

## High demand for local area level statistics – How do National Statistical Institutes respond?

**Marie Haldorson**  
Statistics Sweden,  
Sweden  
E-mail: marie.haldorson@scb.se

Linking people to place is becoming increasingly important, not least because of the 2030 Agenda for sustainable development. Enhancing traditional statistics at the regional level with those at the local area level enriches the way National Statistical Institutes provide information to society. To provide such statistics at local level in a timely manner, the production of statistics needs to be modernised and based on various data sources, such as administrative registers and geospatial information.

The United Nations has recognised the importance of statistical-geospatial integration by creating an Expert Group under the UN Global Geospatial Information Management initiative. The European Statistical System has also paved the way for such an integration by funding the GEOSTAT projects.

For National Statistical Institutes looking to improve their capability to produce statistics at the local level, the Global Statistical Geospatial Framework is a starting point that describes the requirements that need to be fulfilled. For guidance on how to start with a more bottom-up approach, the GEOSTAT 2 project has developed key actions aligned with the Generic Statistical Business Process Model. Statistics Sweden's capability assessment might also provide some inspiration.

**Keywords:**  
statistical-geospatial integration,  
statistics at local area level,  
UN 2030 Agenda for sustainable  
development,  
Global Statistical Geospatial  
Framework (GSGF)

### Introduction

National Statistical Institutes (NSIs) in Europe and across the globe are providing policymakers regional statistics that capture what is happening in the society, economy, and environment. However, how can this be done in a way that enables new policies to be effective and detailed enough to meet targets and improve the situation wherein they are most needed? Territorial policies at national and European levels, together with the UN 2030 Agenda for sustainable development, have put

pressure on European NSIs to produce relevant and timely statistics (Haldorson et al. 2016). The 2030 Agenda aims to leave no one behind; from a geographical perspective, this means, for example, taking the urban–rural dimension into account, as well as issues of accessibility.

It is more important than ever to relate official statistics to the realities people are facing, for example, in deprived neighbourhoods. Furthermore, disaster risk management or measures to improve accessibility to public transport, schools, health clinics, or green areas require local-level statistics. The European Statistical System<sup>1</sup> has recognised this necessity by funding the GEOSTAT projects<sup>2</sup>, such as GEOSTAT 2, thereby providing guidance to NSIs on the fundamentals of a more flexible statistical infrastructure that is point-based (Moström et al. 2017).

Regional statistics are regularly produced by all the countries in Europe, typically using the Nomenclature of Territorial Units for Statistics, NUTS, classification<sup>3</sup> of administrative units in all Member States. The Eurostat database provides harmonised regional statistics, but for users looking for more geographic detail, the situation is far from harmonised. An initial step to improve this has been the introduction of statistical grids, degree of urbanisation, and so on, together with classifying Local Administrative Units as the new geographic outputs as per the revised EU regulation on territorial typologies, Tercet.<sup>4</sup>

Statistical grids are set to become one of the dissemination areas in the next European round of population and housing censuses in 2021.<sup>5</sup> Production of grid-based statistics calls for the modernisation of the way countries conduct their censuses and leads the way to a register-based statistical system.

## Linking people and businesses to place

### Benefits of statistical geospatial integration

Statistics Sweden has used geospatial data and integrated it with statistical data sources, mainly administrative registers, for a long time. It started with census maps, continued with statistics on grids in the 1980s, and has, since then, moved on to using multiple geospatial data sources to produce land accounts; it has a dynamic web tool called the Statistical Atlas for performing advanced geospatial analysis.

<sup>1</sup> Eurostat: European Statistical System <http://ec.europa.eu/eurostat/web/european-statistical-system/overview?locale=fr>

<sup>2</sup> European Forum for Geography and Statistics GEOSTAT Project Descriptions <http://www.efgs.info/geostat/>

<sup>3</sup> NUTS Regulation Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003, on the establishment of a common classification of territorial units for statistics (NUTS), OJ L 154, 21.6.2003

<sup>4</sup> Tercet, Territorial Typologies of the European Union: <http://ec.europa.eu/eurostat/web/nuts/tercet-territorial-typologies>

<sup>5</sup> European Census: <https://ec.europa.eu/eurostat/web/population-and-housing-census/overview>

However, the Statistical Atlas is no longer supported; this means that the website lacks a visualisation tool to convert the statistics to maps. Another problem is that statistics at the local level are mostly sold as commissioned services. Statistics Sweden decided only recently to start publishing such statistics as official statistics accessible to everyone, something that other NSIs have been doing for a long time.

The benefits of statistical geospatial integration are starting to be widely recognised and the UN Committee on Global Geospatial Information Management has created an Expert Group on Integration of Statistical and Geospatial Information (EG–ISGI).<sup>6</sup> Statistics Sweden is taking an active part in the work of the expert group and is currently coordinating the ESSnet project GEOSTAT 3, which focuses on developing a European implementation of the Global Statistical Geospatial Framework (GSGF).<sup>7</sup> An ESSnet project is a network of several ESS organisations aimed at providing results that will be beneficial to the whole European Statistical System, ESS.

The GSGF was adopted in 2016 and directly supports the integration of statistical and geospatial information (UN EG-ISGI 2016). It builds on a national framework developed by the Australian Bureau of Statistics; this has been refined and globalised by the EG–ISGI. Statistics Sweden has started implementing the GSGF nationally and it has proved to be a useful tool for planning and identifying improvements in this aspect of statistics, as described later in this paper.

The GSGF is based on five principles:

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

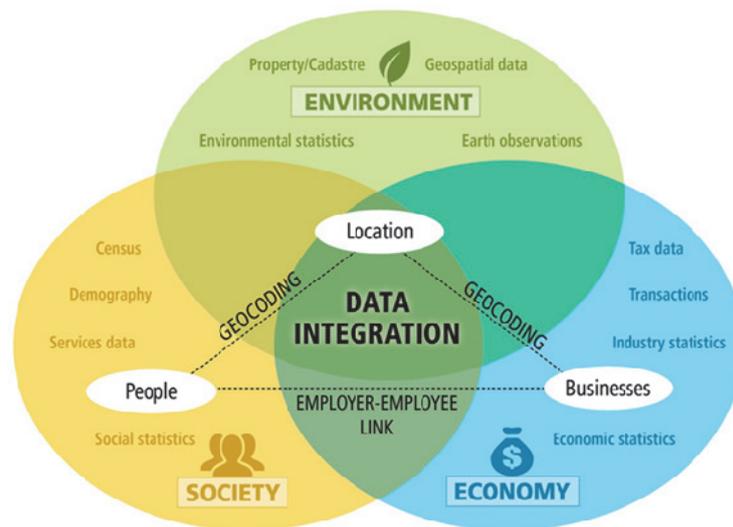
The GSGF helps adding location to statistical data that describes individuals and businesses. It also helps organising geospatial data management and the processes needed to get high-quality statistical output.

<sup>6</sup> UN-GGIM Expert Group on Integration of Statistical and Geospatial Information, <http://ggim.un.org/UNGGIM-expert-group/>

<sup>7</sup> The GEOSTAT 3 Project, <http://www.efgs.info/geostat/geostat-3/>

Figure 1

**Linking society, economy and environment through data integration**



Source: Statistics Australia and Statistics New Zealand.

Furthermore, the GSGF provides a bridge between the official statistical agency and the geospatial profession. Statistical units are geocoded using geospatial identifiers, such as cadastral parcel, address, or building. Data are then aggregated and released for the geographic boundaries in the boundary (or geography) layer. This supports integration of data across diverse sources and provides the bridge between statistical and more traditional geospatial datasets; through use of a common geography and greater flexibility in geographic outputs.

**The five principles of the Global Statistical Geospatial Framework**

The GSGF is a high-level framework, meaning that it is not intended to provide a detailed implementation plan or design; rather, it provides guidance on *what* should be available in countries, providing a lot of flexibility on *how* it is done.

To help countries implement the framework, supporting documents were developed in 2018, both at global level (by the EG-ISGI) and at European level (through the GEOSTAT 3 project). The draft descriptions from GEOSTAT 3 give a good overview of how to understand the principles from a European perspective. As this is still a work in progress, the latest available information can be found on the websites of UN-GGIM and GEOSTAT 3.

***Principle 1: Use of fundamental geospatial infrastructure and geocoding***

The goal of this principle is to obtain a high quality and standardised location attribute or reference. The aim is to agree on the protocol that regulates this attribution of location information to statistics. The framework recommends using existing geospatial information provided by specialised geospatial authorities (e.g. physical address, cadastral parcel, building, or some other location description) in order to assign accurate coordinates and/or a small geographic area or standard grid reference to each statistical unit (i.e. at the micro-data or unit record level).

Time- and date-stamping these locations will place the unit in both time and space. The process for obtaining locations and geocodes should use official and fundamental geospatial data from the National Spatial Data Infrastructures or other nationally agreed sources.

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

The GSGF recommends that the linking of a geocode to each statistical unit record in a data set (i.e. a person, household, business, building, or parcel/unit of land) occur within a data management environment. Persistent and version-controlled storage of high precision geocodes enables any geographic context to be applied when preparing the data to be released in the future (i.e. in aggregating data into a variety of larger geographic units or to adapt to changes in geographies over time). Moreover, geocodes can enable data linking processes that aim to integrate information of different types and from various sources by the so-called linked data techniques. An efficient data management environment should allow linking statistical and geospatial objects at unit record level without compromising privacy of micro data.

***Principle 3: Common geographies for dissemination of statistics***

To enable comparisons across data sets from different sources, the GSGF recommends that a common set of geographies be used for the display, reporting, and analysis of social, economic, and environmental information. While the EG–ISGI recognises the importance of traditional statistical and administrative geographies, it also recommends NSIs to consider the benefits of gridded data as grids provide consistent geography for disseminating and integrating information.

***Principle 4: Statistical and geospatial interoperability – Data, standards, and processes***

Both the statistical and geospatial data communities operate their own general data models and metadata capabilities; however, these are often not applied universally. The statistical community uses the Generic Statistical Information Model (GSIM), the Statistical Data and Metadata Exchange (SDMX), and the Data Documentation Initiative mechanisms. The geospatial community, meanwhile, makes use of the General Feature Model (GFM) and the ISO 19115 metadata standard, plus a num-

ber of application specific standards. A good overview of various standards, frameworks, and models can be found on the website of the joint UNECE/UN–GGIM: Europe Workshop on Standards in Stockholm, November 2017.<sup>8</sup>

Within the statistical community, there is a need to build geospatial processes and standards into statistical business processes in a more consistent manner. Thus, the EG–ISGI has recognised that a top-down approach is required to incorporate geospatial frameworks, standards, and processes more explicitly into the Common Statistical Production Architecture (CSPA) and its components.

UNECE has revised the Generic Statistical Business Process Model (GSBPM)<sup>9</sup> during 2018, including better descriptions of the use of geospatial data and methods in the statistical production process.

***Principle 5: Accessible and usable geospatially enabled statistics***

This principle of the GSGF emphasises the need to identify or, where required, develop policies, standards, and guidelines that support the release, access, analysis, and visualisation of geospatially enabled information. There is a wide range of legislative and operational issues about which organisations should be aware when releasing and analysing information about people and businesses in a spatial context. One important aspect of this principle is to ensure that data can be accessed using safe mechanisms that not only protect privacy and confidentiality, but also enable access to data to undertake various analyses that foster decision-making.

## **Including geospatial information in statistical production**

### **The role of geospatial information**

If the GSGF gives guidance on *what* is required to perform integration of statistical and geospatial information, results from the GEOSTAT 2 project are useful in understanding *how* to incorporate geospatial information in statistical processes.

At first, it is important to establish the role of geospatial information in statistical production. It is helpful to distinguish between the information needed as the *infrastructure* for geocoding data and the geospatial information needed to *create statistical content*.

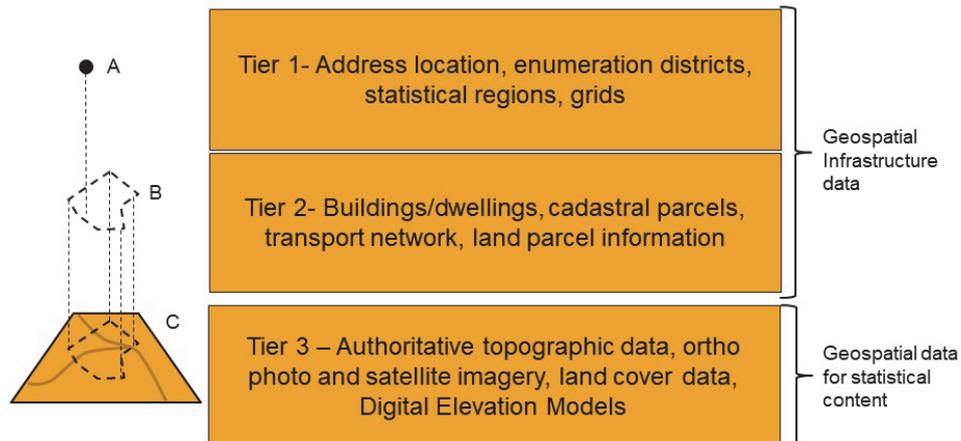
The different roles of geospatial information in statistical production are illustrated in figure 2. A workplace geocoded to an address location (A) can be linked to a cadastral parcel (B), from which land use can be assessed by combining the parcel with a land use map (C). The more consistent the system, the more opportunities for flexible linking of data.

<sup>8</sup> <https://www.unece.org/index.php?id=45404>

<sup>9</sup> <http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>

Figure 2

**Tiers of information in the generic statistical-geospatial data model for the production and dissemination of statistics**



The following are the three different tiers of geospatial information, as identified in the GEOSTAT 2 project:

*Tier 1* consists of geospatial information used exclusively for the purpose of geocoding, geographically representing, or disseminating statistical or administrative information. Examples of information in Tier 1 include address data, census enumeration districts, postal code areas, statistical grids, or other statistical or administrative geographies.

*Tier 2* consists of geospatial information that is used to: 1. geocode, geographically represent, or disseminate statistical or administrative information, and 2. create statistical content. Typical information found in Tier 2 include not only those on buildings, cadastral parcels, and transport networks, but also new data sources, such as traffic sensor information.

*Tier 3* consists of geospatial information that is used only to produce statistical content. Despite its geospatial component, this category of information cannot be used directly to geocode statistical or administrative data. Thus, information in Tier 3 can be regarded as complementary to, and independent from, information in Tiers 1 and 2.

Some examples of data found in Tier 3 include digital elevation models, land use or land cover maps, topographic data, orthophoto or satellite imagery, and other products derived from earth observation data. Typically, data in Tier 3 needs to be combined with data from Tier 1 or 2 to be transformed into statistical information.

The calculation of land area within a NUTS region can serve as an example: data on land mass and topographic maps data (Tier 3) is combined with a data set con-

taining NUTS regions (Tier 1) and the land (and water) area for each NUTS region is calculated. NUTS is the Nomenclature of territorial units for statistics, a hierarchical system for dividing the economic territory of the EU.

### A point-based geocoding infrastructure

The GEOSTAT 2 proposes a point-based approach to flexible statistical production, allowing for all kinds of statistics at the local level. In a fundamental sense, a point-based geocoding infrastructure for statistics can be understood as a production setting where a record holding the X, Y (and Z) coordinates of a location, along with a unique identifier (Id) and time stamp. This process is called the ‘geocoding’ of statistics or other data. The actual purpose of the point-based approach is to assign a single coordinate to each unit record. This point and the associated record can then be flexibly linked to a statistical or administrative data this point is located within.

Figure 3

### The conceptual difference between point-based and area-based geocoding infrastructures

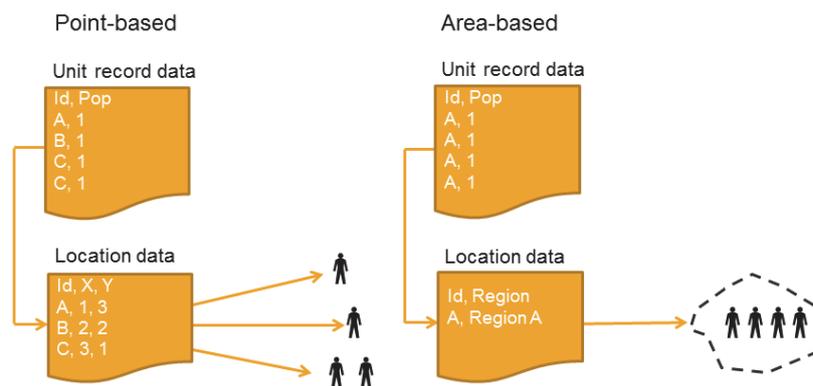


Figure 3 shows the conceptual differences between a point-based and an area-based geocoding infrastructure. In both cases, there is a record with statistical data consisting of four individuals, and a record corresponding to it that contains location data. In the point-based approach, shown on the left, each individual in the statistical data records are linked to a unique dwelling location which has not been aggregated, and is spatially represented by three different point locations. Two individuals have been assigned the same location as they are linked to the same dwelling. In the area-based approach on the right, all four individuals are linked to the same spatial object (Block A); this is because the area-based approach does not support spatial discrimination of their respective dwelling locations.

It should be stressed that the proposed shift from an area-based approach to a point-based one, as described here, only refers to the geocoding infrastructure itself and, hence, to the collection and processing of statistical information. The area-based approach is, and will continue to be, the primary method for the dissemination of statistics in order to protect privacy. The point-based approach can be implemented in the context of traditional census data collection, as well as in that of the administrative database.

### Processes involving geospatial integration

GEOSTAT 2 identified thirteen tasks linked to the first six phases of the GSBPM; these phases are needed to set up and maintain a point-based geocoding infrastructure. Phases 7 and 8 of the GSBPM (Dissemination and Evaluation) are beyond the scope of the project. The tasks are as follows:

1. *Specify needs*: Find out what the users need; promote geospatial statistics and the potential of geospatial information.
2. *Design*: Recognise geospatial data sources; assess data processing capacity; specify geospatial statistics output.
3. *Build*: Create a flexible production set-up; build the geocoded survey frame.
4. *Collect*: Obtain and manage geospatial data.
5. *Process*: Conduct geospatial data quality assessment; assess identifiers to enable correct data linkage; geocode data; prepare geospatial statistics products.
6. *Analyse*: Assess data dissemination constraints.

The GEOSTAT 2 report gives detailed guidance for each of the thirteen activities, underpinned by national case studies from the countries participating in the project.

### Statistics at the local level – Is the terminology clear enough?

Working with statistical-geospatial integration for a number of years, I have come across a number of different terms used to describe statistics at the local area level. Often, they relate to the fact that there is a strong geospatial component: location-based statistics, geographically referenced statistics, location-enabled statistics, or just geospatial statistics. They may also indicate that the NSI is using registers or censuses in the statistical production, linking statistical data to place with a bottom-up approach. In this paper, a number of these terms are used interchangeably.

There are also a number of terms relating to more model-based top-down approaches, such as small area statistics resulting from small area estimation<sup>10</sup>, spatial statistics, or geostatistics. However, at times, one might find that these terms are not

<sup>10</sup> <http://www.nss.gov.au/nss/home.NSF/pages/Small+Areas+Estimates?OpenDocument>

used very consistently; thus, the EG–ISGI needs to keep working on terminology related to geospatial statistics and other related terms.

Let us assume that users, in general, just want to find relevant statistics for different levels of regional or local detail. The subject matter areas should display these levels as clearly as possible in the statistical databases and associated metadata. However, there should also be ways for users to start searching for statistics by providing the name of a place, which could be a municipality, neighbourhood, or locality. Some additional tools are required to help users work with statistics at the local area level related to maps.

The volume of statistics on various regional and local levels will vary, owing to the need for appropriate data sources and confidentiality concerns. Even if it is tempting to start using very specific terms for statistics based on geospatial information, in the long run, our users will benefit more with simple terms, that is, keeping the statistical content in focus, but adding user-friendly tools to better understand local dimension.

While keeping it simple for users, there are many considerations to be taken into account for NSIs embarking on increased production of statistics requiring statistical-geospatial integration. The complexity increases even more when data sources, such as earth observations, are being considered for use.

At Statistics Sweden, regional statistics can be produced simply by supplying the code or the name of the administrative region in one's register or survey. Regional statistics is produced by a large number of statisticians at Statistics Sweden, who have no need for specific geospatial skills to link people or businesses to place. Statistics at the local level, on the other hand, and statistics on geospatial phenomena, such as land use, are only produced by a limited number of experts in geospatial analysis and geospatial tools.

## **The capability to integrate statistical and geospatial information**

### **Statistics Sweden's capability assessment**

The developments and challenges lead to a question about the capability that an NSI needs to develop to increase the volume of statistics at the local level. A capability enables a statistical organisation to undertake one or more activities. The general elements that are required for a capability, according to the Common Statistical Production Architecture<sup>11</sup>, are people, processes, methods, technology, standards, and frameworks.

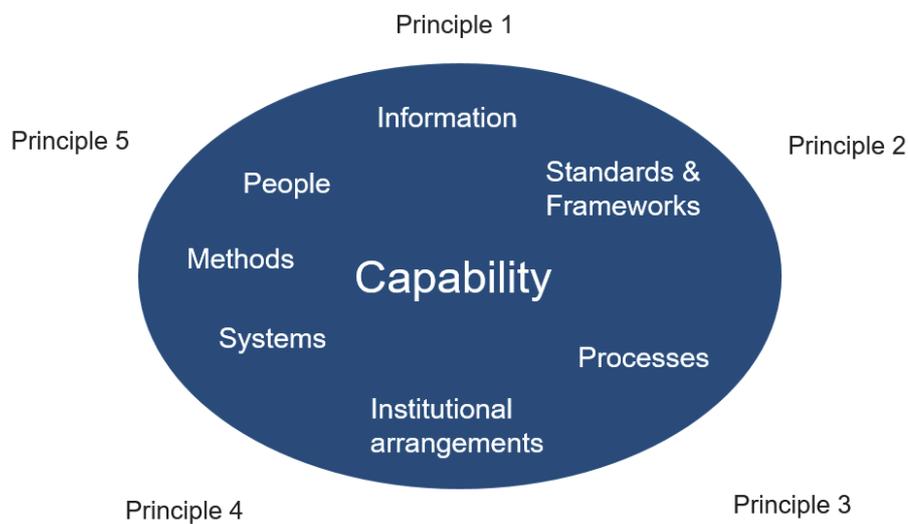
As part of the GEOSTAT 3 project, working on how to apply the GSGF in Europe, Statistics Sweden has performed a national assessment of each of the framework's principles. Looking at the framework from a capability point of view proved

<sup>11</sup> <https://statswiki.unece.org/display/CSPA/Common+Statistical+Production+Architecture>

helpful, as the framework principles need to be underpinned by certain skills, processes, methods, and so on. The assessment showed that in the case of statistical and geospatial integration the elements underpinning the capability may be expanded to include information and institutional arrangements, too (Haldorson–Moström 2018).

Figure 4

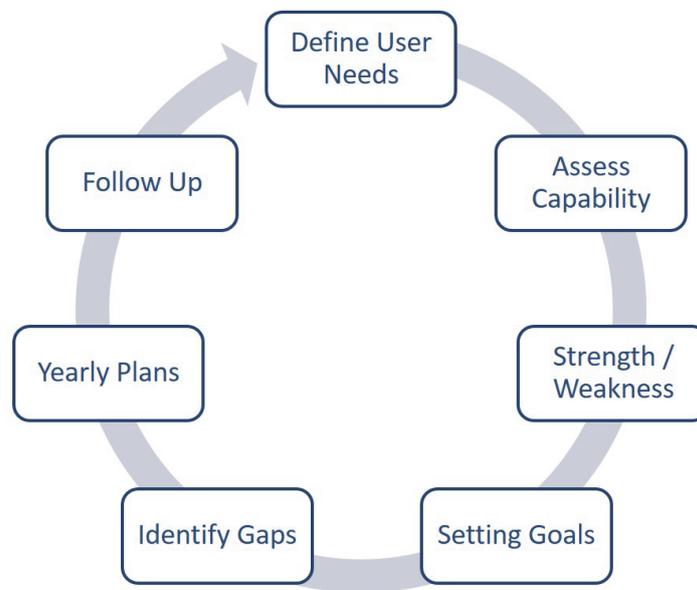
**Elements needed for an NSI to have the capability to integrate statistical and geospatial information**



The assessment exercise can easily be conducted by NSIs: those who have recently started to work on statistical-geospatial integration, as well as those looking for a way to improve it. The steps proposed in the Swedish report are simple and start with user needs.

Figure 5

**The process of identifying how to improve statistical-geospatial integration at Statistics Sweden**



Looking at each of the elements underpinning a capability for each principle in the framework revealed the strengths and weaknesses of Statistics Sweden. Setting goals and deciding which activities to prioritise makes it easier to include actions in yearly plans and to follow up. The actions can be categorised as either new activities that need to be initiated or as those that need to be improved. The assessment has also shown the activities that need to be maintained or reduced.

The following are the results of the assessment for each capability element; they cover the important aspects related to the five principles of the GSGF.

**Information**

This element covers the data needed to perform statistical and geospatial integration activities. Statistics Sweden has a register-based statistical system (Wallgren–Wallgren 2007) which is well suited for producing statistics at regional and local levels. All administrative registers used by Statistics Sweden include information on location, either directly (real estate, address, building) or indirectly (through personal identification number, work place id, etc.). They also include information on administrative geography, such as a municipality or a county.

Statistics Sweden maintains a national register on real estate, with weekly updates, from the Geospatial Agency. The register serves as a full-fledged location data framework, comprising a national address register, building register, and dwelling

register. It includes coordinates and unique ids; this means that the information in the register can be linked to the corresponding cadastral parcels on the cadastral map. To facilitate statistical production, the register stores time series as well as yearly versions.

All references to location are stored in a separate Geography Database (GDB). The GDB is the central hub for all geocoding activities at Statistics Sweden. It is a key database used to connect (link) geospatial locations with data from different administrative registers using a set of identifiers. The general principle is to leave data at its source to avoid the necessity to update geospatial variables in more registers than the GDB.

The GDB currently lacks the geospatial link between real estate and environment and/or land use. However, it can be established by linking the geospatial features of cadastral parcels (polygons) on the cadastral map with the real estate information in the real estate register. The method has been used to produce land use accounts with good results. In the future, cadastral parcels should be included in the GDB.

Authoritative boundary data for administrative geographies are produced, maintained, and provided by the Geospatial Agency as part of the National Spatial Data Infrastructure (Lantmäteriet 2017): Codes are retrieved from Statistics Sweden's database. Boundaries and codes for other non-administrative geographies, such as localities and neighbourhood areas, are produced and maintained by Statistics Sweden.

## People

This element covers the skills, knowledge, and aptitudes required in people undertaking the various statistical activities. Statistical-geospatial integration is mainly carried out by two groups of experts in Geographical Information Systems, GIS, and geospatial analysis at Statistics Sweden, both in the Regions and Environment Department. One group is responsible for land use and real estate statistics, and the other for statistics at the local level.

Statistics on administrative geographies are widely produced in many subject matter units, covering for example statistics on population by age, income, education or household status. As most registers contain references to administrative units (county or municipality), aggregating and disseminating statistics for these units typically does not require the participation of geospatial experts or special methods. For example, the metadata system of the Statistical Database automatically links the national coding system with the NUTS system.

Knowledge of statistical and geospatial models, frameworks, and standards lies with different people in the organisation; thus, greater collaboration is needed. The potential of geospatial analysis is not widespread within the organisation.

## Methods

This element covers the set of specific techniques or algorithms required to undertake the activity. As explained earlier, Statistics Sweden has a register-based system, so the methods are mainly concerned with ways to integrate various data sources by using a bottom-up approach.

Statistics Sweden uses a cadastral parcel as a geocode type, together with addresses, buildings and dwellings. All objects are hierarchically linked. This creates a very flexible system that allows for all kinds of geographic classification and aggregation. All administrative geographies are hierarchical spatial units built on the cadastral parcels (property units) as the smallest spatial element. A change in property structures (property units being merged or split) can affect the administrative borders.

Owing to the comprehensive area coding in the GDB, standard aggregation of statistics into different statistical units is possible without using spatial operations. An even aggregation of grid statistics is based on Structured Query Language (SQL) statements.

There is a need to ensure the use of better integration methods based on the statistical production model. The knowledge of geospatial methods is limited to a few geospatial experts, this being the reason why the rest of the organisation might not recognise the potential of using more geospatial information and tools in statistical production – for example in sampling or editing

Among the new activities identified after the assessment is the need to provide guidance on the use of geospatial data and geocoding practices similarly to the Australian Bureau of Statistics documentation online<sup>12</sup>. This will help to streamline the internal processes at Statistics Sweden, besides helping other authorities responsible for official statistics.

Further, there is a need to find a balance between the obligation to protect confidential information and requests from users to access information with maximum detail. Disclosure control is applied according to a well-established method, but with more flexible output systems, this might need to be changed.

## Standards and frameworks

This element covers the standards, frameworks, guides, and policies required to undertake the activities. At a high level, the enterprise architecture at Statistics Sweden needs to incorporate geospatial elements better.

Relating to more technical standards used for integration, Statistics Sweden uses INSPIRE standards and the Swedish Metadata Profile when providing search, view,

<sup>12</sup> <http://www.abs.gov.au/geography>

and download services through the Swedish Geoportal. INSPIRE is the Infrastructure for Spatial Information in Europe<sup>13</sup>.

As part of the Open Data initiative, the World Wide Web Consortium has suggested the Data Catalogue Vocabulary, DCAT, as a standard that will support the discovery use case of all types of information. Specific application profiles of DCAT to geospatial information (GeoDCAT) and statistics (StatDCAT) are being developed that are interoperable with ISO 19115 and the Statistical Data and Metadata eXchange, SDMX, respectively. Statistics Sweden needs to explore these standards further.

Implementation of international web service standards also needs to improve.

### Systems

This element covers the Information and Communications Technology, ICT, applications, hardware, and platforms required by the organisation to undertake geospatial activities.

The technical infrastructure is based on SQL server as the standard database environment. Microsoft SQL versions prior to 2008 did not support the use of geometric and spatial SQL. With the introduction of spatial SQL, more and more operations have been moved from the desktop environment to the SQL server environment. Production of most grid statistics or the statistics of small areas has been successfully streamlined in Statistics Sweden by using SQL server. However, many processes, such as highly advanced spatial analyses and map-making, still require desktop GIS. The standard desktop GIS software packages at Statistics Sweden are MapInfo, ESRI ArcGIS, and FME. Recently, Q-GIS has also been evaluated for use.

Currently, Statistics Sweden supplies geospatial statistics through traditional file transfer procedures. This also accounts for most of the intake of data from external organisations. Registers and geospatial data from other producers are downloaded through File Transfer Protocol (FTP), or similar secure services and stored as SQL tables or GIS files on servers.

The statistical geographies produced and maintained by Statistics Sweden are accessible under open data licensing and can be searched for and accessed through the national geoportal, as well as from Statistics Sweden's website. The Statistical Database, based on the PX-web tool, is the main dissemination platform at Statistics Sweden.<sup>14</sup> Data in PX-web can be accessed through an Application Programming Interface (API), and more and more applications, both within and outside Statistics Sweden, are using the API to fetch data from the Statistical Database. However, no solution currently exists for the linking of statistical information from the Statistical Database to interactive maps.

<sup>13</sup> <https://inspire.ec.europa.eu/>

<sup>14</sup> <http://www.statistikdatabasen.scb.se/pxweb/en/ssd/>

The Swedish Geodata Portal<sup>15</sup> is used to provide search and view services; links to download services are available. All services are provided as per to the Swedish Metadata Profile.

In 2016, Statistics Sweden launched a new service called REGINA<sup>16</sup>, the online service for territorial units, to enable users to easily track all changes in administrative geographies since 1952. A new feature was added in 2017 to enable users to view older geographies through a map service.

### Processes

This element covers the preferred set of steps or tasks undertaken by the business to perform a particular activity in an efficient and effective manner. These will reference methods, standards and systems where necessary.

Statistics Sweden obtains data on cadastral parcels, buildings and address locations weekly, and data are stored in historicised tables, as well as annual, situational extracts used for regular statistical production. The data transfer process has been automated through a special service.

Geo-referencing and grid data services follow a standardised procedure based on a relatively limited set of core data comprising mainly the GDB for spatial locations and statistical data retrieved from the Population Register, Real Estate Tax Assessment Register, and the Business Register, among others.

Production of geo-referenced official statistics, and/or official statistics without explicit geo-referencing, but where geospatial applications are used, includes processes based on both the core data mentioned above and a wide range of other geospatial data/map data from different agencies.

Statistical production at Statistics Sweden is carried out according to guidance material available in a specific ‘process support environment’; for each process in the production chain, many checklists, methodology support documents, and templates have been developed over the last few years. However, guidance material on geocoding practice or geocoding and management of geospatial data is currently missing. Guidance material concerning the use of geographies for dissemination of data could also be improved.

Further, there is a need for a systematic approach to the geospatial integration of all registers, resulting in improved geospatial accuracy. This will save time in the processes later on, because no additional quality improvements and checking should be required.

The results from the GEOSTAT 2 project show that the GSBPM could be extended to include geospatial data management. There is still a lot of work to be done to include geospatial processes into the statistical process chart for Statistics Sweden.

<sup>15</sup> <https://geodata.se/>

<sup>16</sup> <http://regina.scb.se/>

### Institutional arrangements

This element covers the legal and institutional infrastructure required to conduct and support the activities.

The Swedish Spatial Data Infrastructure, coordinated by the Geospatial Agency, is based on a number of important cornerstones. These are the Swedish Act and Ordinance on Spatial Information, National Geodata Strategy, the Geodata Cooperation, the standardisation work and the technical solution with a national geodata portal, and the links to the European INSPIRE Geoportal. The Geodata Cooperation began in 2011 and the revised agreement in 2015 acknowledged the fact that more and more spatial data is being provided as open data for free. For an annual fee, a joint licence is available that grants unlimited access to the geospatial data.

The legal framework for production of geospatial statistics in Sweden is closely related to the use of administrative sources and registers. Virtually all geocoded statistics are derived from administrative sources – the spatial references and the statistical micro data. Access to administrative sources for statistical production is regulated through the ordinance for official statistics (2001:100), which states that public agencies and authorities are obligated to contribute any data they have, if needed for production of official statistics. The Geospatial Agency provides the Swedish Real Property Register (Cadastre), regulated under the Real Property Register Act (2000:224), to Statistics Sweden.

In order to improve the coordination of the management and use of administrative and statistical geographies within the National Statistical System (NSS), Statistics Sweden as the coordinating body of the NSS needs to address this issue and aim for an agreement between agencies responsible for official statistics. All the geographies used for dissemination of official statistics within the NSS should be disseminated as harmonised services that are well known and used by all producers.

The level of collaboration between Statistics Sweden and the Geospatial Agency is already outstanding, but further improvements are always possible. There is a need for further collaboration with the standards community.

Statistics at the small area level might be sensitive from a privacy and confidentiality perspective; thus, the application of national privacy laws and agreed privacy standards is a top priority to ensure public trust.

The Swedish Geodata Strategy has open data as one of its priorities. Statistics Sweden provides geospatial statistics both as open data (part of official statistics) and as commissioned services. Land use statistics, for example, are published in the statistical databases and illustrated in various articles and through infographics on social media.

There is a strong trend towards more open data and there is a need to make a good business case that highlights the possibility of making detailed and interesting facts at the regional level accessible to society. Starting in 2019, more statistics at the local level will be released by Statistics Sweden as open data.

## The way forward

The 2030 Agenda for sustainable development will not be achieved without finding new ways to measure a nation's society, economy, and environment. Many of the targets need to be addressed at the local level, so that facts on various local levels showing the current situation and changes over time will help in leaving no one behind. For NSIs, this poses the great challenge of collaborating with different data providers, such as agencies with administrative registers, geospatial agencies and space agencies, as well as with academia and others to apply new methods to produce relevant indicators. It also provides opportunities for NSIs to become even more relevant to their users, learning how to describe our society, economy, and environment in diverse ways.

Linking society and economy to a place helps in explaining the local and regional situation and bringing the numbers closer to the citizens. Statistics at the local level helps NSIs to stay relevant to citizens looking for facts showing the differences within a region. However, making statistics at the local level into official statistics also comes with challenges regarding safeguarding privacy and providing easy-to-use visualisation tools.

By implementing the Global Statistical Geospatial Framework in the national context NSIs get support in how to setup a successful production of statistics at local level. Adding a capability aspect to the framework can help in assessing where the NSIs need to improve. . Engaging with users, building the necessary capability, and setting cooperation mechanisms in place are parts of the process. Australia can serve as a very good example with their history of good collaboration between the NSI and the Geospatial Agency. The GEOSTAT 2 project gives useful guidance on how to start with statistical geospatial integration, pointing to the benefits of using the GSBPM.

The way forward is to modernise production of official statistics, so that statistics at the local level becomes part of the statistical infrastructure of more and more countries. Finding some countries to benchmark against is a good way to start, using common frameworks and models is helpful with the goal to build capacity and improve the services to citizens.

Maybe the greatest challenge ahead is not how to increase the use of geospatial information, but how to get access to suitable statistical information from administrative registers to match with the geospatial information. Without regularly updated register-based information on the population qualified/suited to be linked to place, the users will not get timely and relevant statistics at the local area level.

### Acknowledgements

I would like to acknowledge my colleague Jerker Moström at Statistics Sweden, who is behind many of the innovative results shared in this paper. I would also like to thank Martin Brady at the Australian Bureau of Statistics for valuable advice on how to improve the content and make the text more relevant.

## REFERENCES

- HALDORSON, M.–ZACCHEDDU, P.–FOHGRUB, B.–PETRI, E. (2016): Geospatial information management in Europe – responding to the user needs *Statistical Journal of the LAOS* 32 32(4): 481–487. 10.3233/SJI-161010
- HALDORSON, M.–MOSTRÖM, J. (2018): *Implementing the Statistical Geospatial Framework at Statistics Sweden, National Report as part of the GEOSTAT 3 project* [https://www.efgs.info/wp-content/uploads/geostat/3/Implementing-the-GSGF-at-Statistics-Sweden-v0.96\\_Draft\\_for\\_GEOSTAT.pdf](https://www.efgs.info/wp-content/uploads/geostat/3/Implementing-the-GSGF-at-Statistics-Sweden-v0.96_Draft_for_GEOSTAT.pdf) (downloaded: August 2018)
- LANTMÄTERIET (2017): *Swedish Spatial Data Infrastructure and the National Geodata Strategy. Country Report of Sweden* Submitted to United Nations Committee of Experts on Global Geospatial Information Management Seventh Session, New York. [http://ggim.un.org/country-reports/documents/Sweden\\_Land\\_report\\_2018\\_UNGGIM\\_Final.pdf](http://ggim.un.org/country-reports/documents/Sweden_Land_report_2018_UNGGIM_Final.pdf) (downloaded: August 2017)
- MOSTRÖM, J. (ed.) et al. (2017): *A Point-based Foundation for Statistics, Final report from the GEOSTAT 2 project. EFGS and Eurostat* <http://www.efgs.info/wp-content/uploads/2017/03/GEOSTAT2ReportMain.pdf> (downloaded: August 2018)
- UN EG-ISGI (2016): *Background Document on Proposal for a Global Statistical Geospatial Framework Advanced Draft as of 28/07/2016*. <http://ggim.un.org/meetings/GGIM-committee/documents/GGIM6/Background-Paper-Proposal-for-a-global-statistical-geospatial-framework.pdf> (downloaded: August 2017)
- UNITED NATIONS GENERAL ASSEMBLY (2015). *Resolution 70/1. Transforming our world: the 2030 Agenda for Sustainable Development* [https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A\\_RES\\_70\\_1\\_E.pdf](https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf) (downloaded: August 2018)
- VAN HALDEREN, G.–MINCHIN, S.–BRADY, M.–SCOTT, G. (2016): Integrating statistical and geospatial information, cultures and professions: International developments and Australian experience *Statistical Journal of the LAOS* 32(4) 457–470. <https://doi.org/10.3233/SJI-161018>
- WALLGREN, A.–WALLGREN, B. (2007): *Register-based Statistics: Administrative Data for Statistical Purposes* John Wiley & Sons, Ltd Hoboken. <https://doi.org/110.1002/9780470061350>