

Benefits from data sharing - increased use of geospatial information in the statistical production process

Marie HALDORSON

Deputy Director, Regions and Environment Department, Statistics Sweden

The INSPIRE directive aims to create a European Union spatial data infrastructure. This will enable the sharing of environmental spatial information among public sector organisations and better facilitate public access to spatial information across Europe. In Sweden the INSPIRE directive has resulted in a number of activities of which the data and service sharing is of special interest, together with the ambition in the Swedish Geodata Strategy that the organisations should share not only INSPIRE data but all kinds of spatial information. A special Geodata Cooperation Agreement was put in place in 2011 and Statistics Sweden (SCB) has since then been able to improve the statistical production process by integrating new geospatial information. The most evident improvements have been made in land use statistics and in the register system. The benefits that a national statistical institute can gain from a spatial data infrastructure will become more evident in the coming years, as the possibilities to combine data sources and integrating data increase with easy accessible view- and download services.

1 The INSPIRE directive

The European INSPIRE Directive [1] entered in force in May 2007, establishing a common infrastructure for spatial information in Europe to support Community environmental policies and policies or activities which may have an impact on the environment. INSPIRE is based on the infrastructures for spatial information established and operated by the 27 Member States of the European Union. The Directive addresses 34 spatial data themes¹ needed for environmental applications, with key components specified through technical Implementing Rules. This makes INSPIRE a unique example of a legislative ‘regional’ approach.

¹ Listed in Appendix 1

The INSPIRE directive is implemented in various stages, with full implementation required by 2020. The spatial data infrastructure will enable the sharing of environmental spatial information among public sector organisations also on a national level. The National Statistical Institutes are involved in the INSPIRE work as data providers in certain themes, but will also be able to benefit from the infrastructure for statistical purposes as the data sharing gives access to a large amount of data of interest for statistics.

INSPIRE is based on a number of common principles [2]:

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.
- Geographic information needed for good governance at all levels should be readily and transparently available.
- It should be easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

2 Implementing INSPIRE - a pilot study at Statistics Sweden

The impact INSPIRE will have on increased accessibility and use of geospatial information is still not fully recognized, but one thing is clear: INSPIRE promotes cooperation between data producers both on a national and a transnational level. The Mapping Authorities take the lead as national contact points, other authorities with a focus on spatial and environment data have also been involved in the progress of INSPIRE for some time. As the implementation of thematic information according to Annex III is approaching, INSPIRE also involves authorities that perhaps have only a limited amount of spatial data - or authorities like a National Statistical Institute which have a vast amount of data, but what should be regarded as INSPIRE data?

The most obvious themes for most NSI's are Statistical Units [3], defined as 'Units for dissemination or use of statistical information' and Population Distribution/ Demography [4], defined as 'Geographical distribution of people, including population characteristics and activity levels, aggregated by grid, region, administrative unit or other analytical unit' - but also other themes may include spatial data that the NSI need to provide. Implementing rules are being prepared during the summer and early autumn of this year, where the requirements for all themes are defined.

In the process of becoming INSPIRE compliant there are a few steps that needs to be taken, after having identified relevant INSPIRE themes and what should be considered as spatial datasets. As the INSPIRE services require specific competence there is a need to decide whether to handle them in-house or with consultants: Statistics Sweden have the competence needed and will therefore produce the services in-house. You should also find a technical solution for view- and download services that work well with your data storage and with the web servers.

During the autumn 2011 Statistics Sweden carried out a pilot study with the aim to clarify how the authority could fulfil its obligations as responsible for certain INSPIRE themes. The pilot study resulted in recommendations for implementation work during 2012 and 2013 and covered topics as how to optimize view- and download services, what data storage and data portal solutions is required, how to maintain metadata and reuse existing metadata and how to enable payment for download services. One important task during the pilot study was also to benchmark and try to find out if the results from other obligations, like the work with the Census, could be re-used also within INSPIRE.

The study had a technical focus and didn't aim to solve the question what should be regarded as INSPIRE data, as the specifications for relevant INSPIRE themes were in a preliminary version at that stage. The study states however that SCB will continue to use the Swedish Statistical Database as main channel for statistics on municipality and county level. Spatial data at SCB are mainly considered to be data on a detailed regional level, like grids or census districts, and various territorial units like localities - but countries where access to register based statistics is limited might have a different scope for their spatial data. INSPIRE should if possible create added value primarily by making existing spatial statistics available that is not already reported to Eurostat and

published in their Regional databases. It is also important to make sure that if there are existing European datasets they should be used as much as possible, to avoid unnecessary duplication of work.

As the work to get INSPIRE services up and running will be intensified during 2013, with a deadline in December 2013, Eurostat has a role to provide support to the NSI's during the implementation phase. A lot of questions can be foreseen on how to interpret the data models and in which way you could make the INSPIRE services work together with other services at the NSI's. The need for bench-mark and best practise will be evident during 2013 and in the years to come, when the services also should become harmonised.

3 Data and service sharing on national level

While INSPIRE will put the European Spatial Data Infrastructure in place, the Swedish SDI concept has been gradually developed since 1970's. Today the Swedish SDI is based on a number of important corner-stones, such as the 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 [5]. The Geodata Cooperation fulfils the INSPIRE data sharing objectives and a special Geodata Cooperation Agreement was put in place in Sweden 2011.

The agreement is the foundation for a sustainable cooperation within the infrastructure for spatial information and is managed by the National Mapping and Cadastre Authority, Lantmäteriet. Parties in the cooperation are authorities with an information responsibility according to the Swedish Act and Ordinance on Spatial Information, based on the INSPIRE directive, and municipalities, government agencies and other organizations with official duties [6].

The Geodata Cooperation Agreement includes how to handle organization, steering, coordination and responsibilities as well as technical prerequisites, forms of supply and terms of use of spatial data. The parties in the Geodata Cooperation offer each other their spatial data for official use to an annual fee. Available geodata are presented and described in a Product Catalogue. Municipalities, government agencies and other

organisations which conduct official duties can also join the Geodata Cooperation, and thereby get access to all geodata in the Product Catalogue, for official use.

The contents of the Product Catalogue will change over time, with the aim to include as much spatial information as possible from all authorities in the cooperation. The INSPIRE regulation gives a minimum requirement, but in order to fulfil also the goals in the National Geodata Strategy the cooperation has a broader scope: by making as much spatial information as possible available the benefits from sharing information will increase within the public sector.

Data sharing is a cost-effective way for the public sector to use data of high quality for a wide variety of tasks. By making spatial data available as services on the web it is also easier for the private sector to benefit from this infrastructure, as it gives easier access with known conditions and licenses. Statistics Sweden signed the agreement 1 January 2011 and has since then shared spatial information with the other parties.

4 Quality improvements in land use statistics

The new Geodata Cooperation Agreement has resulted in substantial quality improvements in nearly all land use statistics produced since 2011. The data sharing was implemented at a very convenient point in time with regard to the ongoing work on delineating various statistical units at Statistics Sweden, such as localities. This work is undertaken every five years and access to a number of new sets of geodata (in particular the topographic map, the cadastral map and orthophotos) has proven to be particularly useful to obtain a better accuracy.

Previously, revising or creating new boundaries for a locality were solely based on the real estate register (coordinates represented as points on a map). Due to lack of access to data comprising topographical or structural features, the delineation could not take natural boundaries of new building blocks or roads into proper account. Now, thanks to new and extended background information, the GIS-based procedure for delineation of localities is more accurate and the boundaries of all new localities correspond better with ground truth. Some adjustments of old boundaries have also been made to improve the quality.

Following the procedure of defining the geographical features of the localities 2010, a continuous work has been done to analyse their contents and to produce thematic statistics. One assessment has been undertaken regarding buildings and urban structures [7], where the high quality information on buildings, accessed thanks to the data sharing policy, resulted in interesting statistical facts never been possible to produce before. Only a very small share of Sweden could be classified as developed land and the regional differences regarding the distribution and proportions of developed land is significant. Due to the good quality of the in-data used, land use statistics can be presented with high accuracy on very small geographical units. This is assumed to better meet the needs for land use statistics in regional applications.

Another example of results derived from quality improvements widely noted by the media, was the press release on Land and water areas 1 January 2012 [8]. The result proved that the land area of Sweden previously had been overestimated by 3 000 km² or 0,7 percent. The area of inland water (lakes and rivers) had been equally underestimated, while the territorial border remains unchanged.

The enhanced possibility to combine register information with geodata comprises a huge potential. So far this potential has only partly been explored. However it is foreseen that many of the surveys on land and water use undertaken by Statistics Sweden will benefit greatly from this within the coming years. Some applications, such as the delineation of localities, are straight forward while other are more complex and require efforts to develop new methods.

When evaluating the usefulness of spatial datasets available in the data sharing one of the most important factors is the completeness, as most statistics need to cover the area of a whole nation. The spatial datasets provided by Lantmäteriet fulfils this requirement, as well as a large amount of other datasets still to be evaluated in the process to create statistics with higher quality.

5 Quality improvements - using more frequent updates of real estate information

The use of administrative sources is a fundament for the production of statistics in many subject areas at Statistics Sweden. The system of registers is often described as

consisting of three spheres concentrated around three kinds of objects: individuals, companies and real estate. Administrative sources covering individuals and companies have had a high timeliness for a long time, while the real estate information has been updated less frequently. The main reason for the lack of updates of real estate information has been the high cost, which also affected the completeness of the information as only the most important variables were updated frequently. With the Geodata Cooperation Agreement it is possible not only to get more frequent updates, but also to get updates of all relevant real estate information from Lantmäteriet at every occasion.

6 Robust output processes

A National Spatial Data Infrastructure, based on the principles from INSPIRE, put demands on the NSI as provider of spatial statistics - but at the same time gives new opportunities to use spatial information from other authorities, as shown in the examples in this paper. Spatial information may be used in all stages of the statistical production process: to create relevant geographical frames for different surveys, to use in the analysis and evaluation phase and last but not least to present regional statistics and statistics which doesn't follow administrative borders and provide relevant facts for policy makers.

Statistics Sweden has identified three main output processes where spatial information is essential:

- Analyzing statistics with a spatial dimension and presenting results in the Statistics Atlas, based on the web tool Statistics eXplorer and using charts together with maps and stories.
- Making thematic maps using traditional GIS-software, published in reports or directly on the web.
- Producing INSPIRE services and other spatial data services which is possible to view and evaluate through metadata on the Swedish Geodata Portal

To have robust processes in place and make the different tools available on a broader scale within the organisation it has been essential to create a maintenance structure. The overall model for maintenance, used both for statistical products and processes since 2008, is based on a maintenance model where teams get appointed including both

subject matter experts, IT system owner, methodological expertise and other competences needed for the specific product or process. The maintenance team responsible for maps and diagrams is responsible so far for the Statistics Atlas and for the GIS-applications used for thematic maps (MapInfo), when the INSPIRE services are launched next year they will also be included into their responsibility.

The benefits that a national statistical institute can gain from a spatial data infrastructure will become more evident in the coming years, as the possibilities to combine data sources and integrating data increase with easy accessible view- and download services. Statistics Sweden has already been able to improve the statistical production process and get statistical output of higher quality, but there are still a lot of spatial datasets to discover in the broad variety of INSPIRE themes which could be useful in all stages of the statistical production process.

7 References

[1] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)

[2] Information from the European INSPIRE website: <http://inspire.jrc.ec.europa.eu/>

[3] D2.8.III.1 INSPIRE Data Specification on Statistical Units – Draft Guidelines

[4] D2.8.III.10 INSPIRE Data Specification on Population Distribution - Demography – Draft Guidelines

[5] Sandgren U. (2011) Swedish Spatial Data Infrastructure - An efficient network of information resources accessible via the Internet, Serbian geodetic conference 2011

[6] Information from the Swedish geoportal: <http://www.geodata.se/en/>

[7] Statistics Sweden (2012), Statistical report on Localities 2010: Buildings and Urban Structures, MI38

[8] Statistics Sweden (2012), Statistical report on Land and water areas 1 January 2012, MI65

Haldorson M. (2012) Quality benefits from a spatial data infrastructure, Q2012 conference in Athens 2012

Haldorson M. (2012) Implementing INSPIRE at Statistics Sweden and benefits from the INSPIRE data sharing, INSPIRE conference in Istanbul 2012

Appendix 1, The INSPIRE Spatial Data Themes

ANNEX I

1. Coordinate reference systems
2. Geographical grid systems
3. Geographical names
4. Administrative units
5. Addresses
6. Cadastral parcels
7. Transport networks
8. Hydrography
9. Protected sites

ANNEX II

1. Elevation
2. Land cover
3. Orthoimagery
4. Geology

ANNEX III

1. Statistical units (Units for dissemination or use of statistical information)
2. Buildings
3. Soil
4. Land use
5. Human health and safety
6. Utility and governmental services
7. Environmental monitoring facilities
8. Production and industrial facilities
9. Agricultural and aquaculture facilities
10. Population distribution — demography
11. Area management/restriction/regulation zones and reporting units
12. Natural risk zones
13. Atmospheric conditions
14. Meteorological geographical features
15. Oceanographic geographical features
16. Sea regions
17. Bio-geographical regions
18. Habitats and biotopes
19. Species distribution
20. Energy resources
21. Mineral resources