure: This theme includes industrial and utility facilities and the service delivery facilities associated with administrative and social governmental services.
- Population distribution: This theme covers the geographical distribution of people, including population characteristics.
- Transport networks: This theme comprises the suite of road, rail, air, cable and water transport routes and their connectivity.
- Water: This theme covers the extent and conditions of all water features including rivers, lakes and marine features.

A decisive challenge when incorporating geospatial data into statistical production is quality. According to the Handbook on Geospatial Infrastructure in Support of Census Activities (UN DESA, 2009[18]), critical success factors for geospatial implementation in national statistical offices include, among others, well-established quality control and quality assurance procedures. These are particularly important when using existing digital data from external sources and starting to integrate them into established internal processes. Another challenge for integrators of databases is the insufficiency or absence of metadata. If such information is missing, it is difficult to assess the fitness for purpose of the geospatial information. Moreover, missing information about the geographic reference framework could lead to errors when translating between reference systems.

It should be stressed that NSOs, especially in LMICs, should not be restricted to using fundamental geospatial data from governmental organisations. The UN Secretary-General's High-Level Panel of Eminent Persons put it this way in its 2013 sustainable development report: "Better data and statistics will help governments track progress and make sure their decisions are evidence-based. …This is not just about governments. International agencies, CSOs [civil society organisations] and the private sector should be involved. A true data revolution would draw on existing and new sources of data to fully integrate statistics into decision-making, promote open access to, and use of, data and ensure increased support for statistical systems" (High-Level Panel of Eminent Persons on the Post-2015 Development Agenda, 2013[22]).

In the context of data integration, continued advancements in earth observation, GIS and mobile technologies have promising implications, as the panel's statement suggests. New sources of data derived from satellite imagery and aggregated anonymised mobile phone data, for example, can greatly contribute to strengthening the geospatial infrastructure of LMICs. Many data sources are becoming available at granular geographic scale. Not only can these data be used to directly calculate city-level and even greater-scale information, but they can be combined with surveys, census and administrative data, and information from the public services system to model statistical outcomes with tremendous detail and accuracy.

Task 2. Establish the georeferencing infrastructure

The GSGF calls for "all statistical unit records [to] include or be linked to a precise geographic reference (ideally an x- and y-coordinate), and if not, the smallest geographic area possible". Appropriately, georeferencing statistical unit records to a precise (point) location will foster the greatest opportunity for reuse and aggregation of geospatially enabled statistical information.

To provide this precise geographic reference for all statistical unit records, one of two different possible paths are open to NSOs:

- Use an existing data infrastructure with buildings or addresses that can serve as reference data.
- If such data do not exist or are incomplete, build their own reference data from scratch.

Generally speaking, the benefit of the first path is efficiency. Using an existing data infrastructure saves resources. However, if the inventory of fundamental data (task 4.1 of Step 5) finds that useful infrastructure data (buildings or addresses) do not exist, are insufficiently structured, or have serious gaps, the NSO will have to build this infrastructure data on its own. Typically, this will be undertaken as an integrated part of or as a preparatory stage to a population and housing census operation.

Considering the general data scarcity in LMICs, this guide assumes that most LMICs will face the second possible path to ensure a precise geographic reference for statistical unit records.

Some general advice for completing this task includes the following:

- The georeferencing infrastructure should aim to capture the location of the finest possible object level. In principle, this means that for population and housing census operations, the location of each individual building (or centres of small rural settlements) should preferably be recorded.
- The georeferencing infrastructure should preferably be based on a point-based rationale. This means that the finest possible object level should be represented as single point locations rather than coarser areas.
- After completing the georeferencing infrastructure (for the first time), a work plan should be drafted to safeguard a sustainable and long-term maintenance — and possibly improvement — of the infrastructure. Wisely managed, such data can save a lot of resources in future census operations and prove useful for a range of purposes.

Good practice use case: Establishing a georeferencing infrastructure for the 2021 Census in Namibia

The following example is meant to illustrate an important census practice for NSOs that are starting to implement the integration of geospatial data and statistics: establishing a basic georeferencing infrastructure to enable geocoding. The experience of Namibia, a lower middle-income country, offers an interesting option for countries on the path of building their own reference data from scratch.

Namibia is presently conducting a comprehensive census mapping coverage in preparation for the upcoming 2021 Population and Housing Census. Similar to many LMICs, Namibia faces the challenges of an unlegislated National Addressing System (NAS) in both urban and rural areas. Although geographies for urban areas include villages, towns and municipalities, un-standardised street and plot numbers are mostly used as addresses for navigation purposes in the major towns and municipalities. Also, addresses from localities are not harmonised or even available throughout the entire country. There is a vital need to leave no one behind as many names of localities or places used in rural areas are only represented by points without any further disaggregation. Statistics have until now been collected at the enumeration area level and aggregated and reported through smaller scales of administrative units, including the regional and national levels.

The chosen methodology to overcome the lack of structured addresses is depicted in Figure 13. It consists of collecting basic variables on all built-up structures and linking these to basic household data. Basic household and building variables framed within an existing enumeration area are identified with a unique point and linked to their location through dimensional coordinates using an automated process.

The coordinates of buildings are extracted from selected high-resolution satellite images. In this way, a nationwide built-up structure database is created and serves as a basis for the integration of statistics and geospatial data for the enumeration phase of the census operation planned for 2021. This strategy provides each census questionnaire with a precise link to the building using the Cartesian coordinate system and enables the production of statistics for any identified georeferenced spatial object.

Figure 13. Building points in an enumeration area polygon

Source: Namibia Statistics Agency.

The Namibia Statistics Agency has also considered inter-temporal changes of spatial objects and is currently developing a tool to automatically extract new built-up structures and continuously update the built-up structure database. Using a machine learning algorithm, the innovative tool aims to detect changes utilising optical images complemented with direct field verification. Although this is a fit-for-purpose solution for the census preparation, Namibia acknowledges the need to formalise physical addresses and is planning the development of a standardised NAS that will eventually be used to geocode statistics. However, the 2021 Population and Housing Census is envisaged to employ a point-in-polygon enumeration technique using building structures linked to two-dimensional coordinates. This geocoding choice can pave the way for development of a uniform NAS, especially considering the pressure to formalise a comprehensive address system in the midst of a pandemic crisis like COVID-19. In the absence of a proper physical addressing system, geographically disaggregating by x, y coordinates is an alternative to ensure no one is left behind during the census enumeration. Lessons learned from the response to COVID-19 have further spurred demands for the full implementation of an NSDI project as part of the NSS.

5.5 STEP 5 - SET UP A BASIC FRAMEWORK OF GEOGRAPHIES FOR ANALYSIS AND DISSEMINATION OF DATA

In all the previous steps and tasks, basic elements were defined individually to support a foundational geospatial infrastructure, including relationships, people, technologies and data. Now it is time to correlate all those previous steps and put them to work together to produce consented and harmonic outputs.

A key component of a foundational geospatial infrastructure is the existence of updated and harmonised subnational boundaries or geographies. Geographies are spatial representations of the administrative, statistical or functional division of a country. Each country has its own unique setting of administrative, statistical and functional geographies ranging in scale from country parts, provinces, regions and wards to local units such as enumeration areas, mesh-blocks, parishes or neighbourhoods. In some countries, grid geographies also comprise part of the framework of geographies. In the European Statistical System, provision of census data with grid references (1 x 1 km grid cells) is part of the European census legislation.

Many countries in the developing world face significant obstacles to the creation of consistent boundaries due to reasons such as territorial conflicts and a vast diversity of involved agencies at different administrative levels. The practice of using paper maps rather than digital ones creates additional challenges to establishing standardised boundary data across different organisations and government agencies. As geographies play a crucial role in most statistical-geospatial data integration processes, it is recommended that efforts be undertaken at an early stage to set up a coherent framework with at least a minimum structured set of core common geographies.

As in Step 3 with regard to implementation of Principle 1, this step is concerned with the implementation of Principle 3, e.g. counting with common geographies for the dissemination of statistics. In this step, the needed transformation at the NSO aims to advance further in the data integration process, ensuring compatible geographies and spatial units.

This step proposes a combination of two complementary and parallel elements: a technical and methodological element to define and delineate common geographies, on one side, and a participative element, on the other. This reflects the need to involve stakeholders in the creation of geographic classifications, which is crucial to ensure the production and dissemination of information to support informed decision making at all levels of society. The participatory element is especially important where mandates over boundaries are distributed among different official data agencies.

The importance of creating concurrent geographic classifications or interchangeable classifications is that it makes available schemes for dissemination in a predictable and consistent way, facilitating a coherent decision-making process.

The multi-stakeholder project, Geo-Referenced Infrastructure and Demographic Data for Development (GRID3), assists countries in generating, validating and using geospatial data to provide information on population, settlements, infrastructure and boundary issues (GRID3, 2020[23]). With an interesting portfolio of case studies for LMICs, GRID3 catalyses the expertise of partners from government, the UN, academia and the private sector to design adaptable and relevant geospatial solutions based on the capacity and development needs of each country. In this way, GRID3 provides a participatory process that brings together disparate authorities to collaborate on boundary harmonisation.

A particularly illuminating process was for Zambia. Over the course of a fit-for-purpose workshop organised by a GRID3 team to solicit the views of all involved actors, GIS technicians from participating ministries sought to harmonise ward and district boundaries by seeking consensus on official district formalised and non-formalised records, topographic maps, and aerial imagery. Using the co-created output, district boundaries were validated and then shared with provincial and local planners for comments and final adjustments. The endeavour produced 116 district boundaries that were aligned to topography and harmonised with wards at the same mapping scale. Over a period of five months, district shapefiles were endorsed and shared with nine government ministries and with local-level provincial and district planners. The district shapefiles have also been uploaded to a national data platform in Zambia.

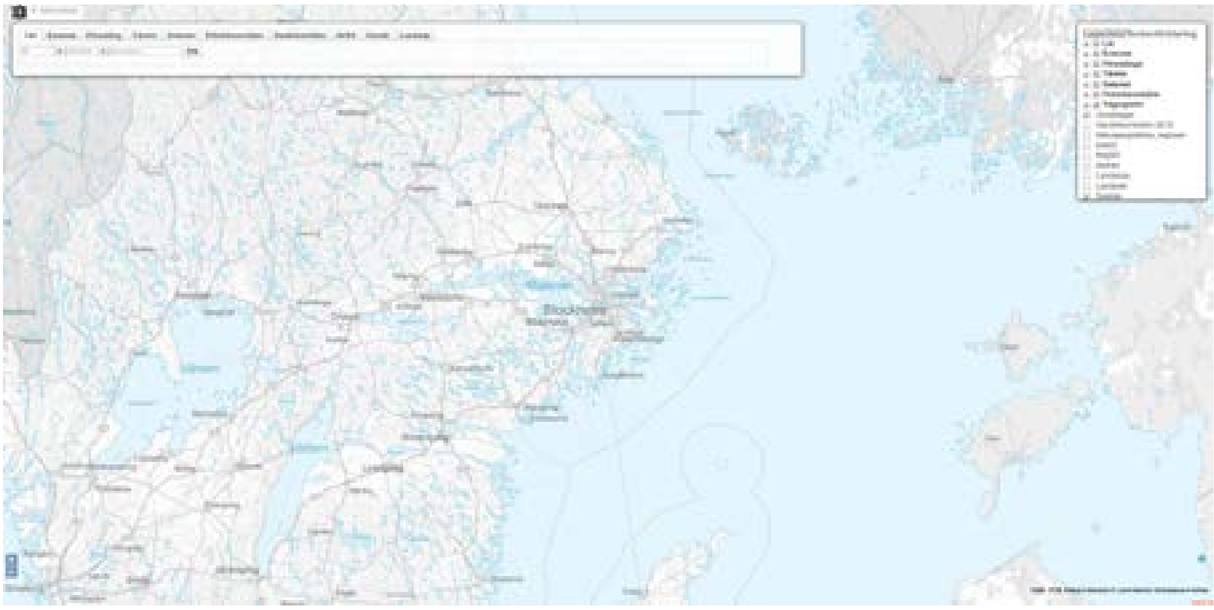
GRID3 has also responded to the challenges faced by countries in which no census has been conducted for more than a decade. In the Democratic Republic of the Congo, for example, the difficulty of creating updated EAs was an important issue due to uncertainty over the relative changes in population for different territories. The GRID3 project's solution to the problem was developing hybrid approaches such as semi-automated delineation to create small geographic units in preparation for census cartography. This method, which calls these "pre-EAs", employs geospatial processing tools to produce boundaries using geospatial datasets on features such as roads, buildings and rivers and high-resolution gridded population data based on plausible assumptions. The estimated population for each small unit is then calculated from high-resolution gridded population data, with units iteratively merged to meet the criteria specified in terms of population size, geographic area and other aspects. The result is a complete set of pre-calculated boundaries that can be used for census planning.

Good practice use case: REGINA common geographies service

In the 1960s, the Swedish government set up a governance structure for administrative geographies whereby the NGIA maintained official boundaries and the NSO maintained official codes of administrative units. This governance scheme is still operational and ensures continuous authoritative information on boundaries and codes for the government, researchers, private companies and the public. The NSO keeps a record of all changes in the administrative division of the country from 1952 to the present. The record is accessible as a web service called REGINA, which any user can query for year-by-year changes in an administrative division (Statistics Sweden, 2021[24]). A visualisation interphase service is connected to REGINA where the user can display the geographies on a map, sorted by type of geography and reference years. This is represented in Figure 14.

While the Swedish governance of administrative geographies was successful in terms of providing universal and authoritative data on the division of the country, there was no integrated service for searching and visualising integrated boundary data and accompanying codes. To address this need, the NSO created the web application, REGINA Web Map (European Forum for Geography and Statistics, 2019[25]), enabling users to concurrently search the REGINA database and display the boundary data. REGINA Web Map is built on an open, web-based interface. Fifteen different types of administrative and statistical geographies are currently available in REGINA Web Map, ranging from counties and municipalities to election districts and functional geographies like urban areas.

Figure 14. User interface of REGINA Web Map



Source: Statistics Sweden.

Specific recommendations

The recommendations offered in this step address both strategic and operational levels. Recommendations for the strategic level are as follows:

- Good governance of boundaries and coding systems is key! NSOs, NGIAs and any other authorities involved need to define and establish clear custodianship roles in relation to the geography framework. Whenever there is a change of the administrative division of the country, it should be clear which organisation is responsible to update and disseminate new data.
- Good governance should build on the agreements between organisations involved and other arrangements related to the data sharing.
- Emphasis should be given to disclosing and formalising uncertain legal status of boundaries and to facing the challenge of obtaining authoritative status of data. There should ideally be one authoritative master dataset for use across government institutions to avoid problems with ambiguous boundaries or confusion of codes or names.
- A continuous consensus among parties is essential in relation to prioritising the content of the geography framework. Priority should be given to those geographies that have the most impact on regional statistics and regional and local decision making.

Recommendations for the operational level are as follows:

- Countries may want to consider the massive scanning of available paper maps to harmonise boundaries of geographies. This practice is usually performed when digital maps are not available for remote or inaccessible areas where only paper maps remain as the geospatial record of the area. Paper maps have to be first converted into a digital format usable by computers. Scanning can convert paper maps into a digital format by capturing features as individual cells or pixels, producing an automated image to be later used with GIS applications to compare existing geographies.
- Technical and methodological issues regarding boundary data and coding systems should be tackled. These should include basic data "hygiene and cleansing factors", always seeking to avoid gaps and overlaps; discontinuity of scales; and ensuring a nested structure of areas and basics of a good coding system, among others.
- Grid geographies should be considered as a complement to traditional geographies. The benefit of grid geographies is the stability, robustness and fitness for geospatial analysis.

5.6 STEP 6 - DESIGN A DATA MANAGEMENT ENVIRONMENT FIT FOR MULTI-PURPOSE USE

Once Steps 4 and 5 provide implementation of Principle 1 and Principle 3 within a process-oriented approach, it is time to implement another important layer according to the GSGF: Principle 2.

Using the georeferencing infrastructure defined in task 2 of Step 4, Principle 2 supports the linking of each statistical unit record to a geographic reference (e.g. a coordinate or small geographic area) that will allow statistics to be applied to any geographic context. The georeferencing infrastructure will support integration or linkage of data from other data sources and mitigate challenges that arise with new geographies or changes in existing geographies

There are broadly three methods to georeference a statistical unit record: inclusion of a direct location reference in the form of x- and y- coordinates; reference to a physical address and/or building; or references to another form of geography (EAs, etc.). From a technical point of view, the means and methods for georeferencing statistical information are defined by the infrastructure provided through task 2 of Step 4.

If high-quality physical address or building data exist, geocoding can be used as the main method to link each statistical unit record to a geographic reference. Strictly speaking, geocoding is the process of linking unreferenced location information, often in the form of a text string that is associated with a statistical unit, to a geocode. Alternatively, the geocode can be directly incorporated into the statistical unit record. Simply put, it is transforming addresses (such as a street address usually composed of a number and a name) into geographic coordinates (such as latitude and longitude).

As mentioned under task 2 of Step 4, many LMICs will not be able to rely on existing data infrastructures comprising address systems of building information. In such cases, the georeferencing infrastructure will have to be built concurrent with census work, where location (x- and y-coordinates) is recorded as part of, or as a preparatory stage to, the surveying operations. Regardless of the methodology, best practice is to always georeference data to the smallest possible geographical level. Best practice is also to carefully document the method used to georeference data, the accuracy of the location and the reference system used for location data capture. Such information is important for quality assurance of the georeferenced output. Doing so will in turn promote the use of standards and good practices of geocoding to make data transparently available and with easily accessible information, unlocking at the same time new insights and relationships between data of different nature.

Once a georeference is associated with each unit record, this enables the statistics generated from these datasets to be produced for a wide range of geographic contexts — for example, various administrative and statistical geographies as well as grid systems — and supports future aggregation of statistical data into new geographical units or adaptation to changes to existing geographies over time.

A key benefit of storing geo-enabled statistical microdata is that these data can be easily incorporated into geospatial analyses such as measures of proximity of households to public open space or to health or other services and activities. Another benefit is that it enables data linkage processes that are dependent on, or enhanced by, having precise location information — for example, in matching census and administrative records based on location and other demographic characteristics.

Storing of geospatially enabled microdata (in a data management environment) relies on a technically sound, safe and capable environment. Storing of data including geocodes, coordinates and geometries will require significant storage capabilities including, as discussed, considerations of privacy, security and flexible disaggregation.

At this point in the transformation process, NSOs in LMICs may be at a critical stage, especially if they have been struggling to establish a reliable georeferencing infrastructure and have lagged behind in the use of, for instance, server-based storage or databases capable of handling geospatially enabled data. Therefore, the NSO's general information technology (IT) capabilities play a central role. Strong emphasis should be placed on tackling the lack of a sound IT infrastructure for successful implementation of Principle 2.

This step should incorporate, among other angles, a multi-purpose perspective so that data are not organised and stored in a way that prevents their flexible use. It is important that geospatial aspects are considered when planning for and organising data from census operations.

Step 6 can be broken down in the following four tasks:

Task 1. Diagnose the technical infrastructure

A thorough diagnosis should be made of the storage conditions situation, including proper assessment of all storage capacities within the NSS and NSDI systems in order to organise a multi-purpose infrastructure and avoid, among other risks, duplication of capacities to take advantage of scale economies while ensuring the formal commitment of all parties.

Such diagnosis also should include an assessment of how data management tools are being used in all relevant institutions, the techniques and methods being used to store, and the standards and good practices these institutions would consider following to facilitate the linking and management of geocodes within statistical datasets. A comprehensive study, once it is made available, will be useful to ensure that in the next step, privacy and confidentiality requirements are correctly managed for the released data.

Task 2. Describe and document georeferencing methods

A detailed study and documentation of viable georeferencing methodologies should be completed once the storage diagnosis is made. This is important to identify the state of the art in geocoding techniques and compare them to the actual situation.

Task 3. Produce internal supporting material

Manuals and handbooks should then be produced, ideally through co-creation exercises that include the definition of key roles for georeferencing as well as a clear definition of technical, supervision and custodianship roles for the full scope of designing a data management environment fit for multi-purpose use.

Task 4. Conduct training

Finally, after completing the inventories of techniques, methods, good practices and documentation for enabling an adequate storage infrastructure, an immersion workshop and training programme are recommended for all involved stakeholders, both at the strategic level and for those in technical positions.

Issues to be covered in these trainings should include, among other topics, statistical microdata good practices; simplification of the aggregation of data for larger geographies; techniques to design unique identifiers or coding for small area geographies as well as standard grid cell management; adaptation of changes introduced by the previous step of existing geographies; principles for protection of privacy and confidentiality management; and consistency tests for geocoded information and metadata, including their maintenance.

Good practice use case: Microdata management using Redatam

Redatam — Microcomputer Data Recovery for Small Areas (REcuperación de DATos para Áreas pequeñas por Microcomputador in the original Spanish) — is an in-house project of the Latin American and Caribbean Demographic Centre, which is the population division of the UN Economic Commission for Latin America and the Caribbean (United Nations, n.d.[26]). The division has supported member and non-member countries in the processing and dissemination of statistics and censuses, particularly the Population and Housing Censuses and their use in the current census decade. There are versions of Redatam in 33 Latin American and Caribbean countries, five African countries, and even in Asia. The project also supports needs of developed countries, and about 1 600 users and institutions from 80 nations are registered users.

Redatam uses a highly compressed, hierarchically structured database with millions of records on people, houses, city blocks and administrative divisions in a given country drawn from data obtained from any combination of censuses, surveys or other statistical sources. It allows users to define, from a database, any geographic area of interest (using city blocks) or combinations of areas to create new variables; to obtain several types of tabulations swiftly; and to export outputs to other software such as digital mapping software.

Redatam's capabilities include socio-demographic analysis. This development is interoperable with the QGIS software referred to in Step 3, extending the possibilities of the system through several applications and plugins. QGIS software enables the use of geoprocessing tools, including union tables of attribute, which allows the creation, in turn, of detailed maps linking information from population censuses and local administrative units in the territory.

Specific recommendations

- The identification and optimisation of storage capabilities, once legal instruments and data regulations of the country are revised, should consider not only server-based services but also cloud storage possibilities with an open attitude.
- Recognising the actual technological capabilities of the NSO should not be taken as a humbling or diminishing exercise. Instead, it is preferable to start from a realistic point of departure to try to build capabilities according to the possibilities.
- Being pragmatic to adopt and adapt solutions coming from other developing countries can mostly help to understand how the process of geocoding and the inclusion in the NSO's own processes could be done.

5.7 STEP 7 - DEFINE GOALS FOR DISSEMINATION AND PROCURE THE NECESSARY TOOLS FOR IT

This step comprises the implementation of Principle 5 of the GSGF: "the need for data custodians to make geospatially enabled statistics accessible and usable according to agreed standards and good practices" (UN DESA, 2019[1]). In simple terms, this step requires that the objectives of the dissemination of geospatially enabled statistics are in line with the requirements of the end users.

At this point, through the implementation of all preceding steps, geospatial data and statistics should be at the stage of virtual integration. In undertaking Step 7, it is important for NSOs and NGIAs to be aware of any sensitive confidentiality, legal or operational issues that may arise in analysing information and releasing it to the general public. Implementation thus includes assuring that data can be accessed using safe mechanisms that protect privacy and confidentiality while also enabling the analysis of data to support data-driven, evidence-based decision making. Also, data quality must be assured in its different dimensions such as reliability, timeliness and relevance. Finally, access should be provided to analysis, dissemination and visualisation capabilities.

A key benefit of the GSGF(UN DESA, 2019[1]) is that it allows the NSO and its ecosystem more flexible production systems to respond to new and emerging user demands. It should be possible to achieve this without having to redesign data collection methods or run costly changes in data production systems. Another advantage is the resulting spatial granularity of statistics that enables the enrichment of thematic contents and more flexible geographies for target beneficiaries. At Step 7, the user is at the centre of the entire transformation process.

Step 7 can be broken down in the following three tasks:

Task 1. Involve users to define needs for data

Run a user needs assessment exercise with targeted users and elicit their needs and proposals through rapid appraisal methods, focus groups and/or quick surveys. A clear understanding of all users' needs will greatly help develop a solid strategy for dissemination, including the needs of data communities from the general public, all levels of government, the private sector, international organisations, philanthropic bodies, NGOs and academia, among others. The user needs assessment should seek to identify not only what kind of data the users need but also how they want to access the data.

Task 2. Assess the solutions needed to meet the user needs

Once the needs of the users are identified, solutions and capabilities to meet these needs should be assessed. For many NSOs, this task will entail carefully balancing the demands of the users against the NSO's available resources and capabilities. One challenge that most NSOs will face is how to meet the needs of users with varying levels of maturity. Non-expert users may prefer more readily packaged data and simple atlases for data display, whereas expert users may prefer access to "raw data" and geospatial services for use in their own fit–for-purpose applications.

Task 3. Define dissemination goals

On the basis of task 1 and task 2, the dissemination goals should be defined. The goals should preferably be defined through a consensus-seeking process, including explaining the priorities, mapping the actors that will collaborate, setting a schedule for release of the resulting integration of data and, last but not least, revealing funding sources.

Finally, a working dissemination strategy should always aim to call on the power of storytelling behind the available data. Storytelling, in this context, is about communicating insights effectively and giving data a "voice". Some geospatial dissemination solutions combine data science techniques in order to generate compelling visualisations, although the narrative sometimes fails to use simple language to “tell a story” with the data. Once data emerge from statistical production processes, NSOs need to find this voice if they are to be accepted by various users and citizen groups and enrich an evidence-based public dialogue. The value added NSOs can provide to the data ecosystem is a compelling final product of enhanced sets of data.

Since dissemination is not limited to the technical operation of making data available and allowing the extraction of data, there are certain analyses that can be used to inspire narratives and good stories. For LMICs with an as-yet-to-be empowered civil society, an academic system that has little experience in publishing or philanthropic organisations with limited reach, there may be opportunities for NSOs to promote the development of a storytelling role among existing data user communities. This does not mean the NSO should seek to exert political influence; rather, it could consider how to reimagine the data in such a way that the context and possible messages can be widely understood, adopted and shared.

Good practice use case: The Paraguay Gender Atlas

The Gender Atlas is a platform developed by the Paraguay General Directorate of Statistics, Surveys and Censuses Paraguay (DGEEC) in collaboration with UN Women as part of the activities of the gender statistics working group of the Statistical Conference of the Americas of the UN Economic Commission for Latin America and the Caribbean (Paraguay General Directorate of Statistics, Surveys and Censuses, 2018[27]).

The motivation to map gender statistics stems from the fact that they are able to capture regional differences and inequalities in the situations of women and men with regard to diverse issues such as education, health and income. This project helps build awareness of these differences among data users, facilitating changes in communication patterns, promoting citizen participation, structuring constructive public debates and providing evidence for decision making.

Based on the source maps of the OpenStreetMap community, the cartographies pictured in the Gender Atlas display characteristics of women and men throughout Paraguay. These are compared and contrasted across several topics to identify gender gaps and inequalities and focus on the uniqueness of groups such as indigenous communities. Issues like these are central to the monitoring of the SDGs, especially given the ambition to leave no one behind.

The tool also features relevant indicators that allow the identification of social, economic and demographic gaps and the mapping of existing differences in Paraguay through online navigation. At this stage, it disseminates data on the general population that can be searched by gender, area of residence or department. An example is provided in Figure 15, which shows how the platform tabulated and mapped school absenteeism rates by gender of children 6-14 year-olds in indigenous communities.

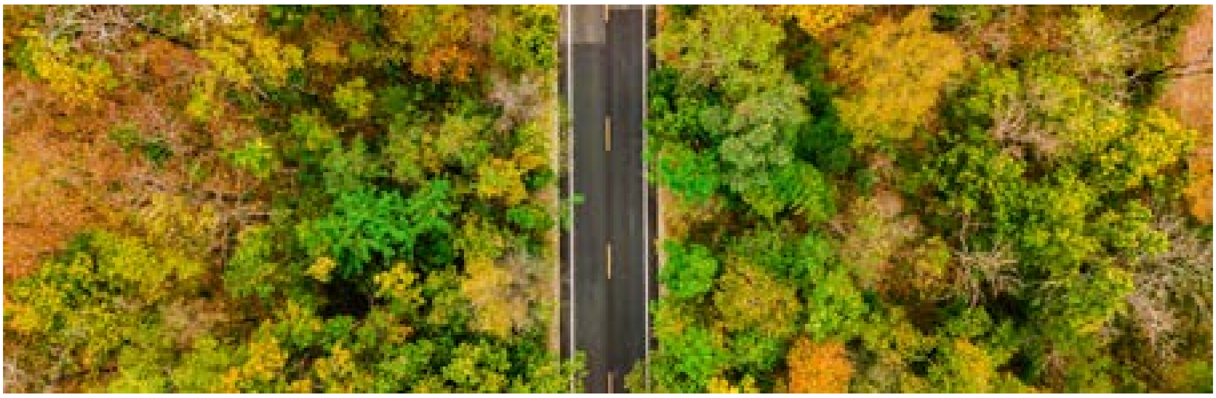
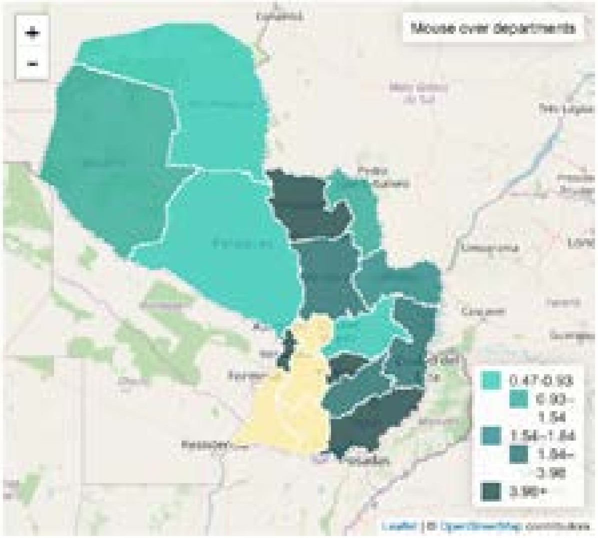


Figure 15. Example of tabulated statistics projected on maps using common administrative units

Table | Metadata

2002	Men	Women	Gap
Total			
Total Country	29.76	30.36	0.60
Area			
Urban	26.18	23.58	2.60
Rural	30.08	30.95	0.87
Department			
Assumption	40.00	38.30	1.70
Conception	39.11	34.98	4.13
Saint Peter	18.67	22.47	3.79
Guaira	44.30	48.39	4.08
Caaguazú	29.89	29.36	0.53
Caazapá	47.67	49.52	1.85

Metadata Download



Note: Translation of the Spanish subtitle of this figure is “Absenteeism rates for a formal educational institution from the indigenous population for children of 6 to 14 years old”.

Source: Paraguay General Directorate of Statistics, Surveys and Censuses (2018[27]), The Gender Atlas platform (webpage), https://atlasgenero.dgeec.gov.py/index.php.

The tabulated data are also projected on maps within the common geography of administrative units of departments using the OpenStreetMap base maps in tandem with the Leaflet application. This is an open source JavaScript library that can generate mobile-friendly interactive maps and focuses on rapid rollout of interactive cartographic data (Leaflet, 2021[28]) and also allows the use of pop-ups, choropleth mapping and dynamic layer control.

An aim of the Gender Atlas platform was to simplify access to information for Paraguay’s citizens, and it has succeeded. It also provides full details of the methodology used for calculating the indicator in a metadata tag. Although atlases are usually perceived as publications that are somehow “static”, with little margin to undertake the crossing of variables, the open data information offered in the Paraguay example through Excel files offers potential flexibility by combining them using pivot tables. Indeed, this project implicitly sets out the dissemination ambitions of the DGEEC for future endeavours.

The commendable effort of DGEEC aligns with the NSDS frameset, which suggests there is potential for statistics dissemination in partnership with third parties when the government’s budget is insufficient. One of its strengths is that it makes available reusable data through widely used software. One final caution about this remarkable initiative, however, is that geocoding is undertaken for only one level of geography, the departments. Although this has been a valuable disaggregation effort for an LMIC, the effort should be framed within Principles 1 and 2 in order to advance the consolidation of geocoding practices at a larger scale. It should also be linked to Principle 3 by making available a set of common geographies of different administrative levels.

Specific recommendations

For the satisfactory fulfilment of Step 7, implementers must carefully consider issues of privacy, confidentiality, legality and good practice. With a comprehensive supervision of the dissemination of data, data custodians will be able to release information with confidence so that user communities can easily discover and access geospatially enabled statistics to reuse, inform, visualise, analyse and use in decision making to aid all kinds of beneficiaries.

The following recommendations should guide NSOs towards a sound and safe dissemination:

- Employ safe mechanisms to protect privacy and confidentiality to enable the analysis of data to support informed, data-driven decision making.
- Employ standardised metadata, inspired by models such as ISO/EC 19115:2003 (ISO, 2014[29]), which specify when the information was last updated and published and when it is expected to be updated for the next release.
- Use an internal certification of the quality of the data, without prejudice to other external certifications such as ISO/EC 19113:2002 (ISO, 2013[30]), especially in relation to their reliability, timeliness and relevance.
- Seek to offer compelling visualisations adapted to a broad audience that enable influential data analysis.
- Consider policies and standards beyond metadata, including good practices, and state-of-the-art technologies to make geospatial data readily available and ensure broad and flexible accessibility to users.
- Provide access to layers of data at the smallest geospatial unit available, as permitted by any disclosure regulations mandated to the NSO.
- Consider usability for both technical and non-technical users to support policy and decision making and ensure that outputs are easily understood by all target data community users.

5.8 STEP 8 - USE GOOD PRACTICES AND STANDARDS TO OBTAIN INTEROPERABILITY

As explained, the ordinal numbers of the principles are only loosely associated with the order in which they should be implemented, which will depend on the particular situation of each country. At some NSOs, it could be more important to make available common geographies than to boost geospatial infrastructures or mainstream geocoding. Others may have a well-functioning dissemination strategy but may be lacking geocoding capabilities or an effective data management environment.

However, the reason this step-by-step guide leaves implementation of Principle 4 (UN DESA, 2019[1]) for the final step is because many NSOs and NGIAs in LMICs may need to focus on consolidating the previous steps before approaching the rather complex, multi-layered issues related to interoperability. However, in the logics of both the GSGF and in this guide, interoperability does not represent in any strict sense an independent step or task. Rather, it is a perspective that should be mainstreamed in all other steps. In this context, it is a plausible argument that interoperability might demand a greater effort in the long run based on the other steps elaborated in this guide and for that reason, it is listed last.

Moreover, the path towards the establishment of a full-fledged infrastructure for geospatially enabled statistics using a step-by-step logic first encounters the need for organisational transformation in Step 3, after defining the "people resources" that are addressed in Steps 1 and 2. From this point of view, Step 8 represents the final of all possible transformations, comprising the integration of several layers that encompass a comprehensive management of geospatial and statistical data: the organisational layer, the metadata generation layer, and also semantical and technical interoperability.

The underlying concept of transformations can be associated with the general idea of digital transformation used in the field of modern management (Salesforce, 2021[31]). Such transformation is conceived as the adoption of digital technology to transform services and businesses through replacing non-digital or manual processes with digital processes or through replacing older digital technology with newer digital technologies.

The European Interoperability Framework offers comprehensive guidance on the different dimensions of interoperability, including organisational, technical, semantic and legal interoperability. This framework also gives specific guidance on how to set up interoperable digital public services (European Commission, 2017[32]). Task 1 and task 3 of Step 5 cover the two first elements of this framework. Semantic interoperability ensures that the precise format and meaning of exchanged data and information are preserved and understood, including syntaxes referring to the terminology used to describe concepts, and also that the format of the information is precise. Finally, considering legal interoperability enables organisations to operate under different national legal frameworks, policies and strategies and to work together and avoid any obstacles to co-operation.

As full implementation of interoperability means a multi-layered and more evolved stage of integration of geospatial data and statistics, it requires the following tasks:

Task 1. Reassess existing models and processes and modify if needed

Already established co-operation models and processes, as per the guidance offered for Steps 1 and 4, need to be continuously assessed and modified in order to facilitate the integration and use of data. Increase of interoperability may require more calibration of arrangements regarding data sharing when enabling geographic disaggregation so that these do not hamper the integration and use of data.

Task 2. Consolidate the organisational interoperability

The organisational level for interoperability may require documenting the involved data business processes so they are comparable and mainstreamed with other data producers, detailing the different steps of the statistical production based on a quality assurance scheme like ISO/EC 9001 (ISO, n.d.[33]) and also statically generally accepted principles. For this task, the answer to the question of which existing statistical process should be linked or updated regarding geospatial data is crucial.

Task 3. Consolidate the technical interoperability

Once the organisational aspects are covered, it is time to focus on the technical level. This task involves considering the use of established standards for data collection, reference systems and metadata to develop common solutions among actors of the interoperable geospatial infrastructure so that data and tools can be reused, avoiding duplication of efforts.

With greater progress and results regarding data integration in the transformed NSO, it will become more evident that increased interoperability between statistical and geospatial data and metadata standards is indeed the final step in the journey. Once the organisation removes the structural, semantic and syntactic barriers between data and metadata from different origins and providers, results will show that interoperability has been reached.

The excellence level of this organisational and technical characterisation in the NSO will be evidenced when the discovery, access and use of geospatially enabled statistical data improve for all parties. The enhancement of interoperability improves the adapted geo-enabled statistics for more flexible use in a range of applications and data management systems, including modelling with data and production planning.

Good practice use case: Geospatial Reference Architecture in Statistics Finland

Statistics Finland recognised a core problem for its data integration endeavour was in the processing of geospatial information by the NSO using solutions that generated duplication of outputs and overlaps, leading in some cases to non-uniform results and redundancies in the results.

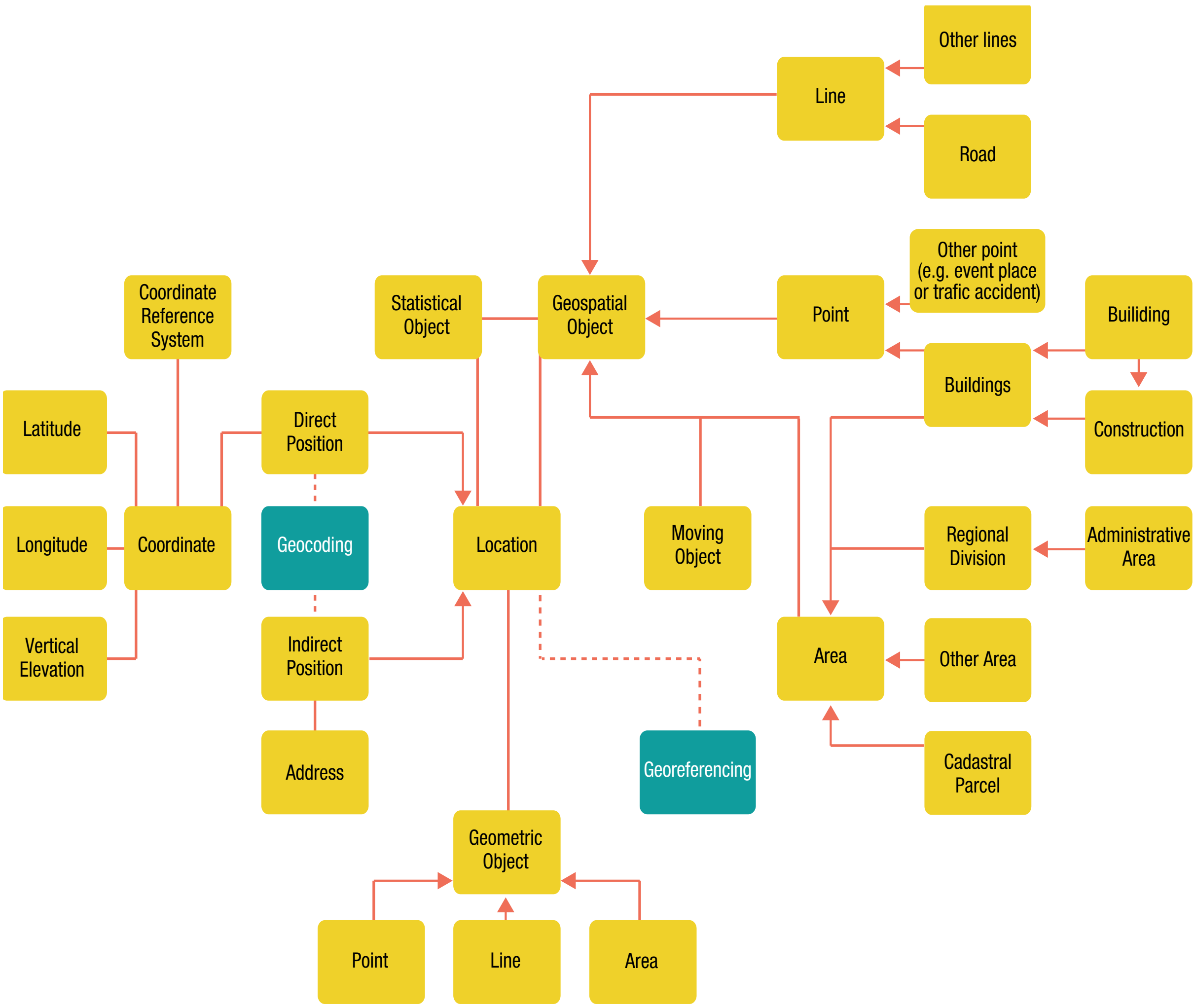
To address the issue, the implementers of the transformation designed a reference architecture to define their interoperability ambitions.

The proposed framework considered the role of the actors so that the processing of geospatial information would be used and produced in a uniform way within the statistical processes of Statistics Finland (European Forum for Geography and Statistics, 2019[25]). Its key innovation was to improve interoperability within the NSO before dealing with interoperability in relation to external stakeholders of the geospatial ecosystem.

The design strategy first addressed the organisational interoperability issues through a strategy map that identified drivers behind the desired transformation, strategic goals and detailed objectives.

In a second stage, Statistics Finland approached technical interoperability through the geospatial data conceptual model scheme reflected in Figure 16.

Figure 16. Geospatial data conceptual model scheme



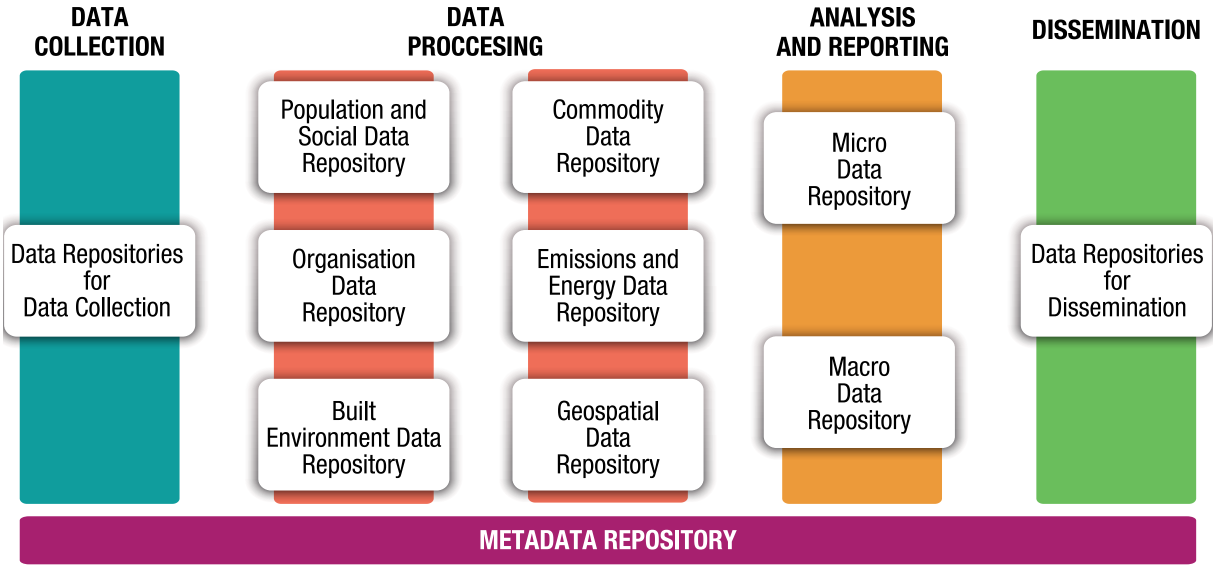
Source: Statistics Finland.

The anticipated results of this strategy are promising: The conceptual model covers not only statistical production concepts but also geospatial information concepts and places geographical features at the centre of the model, hence allowing a swift interoperability that is both organisational and technical.

Statistics are conceived as a (territorial) object to be identified with specific locations, following the logic of Step 6. Thus, the concept of locations is the element upon which the data model defines geocoding. Locations, depending on the nature of the statistical measurement, are sometimes defined in a direct position and in some other cases, are defined as indirect positions.

One of the especially valuable results from the attained interoperability is that the inputs that feed the conceptual data model come from the logical geospatial data repository streaming from the processing stage, linking geospatial data to other, non-geospatial logical data repositories and mapping the location information of statistical objects. This is represented in Figure 17.

Figure 17. Statistics Finland's logical data warehouses



Source: Statistics Finland.

Thus, the specifications of the conceptual model are carried out in such a way that they are compatible with the geospatial data logic warehouse, establishing a centralised data repository for geospatial information. The geospatial data logic warehouse is built in such a way as to be naturally integrated with the warehouses and services of other geospatial data producers.

Specific recommendations

The transformational step of consolidating interoperability calls for addressing the organisational transformation and the technical transformation successively to assure a fluid integration of available repositories of data based on data modelling and standardisation. After this, semantic and legal interoperability should be ensured. In this regard, the specific recommendations to enable a smooth integration are the following:

- It is crucial to consider the use of established standards and good practices for the entire statistical production process regarding data collection, reference systems or metadata. Developing and spreading common norms and solutions across all organisations will allow data and tools to be reused in an easier way, avoiding duplication and overlapping efforts.

- NSOs and implementers should be aware that most databases collected by different organisations may need to undergo an iterative cleansing process for uniformity before they are exchanged and enabled with interoperability.
- Raising awareness about the needed generalisation of metadata and harmonisation of semantics for all actors involved in the geo infrastructure, especially when overlapping and duplicated databases are identified, is also important.
- Legal interoperability should be achieved by mapping all aspects of laws, regulations and mandates regarding the collection, processing, organisation and dissemination of data, either geospatial or coming from the statistical realm.
- In some countries, it will be necessary to make sure that the physical addressing scheme from cadastres is harmonised with the classical national geospatial frameworks found in NSOs, especially in those NSSs created in the Hispanic tradition.

6

6. DEFINITIONS AND ADDITIONAL RELEVANT AND RESOURCES

While this guide is designed as a standalone document that explains concepts and frameworks, this section provides complementary information to facilitate comprehension of this material, starting with key definitions and concepts (highlighted in boldface text by the authors) followed by a guide to further guidance material and relevant frameworks, guidance on census operations and registers, and handbooks. These lists are not exhaustive.

6.1 DEFINITIONS AND TERMINOLOGY

"Data and statistics are not the same thing. While the terms are frequently (and incorrectly) used interchangeably or synonymously, they are in fact two different things. Data are basic elements or single pieces of information. Statistics are numerical data that have been organised through mathematical operations in line with conceptual frameworks." Although the two concepts share the characteristic of being observations of reality, geospatial data constitute a set of organised and related pieces of information that usually have geocoding or coordinates to represent relationships of objects in the space. Statistics, instead, result from calculations that follow a certain conception of social, economic and environmental issues represented by numerical, and generally scalar, representations. See https://www.researchgate.net/publication/333528323_You_say_you_want_a_data_revolution_A_proposal_to_use_unofficial_statistics_for_the_SDG_Global_Indicator_Framework

As "the ultimate objective of any statistical activity is to generate statistical data", statistics are characterised as the information that has "been collected for statistical purposes, or processed from non-statistical sources, to contribute to the production of official statistics". See http://unstats.un.org/unsd/dnss/QualityNQAF/nqaf.aspx

"Geospatial data is data about objects, events, or phenomena that have a location on the surface of the earth. The location may be static in the short-term (e.g., the location of a road, an earthquake event, children living in poverty), or dynamic (e.g., a moving vehicle or pedestrian, the spread of an infectious disease). Geospatial data combines location information (usually coordinates on the earth), attribute information (the characteristics of the object, event, or phenomena concerned), and often also temporal information (the time or life span at which the location and attributes exist)." See https://www.sciencedirect.com/topics/computer-science/geospatial-data

"The integration of geospatial and statistical data or data integration is one of the most promising paths to provide more timely, reliable, relevant and detailed information. Data integration in this context should be understood as the combination of geospatial and statistical information that can result in new insights that we could not otherwise gain". See https://ec.europa.eu/eurostat/documents/4031688/10158240/KS-03-19-423-EN-N.pdf/c8f75ee1-2181-288c-1efa-1622c5abb980

Stakeholder analysis is used to identify key actors and works by differentiating and studying "stakeholders on the basis of their attributes and of criteria appropriate to a specific situation. Attributes may include the interest of each stakeholder, the influence and importance of the stakeholders, and the networks and coalitions to which they belong. A list of stakeholders, eventually categorised according to certain criteria, is input into a series of stakeholder tables and matrices that are used to summarise and visualise information about stakeholders' attributes". See https://webapps.itc.utwente.nl/librarywww/papers_2003/tech_rep/groenendijk.pdf

"Power mapping is simply a way to identify who has power in the community, and to figure out what will move those individuals or institutions to do whatever it is you want them to do." See https://neaedjustice.org/power-mapping-101.

A needs assessment survey is “a way of asking group or community members what they see as the most important needs of that group or community. The results of the survey then guide future action. Generally, the needs that are rated most important are the ones that get addressed”. See https://ctb.ku.edu/en/table-of-contents/assessment/assessing-community-needs-and-resources/conducting-needs-assessment-surveys/main

“Geocoding is the process of transforming a description of location or location information (such as an address, name of a place, or coordinates) to a location on the earth’s surface. In other words geocoding is a way to ensure data know where they are. For the purposes of the Global Statistical Geospatial Framework, geocoding is generally defined as the process of geospatially enabling statistical unit records so that they can be used in geospatial analysis”. See https://www.efgs.info/information-base/introduction/terminology/

“Georeferencing is a set of broad processes that includes geocoding. Georeferencing, or geospatial referencing, is the process of referencing data against a known geospatial coordinate system, by matching to known points of reference in the coordinate system (e.g. image rectification to survey points or addresses linked to parcel centroids), so that the data can be viewed, processed, queried and analysed with other geographic data.” See https://www.efgs.info/information-base/introduction/terminology/

Geographies are spatial representations of the administrative, statistical or functional division of a country. Each country has its own unique setting of administrative, statistical and functional geographies, ranging in scale from country parts, provinces, regions and wards to local units such as Enumeration Areas, mesh-blocks, parishes or neighbourhoods. In some countries, grid geographies also comprise part of the framework of geographies”. See Section 5.5, Step 5

“While some common geographies refer to a single geographically stable target, others refer to clusters of interdependent targets bound together by virtue of sharing common sources.” See https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6296761/

“Interoperability [is] the ability to integrate datasets and related services of different types and from different sources”. See http://ggim.un.org/documents/Standards%20Guide%20for%20UNGGIM%20-%20Final.pdf

”Data storytelling is the practice of building a narrative around a set of data and its accompanying visualizations to help convey the meaning of that data in a powerful and compelling fashion.” See https://tdwi.org/portals/what-is-data-storytelling-definition.aspx

Other material on terminology material can be found at https://www.efgs.info/information-base/introduction/terminology/

6.2 RESOURCES AND FURTHER READING

Frameworks and implementation guidance documents

The Global Statistical Geospatial Framework is available at http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf

Information on the Integrated Geospatial Information Framework is available at https://ggim.un.org/IGIF/

A description of the GEOSTAT 2 project, a point-based foundation for statistics, is available at https://www.efgs.info/geostat/geostat2/

For information on the GEOSTAT 3 project, a statistical geospatial framework for sustainable development and implementation guide for the GSGF in Europe, see https://www.efgs.info/geostat/geostat-3/

The GEOSTAT 4 project to complete the GSGF Europe and support its implementation, which is in progress, is described at https://www.efgs.info/geostat/geostat-4/

"Geospatial information services based on official statistics", the first in the United Nations Economic Commission for Europe's Working Paper Series on Statistics, can be found at https://www.unece.org/fileadmin/DAM/stats/publications/2016/Issue1_Geospatial.pdf

Harmonising Subnational Boundaries, a report by GRID 3, is available at https://www.grid3.org/content/uploads/2020/09/GRID3_BoundariesWhitePaper_Finalized.pdf

The European Forum for Geography and Statistics publishes an introduction to spatial statistics glossary of terms at https://www.efgs.info/information-base/introduction/terminology/

Guidance on census operations and registers

United Nations Population Fund: UNFP Strategy for the 2020 Round of Population & Housing Censuses (2015-2024), https://www.unfpa.org/publications/unfpa-strategy-2020-round-population-housing-censuses-2015-2024

United Nations Population Fund: The Value of Modelled Population Estimates for Census Planning and Preparation, https://www.unfpa.org/resources/value-modelled-population-estimates-census-planning-and-preparation

United Nations Economic Commission for Europe: Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses, http://www.unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20184.pdf

United Nations Department of Economic and Social Affairs Statistics Division: Handbook on Geospatial Infrastructure in Support of Census Activities, https://unstats.un.org/unsd/publication/seriesf/Seriesf_103e.pdf

Handbooks

Insee/Eurostat/European Forum for Geography and Statistics: Handbook of Spatial Analysis, https://www.efgs.info/wp-content/uploads/informationbase/introduction/Handbook_of_Spatial_Analysis_INSEE_EUROSTAT_2018.pdf

UN Habitat: GIS Handbook for Municipalities, https://unhabitat.org/sites/default/files/download-manager-files/GIS%20Handbook%20for%20Municipalities.pdf

Statistics Portugal/United Nations Committee of Experts on Global Geospatial Information Management-Europe Working Group: The Territorial Dimension in SDG Indicators: Geospatial Data Analysis and Its Integration with Statistical Data, https://un-ggim-europe.org/wp-content/uploads/2019/05/UN_GGIM_08_05_2019-The-territorial-dimension-in-SDG-indicators-Final.pdf

United Nations Task Team on Satellite Imagery and Geospatial Data: Earth Observations for Official Statistics, https://unstats.un.org/bigdata/taskteams/satellite/UNGWG_Satellite_Task_Team_Report_WhiteCover.pdf

Food and Agriculture Organization: Handbook on Remote Sensing for Agricultural Statistics, http://www.fao.org/3/ca6394en/ca6394en.pdf

REFERENCES BIBLIOGRAPHY

Carranza Tresoldi, J. (2009), *Metodología de Taller de Presentación y Difusión de Infraestructuras de Datos Espaciales para Honduras [Workshop Methodology for the Presentation and Dissemination of Spatial Data Infrastructures for Honduras],* https://www.academia.edu/218876/Metodolog%C3%ADa_de_Taller_de_Presentaci%C3%B3n_y_Difusi%C3%B3n_de_Infraestructuras_de_Datos_Espaciales_para_Honduras [12]

Carranza, J. (2020), "From COVID-19 to the data revolution: What are geospatial data partnerships organized by NSOs suggesting?", *UN World Data Forum blog,* https://unstats.un.org/unsd/undataforum/blog/from-covid-19-to-the-data-revolution/ [3]

Central Statistics Office of Mauritius (2007), "The Central Statistics Office Cartography Unit, Mauritius (PowerPoint presentation)", *United Nations Regional Workshop on Census Cartography and Management*, https://unstats.un.org/unsd/demographic/meetings/wshops/Zambia_8Oct07/docs/Countries_presentations/Mauritius_presentation.ppt [13]

Esri (n.d.), *Introduction to Geostatistical Analyst: What is geostatistics? (webpage)*, https://pro.arcgis.com/es/pro-app/latest/help/analysis/geostatistical-analyst/what-is-geostatistics-.htm [6]

European Commission (2017), *European Interoperability Framework (EIF)*, https://ec.europa.eu/isa2/publications/european-interoperability-framework-eif_en [32]

European Forum for Geography and Statistics (2019), "Annex 2: Good Practice Cases", in *GSGF Europe - Implementation Guide for the Global Statistical Geospatial Framework in Europe, Eurostat, Luxembourg*, https://www.efgs.info/wp-content/uploads/geostat/3/GEOSTAT3_GSGF_EuropeanImplementationGuide_Annex-2_GoodPractice_v1.pdf [25]

Ghana Statistical Service (2014), *Country Experiences in Integrating Statistical and Geospatial Information (PowerPoint presentation)*, http://ggim.un.org/meetings/2014-IGSI_Beijing/documents/09_Ghana_UN%20CHINA%20PRES.pdf [7]

Government of Bangladesh (2014), *Bangladesh Country Experience in Integrating Statistical and Geospatial Information Using Geography: Building on Past Tradition, Recent Achievements and Looking Forward to Future Possibilities, Statistics and Informatics Division, MInistry of Planning, Bangladesh*, http://ggim.un.org/meetings/2014-IGSI_Beijing/documents/01_Bangladesh%20presentation%2009.06.14.pdf [2]

GRID3 (2020), *Harmonising Subnational Boundaries*, https://www.grid3.org/content/uploads/2020/09/GRID3_BoundariesWhitePaper_Finalized.pdf [23]

High-Level Panel of Eminent Persons on the Post-2015 Development Agenda (2013), *A New Global Partnership: Eradicate Poverty and Transform Economies Through Sustainable Development,* United Nations, New York, https://www.post2020hlp.org/wp-content/uploads/docs/UN-Report.pdf [22]

ISO (2014), ISO 19115:2003 - *Geographic information: Metadata*, International Organization for Standardization (ISO), Geneva, https://www.iso.org/standard/26020.html [29]

ISO (2013), ISO 19113:2002 - *Geographic information: Quality principles, International Organization for Standardization (ISO), Geneva*, https://www.iso.org/standard/26018.html [30]

ISO (n.d.), *Popular Standards: ISO 9000 Family - Quality Management (webpage),* International Organization for Standardization (ISO), Geneva, https://www.iso.org/iso-9001-quality-management.html [33]

Leaflet (2021), *Leaflet: An open-source JavaScript library (webpage)* https://leafletjs.com [28]

Luaces, M. et al. (2004), *A Generic Architecture for Geographic Information Systems*, Database Laboratory, University of Coruña, Spain, https://lbd.udc.es/Repository/Publications/Drafts/AFratoCre.pdf [19]

Massachusetts Institute of Technology (2012), *Spatial Statistics Using GIS: An Introduction (PowerPoint presentation),* https://libraries.mit.edu/files/gis/spatialstatpresentation_iap2012.pdf [20]

Moström, J. (ed.) (2019), *GSGF Europe – Implementation Guide for the Global Statistical Geospatial Framework in Europe: Proposal from the GEOSTAT 3 Project*, European Forum for Geography and Statistics/Eurostat, Stockholm/Luxembourg, https://www.efgs.info/wp-content/uploads/geostat/3/GEOSTAT3_GSGF_EuropeanImplementationGuide_v1.0.pdf [4]

OmniSci (2020), *Data Integration Definition (webpage),* https://www.omnisci.com/technical-glossary/data-integration [5]

Paraguay General Directorate of Statistics, Surveys and Censuses (2018), *The Gender Atlas platform (webpage),* https://atlasgenero.dgeec.gov.py/index.php [27]

PARIS21 (2018), *NSDS Guidelines: Guidelines for the Elaboration of a National Strategy for the Development of Statistics,* https://nsdsguidelines.paris21.org/NSDS-GUIDELINES-full-lang-en.pdf [10]

Salesforce (2021), “What is digital transformation?”, https://www.salesforce.com/products/platform/what-is-digital-transformation/ [31]

Statistics Sweden (2021), *Welcome to REGINA - The retrieval system (database),* http://regina.scb.se/ [24]

Statistics Sweden (2014), *Positioning for the future: Getting statistics and geography together (PowerPoint presentation),* http://ggim.un.org/meetings/2014-IGSI_Beijing/documents/02_Sweden_Beijing%202014%20Session%205%20Haldorson%20Sweden.pdf [8]

UN DESA (2021), *Integrated Geospatial Information Framework (IGIF) (webpage),* United Nations Department of Economic and Social Affairs, Statistics Division, New York, http://[11] https://ggim.un.org/IGIF/ [9]

UN DESA (2019), *The Global Fundamental Geospatial Data Themes*, United Nations Department of Economic and Social Affairs, Statistics Division, New York, http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/Fundamental_Data_Publication.pdf [21]

UN DESA (2019), *The Global Statistical Geospatial Framework,* United Nations Department of Economic and Social Affairs, Statistics Division, New York, http://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/The_GSGF.pdf [1]

UN DESA (2009), *Handbook on Geospatial Infrastructure in Support of Census Activities*, United Nations Department of Economic and Social Affairs, Statistics Division, New York, https://unstats.un.org/unsd/publication/seriesf/Seriesf_103e.pdf [18]

United Nations (n.d.)*, About REDATAM (webpage),* United Nations Economic Commission for Latin America and the Caribbean, Santiago, Chile, https://www.cepal.org/es/temas/redatam/acerca-redatam [26]

United Nations Economic Commission for Europe (2019), *Generic Statistical Business Process Model GSBPM, Version 5.1,* https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1 [17]

United States Census Bureau (2017), *Tool for Assessing Statistical Capacity (TASC),* https://www.census.gov/data/software/tasc.html [15]

Vale, S. (2018), *Abstract for EFGS 2018: The Review of the GSBPM and Other Statistical Standards*, European Forum for Geography and Statistics, Stockholm, https://www.efgs.info/wp-content/uploads/conferences/efgs/2018/abstracts/Vale_GSBPM_abstract_The-review-of-the-GSBPM-and-other-statistical-standards.pdf [16]

Van Halderen, G. et al. (2016), “Integrating statistical and geospatial information, cultures and professions: International developments and Australian”, *Statistical Journal of the IAOS*, Vol. 30/4, pp. 457-470, https://content.iospress.com/download/statistical-journal-of-the-iaos/sji1018?id=statistical-journal-of-the-iaos%2Fsji1018 [11]

Weisberg, A. (2017), “How to Conduct Needs Assessment Part 1: What is it and why do it?”, *Industry Expansion Solutions blog,* https://www.ies.ncsu.edu/blog/how-to-conduct-needs-assessment-part-1-what-is-it-and-why-do-it/ [14]


SCB
PARiS 21
Partnership in statistics
for development
in the 21st century