Ingegerd Jansson
Jens Malmros
SCB

Statistics Sweden Scientific Advisory Board

May 4, 2023

Attending Board Members

Jan Bjørnstad, University of Oslo
Barteld Braaksma, Statistics Netherlands
Xavier de Luna, Umeå University
Steve Heeringa, University of Michigan
Anders Holmberg, Australian Bureau of Statistics
Per Johansson, Uppsala University
Annette Jäckle, University of Essex
Sune Karlsson, Örebro University

Attending Statistics Sweden staff

Mats Bergdahl-Kercoff
Marie Haldorson
Ingegerd Jansson, secretary
Lilli Japec
Thomas Laitila
Jens Malmros, secretary
Magnus Sjöström
Kristina Strandberg
Gustaf Strandell
Joakim Stymne, chair

Current Issues at Statistics Sweden

Joakim Stymne welcomed the members of the Board.

On budget issues (see recommendations from last meeting). SCB shares the concerns of the Board. SCB has asked for 3,5 million euros for 2024 and 6,5 million euros in 2025. SCB also asked for additional funding for data collection and data processing but is not optimistic on new funding. There are new costs due to new EU regulations, for example on open data. In an interview for a politically independent online news portal, Joakim expressed concerns that were acknowledged by the ministry.

On the consequences of priorities. SCB made a careful analysis of the situation last year, to ensure resources for production and development. Budget cuts were imposed mainly on administration and communication, and some surveys were deprioritized. Two examples are the Party Preference Survey that will be carried out once a year instead of twice a year, and the Name statistics that will no longer be produced. The economic situation for this year looks better.

On the Swedish EU presidency, first half of 2023. The head of the statistical agency is also head the working party. An important issue during this period is population and housing statistics. There is a proposal on which the member states should agree. The proposal treats for example the definition of population base and sharing of data and is a compromise between costs and flexibility.

Another important issue is access to privately held data and the Commission legislation act 223.

New premises in Örebro from September. The office in Örebro is moving into a new building in September this year. Joakim invited the Board members to Örebro for the next meeting with the SAB in November 2023.

Comments from the board

- Why cut the budget on communication? Good communication is important, as the example of the interview with Joakim shows. Reply: the cuts have a focus on organisation. It was possible to make budget cuts and reorganize in a way that will improve efficiency. It does not imply that communication is of less importance.

Statistical leadership in the emerging digital societies

Marie Haldorson presented a summary from the high-level meeting with the Directors-General from all European statistical agencies that was held in Stockholm in April.

The theme of the meeting was statistical leadership. The general view from the meeting, as expressed by Eurostat's Director General Mariana Kotzeva, was that statistical leadership is a good way of framing what we are doing even though the content is not new.

Statistical leadership has also been lifted within ASPIRE (presented to the Board in December 2022) and fits with the strategic goals of SCB (presented to the Board in May 2022).

Susan Linacre, former Deputy Director General of the Australian Bureau of Statistics, was the keynote speaker. Marie presented highlights from Susan's presentation.

Statistical leadership has been part of SCB's strategy for a while. This is now re-enforced and supported by the keynote, together with a citizen-centred focus and an outreach to policy makers. An important goal is to have better communication with government, to show Statistics Sweden's value.

Comments from the board

The Board agreed that this is important work and a good way forward. The meaning of statistical leadership and related issues were further discussed:

- Reliability and availability of data is important, as is a close and active communication with the users of the data. Analysis is the core of statistical thinking. SCB needs to understand the analyst as a user and what they are doing.
- Statistical leadership is more than analysis, for a statistical office it also includes taking a leading role, and identifying new needs, find data or define concepts. The leading role can be taken before there is funding, the agency can take the initiative.
- It is important for a statistical agency to make the public trust them. Show cooperation and achievements, go from being discrete to being out there. If the statistical system is highly decentralized, like in the US, this is even more of a challenge, and the public is sceptic. SCB is in a better position.
- Stakeholder engagement is key to staying relevant, networking and engaging is very important.
- Citizens can also be involved, they are aware of the power of data, they can challenge decisions and do their own analysis
- It is important to motivate the people within the organization. Individuals must be motivated to change and there must be incentives to get better. Good people are necessary for the ability to produce relevant, accurate, and timely statistics.
- There is an innovation network at EU-level that fits well with these thoughts. SCB is involved. There are also other fora where similar thoughts are discussed, and it is important to work together.

Joakim concluded that the discussion was very interesting and revealed somewhat disparate views on what statistical leadership is.

Reply to recommendations

Jens Malmros and Lilli Japec presented SCB's replies to the recommendations given by the Scientific Advisory Board (SAB) at the December 2022 meeting.

Progress on editing

Magnus Sjöström and Mats Bergdahl-Kercoff presented how the work on editing has proceeded since May 2022.

Comments from the board

The Board congratulated SCB on the progress. Some questions:

- Editing is not one process, but several, and there is heterogeneity between different types of data, e g surveys, administrative data. For example, there may be a separate process how to treat outliers. How are the different processes handled?
    - In practice, this is not one process, rather there are different activities taking place during production. But the name "process for data editing" was coined from the start and was helpful when communicating.
- There is a difference between soft and hard errors, where soft errors are unlikely or inconsistent values that might be true. How is that handled?
    - This is handled at the product level, there is no common approach.
- It is important to take respondents seriously and involve expert on survey measurement in the design of surveys and questionnaires.
    - Cognitive experts are involved at product level and in the work with a common web collection tool.
- There are cultural differences concerning how the staff at SCB has accepted the new approach to editing.
- How is the decrease in manual editing measured?
    - Measured as reduction in hours worked on editing. A lot of work has been done on transforming the savings into efficiency improvement.
- Process data is difficult to employ in the editing process. IT-systems are not tailor made for process data, and there is probably not one way of using the data.
- A general question on the ML-process. How do you in general handle ML, models, and prediction at a statistical agency?
    - The roadmap for ML states that ML at SCB will be used for editing, coding, and imputation. There is currently a focus on models for that. A platform is being developed for sharing models, code, etc. There will also be some options for different degrees of data sharing.
- Automatic editing is not really applied in the approach, it turned out to be more difficult than anticipated. Rather, identify what parts of manual editing add real value, and what part does not. Manual editing is much about calling back to

respondents, it is very labour intensive and add on respondent burden.

- It is truly a necessity to make the editing process more efficient which has also been raised in earlier evaluations at SCB. However, it was stated a challenge to assess effects on the quality of resulting statistics. In evaluations made on the manual editing, comparisons are made with earlier data collections and edited data sets. This is a problem because such evaluations rests on at least two assumptions. Earlier edited data sets are correct, meaning the edits detects all the important errors, and the edits will detect all the important errors in future data sets. Both assumptions are not realistic, of course.

  A suggested process implying a drastic reduction of manual editing which also improves the quality of statistics is the following: Presently SELEKT is used in several products for error detection. The principle is to compare an observed value with an estimate. Now, if SELEKT signals a potential error the suggestion is to replace it (imputation) with the estimated value. Thereafter a probability sample of imputed values are manually scrutinized and checked with the respondents. Several obvious gains are obtained with such a procedure. The most important is the ability to assess the quality of statistics on dimensions not possible with the present process or the one it is replacing. Some level of manual editing is always necessary even if its effect is minor. The reason is we cannot draw a conclusion on the results of manual editing from one data set to another. The importance of official statistics in society demands us to say, “we know” and not “we believe”.

Concluding words

Joakim Stymne closed the meeting.