Ingegerd Jansson
Jens Malmros
SCB

Meeting with the Scientific Advisory Board of Statistics Sweden

8-9 May 2025

Attending board members

Jan Bjørnstad, SSB and Oslo university
Barteld Braaksma, CBS
Steven Heeringa, University of Michigan (online)
Anders Holmberg, ABS
Annette Jäckle, University of Essex (online)
Sune Karlsson, Örebro university
Johanna Laiho-Kauranne, CSC IT Center for Science
Thomas Laitila, SCB
Bella Struminskaya, University of Utrecht

Attending Statistics Sweden staff

Eva-Lo Ighe, Director General, chair
Lilli Japec, deputy chair
Marie Haldorson, head of methodology and architecture unit
Anna-Maria Kling, head of methodology section, support unit
Kristina Strandberg, head of methodology section, data management
Mikaela Järnbert, head of methodology section, social statistics and analysis
Jenny Hjort, head of methodology section, economic statistics and analysis
Jens Malmros, secretary
Ingegerd Jansson, secretary

SCB

Welcome

Eva-Lo Ighe welcomed everyone and gave a special welcome to Bella Struminskaya who attended the meeting for the first time. Eva-Lo introduced herself as she is the new director general of Statistics Sweden. After that followed a round of introduction of the board members and other attendants at the meeting.

Current issues at Statistics Sweden

To be successful in the data-driven world of today, strong governance and coordination in fields like legal frameworks, technology, quality assurance, security, and methodology is necessary. Recognizing this need, the European Commission has revised the basic legal act for Statistics. This revision allows national statistical agencies to access new data sources and to take a more active role in managing data issues in their countries, becoming more of a data steward and taking a leading role in national data governance.

The new act emphasizes the importance of leveraging the skills and experiences at the national statistical institutes. By utilizing the expertise in data interoperability and data quality across various domains, NSIs are well positioned to contribute to the creation and implementation of a national data governance framework. This enables active managing, sharing, and protection of data within the countries.

Statistics Sweden has been tasked by the government to explore the concept of data stewardship further. We are examining the national context and will provide our view on whether and how our agency can assume a new role within the broader data ecosystem.

Reply to recommendations

Lilli Japec gave a review of the recommendations for the topics presented at the previous meeting in May 2024. Three topics were discussed at the previous meeting.

Towards the Vision for Coding: An Accurate and Efficient Coding Process. Stat Sweden was recommended to revisit the vision for coding and align it with the AI-policy. This is now work in progress.

It was recommended to continue to invest in new coding methods and support experts. Many of the recommendations concerning this were included in the report from the project.

An ongoing project is aiming to develop ML-models and acceptable error rates, which was also a recommendation from the Board

Quality Assurance of the Re-coding to NACE Rev. 2.1, Combining Model and Manual Coding. The Board pointed out the importance of

communication and interaction with stakeholders and users. This work will be intensified, especially with external users.

The importance of a feedback loop was further pointed out. SCB's iterative approach in developing the coding algorithm allows continuous method evaluation.

Comments on the suggested model for recoding were for example to validate the quality of the business register and to be aware of model degradation. The Board recommended not to use the suggested imputation model but to consider alternatives. SCB will evaluate alternatives.

In order to prioritize, choices should be made explicit and documented, and prioritizations should be made together with users and stakeholders. The project is implementing a continuous improvement process, and all choices are documented.

Household Consumption. Estimation of product composition. The project has not continued during the last year, due to lack of data. The plan is that the work will start again, eventually.

Topic 1 Data collection in the Swedish LFS
Sara Frännlid introduced the topic.

Attendants from Statistics Sweden: Elisabet Andersson, Mats Bergdahl-Kercoff, Annika Nordström, Krister Näsén, Melul Sevim, Frida Videll, and Matilda Wedtström Kjerf.

The response rate in the Swedish LFS has declined for several years. Several measures have been introduced to mitigate the decline in response rates; however, the joint effect of these measures has not been previously studied. In addition, the overall result of the data collection has not been evaluated. Consequently, there is a need for a thorough evaluation of the data collection in the Swedish LFS.

Statistics Sweden has received EU funding for a project to map the data collection, suggest areas for improvement, and implement the suggested improvements. The SAB topic presents the result of the data collection mapping and the suggested areas for improvement.

Discussion
Steven Heeringa opened the discussion.

Summary of comments from the discussion:

- The response climate will get worse. SCB needs to be aware of and pay attention to all "dangers" in society now, i e frauds, blocking, etc. The possibility to use all sorts of devices, apps,

etc does not stand out anymore, and does not necessarily attract positive attention. Is it worth the effort? Instead, use existing information, for example other sources.

- It is important to focus on communication, it helps to some extent. Older strategies might still help but not as much as it used to. SCB needs to find what works today. How does the LFS appear in the eyes of the users? Do they find the statistics important, what do they want to know about and how? Investigate other ways of communicating to motivate the public. How can we sell what we do and how we do it?
- Legislation does not permit all ways of communicating and reminding selected persons. Revisit the legislative part again, i.e., is there any chance of allowing private email and text messaging?
- The dependence on the self-administered mode is quite low in the proposed implementation. Possible improvements include offering an option for postal contact, offering the option for web in telephone interviews, and proper adaptive design.
- How to handle and deal with nonresponse is the most important topic for all statistical agencies. Ethics and professional integrity are important aspect of basing estimates on data with high nonresponse. What are statistical agencies willing to do to reduce and handle the nonresponse? Are we interested in doing a major change or only minor changes?
- Response rates vs quality or bias correction. High response rates used to be the only quality measure, without any regard to bias. Now response rates are low, but you still get a lot of data from large surveys like the LFS. The question is if it is of good enough quality. If you deem the quality to be good at 43 % response rate, can the same quality be reached at 25 % response rate?
- There was disagreement on the possibility to reduce sample sizes. On the one hand, it could be possible to reduce the sample since you still get a lot of data, and instead make sure that what you get is of good quality. Efficient data collection is good, and pressing for more data will only increase measurement error. However, with reduces sample sizes, precision will be an issue as well as bias.
- Look more into what´s going on. The No answer-category is large in the LFS, SCB should look for the reasons and characteristics of different groups. An experiment in the survey Understanding society was mentioned, where it has been further investigated when people want to be contacted.
- Costs. LFS is a complex survey with large sample size. There are no low hanging fruits, except increasing the web portion. It is already a short questionnaire, and some categories in the sample are not interviewed every time in the panel. Can you see

what is the most effective of the things you do, i.e., reminders, the call schedule, etc? It would show where to put the efforts and what would be most effective to get people to do the online survey. Maybe you could have an approach where respondents contact you, so that you find the persons at the right time.

- Mandatory participation. The one thing that would work best for increasing response rates is to make the survey mandatory. It has worked in Norway and in Australia, where all social surveys are mandatory. In Sweden, the only mandatory survey was the traditional census, carried out for the last time in 1990.

Topic 2 Searching for a new design of the economical-statistical system with access to new data sources

Susanne Svartengren and Gustaf Strandell introduced the topic.

Attendants from Statistics Sweden: Per Andreasson, Daniel Lennartsson, and Cecilia Wass.

During the autumn of 2024 a working group at Statistics Sweden was given the task to formulate a vision for the production of economic statistics using new data sources that will emerge in society within a few years, in combination with sources that Statistics Sweden already works with in one way or another.

With new data sources, opportunities emerge for the production of official statistics. In Sweden, several new data sources will be established regardless of whether SCB choses to act today or not, but there might be a possibility to influence the content and the final implementing act and thus the future use of these sources. In order to realize this SCB will have to team up with, e.g., Swedish authorities, other NSIs, and Eurostat. For some of the incoming data sources, the final design of the actual reporting obligations may be crucial for the potential these new data sources will have for statistical purposes.

In the conceptual proposal of EKSTAT 4.0, the administrative sources *Standardised Accounting Statement*[1] *(SRU)*, *Digital Annual Reports*[2] *(DIAR)*, *Value Added Tax (VAT/OSS)*, and *Monthly PAYE Return (AGI)* will constitute the principal sources for establishing stable levels in the production of economic statistics for the non-financial corporate sector per year, quarter, and month. For the breakdown of statistics at a more detailed level, e.g., by product groups, new data sources, for example,

---

[1] Standardised Accounting Statement is an annex to enterprises tax return.

[2] Annual Reports in structured format (inline XBRL) received by Swedish Companies Registration Office

Transaction-By-transaction-level-Reporting[3] (TBR) and *detailed Accounting Data according to a Standard Chart of Accounts (BiSK),* will play an important role. In addition, the Swedish central bank (the Riksbank) aims to produce a database covering *all payment transactions in society (Payment Data*), which will be a valuable complementary source.

By utilizing these data sources, SCB will be able to reduce the data collection directly from enterprises and organizations, compared to the current situation. The remaining direct data collection will focus on areas not captured in administrative data, such as qualitative questions and hard-to-capture phenomena in the economy, e.g., contract processing and merchanting. Hence, direct data collection could be minimized in a future scenario utilizing several new data sources, especially in case of TBR implemented in Sweden according to an optimal scenario from SCB: s point of view.

Discussion
Barteld Braaksma opened the discussion.

Some comments from the discussion:

- There will be large volumes of data. Statistics Sweden will need to know what is required, what are the risk, which competences are required, and what is the data policy at federal level. Data management policies are necessary. There will still be large costs, but they will be on the statistical office instead of the businesses. The question is if it really saves costs for the society? It is a potential IT- cost bomb, and it remains to prove that it is cheaper than what is done today.
- There are several issues to be resolved in working with new data sources. The quality is not always fit for statistical purposes and stability over time is not guaranteed. Other issues include different classifications, different concepts, different units, and insufficient timeliness.
- Can the necessary work be done in cooperation with data holders? Is it necessary to duplicate the data and store it at Statistics Sweden? Copying data make them prone to changes at the supplier. Statistical agencies are used to doing all the work by themselves, but cooperation might improve sustainability, as well as the production of high-quality statistics.
- The combined information may have risks. You have to consider the responsibility of your organisation. There is also

---

[3] Digital reporting requirement based on e-invoices, part of the EU reform package, VAT in the Digital Age (ViDA).

legislation, e.g., the digital services act, which may be applicable.
- There will be conceptual difference in accounting data compared to what Statistics Sweden wants. It is necessary to sort out those differences.
- Statistics Canada encountered issue with privacy when they wanted to use transaction data. It was necessary to analyse what data were needed for making the statistics, and it turned out that data could be sampled without compromising quality. It is something that Statistics Sweden should pursue, but you need to be able to tell the users what happens. There are solutions for sampling from large data sets, it is an approach that has been used before to improve timeliness. You should consider the necessity and proportionality framework of Stats Canada.

Key areas for methodology development at SCB

Kristina Strandberg, Mikaela Järnbert, Anna-Maria Kling, and Jenny Hjort presented the key areas for methodology development.

Statistics Sweden's strategy aims to ensure impartial, high-quality statistics. The focus is on an innovative statistics production, particularly the development of the application of new statistical methods to leverage new technology and new data sources. To support these ongoing developments, the heads of the methodology sections at Statistics Sweden were appointed to describe the key areas for methodology development. The proposed areas are *data capture and use of data sources*, e*fficient and modern design in statistical production*, and *producing statistics.*

Data capture and use of data sources consist of two subareas, *evaluating data sources*, and *uniform and efficient data collection and data management*. To improve evaluation of data sources, new and existing data sources for statistics production need to be evaluated and described regarding the data generation process and data quality. The possible streamlining of data collection using AI, efficient coding, editing, and imputation using models as well as efficient data processing can be investigated.

Efficient and modern design in statistical production focusses on *design and re-design, quality in statistics,* and *handling of missing data*. Work on design should focus on subject design, increasing methodological competence, and increasing opportunities from new data sources. To improve quality in statistics, the description of quality in multi-source statistics and of non-sampling errors can be improved.

The production of statistics concerns *timely and relevant statistics, statistical models in the production process,* and *protection of data.*

Examples of possible improvements include methods for high-frequency time series, developing the use of machine learning in statistical production, and development of disclosure control methods.

Some comments from the board:

- With deterioration and reduced effectiveness of surveys we need to think carefully about using models in production. We should be thinking about designing surveys that improve our ability to model processes. Survey-assisted modelling can be an addition to this area.
- The main objective should be to expand the capability frontiers. You must have an end-to-end perspective from input to output, otherwise you risk becoming too thin. Process quality is important in this respect.
- The data stewardship approach could be considered, where that the data providers are at one end and the users at the other end. Competence development also in the organisations providing the information is of interest.
- It is useful to have a flexible design approach, but it may have an impact on the microdata availability which may affect researchers.
- If you should handle missing data, you need to model it, for which you need competence to do proper modelling. Handling of missing data and statistical models should be put together in an estimation process.
- The best way to handle missing data is not to have it in the first place. Solve it at the design stage.
- Editing is important in economic statistics, but it does not appear in the plans. Statistics Sweden has already put a lot of resources into improving editing.
- A comparison was made with the history and development of Ford automobiles; you don´t want to make faster horses, you want to make cars. Statistics Sweden could be more forward looking.

Topic 3 Scientificity in Data Management

Sophia Lennartsson introduced the topic.

Attendants from Statistics Sweden: Natalie Jansson and Cecilia Wass.

Scientific rigor is important in the production of official statistics. However, how to adhere to scientific methods in data management is less clear. Statistics Sweden has therefore started to investigate how a scientific approach should be designed, implemented, and maintained in data management. The present project encompasses the methodological principles, from which they set out from, and further investigations in data management issues with respect to scientificity.

Discussion

Johanna Laiho-Kauranne and Sune Karlsson opened the discussion.

Some comments from the discussion:

- Research is about understanding the world and includes following good practice. The production of statistics focusses on the results and should be guided by science. The present work is leaning too much towards the research world, which is not necessarily applicable to the real world.
- It is important to consider existing frameworks on data management. They highlight the professional standards and how to adhere to scientific principles. It might be enough to work with existing frameworks which will provide direction to the present work. Map your terminology to existing frameworks and focus on what you think is missing.
- Examples of frameworks to consider include the FAIR principles, the GSBPM, and various ISO standards. It can be of interest to consider practices at other NSIs.
- This topic is becoming more important in taking new roles as data stewards. We should know our position.
- The presented dimensions are relevant. A dimension about ethical standards and responsible data use could be added. It is important when dealing with other parties that we are explicit with what is acceptable to us.
- This is about quality assurance, doing the right things. At ABS, we talk about our methods being scientifically defensible. That does mean published but built on best practice. If you could explain that what you are doing follows the principles of methods, you have come a long way.
- You may argue that knowledge is produced at NSIs, but the focus on the what not the why. The understanding comes when we give our numbers to others in society, therefore our process is very similar to a scientific process.
- We should talk more about these matters. Statistics should be based on scientific results - then we know that the results work.
- Explainability is important but not straightforward, in particular in connection with AI. It is a black box.
- The standards are its own thing and should be kept apart from practical factors. Standards will not change, but practical factors will.
- The terminology in the discussion paper is not standard, especially the word "scientificity". A proposal is to replace scientificity with "evidence-based".
- It is not clear what constitutes scientific data collection. The point is to use scientific principles when handling the data. The point is good data. It is a quality thing more than science.

- The scientific approach is globally threatened. We must make a stance and defend the scientific values at various levels. Hence, discussing similar questions are important.

Concluding words

Eva-Lo Ighe thanked the board for the valuable discussions.

Next meeting

The next meeting will be online on November 20 and 21.