**More and better – Improvement of official statistics through the Swedish Geodata Cooperation**

Jerker MOSTRÖM

Senior Advisor, Regions and Environment Department, Statistics Sweden

Since 2011 Statistics Sweden has been member of the Swedish Geodata Cooperation, an initiative aiming to create a Swedish infrastructure for geodata and increase access to, and use of geodata within the domain of government agencies and other public bodies in Sweden. The Geodata Cooperation was launched as a response to the INSPIRE directive in Sweden. For Statistics Sweden the Geodata Cooperation has encompassed a paradigmatic shift of the production of land use statistics, and other geospatial statistics, due to the widely extended access to geodata. Through a set of practical examples, this paper aims to demonstrate how increased access to geodata can improve the quality of official statistics but also generate new statistical applications, through the extended opportunities to combine register data with geodata.

**1 The Swedish Geodata Cooperation**

While INSPIRE will put the European Spatial Data Infrastructure in place, the Swedish spatial data infrastructure (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 [1].

The Swedish Geodata Cooperation Agreement is the national foundation for a sustainable cooperation within the SDI. The cooperation was launched in 2011 in order to fulfil the obligations of the INSPIRE directive, yet it goes beyond the aspirations of INSPIRE as it reaches further than the scope of environmental information. As of today, the Geodata Cooperation offers a pool of data exceeding 400 geodata products from 19 different providers. The very basic concept of the cooperation is that parties sign *one* agreement and pay *one* annual fee but get access to geodata products from all data providers. This concept stands in sharp contrast to the situation before 2011, where data acquisition entailed complex business models and expensive agreements.

The cooperation is managed by the National Mapping and Cadastre Authority (NMCA), Lantmäteriet. Parties to the cooperation are authorities with a data management responsibility according to the Swedish Act and Ordinance on Spatial Information, based on the INSPIRE directive, and municipalities, government agencies and other organisations with official duties [2].

The Geodata Cooperation Agreement includes how to handle organisation, 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 public 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 available as much spatial information as possible the benefits from sharing information will increase within the public sector.

Until recently, universities and research institutions did not take part in the data sharing. But in May 2014 a special agreement was signed, giving the research community access to a substantial part of the data pool of the Geodata Cooperation.

**2 General benefits of the Geodata Cooperation**

In general terms, data sharing is a cost-effective way of enabling the entire public sector the use of high quality data 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.

One of the most obvious results of the cooperation is reduced administration regarding licensing and data acquisition. The simple and straight forward business model makes it easy to budget future (next years) costs for Spatial Information. The situation before the Geodata Cooperation suffered from multiple agreements between data providers and users and complex terms of use obstructing flexible use of data and creation of services.

**3 Geodata and geospatial applications at Statistics Sweden**

Statistics Sweden has a longstanding tradition of geographical applications in the production of official statistics. The first steps towards usage of georeferenced information as a regular component of the statistics portfolio were taken already in 1980s. In late 1980s, real property coordinates together with data from the population register were used to produce the first machine generated population grid of Sweden. Today, GIS and geospatial information, are integral parts of the production chain in many statistical products, especially in the field of land use statistics.

The use of geospatial information within Statistics Sweden can be broadly divided into two different categories depending on the purpose of the usage and the properties of the end-use product.

1. Production of geospatial statistics, such as gridded statistics or other small area statistics, where the geospatial statistics itself is released as the end-use product or at least forms an essential part of the result. This category also includes delimitation of localities etc. In general, geospatial statistics is not part of the official statistics in Sweden. Consequently, this category does not encompass official statistics.
2. Production of official statistics where geospatial information and/or geospatial processing is involved at some stage of the production chain but not essentially part of the disseminated result. Stages of production may concern design of surveys, sampling, data collection, processing, analysis and dissemination.

In a broad sense, most statistical products retrieved from administrative records have a geospatial component as many of them rest upon an underlying framework of georeferenced records. However, in terms of production setting, such as data sources used, tools, methods for data processing and analysis, the most “geography intense” field of statistics is land use statistics. Accordingly, the impact of the Geodata Cooperation has been most significant in the production of land use statistics.

As of today, Statistics Sweden is responsible for some twenty official products concerning use of land and water, comprising statistics on land use, land cover, land ownership, protected nature, urban green areas, coastal settlement and development, urban development etc [3]. In all of them, external geodata and geospatial processing plays a key role. For Statistics Sweden the Geodata Cooperation has encompassed a paradigmatic shift of the production of land use statistics due to the widely extended access to geodata. Below, a few practical cases on quality improvement in current production as well as emergence of new statistical application will be described in brief.

**4 More and better**

The production scheme of virtually all official products in the field of land use statistics has changed since 2011 due to the new and richer data situation. One of the single most significant elements of change encompasses the introduction of the Cadastral map in the production chain. Before Statistics Sweden joined the Geodata Cooperation, the Cadastral map was far too expensive to acquire. Today, a full copy of the map database is acquired once a year. The Cadastral map contains not only geographical representations of real estate parcels (covering the entire country) but also land use regulations and topographical features, such as land use, water bodies, buildings, roads, extraction sites, airports etc. in scale 1:10 000. The geographical representations of the parcel property units has id’s corresponding to records with administrative information in the Real Property Register, concerning ownership, taxation etc. Statistics Sweden had access to the Real Property Register, including coordinates, even before the Geodata Cooperation but due to the lack of physical delimitation of the real estate parcels, the use of the register information was limited in a geospatial sense.

By combining different thematic components of the map together with corresponding information from the Real Property Register the land use context and ownership can be assessed at the level of individual parcels. Before accessing the Geodata Cooperation, Statistics Sweden’s calculations of built-up land were entirely based on administrative figures from the Real Estate Tax assessment register. By combining the Real Estate Tax assessment register with the Cadastral map, thus increasing the accuracy, the estimated area of certain categories of built-up land has been reduced by 50 percent [4].

The increased accuracy of area estimations is not only limited to calculation of built-up land. Another typical example of quality improvements, widely noted by the media, was the press release on Land and water areas in 2012. The new method employed, based on real estate boundaries and water bodies from the Cadastral map, “shrinked” the land area of Sweden with some 3,000 square kilometres or 0.7 percent. The area of inland water (lakes and rivers) had been equally underestimated, while the territorial border remains unchanged [5].

One example of new statistics following in the wake of the Geodata Cooperation is building footprints, i.e. the total ground space of physical buildings. In conjunction with the launch of the Geodata Cooperation in 2011, the first complete version of the national building register was released by the NMCA. The register is a part of the Real Property Register [6] containing some 7,2 million georeferenced buildings including information on building type etc. The building register was used by Statistics Sweden to make the first accurate calculation of the total building footprint in the country [7]. The press release yielded great media response as it showed that the footprint amounted to no more than 93.000 hectares, an area that equals two thirds of the small Baltic sea Island of Öland, a well-known place to Swedish holidaymakers. The building footprint figures have also been frequently requested by researchers and students trying to make accurate assessment on roof space potentially available for solar panel installations [8].

*Figure 1. Infographics from Statistics Sweden’s Facebook account illustrating the total building footprint of Sweden.*



**5 Some challenges**

This paper has focused on the benefits of data sharing and the quality improvement in official statistics as a result of the increased access to geospatial data. But a growing usage of geospatial data from a variety of producers also brings about challenges. As the bulk of data from external producers in the production process increases, the harder it is to overview the quality, accuracy and coverage of the data sources involved. Usually metadata typically describing “map products” does not satisfy the needs from a statistical point of view. This is why it is important to invest time to get acquainted with the data producers, to understand the underlying conception of their production as well as the history and original purpose of the information. Recent initiatives within the UN (UN-GGIM) and the EU (ESS), aiming at a better integration between statistics and geospatial data, put an emphasis on the need for a closer cooperation between statistical institutes and mapping agencies. This is a necessary step towards a mutual understanding between the statistical and mapping communities, but eventually such cooperation must include also other data sharing institutions.

Another challenge has to do with the understanding and recognition of the needs associated with geospatial activities within the statistical offices. Geospatial processing typically demands a technical environment beyond the standards of many statistical offices in terms of software, hardware and storage but also in terms of competence. If there is little understanding regarding these special needs, the prospects of a successful integration of geospatial activities in the production chain is low.

**6 References**

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

[2] Information from the Swedish geoportal: http://www.geodata.se/en/

[3] Official land use statistics products at Statistics Sweden:

http://www.scb.se/en\_/Finding-statistics/Statistics-by-subject-area/Environment/

[4] Land Use in Sweden 2005 and 2010: http://www.scb.se/mi0803-en

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

[6] <https://e-justice.europa.eu/content_land_registers_in_member_states-109-se-en.do?member=1>

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

[8] Kamp S. 2013. Sveriges potential för elproduktion från takmonterade solceller. Teoretisk, teknisk och ekonomisk analys. Uppsala university.

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

Haldorson M. (2012) Benefits from data sharing - increased use of geospatial information in the statistical production process, DGINS conference in Prague 2012