Marie Haldorson, Jerker Moström

Implementing the Statistical Geospatial Framework at Statistics Sweden

Note: This is a draft describing the implementation of the GSGF at Statistics Sweden. The report will be disseminated as part of the GEOSTAT 3 project. The final content of the report may change.

Content

Summary

The purpose of this report is to assess the Global Statistical Geospatial Framework from a Swedish perspective. When a statistical organisation want to improve its capability to integrate statistical and geospatial information the five principles of the framework are very helpful. The assessment shows where there are room for improvements and where there is a need for new tasks performed by Statistics Sweden. To understand the need for increased data integration the framework is introduced as part of the solution for delivering new statistics as input for measuring the Sustainable Development Goals.

Integration of statistical and geospatial data has been done since the 1980's, evolving over time and becoming more advanced as the tools and the skills allow for more geospatial analysis. Geospatial data, methods and processes are despite this not commonly known at Statistics Sweden and not well described in guiding documents.

In recent years Statistics Sweden has increased its international engagement, both in the UN Global Geospatial Information Management, UN-GGIM, and its global expert group on integration and through the coordination of EU-funded projects on integration of statistical and geospatial information. The adoption of a Global Statistical Geospatial Framework gives the right platform to describe and improve the way Statistics Sweden work with integration.

The report includes a number of ideas for new tasks and improvements concerning Statistics Sweden's capability to integrate statistical and geospatial information. It also identifies where it is good enough to maintain the current situation and where it could be possible to reduce efforts. These activities needs to be included in the yearly work plans, described more in detail with deadlines and who does what.

User needs involving the use of new data sources such as earth observations in statistical production also needs to be explored. The 2030 Agenda for sustainable development challenge the national statistical institutes on both how to integrate statistical and geospatial information and how to use earth observations. Further work in close collaboration with experts on earth observations will be essential to identify Statistics Sweden's need for capability also in this area.

Introduction

User needs

Statistics Sweden has decided to improve its capabilities to be able to stay relevant and cost-effective, aiming to reach goals set out in its "2020 Strategy". The vision statement in the strategy says, "Statistics are fundamental to a democratic society. Statistics Sweden meets the current and future needs for reliable statistics as a basis for analysis, discussion and decision-making"[1]. Geospatial information, including earth observation data, is an important data source for statistics. However but there is a need to improve the organisation's capabilities to raise awareness of how to unleash its full potential.

To answer the question of "where" geospatial data need to be integrated with other data sources, like administrative registers, and processed, analysed and disseminated in a smart and user-friendly way. Location is often crucial in policy making, impact assessment and planning. By dissemination of statistics on maps it is possible to analyse spatial patterns, but users might also need Statistics Sweden to provide services like land accounts, geospatial analysis based on a large number of data sources.

Statistics Sweden is also together with Lantmäteriet, the Swedish National Mapping and Cadastral Agency, and other geospatial data providers working towards the goals in the national Geodata strategy aiming for "Advanced cooperation for open and usable geodata via services"[2].

Future needs for statistics relying on geospatial information will emerge responding to the 2030 Agenda for Sustainable Development. The Agenda specifically demands the need for new data acquisition and integration approaches to improve the availability, quality, timeliness and disaggregation of data. Earth observation data and geospatial information are specifically mentioned as data needed to support and track progress towards the Sustainable Development Goals[3].

Statistics Sweden will need to look beyond traditional statistical methodologies to the methods and technologies used in other disciplines, such as those used in the geospatial community, to make the most of these new data sources. This approach will also require

---

[1]Statistics Sweden's Strategy 2020:
[2]The Swedish Geodata Strategy:
https://geodata.se/globalassets/dokumentarkiv/styrning-och-uppfoljning/geodatastrategin/national_geodata_strategy_2016-2020.pdf
[3]United Nations General Assembly. Resolution 70/1. Transforming our world: the 2030 Agenda for Sustainable Development.
https://sustainabledevelopment.un.org/post2015/transformingourworld. 2015.

some investment, or collaborative use of resources, in new technologies and methods. The focus of this report is on how to integrate more traditional data sources, but the next step should be to look more deeply into capability needed for new data sources.

Statistics Sweden has been assigned by the government to propose the national statistical "base line" for the 2030 Agenda including new metrics[4], some of which need geospatial information. The government has also appointed a Swedish Delegation for the 2030 Agenda, tasked to support the actual achievement of the goals. The Delegation has delivered a first proposal for a Swedish action plan in June 2017[5]. They identified six areas where Sweden have challenges and where there are good possibility for improvement:

1. A gender equal and overall equal society
2. Sustainable cities
3. A circular economy with social benefits
4. A corporate sector with sustainable business models
5. Sustainable and healthy food
6. Strengthened knowledge production and innovation

Three of these areas are also described in the Swedish Geodata Strategy (2, 4 and 6) as areas where geospatial information is crucial.

National Data Supporting SDGs

In an article by Greg Scott and Abbas Rajabifard[6] they propose a strategic framework for integrating a global policy agenda into national geospatial capabilities. This could be regarded as a "high-level framework" to which the Statistical Geospatial Framework is one of many enabling structures (framed in red in Figure 1).

In the article, they state that the SDGs have set the requirements for information from the global, regional and national level. The ideal situation is to have national data integration "where the inputs from the National Information Systems are integrated, aggregated and disaggregated against the required SDG metrics. While the policy process, through National Statistical Offices, has been determined the means to compile and disaggregate the data have not".

Statistics Sweden has compiled two reports including a proposal for the national, statistical data reporting and a baseline for a large

[4]Report to High-level Political Forum
https://www.scb.se/contentassets/4861167095b54ea686074be19d75a5cb/agenda_2030_eng.pdf
[5]Swedish Delegation for the 2030 Agenda, report (Swedish only):
https://agenda2030delegationen.se/uppdraget/publikationer/
[6]Scott, G. & Rajabifard, A. (2017): Sustainable development and geospatial information: a strategic framework for integrating a global policy agenda into national geospatial capabilities, Geo-spatial Information Science, 20:2, 59-76.

number of the global indicators. The data to support these indicators come from various agencies responsible for official statistics as well as other authorities, all in all more than 30 different government agencies. The reports also identify a number of indicators, which are relevant for Sweden, but where no data is available yet. The Swedish Government has got a number of proposals on how to address these data gaps in the second report which was ready 15 October 2017.

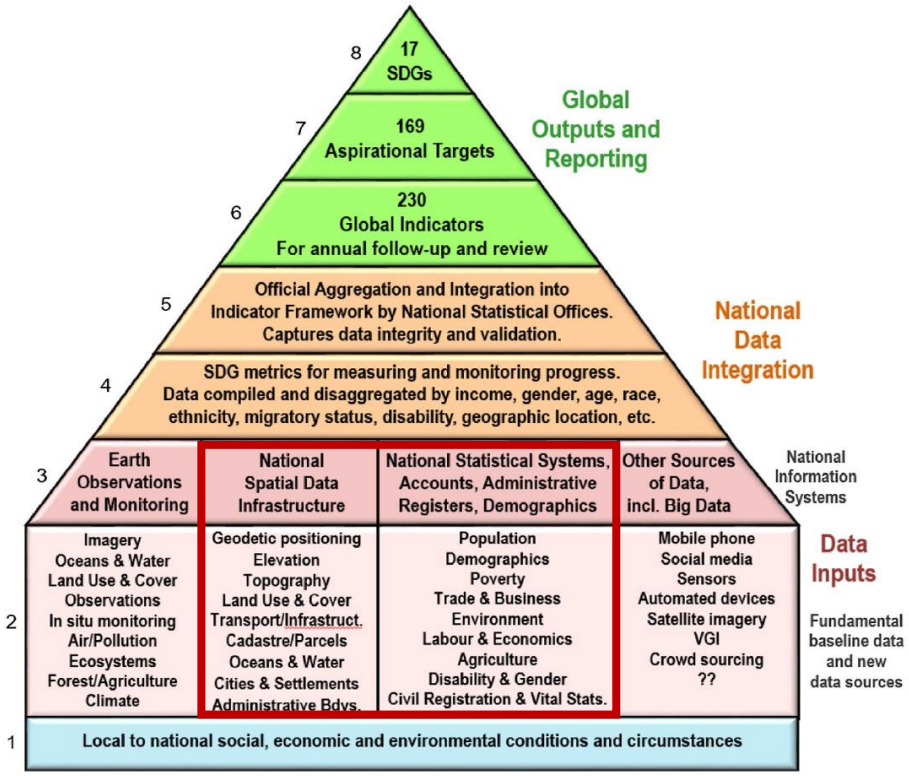

Figure 1. A general national information systems sustainable development 'data flow' framework that provides the building blocks and processes for any given country to measure and monitor the SDGs from local real-world conditions through to global harmonized reporting.

Swedish authorities are supposed to assist each other when fulfilling their official duties, so the data access should be possible to solve. What might be more challenging is to develop totally new metrics, which demand research efforts and new methods. This is a good opportunity for Statistics Sweden to work together with the research community and also to learn how to use new data sources.

One data source that is challenging for Statistics Sweden is earth observation data (EO). Advanced use of EO data for statistical purposes require remote sensing expertise and specialised software packages. For an institution like Statistics Sweden, with a small group

of geospatial experts involved in very diverse business, it is difficult to host such specialised resources in-house.

Nevertheless, Statistics Sweden has successfully been able to integrate use of EO data in production of official statistics, namely for statistics on urban green space and public green areas in the larger cities. To conduct this production, external expertise (a consultant) had to be contracted to perform the initial classification of the EO data according to a model developed jointly by Statistics Sweden and the consultant. The post-processing and integration of classified EO data and register data could be done in-house by the regular geospatial analysts.

Use of EO data in official statistics shouldn't be limited to green space statistics. A more thorough investigation will be needed to identify additional applications. Benchmark with Australia might be one way to do this, where the Australian Geoscience Data Cube[7] is an interesting solution.

The framework also covers the circumstances within individual countries (coloured in blue) which is where the real measures of progress and success need to be ultimately determined. Sweden will work on achieving the 2030 Agenda both related to targets directly from the global Agenda and related to national targets that will help the situation in Sweden. The National Delegation has set the top priorities as described above. Indicators dealing with accessibility (connected to sustainable cities) will demand integration of statistical and geospatial information.

The Global Statistical Geospatial Framework

Statistics Sweden has a longstanding history of using geospatial data and integrate it with statistical data sources, mainly administrative registers. It started with census maps, continued with statistics on grids in the 1980's and has since then moved on to using multiple geospatial data sources to produce land accounts, having a dynamic web tool called the Statistical Atlas and making advanced geospatial analysis.

With the emerging focus on this integration, manifested by the creation of the UN Committee on Global Geospatial Information Management and its Expert Group on Integration of Statistical and Geospatial Information[8], Statistics Sweden is taking active part in various international activities. Statistics Sweden is also coordinating the ESS-net projects GEOSTAT 2 and GEOSTAT 3, the latter focusing

[7]Australian Geoscience Data Cube: http://nci.org.au/services/virtual-laboratories/australian-geoscience-data-cube/

[8]UN-GGIM Expert Group on Integration of Statistical and Geospatial Information, EG-ISGI: http://ggim.un.org/UN_GGIM_Expert%20Group.html

on developing a European version of the Global Statistical Geospatial Framework, GSGF[9].

The GSGF was adopted 2016 and provides for the first time a structured way to assess national capabilities regarding integration of statistical and geospatial information[10]. Statistics Sweden is already collaborating with Lantmäteriet and other geospatial and earth observation data providers, and by implementing the GSGF we will hopefully identify more areas for development and improvement.

The GSGF consists of 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

5 Accessible and usable geospatially enabled statistics

The GSGF helps to add "location" to statistical data describing individuals and businesses. It also helps to organise the geospatial data management and the processes needed to get high quality statistical output.

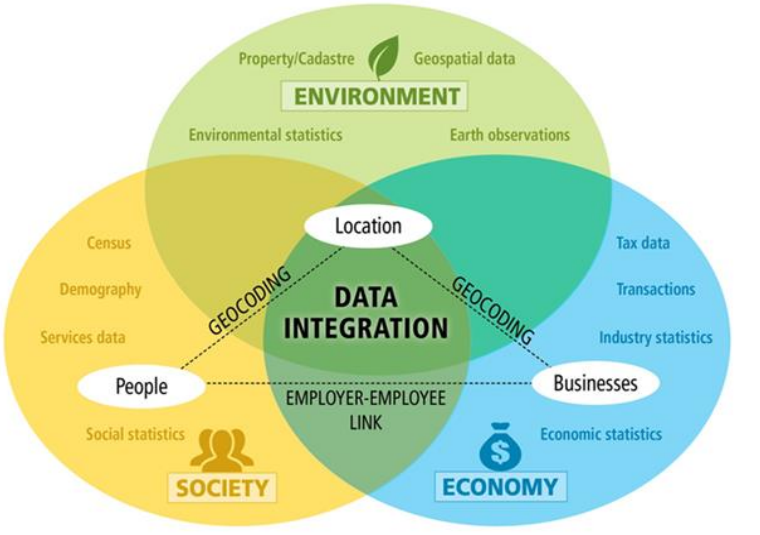

Figure 2: Linking society, economy and environment through data integration.

[9]Information on the GEOSTAT projects on the website of the European Forum for Geography and Statistics (EFGS): http://www.efgs.info/geostat/

[10]Global Statistical Geospatial Framework: http://ggim.un.org/meetings/GGIM-committee/documents/GGIM6/Background-Paper-Proposal-for-a-global-statistical-geospatial-framework.pdf

The GSGF also provide a bridge between the official statistical and geospatial profession. Statistical units are geocoded using geospatial identifiers, such as cadastral parcel, address or building. Data is then aggregated and released for the geographic boundaries in the boundary layer. This supports integration of data across diverse sources and provides the bridge between statistical datasets through use of a common geography[11].

[11]Van Halderen, G. et al. (2016): Integrating statistical and geospatial information, cultures and professions, Statistical Journal of the IAOS 32 (2016) 457–470

The GSGF applied at Statistics Sweden

Principle 1: Use of fundamental geospatial infrastructure and geocoding

The goal of this principle in Statistics Sweden is to sustain and develop our point-based geocoding infrastructure based on high quality and authoritative physical address, building and cadastral parcel locations, or other location description, in order to assign accurate and unambiguous locations to each statistical unit.

The location data used is entirely based on information jointly collected by Lantmäteriet and the municipalities and provided to Statistics Sweden by Lantmäteriet as part of the National Spatial Data Infrastructure, NSDI. Figure 3 below, shows how the information flow is organised and regulated.

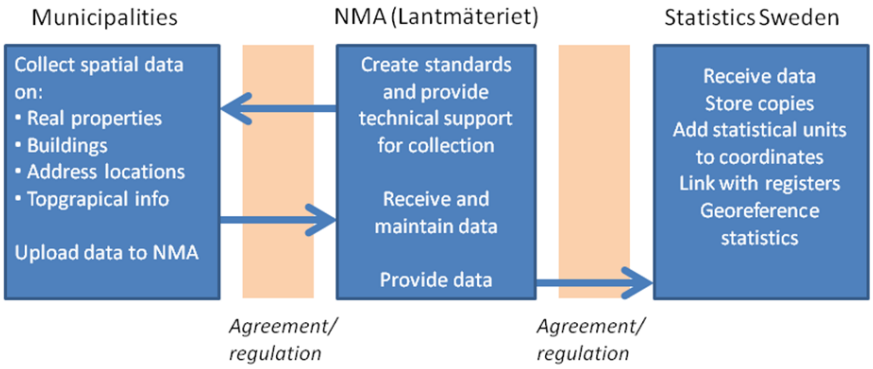

Figure 3: Data-flow for spatial data on address locations, buildings and real properties

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

Geocoded unit record data should be in a data management environment in order to ensure that all statistical data is consistently geospatially enabled. Statistics Sweden uses a Geography database, GDB, to store its point-based infrastructure, enabling integration with statistical micro data, for individuals, business and real property.

Statistics Sweden prepares an annual version of the GDB for use in regular statistical production, which is typically provided annually for a fixed reference time (December 31st). However, the register information (population register etc) is continuously being updated via automatic daily notifications from Tax Administration. Each unit record is stamped with time and changes are recorded and kept via historization.

An important process is to create situation extracts of unit record data that corresponds temporally with the situational extract of location

data. In some cases, there can be temporal discrepancies between location data and unit record data due to delays in information flows.

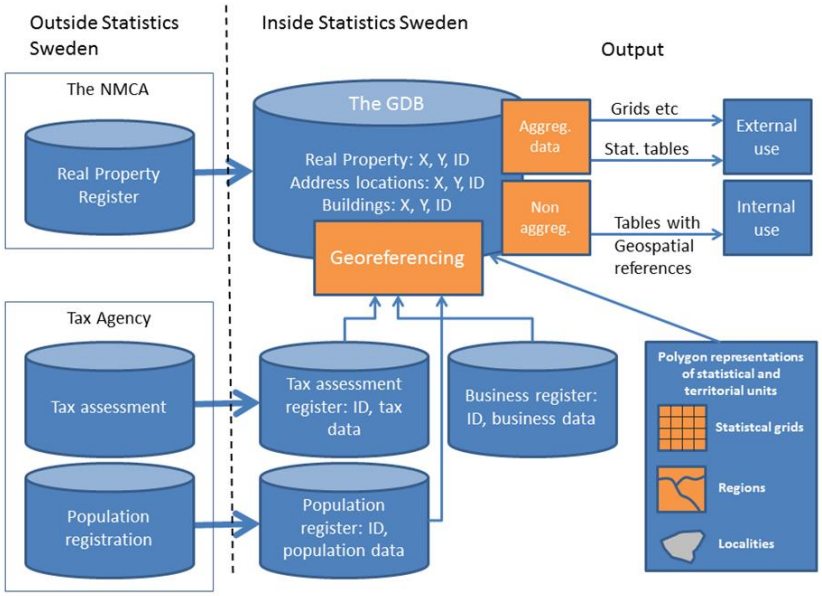

Figure 4: The Geography Database (GDB) and its relations to unit record data and other data sources

Principle 3: Common geographies for dissemination of statistics
A common set of geographies ensures that all statistical data is consistently geospatially enabled. This enables users to discover, access, integrate, analyse and visualise statistical information seamlessly for geographies of interest.

Statistics for most of the geographies can be aggregated from real property level. There are relatively few purely statistical geographies in Sweden. The main geographies used are administrative or functional geographies (grids). Recently Statistics Sweden has launched a new statistical geography for sub-LAU level, called DeSO (Demographic Statistical Areas). DeSO replaces an older system used to disseminate small area statistics from 2018.

Since many years, Statistics Sweden and Lantmäteriet have a joint responsibility to maintain the administrative geographies of Sweden. Statistics Sweden is responsible for the coding system and to keep track of all changes and Lantmäteriet is responsible for maintaining the boundaries of the geographies. In 2016 Statistics Sweden launched a new service (REGINA) to enable users to easily track all changes in administrative geographies from 1952 until today. A new feature was added in 2017 to enable the user to display older geographies in a map service.

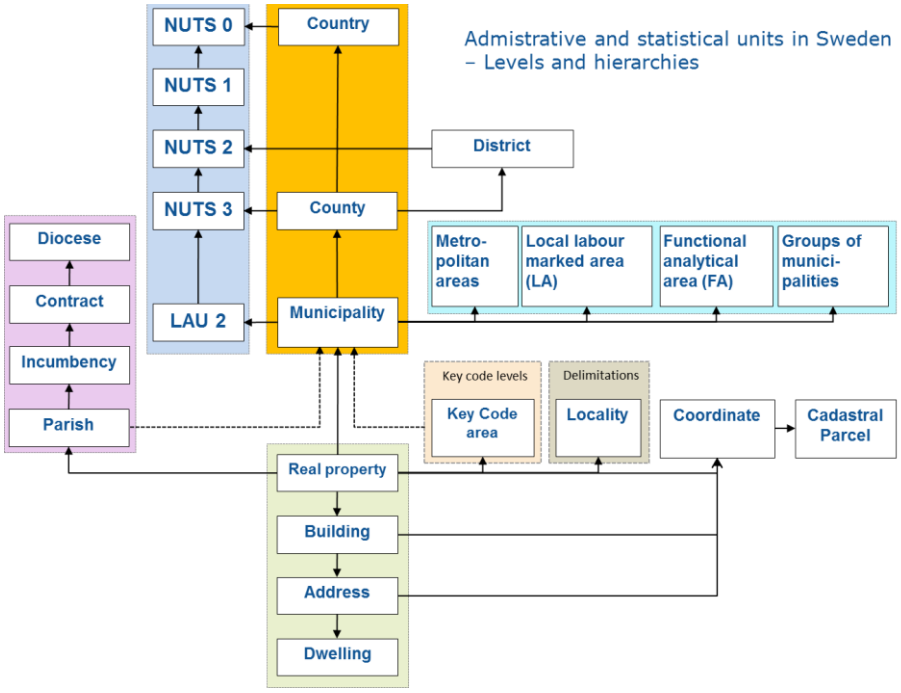

Figure 5: Regional divisions and administrative and statistical geographies in Sweden

The Swedish statistical system is decentralised, thus some 27 agencies share the responsibility for official statistics. So far no coordination efforts have been made to make sure that all agencies have a proper use of geographies for production and dissemination of statistics. Besides Statistics Sweden's REGINA service, there is currently no service to access all administrative and statistical geographies used in official statistics. Up-to-date data on administrative geographies can be accessed through INSPIRE services, but this service does not include historical boundaries.

Principle 4: Statistical and Geospatial Interoperability – Data, Standards and Processes

Principle 4 aims to improve interoperability between statistical and geospatial data and metadata standards, from cataloguing to data interchange. Overcoming structural and syntactic barriers between data and metadata from different communities and providers will enhance the efficiency of discovery, access, and use of geospatially enabled data. Within the statistical community, there is also a need to understand how geospatial frameworks and standards can be included into statistical modernisation models, developed under the

UNECE High-Level Group for the Modernisation of Official Statistics[12], in a more consistent manner.

Statistics Sweden need to handle this principle through developing the geospatial aspects in our statistical production architecture and implement geospatial components in our information model as well as in our business processes. The processes will need to be underpinned by technical standards.

From Statistics Sweden's perspective, two main challenges can be identified related to principle 4:

- To better recognise the role of geospatial data in statistical production and to develop a more structured approach to the use and management of geospatial data in the production process.
- To create better conditions for flexible linking of statistics published in the Statistical Database with statistical and administrative geographies.

Geospatial information has been part of Statistics Sweden's core business since the 1980s. However, use of geospatial information has been (and still is) a concern mostly for a limited number of geospatial experts. The general awareness is higher today then before but so far, few attempts has been made to streamline geospatial issues more systematically in the generic production model (from design to output). The lack of integration of geospatial issues sometimes leads to undesirable side effects such as quality problems and inefficiency.

Principle 5: Accessible and usable geospatially enabled statistics

Principle 5 gives advice on how to make the statistical output accessible and usable. It stresses the need to identify or, where required, develop policies, standards and guidelines that support the release, access, analysis and visualisation of geospatially enabled information.

Statistics Sweden lacks general map services and visualisation tools including maps on the web. The users need to go to the national Geodataportal to find some search, view and download services, but they only cover a very limited amount of the statistics that would be interesting to make accessible to users. The largest obstacle to make small area statistics accessible in tables, graphs and maps is the current funding model. All small area statistics are chargeable services. This is about to change, as the demand for open data grows.

[12] https://statswiki.unece.org/display/hlgbas/High-Level+Group+for+the+Modernisation+of+Official+Statistics

Capability to integrate statistical and geospatial information

While working with the European version of the SGF (ESS-SGF) in GEOSTAT 3, it has also been important to assess the current situation and to identify gaps and actions needed from a domestic perspective.

Each principle is assessed from a capability perspective, influenced by the CSPA[13]. A capability enables a statistical organisation to undertake one or more activities. General elements that are required for a capability are people, processes, methods, technology, standards and frameworks. In the case of statistical and geospatial integration the elements underpinning the capability have been expanded to include *Information* and *Institutional arrangements*.

The capability perspective presented in this report needs to be further developed and refined by the enterprise architects at Statistics Sweden. One question is if the Statistical Geospatial Framework can be supported by one capability (expressed as "the capability to integrate statistical and geospatial information in the production process"), or if it needs more than one capability.

This report will be used as input for a workshop at Statistics Sweden with geospatial experts, enterprise architects, process owners and others. One expected outcome from the workshop is to get the capability elements assessed and coloured with green (good), yellow (ok), red (action needed). Another outcome is to get a common understanding within Statistics Sweden on how geospatial information can be used in statistical processes.

Capability	**Information**: The data needed to perform statistical and geospatial integration activities.
People: The skills, knowledge and aptitudes required from people to undertake the activity.	**Methods**: The set of specific techniques or algorithms required to undertake the activity.
Systems: The ICT applications, hardware and platforms required by the organisation to undertake the activity.	**Standards and Frameworks**: The standards, frameworks, guides and policies required to undertake the activity.
Processes: The preferred set of steps or tasks undertaken by the business, with reference to methods, standards and systems where necessary, to perform the activity in an efficient and effective manner.	**Institutional arrangements**: Legal and institutional infrastructure required to conduct and support the activity.

Figure 6: Capability elements

[13]https://statswiki.unece.org/display/CSPA/User+Guide%3A+Investment+and+Capability+Catalogues

	Information This element covers the data needed to perform statistical and geospatial integration activities.
Principle 1	The NMCA provides the Swedish Real Property Register (Cadaster) to Statistics Sweden. The Cadaster does not only serve as a national land registry, but as a full-fledged location data framework, comprising a national address register, building register and dwelling register. All object types are spatially represented by coordinates. By means of unique ID:s, the information in the register can be linked to corresponding cadastral parcels in the cadastral map. A full copy of the Cadaster, including data for statistical content and coordinate references, is stored in the Real Property Register. The Real Property Register is updated weekly and changes are stored in historicised tables. Hence, the content of the register can be rolled back to any given date. Besides historicised tables, yearly versions of the content of the register is stored to facilitate statistical production A subset of data from the Real Property Register (coordinates, id:s and references to administrative geographies) is passed on to the Geography database as of 1 of January every year, where the tabular coordinate information is spatialised into point geometries.
Principle 2	Statistical data management is based on the three main object types; individuals, businesses and real estate units. Real estate information is stored in a statistical version of the Real Property Register, while all references to location is stored in the Geography Database (GDB). The GDB is the central hub for all geocoding activities at Statistics Sweden. The GDB is basically a key database used to connect (link) spatial locations with data from different administrative registers using a set of identifiers. All real property units, buildings and address locations have unique identifiers which correspond to identifiers used in unit record data. References to grid cells (national and European) and statistical geographies are added to each spatial object in the GDB. This is conducted by using simple point-in-polygon arguments in SQL server. The spatial link between real estate and environment and/or land use is currently lacking in the GDB. However, by linking the spatial features of Cadastral parcels (polygons) from the Cadastral map with the property information in the Real Property Register this link can be established. Several attempts have been undertaken to use this method to produce land use accounts with good results. In the future, cadastral parcels should be included in the GDB. All administrative registers include information on location, either directly (real estate, address, building) or indirectly (through personal id

number, work place id etc). They also include information on administrative geography, like municipality or county.

Principle 3 Authoritative boundary data for administrative geographies are produced, maintained and provided by the NMCA as part of the NSDI. Codes are retrieved from Statistics Sweden's database. Boundaries for additional, non-administrative geographies such as DeSO, localities, smaller localities are produced and maintained by Statistics Sweden.

Principle 4 The general principle is to leave data at its source in order to prevent unnecessary copies with actuality problems as an unwanted result. Currently the information model used in Statistics Sweden will need to be expanded to include more geospatial elements and links to location.

Principle 5 Statistics Sweden provide geospatial statistics both as open data (part of official statistics) and as commissioned services. The trend towards more open data is strong and there is a need for a good business case showing the possibilities of making interesting facts on a detailed regional level accessible to the society.

Land use statistics is published in the Statistical databases, illustrated in various articles and through Infographics on Social media.

People
This element covers the skills, knowledge and aptitudes required from people to undertake the activity.

Principle 1 The responsibility to obtain and maintain data from the Cadaster is divided between two units (= two register expert groups) within Statistics Sweden, one responsible for the Geography Database (GDB), and one responsible for the Real Property Register. There are also GIS-experts in these units involved in quality checks, giving input as the main users of the geospatial data. The necessary resources and knowledge to manage data is in place but currently a process has started to integrate the GDB and the Real Property Register into one single register in order to obtain a more efficient and flexible maintenance.

Principle 2 Statistics on administrative geographies are produced widely in many subject matter units, while statistics on functional areas or grids are produced mainly in the unit responsible for maintenance of the Geography database.

Principle 3 As most registers contain references to administrative unit (county, municipality), aggregating and disseminating statistics for these units typically does not require participation from geospatial experts or methods. For example, the meta data system of the Statistical database automatically links the national coding system with the NUTS system.

Principle 4	The knowledge on statistical and geospatial models, frameworks and standards sits with different people in the organisation, further collaboration is needed.
Principle 5	GIS-experts are responsible for the web services, together with the Communication department. GIS-experts are doing geospatial analysis, it is not spread more widely in the organisation.

	Methods This element covers the set of specific techniques or algorithms required to undertake the activity.
Principle 1	Data is transferred automatically to Statistics Sweden through weekly file notifications. Data is run through a number of quality checks and organised in MS SQL databases.
Principle 2	Statistics Sweden uses cadastral parcel as a geocode type, together with addresses, buildings and dwellings. All objects are hierarchically linked. This creates a very flexible system, which allows for all kinds of geographic classifications.
Principle 3	All administrative geographies are hierarchical spatial units built on the cadastral parcels (property units) as the smallest spatial element. If property structures are changed (property units merged or split) this can affect the administrative borders. Accordingly, any cross-border changes in property structures must be approved by local and regional authorities. Once such decision is taken, Statistics Sweden will be notified and changes will be recorded in the database on regional codes. Concurrently the NMCA revise the boundaries. Due to the comprehensive area coding in the GDB, standard aggregation of statistics to different statistical units is possible without using spatial operations. Even aggregation of grid statistics is based on SQL statements.
Principle 4	There is need for improvement on how to enable better integration methods based on the statistical production model. The knowledge of geospatial methods is limited to a few geospatial experts.
Principle 5	Need to find a balance between the obligation to protect confidential information and the 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 challenged.

	Standards and Frameworks This element covers the standards, frameworks, guides and policies required to undertake the activity
Principle 1	The municipalities use an official national address standard to create address information, which is then provided to the NMCA. Based on the standard, the Swedish standards institute provides a handbook for addressing.
Principle 2	There is need for better guides to support the principle.
Principle 3	There is need for better guides to support the principle.
Principle 4	The Swedish enterprise architecture including the different supporting elements of the information model, process model and activity model needs to incorporate the geospatial elements better. Relating to more technical standards used for integration Statistics Sweden use INSPIRE standards and the Swedish Metadata Profile when providing search, view and download services through the Swedish Geoportal. As part of the Open Data initiative, the World Wide Web Consortium (W3C) has suggested Data Catalogue Vocabulary (DCAT) as a standard supporting the discovery use case of all types of information. Specific application profiles of DCAT to geospatial information (GeoDCAT) and statistics (StatDCAT) are being developed which are interoperable with ISO19115 and SDMX respectively. Statistics Sweden needs to explore these standards further.
Principle 5	Implementation of international web service standards needs to improve.

	Systems This element covers the ICT applications, hardware and platforms required by the organisation to undertake the activity.
Principle 1	Data obtained from the Cadaster is stored in MS SQL databases.
Principle 2	The technical infrastructure is based on SQL server as the standard database environment. MS SQL versions prior to 2008 did not support use of geometry and spatial SQL. With the introduction of spatial SQL more and more operations have been moved from desktop environment to SQL server environment. Production of most grid statistics or small areas statistics has been successfully streamlined in Statistics Sweden, using SQL server. However, many processes still require desktop GIS, e.g. for more advanced spatial analyses and map-making. Standard desktop GIS software packages at Statistics Sweden are MapInfo, ESRI ArcGIS and FME. Recently also Q-GIS has been evaluated for use.

Currently, Statistics Sweden does not have the ability to provide spatial statistics as georeferenced data warehouses. The most common way of data exchange is still traditional file transfer procedures. This accounts also for most of the intake of data from external organisations. Registers and geospatial data from other producers are downloaded through FTP services or similar and stored as SQL tables or GIS files on servers.

The current standard for the Geography Database is to store coordinates in SWEREF 99, the official national projection.

Principle 3 The statistical geographies produced and maintained by Statistics Sweden are accessible under open data licensing and can be searched and accessed through the national geoportal and from Statistics Sweden's website. Currently there are no services set up to link data from the Statistical Database (PX-Web), main tool for dissemination of statistical data, to map services holding statistical geographies. In principle, this could be done by means of APIs.

Principle 4 The Statistical Database based on PX-web is the main dissemination platform in Statistics Sweden. Data in PX-web can be accessed through an API and more and more application, both within and outside Statistics Sweden use the API to fetch data from the Statistical Database. However, currently no solution exist as to linking of statistical information from the Statistical Database to interactive maps.

Statistics Sweden provides search, view and download services to the national Geoportal. The REGINA application includes view services.

Principle 5 The Swedish Geodata Portal is used for search and view services, there are also links to download services. All services are provided according to the Swedish Metadata Profile.

In 2016 Statistics Sweden launched a new service (REGINA) to enable users to easily track all changes in administrative geographies from 1952 until today. A new feature was added in 2017 to enable the user to display older geographies in a map service.

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

Principle 1 Statistics Sweden obtains data on cadastral parcels, buildings and address locations weekly, and data are stored both in historicised tables and as annual, situational extracts used for regular statistical production. The data transfer process is automated through a special service ("tratten").

Principle 2 Geo-referencing and grid data services follows a standardised procedure based a more limited set of core data mainly comprising the Geography

database for spatial locations and statistical data retrieved from the Population Register, Real Estate Tax Assesment Register and the Business Register etc.

Production of geo-referenced, official statistics and/or official statistics where geospatial applications are used include processes based on both core data mentioned above and a wide range of other geospatial data/map data from different agencies.

The statistical production at Statistics Sweden is done according to guidance material available in a specific "process support environment": for each process in the production chain a lot of checklists, methodology support documents and templates have been developed over the last years. Guidance material concerning geocoding practice or geocoding and management of geospatial data is however currently missing.

Principle 3 The statistical production at Statistics Sweden is done according to guidance material available in a specific "process support environment". Guidance material concerning use of geographies for dissemination of data could be improved.

Principle 4 The results from GEOSTAT 2 project show that the Generic Statistical Business Process Model, GSBPM, could be extended to include geospatial data management. There is still work to be done to include geospatial processes into the process chart of Statistics Sweden.

Principle 5 Statistics Sweden has a new maintenance structure with a team for communication and providing digital services where all web services are coordinated.

Institutional arrangements
This element covers the legal and institutional infrastructure required to conduct and support the activity.

Principle 1 The Swedish Spatial Data Infrastructure is based on a number of important corner-stones, such as the Swedish Act and Ordinance on Spatial Information (based on the INSPIRE directive), 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 started in 2011 and the agreement was revised in 2015 and adopted to the fact that more and more spatial data is provided as open data for free. A joint licence is available at a yearly cost granting unlimited access to the geospatial data that still is charged for.

The NMCA provides the Swedish Real Property Register (Cadaster) to Statistics Sweden, regulated under the Real Property Register Act (2000:224).

Principle 2 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 is derived from administrative sources, both 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), stating that public agencies and authorities are obligated to contribute with any data they have, if needed for production of official statistics.

Principle 3 In order to improve the coordination of the management and use of administrative and statistical geographies within the National Statistical System (NSS), the coordinating body of the NSS needs to address this issue and aim for an agreement between agencies responsible for official statistics. All geographies used for dissemination of official statistics within the NSS should be disseminated as harmonised services and well known and used by all producers.

Principle 4 Collaboration between Statistics Sweden and the NMCA, as well as with the standards community can be improved.

Principle 5 Statistics on small area level might be sensitive, why application of national privacy laws and agreed privacy standards is top priority to ensure public trust.

The Swedish Geodata Strategy has open data as one priority. Statistics Sweden still charge for statistics on grids and other small areas.

Actions to increase the capability to integrate statistical and geospatial information by 2020

For each principle, a number of actions have been identified through the capability assessment. They have been categorised as new activities that need to start or as activities that need to be improved. The assessment has also showed what needs to be maintained or reduced. The action list needs to be prioritised and incorporated in yearly action plans at Statistics Sweden.

The activities are numbered linking to the principles

P1: Use of fundamental geospatial infrastructure and geocoding
P2: Geocoded unit record data in a data management environment
P3: Common geographies for dissemination of statistics
P4: Statistical and Geospatial Interoperability – Data, Standards and Processes
P5: Accessible and usable geospatially enabled statistics

Need to start new activities

The following actions will lead to improved capability in statistical and geospatial integration. They need to be prioritised for 2018-2020.

P1	1	Providing guidance on geospatial data and geocoding practices, similar to the ABS documentation online[14]. This will help streamline the internal processes at Statistics Sweden as well as help other authorities responsible for official statistics.
P2	2	A systematic approach to the geospatial integration of all registers with improved geospatial accuracy as a result. This will save time later in the processes, as no additional quality improvements are necessary.
P3	3	Providing guidance to other producers of official statistics on how to use geographic boundaries.
P3	4	Providing all statistical geographies as open data.
P4	5	Geospatial data and metadata embedded into Statistics Sweden's statistical production architecture. This will allow better integration between statistical business process, data management, metadata and infrastructure.
P4	6	Concrete targets for desirable geographical quality in registers (e.g. the share of worksplaces in the business register to be successfully geocoded).
P5	7	Launch of interactive map services on scb.se.
P5	8	A number of core indicators on small area statistics geographies released as open data.
P5	9	Workshops on geospatial analysis.

Need for improvements

The following actions will lead to improved capability in statistical and geospatial integration. They need to be prioritised for 2018-2020.

P1	1	Feedback from Statistics Sweden to Lantmäteriet on quality issues in the data deliveries.
P1	2	The use of agreed terminology and data quality declarations.
P1	3	Support Lantmäteriet with arguments in its efforts to better involve municipalities in improved data collection.
P1	4	Increased use of SQL server spatial to conduct spatial operations.
P2	5	Improve address information in the Business register, so that all workplaces can be properly georeferenced.

[14] http://www.abs.gov.au/geography

P2	6	Maintenance teams and a solid structure for updates of tools and geospatial information.
P2	7	The use of geospatial information through the whole production process and increased support by the GIS-experts to the rest of the organisation.
P3	8	Availability of all administrative and statistical geographies used by any agency within the national statistical system as open data through a web service. As far as possible, also historical data should be provided to ensure that historical statistics can be properly mapped onto accurate geographies.
P4	9	Making more geospatial data accessible to the whole organisation, good metadata descriptions and guidance documents.
P4	10	Organised training courses and workshops in how to use geospatial data for your statistical analysis.
P5	11	Improved confidentiality methods to take into account the more flexible way that users download and combine various datasets (geographic differencing).
P5	12	Visibility on the web of various services that can be ordered from Statistics Sweden based on geospatial analysis.
P5	13	GIS- and spatial data competence with both experts and more widely spread knowledge on a need-to-know basis on how to use geospatial information in the production process.
P5	14	Provide a geographical (via map) entry to the Statistical Database (as a complement to the tabular entry that currently exists).

Maintaining what works well

The assessment also shows what is already working well. Statistics Sweden has a good cooperation with Lantmäteriet and other geospatial data providers. The register-based statistical system with administrative data sources is the foundation for integration.

The REGINA service is a valuable improvement in accessibility to the historical changes in administrative geographies. It has been improved with view services, but will need to be managed and improved to include more historical administrative geographies.

Expertise in seach, view and download services required by INSPIRE is valuable when working on how to increase our web services.

There are automatic confidentiality processes working well, making production of small area statistics efficient reducing disclosure risk. Statistics Sweden has a Delivery Portal for statistics on demand, which is a safe delivery system for commissioned services.

Reducing inefficiencies

Assessing capability elements it is evident that a lot could be gained by reducing inefficient processes regarding quality checks and duplication of work. Cumbersome ad-hoc correction of data and quality controls late in the production process should be avoided.

There is also inefficient data storage and duplications of databases. Statistics Sweden no longer stores historical versions of geospatial information. They are stored in a national data storage that makes the production more complex.

There is unnecessary confusion about different versions of geographies (scale, boundary changes etc) which need to be reduced.

Moving towards more open data will reduce the need for chargeable services, which will make small area statistics more accessible and more used in society.

Conclusions

Geospatial information is still considered something that mostly concern a small group of GIS-experts at Statistics Sweden. This assessment is targeting where new initiatives and improvements are necessary to get an increased uptake of geospatial data and processes in the whole organisation.

For each principle, a number of capability elements have been assessed resulting in proposals on what we need to start, improve, maintain or remove. The next step will be to include activities in the yearly work plans to reach a more mature stage by 2020, with geospatial data and processes included and utilised by the whole organisation.

The assessment exercise is warmly recommended to any NSO wanting to improve its geospatial capability. By implementing the Statistical Geospatial Framework with a focus on relevant capability elements the NSO can prioritise where to invest in skills, processes and tools. The UN-GGIM Expert group on integration has planned a Knowledge Database with support for implementation of the framework, it will be important to follow these developments closely and evaluate how to implement coming proposals.