

Strengthening Official Statistics as a strategic part of an Evolving Data Ecosystem

Marie Haldorson, Statistics Sweden, marie.haldorson@scb.se

Kristina Strandberg, Statistics Sweden, kristina.strandberg@scb.se

Johan Erikson, Statistics Sweden, johan.erikson@scb.se

Abstract

The concepts of “Data Stewardship” and “Data Governance” are getting a lot of attention in the official statistical community. Depending on the national context there are different paths forward to strengthen official statistics as a strategic part of the national data ecosystem. All National Statistical Institutes, NSIs, share the same challenge; how to harness the potential from all kinds of data generated in society – and how to guard, grow and give it back to society as official statistics of high quality. For example, increasing use and re-use of our statistics as open data, fulfilling legal obligations regarding the Once Only Principle (reducing response burden for businesses), being able to provide new and timely statistics and sharing statistical knowledge and tools.

This paper describes Statistics Sweden's role as coordinator of the Swedish National Statistical System (NSS). The NSS is an important part of the national data ecosystem and an important rationale behind our efforts to strengthen Statistics Sweden's data governance capability is to be able also to strengthen the NSS as a whole. More specifically the paper gives a concrete example on how to do this by sharing the so-called Statistical Production Support, SPS, a knowledgebase and quality assurance hub based on a national version of the Generic Statistical Business Process Model (GSBPM).

The idea to share the SPS is not new. It was recommended already in a previous peer review (2014) within the European Statistical System (ESS), that Statistics Sweden should share its methods and tools with other statistical agencies responsible for official statistics. Since then, this has been included in the action list supporting joint Strategic Objectives of the Official Statistics of Sweden: "We do it better together". By sharing methods, tools, and processes, it is easier to secure cost efficient production of official statistics with quality that is fit for purpose.

Keywords: Data ecosystem, data governance, National Statistical System, quality assurance, statistical production process, GSBPM, official statistics, national statistical institute, fit for purpose

1. Introduction

Data Stewardship and Data Governance

The concepts of “Data Stewardship” and “Data Governance” are getting a lot of attention in the official statistical community, especially in EU as the European Data Strategy together with new legislation is changing the European data ecosystem. The main legislations of relevance here are the Open Data Directive with implementing acts on High Value Datasets, Data Governance Act and Data Act. Depending on the national context there are different paths forward. All national statistical institutes share the same challenge; how to harness the potential from all kinds of data generated in society – and how to guard, grow and give it back to society as official statistics of high quality.

In order to better understand how Statistics Sweden can position itself as data steward, within our existing legal mandate, we compared our strategic goals with characteristics associated with data stewardship – such as to provide data, curate data, facilitate data dissemination, establish trust, encourage data use and create regulations (UN, 2022).

The conclusion was that Statistics Sweden can strengthen its position as coordinator for the National Statistical System (NSS), and by that take on more of a Data Steward role in all steps of statistical production (input, throughput and output). We can also be an active partner to the Agency for Digital Government (DIGG), as they have been assigned to coordinate the joint digital infrastructure for government, including building basic data domains and establishing data spaces in line with the EU Data Strategy.

The National Statistical System

Official statistics in Sweden are regulated according to the Official Statistics Act (2001:99) and the Official Statistics Ordinance (2001:100). The ordinance regulates which government agencies that are responsible for official statistics covering different subject areas, currently 29 agencies with Statistics Sweden as both producer of official statistics and coordinator of the NSS (SCB, 2022).

The statistics themselves are divided into subject areas, statistics areas and products. At the moment there are 23 subject areas and 115 statistics areas. Each

agency is responsible for one or more statistics areas. This also means that there can be more than one responsible agency within a subject area. Statistics Sweden is responsible for 45 statistics areas within 13 subject areas.

ESS Peer review

Already in 2014, the ESS peer review process pointed out that to strengthen the Swedish system of official statistics, it would be favourable to broadening the knowledge base about the statistical production process across the whole system. Statistics Sweden was recommended to share its Statistical Production Support, SPS (described in more detail in chapter 2 below), with the other NSS agencies. This factor became one of the inputs into a review of the SPS starting in 2019.

In 2019 a joint strategic objective for the NSS was developed (SCB, 2019), endorsed by the Council for Official Statistics. Common goals related to input, throughput and output needed to be underpinned by sharing of knowledge and tools, which has resulted both in training in the statistical production process for all statistical agencies and a decision to start sharing the SPS.

In this paper we describe the importance of strengthening the National Statistical System as part of a broader data ecosystem and some of the challenges we encountered when creating a SPS accessible to other agencies in the NSS.

2. About the study: Generic processes – generic statistical production support

Benefits of generic processes

Statistics Sweden's Process Model (in the following referred to as the process model) was developed in 2007, based on the work by Statistics New Zealand (Pearson and Savage, 2007), in parallel with (although independent from) the development of the GSBPM (Erikson, 2020). It was created as part of a large modernisation and standardisation project at Statistics Sweden, initiated when it was decided to introduce a process-based way of looking at the statistical production chain.

Since the first version of the process model, there has only been minor revisions to it. To support the standardisation, the Statistical Production Support (SPS) was created

to contain all the decided standards used in the statistical production process at Statistics Sweden. Besides process descriptions and links to IT tools, it also contains standard routines and support in the form of checklists, templates, and other things necessary to follow the standards. It can be seen as a web-based handbook on how to carry out the statistical production process from start to finish. While it is not a production system in itself, it does contain lots of helpful links to production systems. From the beginning the SPS has had a rather simple structure, built on a clickable version of the process model (see figure 1).

Figure 1. The Statistical Production Support (start page). Source: SCB



Statistikproduktionsstödet - introduktion

Statistikproduktionsstödet kan beskrivas som en handbok i statistikproduktion, och beskriver SCBs gemensamma metoder, verktyg och arbetssätt upplagt så att det följer SCBs processkarta för statistikproduktionsprocessen. [Processkartan »](#), Det är uppdelat så att varje delprocess har en beskrivningssida uppdelad i

- Input (vad behövs för att kunna starta delprocessen),
- Output (vad som kommer ut ur delprocessen).
- Genomförande (mer detaljerad information)
- Dokument och länkar att använda

Inom genomförandedelen finns det praktiska stöd som behövs för att genomföra delprocessen, såsom tillvägagångssätt och extra stöd. Aktivitetslistor, checklistor, mallar med tillhörande anvisningar och andra fördjupande bakgrundsdokument samlas under Dokument och länkar att använda.

SCBs Statistikproduktionsstöd kommer successivt att tillgängliggöras för andra Statistikansvariga myndigheter. Just nu är följande delprocesser tillgängliga för SAM:

- Process 2 - Designa och planera
- Process 3 - Skapa och testa

Each main process and sub-process (which in turn can be divided into more detailed sub-sub-processes or additional pages when necessary) has a page divided into four parts; introduction, input, output and realisation:

- A short description of the (sub) process and the purpose of it.
- Input: things that need to have been done before starting this process, and where the information needed comes from (with links to previous sub-processes when applicable).
- Output: The results coming out of the process.
- Realisation: a detailed description of how to carry out the process, including links to appropriate tools and descriptions of standards to follow (including templates, checklists, et cetera).

The order of the parts has previously had realisation before output, but the current order makes the order more logical and the new order is now being implemented over time.

The SPS has a somewhat different role than the process model itself; the main purpose is to present the normative standards, decided upon by Statistics Sweden, that should be followed by all staff involved in statistical production. Compliance with the standards have been followed through quality revisions up until 2021, it is now under discussion if and how to follow up compliance in the future.

Challenges with a standardised approach

The main challenges encountered when implementing the process support system were the following (Erikson, 2020):

Acceptance

Standardisation always comes with acceptance challenges. Some people question why they should change something that is working. A standard procedure might be good for the NSI as a whole, but not always for every single user. The need for standardisation is questioned by experienced people who know current routines by heart and from experience. Standardisation might be seen as stifling creativity, and the value of implementing them (for example to gain a certain certification) is also questioned.

Finding relevant information

The process model and the SPS contain a lot of information, and not all standards are applicable for all types of surveys. For example, some standards are only applicable to sample surveys with primary data collection, while others apply to surveys based on administrative data. There are a lot of standard routines connected to different modes of data collection, and surveys differ depending on which modes they use. There are strict rules to follow for mandatory surveys that are not applicable

to voluntary surveys. There are differences between collecting data on individuals and from enterprises, et cetera.

This means that for many sub processes, it will be necessary for the production teams to find the standards that are applicable to them and use only those standards. This can be a difficult task sometimes; one solution is to include generic parts in a more tailormade quality assurance system for daily operations. This must go hand in hand with routines for updates though, otherwise production teams risk missing important information.

Keeping the information up to date

It is very important that everything in the support system is up to date and still relevant. This has two implications. First, when routines or standards change, it is important that the information in the SPS is updated accordingly. Second, not all standards actually change over time, at least not very often. This means that some standards developed five or even ten years ago are still relevant today. However, encountering a document, describing a process, which was created and last updated in 2013 can make users uncertain as to whether the information is still relevant.

Review of the SPS

Since the start in 2007, different aspects of the SPS have been reviewed and re-worked several times. In 2019, it was apparent that a larger review of the SPS was necessary and a small task team with experience from different parts of the organization was formed. Previously conducted reviews had all been focused on usability – for specific user groups or for Statistics Sweden as a whole [reference]. This time around, the review team had two major instructions:

1. Create a functional statistical production support system that clearly supports Statistics Sweden's strategy. The rationale behind this being that a SPS firmly grounded in Statistics Sweden's strategy would help the transition toward important strategic goals.
2. Share the support with other agencies responsible for official statistics, as a move to strengthen Statistics Sweden's position as coordination for the NSS, and by that take on more of a data steward role in all steps of statistical production.

An initial evaluation of the SPS showed that the results from previous reviews had never been fully addressed – most importantly, usability was still an issue. Long texts with passive information, too many clicks to find the right information, a bad search engine, a mix of formats (html, word, pdf and excel) and ambiguity in what was rule and what was advice made the support hard to use.

It was also clear that the content had grown over the years. The support had become hard to manage and keep up to date. The scope of the review therefore widened, and the questions raised got broader: “What is the best use of the SPS? How do we handle the maintenance of the support?” After careful consideration and user input, the review team presented three possible directions for the SPS. All three options were deemed possible, even though cost and work effort varied considerably between them.

- **MINI:** A minimal support solution that basically boiled down to a handbook, with the most essential descriptions of the statistical production process. Upkeep and management of the support would be a minimum.
- **MIDI:** A middle-of-the-road solution which included a complete update of the current support. Many of the texts would be rewritten with more focus on forward-leaning methods for data collection and data management. While the initial work would require a lot of work, upkeep and management would be similar as for the current support.
- **MAXI:** A new support system on a new technical platform. The support would have an interactive interface, where the content would be tailored to the user need, depending on his or her function within the production team. The rebuild would be substantial and require new development. Maintaining the support would also be very work intense.

Another important find in the review was the different structure of the separate processes. Over the years, through different models for management support, the processes had been managed by separate teams. As a result, the support lacked a common structure, tying the whole system of process pages together. A suggestion designed to help create a more cohesive support was to establish an editorial team, responsible for the overall content of the support.

In the end, the second goal of the review (to share the support with other NSS agencies) together with cost considerations pointed us toward the MIDI-option. A re-write and an update of the existing content seemed the most efficient way of creating a functioning support for production of statistics within the NSS. An editorial team was appointed, responsible to deliver on the findings from the review.

3. Results – Sharing the Statistical Production Support

With a designated direction established, and an editorial team in place, it was time to deliver on the conclusions of the review. Initial work focused on “low-hanging fruit”, such as an update of the design, the removal of old or incorrect information or links and the transfer of information from external documents to the html-pages. An effort to improve on the search engine was also successful.

Simultaneously to the “cleaning” of the existing support, an investigation into how the support should be exposed outside the Statistics Sweden intranet was initialized. Two questions had to be answered; first whether it was needed to divide the support into general content that could be shared and content specific to Statistics Sweden and second, which technical platform that should be used for the support.

An ambitious choice would be to divide the support in a way that would make general information publicly accessible while information specific to Statistics Sweden would be restricted and accessible only to internal users. This would be achieved by designing the pages in two “layers”, where the first layer would be accessible to all and therefore only would contain general information. Internal users would then have access to the next layer, where information specific to Statistics Sweden would be kept. Since this would be a completely new way to present the material, it would mean that all texts in the SPS would have to be completely rewritten.

A much simpler approach would be to share the support “as is”, i.e., letting the external version be an exact copy of the internal version. Not only would this easier solution mean less re-writing and easier maintenance, it would also mean a simpler technical solution since it would dodge the need to handle access related questions.

Table 1: Alternatives for sharing, pros and cons.

Alternative	Pros	Cons
Alternative 1: "As is"	<ul style="list-style-type: none"> • Less work, minimal re-writing. • Only one version to manage and keep updated. Cost for maintenance would be lower. • Transparency outside of Statistics Sweden • Quicker to get up and running. • Simpler technical solution. 	<ul style="list-style-type: none"> • Lots of information, not everything relevant outside of Statistics Sweden. • Templates with Statistics Sweden trademark would have to be handled. • Could potentially generate lots of questions from other agencies.
Alternative 2: Two versions	<ul style="list-style-type: none"> • Content suited to different user groups. • No need to share all information. • Cleaner look, both internally and externally. • Easier to use, internally and externally. 	<ul style="list-style-type: none"> • Work intense to set up first version. • Two versions to manage and keep updated. • Templates with Statistics Sweden trademark would have to be handled. • More clicks for internal users. • Need to create a way to handle access for external users.

After looking at the pros and cons for the two alternatives (table 1), the most pragmatic choice was made. To share an exact copy of the internal support would be the easiest to accomplish from a technical standpoint. It would also mean very little extra maintenance work since the external page could be maintained through a time-controlled automated copying process. In addition, one of the deciding factors was the fact that the SPS first and foremost is set up for the benefit of Statistics Sweden. Creating a more complicated structure would most likely take away from the internal usability of the support, which was not an option. Sharing an exact copy would allow us to custom make the SPS to the needs of Statistics Sweden and still offer support and advice to other agencies producing official statistics.

Technical solution

Exactly how the technical solution for the external pages would be set up had to be a joint effort between the SPS editorial team, the security section of Statistics Sweden and the IT-support crew. After careful consideration, it was decided that the SPS would remain on the SharePoint platform. A server would be set up outside of the secure shell and a “mirror image” of the SPS would be automatically copied to this server on a pre-set time interval. This way, only one version of the SPS would have to be managed and the external copy would always be identical to the internal copy. Even though this appeared to be a “simple solution”, the setbacks were several and the fixes not always easily achievable. At Statistics Sweden, there is a lot of information stored on the joint SharePoint platform, not only the pages that we refer to as the SPS. Careful mapping of the content of the whole SharePoint system was necessary to decide which parts should be copied onto the external server. To make the split possible, major rearrangement of the libraries (folders) in the SharePoint system was necessary. Even the copying process itself was not as easy as it first appeared since it was somewhat limited by internal functionality of SharePoint.

Management of the folders on the SharePoint platform and of all the content is the responsibility of the SPS editorial team. A new structure was adopted, that allowed for a split between folders that should be copied to the external SharePoint platform (on a server outside of the secure shell) and the rest (which is the bulk of the content). With this new structure in place, it was possible for the it-support team to create a script that automatically copies the designated folders to the external platform every night. The scheduled copying is possible through the program Sharegate.

4. Discussion

The new task for the SPS editorial team

When the SPS editorial team was put together, its main purpose was to deliver on findings of the last review of the SPS and to create a support accessible to other agencies producing official statistics. Any further need of the editorial team, after this work was done, would have to be discussed after the completion of those tasks. But, as they say, things change.

In 2021, Statistics Sweden was reorganized to strengthen the line organization. Maintenance teams for IT-systems were kept but previous process and maintenance roles, important for the content of the SPS, were discharged. In addition, there were important changes in the functions responsible for the quality management. Statistics Sweden had already (January 1, 2021) left the ISO-certification standard and relied instead on the SPS for guidance regarding quality standards. Now the procedures for internal quality review would change, making the SPS even more important as a hub for quality management and documentation.

In the new organization, the responsibility of the statistical production process was concentrated to a small unit called Methods- and Architecture Governance, placed at the Executive Office, together with the SPS editorial team. Through this change, the SPS editorial team was permanently dedicated to take the overall responsibility of the SPS. Not only was the editorial team now responsible for the changes in the SPS, but it was also the entity responsible for the overall management and upkeep of the support.

An important new task was identified with new working methods for the upkeep of the SPS. It was clear that the only way to keep the support up to date was to rely on subject matter expertise in the line organization. The SPS editorial team is responsible for the operational decisions regarding the whole SPS, such as design, level of detail in the text, and the administration of user questions/enquiries about the support. The editorial team is also in charge of the external support and all administration around the SharePoint solution. However, the actual content of the pages, i.e., the texts describing procedures and methodology, is left to experts in each field. With that solution, not much would change from the old organization, except that one entity would take sole responsibility for the overall content of the SPS – completely in line with the recommendations of the review.

External version of the SPS

A major undertaking for the editorial team in the new organisation was to launch the external version of the SPS. As described, it did get more complicated than first assessments suggested. Finally, in the spring of 2022, the first process “Design and Plan” was released on the external platform, easily accessible to the NSS through “SAM forum”, a webpage for coordination and support for the Swedish official

statistics (SAM forum, 2022). A careful evaluation of the technical solution showed that everything was working according to plan. Since then, the process “Build and Test” has been released and the next process in the pipeline, “Collect”, is scheduled during the summer.

Keeping the SPS updated and relevant is vital for quality management at Statistics Sweden – exposing it to other agencies in the NSS, and to the public, adds to that importance. On the other hand, it opens a new channel for discussion and dialogue within the NSS. By being transparent with the SPS, Statistics Sweden hopes to ignite a more involved dialogue between the agencies within the NSS regarding tools and methods involved in the production of statistics. In the SPS, Statistics Sweden has a useful tool in the quest of statistical leadership. The dialogue has just started, and we hope that the SPS can help us build for the future.

Further development of the process model and the SPS

Many factors drive the need to further modernize the statistical production process. For example, the possibility of using new types of data sources combined with the increasing difficulty to obtain data through traditional direct data collection. Furthermore, while we see increasing user needs regarding both detail and coherence, we also face the need to reduce costs while maintaining quality.

In short, one of the main ingredients of a modernized statistical production process is the cohesive design of whole statistics areas / clusters of statistical products together, rather than designing separate surveys, resulting in coherent statistics fulfilling the needs of the users. This can be achieved using a combination of surveys, registers, and new data sources, together with automated processing.

The modernized process is design driven, data driven and metadata driven. Combining various data sources, with registers and data from the administrative systems of businesses first and foremost, puts the emphasis on somewhat different steps in the production process than before. Much of this has to do with finding, evaluating and using data sources.

Statistics Sweden has developed an updated version of our process model, taking these things into account. The updated model is also built on our thoughts of structured data storage supported by metadata in designated “data exchange points”

in the process. The updated model supports maximum reuse of data between statistical programs, data capture by system and periodicity rather than by product and coherent design of statistics areas. With the emphasis on data and metadata, the updated way of working will also better support the idea of Statistics Sweden as an actor in the national and international data eco systems.

5. Future perspectives: Statistics Sweden as data steward for the National Statistical System

Late 2021 Statistics Sweden was peer reviewed again as part of the 2021-2023 round of ESS peer reviews. Among the recommendations, to improve beyond compliance with the Code of Practise, the peer review team proposed that *„Statistics Sweden and the government should better define the role of Statistics Sweden in the national data ecosystem; the role of national data steward could be considered”*.

Statistics Sweden will use this recommendation to strengthen its role mainly as data steward for the NSS, in close collaboration with DIGG, data holders and the Government (as illustrated in figure 2).

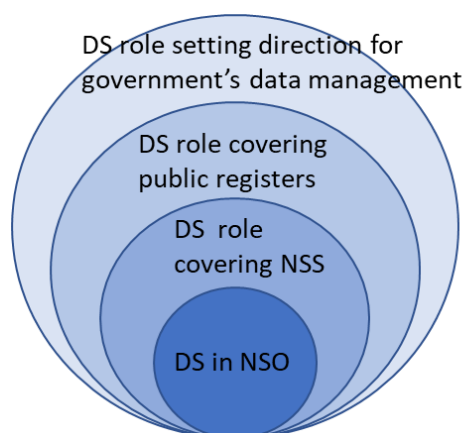


Figure 2: Different levels of data stewardship for National Statistical Offices, NSOs, as illustrated in a draft report (UNECE, 2022).

Data steward in NSO: Statistics Sweden has established a Data Department since September 2021, as part of its new organisation and with the aim to strengthen an internal data stewardship function.

Data steward role covering the NSS: The coordinating role for the NSS can be broadened to a data stewardship role, including data governance issues (data

capture from a NSS perspective), data processing (joint methods, tools, SPS-support etc) and data/ statistics dissemination (open data, PX-Web etc). Some of these services can be developed free of charge for other agencies, others can be part of commissioned services.

One example of services that will be free of charge is an ongoing project with funding from VINNOVA which will deliver an open platform for data processing and joint analysis, including how to work with open-source tools like R and Python. Statistics Sweden can be the link between the NSS and academia through strategic partnerships with Örebro University and AI Sweden.

Data stewardship role to cover public registers: DIGG is responsible in collaboration with agencies responsible for basic data domains within Ena, (DIGG, 2022). Ena means „unify” and is the name of the national digital infrastructure. Statistics Sweden has a supporting role as basic data provider and has been assigned by the government to develop Ena together with DIGG and other agencies.

Data stewardship role setting direction for government’s data management:

The government has developed a national data strategy in 2021 and assigned a number of agencies with specific tasks to achieve the strategic goals, Statistics Sweden is responsible for one assignment on „Data for smart statistics”. The assignment has a deadline in March 2023 and will deliver results relevant for the whole NSS regarding quality criteria for new data sources, statistics based on mobility data, rules for official statistics as open data and ideas on how to better visualise statistics.

In conclusion: Statistics Sweden will aim to strengthen its role as data steward for the NSS, through close collaboration with DIGG and other agencies – benefiting from results coming from the assignment on “Data for smart statistics”. This is also the response to the peer review, including the following three actions:

1. Proposal to establish a “building block” (joint capability) for official statistics in Ena, with frameworks, standards, tools etc. This action aims to strengthening official statistics as an important use case for the national digital data infrastructure. The proposal needs to be formally approved by DIGG.
2. Deliver results from government assignment “Data for smart statistics”.

3. Develop a business case for extended commissioned services regarding data capture and processing for other Government Agencies.

Sharing the SPS with other agencies has paved the way to continue strengthening the NSS as part of the national data ecosystem. Collaboration and strategic partnerships will be crucial for producers of official statistics, for example to make it mandatory for providers of administrative registers to inform statistical agencies of any changes in their registers or making joint agreements with private data owners. Statistics Sweden's journey as data steward, together with developing our statistical leadership, will help us achieve our vision to provide society with useful and trusted statistics.

6. List of references

DIGG, 2022: Ena, the national digital infrastructure. Available at:

<https://www.digg.se/ledning-och-samordning/ena---sveriges-digitala-infrastruktur> (accessed at 22 june, 2022).

ERIKSON, Johan, 2020: Le modèle de processus statistique en Suède. Courrier des statistiques N4 – 2020, Insee, France. Available at

<https://www.insee.fr/fr/information/4497085?sommaire=4497095> English version available through the author.

PEARSON, John and SAVAGE, Tracey, 2007: Statistics New Zealand's Business Model Transformation Strategy: Creating a new Business Model for the 21st Century National Statistical Office. June 18-21 2007. American Statistical Association (ASA), International Conference on Establishment Statistics ICES-III, Montréal, Québec, Canada, pp. 1150-1157.

SAM forum, 2022: Samordning och stöd för den svenska officiella och den europeiska statistiken samt internationellt statistiksamarbete. Available at

<https://www.scb.se/sam-forum/hem/> (accessed at 22 june, 2022).

SCB, 2019: Strategic objective of Official Statistics of Sweden. Link:

https://scb.se/contentassets/15740059c3c74fde9310ff95126a9f21/malbild-sveriges-officiella-statistik_en.pdf (accessed at 22 june, 2022).

SCB, 2022: Description of the National Statistical System. Link:

<https://scb.se/en/About-us/official-statistics-of-sweden/> (accessed at 22 june, 2022).

UNECE, 2022: Task Force on data stewardship. Link:

<https://unece.org/statistics/task-force-data-stewardship> (accessed at 22 june, 2022).

UN, 2022: UN Statistical Commission, 53:rd session in March 2022, Background document from the UN Working group on Data Stewardship: Preliminary results from the Working Group on Data Stewardship.