Reducing Manual Editing at Statistics Sweden – an Agency Wide Approach

Jenny Hjort[1], Mats Bergdahl-Kercoff[1], Magnus Sjöström[1]

[1] Statistics Sweden, Sweden

Abstract

Like most NSIs, Statistics Sweden allocates a significant proportion of its resources regarding business statistics for the manual editing of data and output. Over the years, Statistics Sweden has made several attempts to make the editing process more efficient, with some success for some specific surveys but with limited total effect.

In late 2021. Statistics Sweden's top-management team decided that another approach was needed to drastically reduce manual editing so that resources could be used for much needed development work in other areas. The approach was labelled, "The New Editing Process", with the main focus on ensuring that the data coming to Statistics Sweden was correct by means of well-developed respondent checks as well as a clear focus on macro level editing to determine questionable values. The importance of editing was also clearly stressed if justified based on the stated purpose of the statistics, although not with the aim to catch every single value that was not fully correct.

The statistics were divided into three categories based on previous knowledge of resources spent on editing. There were around 20 statistical products in category 1 and 2 respectively which were the focus of the central implementation efforts as they together comprised about 80 percent of the data editing efforts. It was a clearly stated goal for these products that manual editing should be reduced as much as possible and that only clearly justified editing should remain. The responsibility for making changes were placed with the line organisation, but a central implementation organisation was also put in place to coordinate, provide support and to make sure that common development needs were taken care of.

Initially, training regarding the quality concept for official statistics was carried out as well as workshops to better understand and define user needs. Resources, including expert support in methodology and measurement issues, were allocated to the statistical products to assess the current editing approach and develop implementation plans. The approach was initially met with much scepticism but was also considered long overdue by many. This meant that progress was quick for some statistics but significantly slower for others. Two years on, Statistics Sweden has reduced resources spent on manual editing by about 40 percent and there is indication that more potential exists for a few statistical products.

Our presentation will describe the overall approach, results, the main features of the new editing process as well as the identified success factors.

Keywords: Manual editing, Efficiency, An Agency Wide Approach, The New Editing Process, User needs

1. Introduction

Statistics Sweden 's management team has been working over the past few years to ensure a long-term balance in both the economy and operations, aiming to facilitate necessary development. Ahead of 2022, it was assessed that expected efficiencies were insufficient to maintain both ongoing production and development work.

Therefore, in the autumn of 2021, a study was initiated with the directive to phase out all manual editing at Statistics Sweden of directly collected data by the end of Q1 2022.

The editing study resulted in a new editing process and, based on experiences during the work, an implementation plan was developed. While maintaining high ambitions for reducing manual editing, the ambition on time was significantly lowered. For instance, the new timeline meant that 20 percent of the editing would be phased out during Q1 2022. The combination of the assessment that significant potential remained to reduce manual editing existed and a pressing need to streamline operations in 2022 led the management team to focus on reducing manual editing.

2. The New Editing Process

The new editing process is based on the international model, the Generic Statistical Data Editing Model (GSDEM), consistent with other international models used by Statistics Sweden.

The purpose of the statistics and associated quality requirements form the basis for the editing. The focus is on aspects relevant to the quality of the statistics. Data providers are assisted in providing correct information, thereby reducing both recontacts and the need for changes. Manual editing (with machine support) is only used when well-motivated. The macro perspective also entails selective micro-level editing.

Although the editing process is not entirely new, it is clearly guided by four principles:

1. A pre-designed editing that efficiently ensures the quality justified by the purpose of the statistics.
2. Preventive work to ensure that data coming to Statistics Sweden is of such quality that micro-editing are only necessary in specific cases.
3. A micro-editing with a macro perspective, aiming to only identify potential errors crucial to the final quality of the statistics. The editing study also pointed out that these errors should primarily be handled with automated measures rather than manual recontacts/investigations.
4. Primary focus on macro-editing (rule-based), i.e., a structural assessment of the final statistics and whether they deviate in an unacceptable manner, requiring

investigation of causes and correction of underlying values. This assessment should be based on predefined rules/parameters.

The central work has therefore focused on supporting implementation in the following areas:

- Data provider editing
- Automatic changes
- Selective micro-editing
- Macro-editing
- Evaluation and adaptation of the editing

2.1 Data Provider Editing

Providing data providers with good opportunities to quality-assure their data before submitting it to Statistics Sweden leads to both reduced need for manual editing and reduced burdens on data providers. It has therefore been essential in the implementation of the new editing process to provide support to data providers in terms of information about what Statistics Sweden request and well-designed editing checks in the electronic forms.

The central support within Statistics Sweden has gradually been developed and supplemented by developing Guidelines for Data Provider Editing in web forms and files, as well as a Checklist for evaluating these. In addition to this, the introduction of so-called fixed response options has occurred, where the data provider specifies the reason why a value is correct even if it falls outside the limits set by Statistics Sweden's controls. This contrasts with the previous requirement of providing a justification in free text, which then needed to be handled manually by Statistics Sweden. The fixed response options include the most common reasons applicable to the product as well as a free text option.

2.2 Automatic Changes

During the work on phasing out manual editing, a principle decision was made early on to focus primarily on the possibilities provided by the current production environment and not to rely on development work that would not yield effects in the near future.

The focus in the work on automatic changes has instead been on developing functionality for identifying and managing unit errors. The possibilities for more general short-term solutions were assessed to be significantly better in this area. Various alternative methods for identifying and managing unit errors were investigated, and the decision was made to proceed with the so-called Position Method.

2.3 Selective Micro-Editing

Manual micro-editing will continue to be an important part of the toolbox for quality-assuring statistics, even though the focus is on minimizing it as much as possible. To ensure that only well-justified editing is conducted, a selective perspective must be applied, meaning that the goal is to only identify and address errors that have a significant impact on the final statistics. Finding the right level of ambition for selective editing has been a central part of all products' work in implementing the new editing process and has resulted in everything from having no manual micro-editing at all to relatively minor adjustments.

It is also important to understand that selective micro-editing has been the norm at Statistics Sweden for a long time, and significant efforts have been made to implement both methodological and IT support for it over time. However, this has often not been done with such a clear focus as now.

2.4 Macro-Editing

Macro-editing entails the control of the results of statistical production, i.e., editing the statistics to be delivered or published. Macro-editing is necessary to identify measurement errors with significant effects that have not been previously detected in the editing process and to detect any errors in processing. Therefore, the focus in the implementation work has been on achieving effective macro-editing designed according to what is considered good and well-established practice at Statistics Sweden.

The need for well-functioning and efficient tools for macro-editing has been emphasized from many quarters, and at Statistics Sweden, SAS VA and Power BI are widely used for macro-editing.

2.5 Evaluation and Adaptation of the Editing

The design of a survey is determined based on the purpose of the statistics, which in turn should reflect the prioritized user needs. To maintain a relevant design over time, manage changes, meet operational goals, and seize opportunities, the design needs to be regularly evaluated and adapted. This also applies to the new editing process. Therefore, at the outset of the implementation of the new editing process, considerable resources were devoted to disseminating information and creating insight into the concept of quality and how it is operationalized in such a way that the editing aligns with this.

It was also noted that relatively few products had regular follow-up of the editing's effects, both regarding controls accuracy and the controls the contribution to the entire editing process. One contributing factor to this is that access to process data from the collection has been insufficient, and there have been uncertainties about the basis for comparing estimations made on edited versus un-edited data. During the work, the existing guidance for retrieving process

data has been updated and supplemented with, among other things, code examples for accessing and processing data.

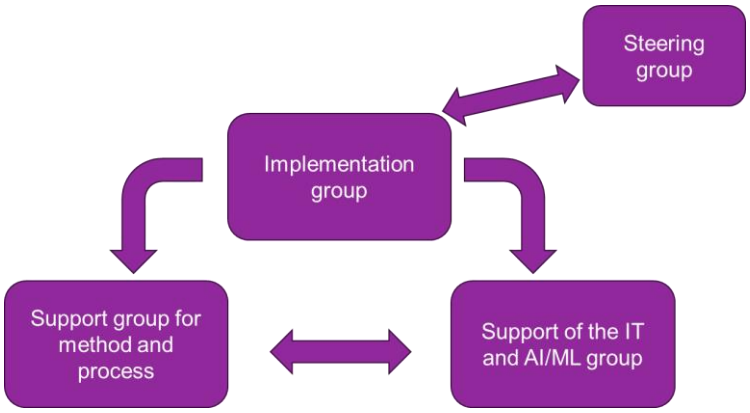
3. Implementation

The responsibility for implementing the new editing process has been placed with the line organisation, but support for the work and control of development efforts have been provided by a central implementation group. The starting point in the implementation work has been to achieve quick effects, and therefore, major investments in systems development have been avoided.

3.1 Implementation Organization

Based on the proposal in the editing study from the autumn of 2021, an implementation organization was formed according to Figure 1.

Figure 1: Implementation organization



The task of the implementation group was to lead, plan, coordinate, and drive the implementation work of the new editing process at an overarching level.

The members of the support group for method and process were experts in work procedures, methods, and tools in the new editing process. Their task was to provide support to different products when needed and to work on developing various central instructions and common support documents.

Initiatives on IT for common solutions have been coordinated and planned with the support of the IT and AI/ML group, while further development of product-unique components has been handled as previously. Parts of the group have collaborated with the method and process support group in some initiatives, such as investigating ML methods in certain products.

The implementation group chose early on to establish a steering group consisting of members of Statistics Sweden's top-management team who were most directly affected by the implementation.

3.2 Plans and Follow-up

Products were divided into three categories based on how much resources were allocated to manual editing based on the mapping done in connection with the editing study. The implementation group mainly focused on products in category 1, as it was assessed that there was the greatest potential for efficiencies in these products, but also the second category received attention.

The work was planned with implementation plans for each product. The formal and practical responsibility for implementing the new editing process lay within the line organization. This meant that the responsibility for developing implementation plans as well as implementing the activities lay there.

A template based on the new editing process was available to plan activities for the product and to add other product-specific activities that would reduce manual editing. The implementation plans were updated over time and were finally established in the autumn of 2023 for products in category 1 and category 2. The implementation plans were established through a procedure for consultation involving relevant section heads in the subject matter, data collection, and methodology areas. The final decision was the taken by the relevant department heads.

3.3 Communication

Considering the high level of ambition and that large parts of Statistics Sweden were both directly and indirectly affected by the work, communication was an important area. This applied mainly internally throughout the work but also externally towards users, customers, and other stakeholders. Internally, a clear distinction was made between those directly affected and with concrete tasks in the implementation of the new editing process and others. The first group needed to receive all the information and knowledge required for the work but also be given opportunities to benefit from work done on other products with good results. It was also considered important to give them good opportunities to raise questions, both of a concrete nature and more generally to the implementation organization. Other employees at Statistics Sweden primarily needed information about the progress of the work based on set goals and to channel their questions and concerns, mainly through their immediate supervisors.

Externally, communication was initially quite intensive to provide a clear picture of what Statistics Sweden intended to do, why, and how Statistics Sweden would ensure the necessary quality during the work.

The format and content of communication have varied during the period to meet the assessed needs. Communication was initiated early via the intranet, both in the form of articles and by the internally editing blog that was started to reach a wide audience at Statistics Sweden.

A large number of different information meetings have also been held with various target groups based on the state of the work at the time. Initially, information meetings were directed at section heads and employees working with products in category 1. Afterward, many meetings have also been held with other affected managers and employees in the organization.

4. Results and Conclusions

During the period 2021-2023, the editing is estimated to have decreased by at least 40%, which is approximately 19,000 hours.

Taking into account that measures taken in 2023 in some cases do not fully take effect until 2024 and concrete plans for reduced editing in 2024, the reduction amounts to at least 47% for the period 2021-2024.

It is generally difficult to directly measure the effects of various changes on the quality of statistics, and this also applies to the work that has now been carried out. It is reasonable to assume that the reduction in manual editing for some products has led to some increased uncertainty in the statistics due to increased measurement errors, but there is nothing to suggest that this has led to a significant deterioration in relation to the users' purposes for the statistics.

The implementation work has been carried out with great perseverance, based on a conviction that significant reductions in manual editing could be achieved. This conviction has gradually gained greater acceptance as the implementation work has progressed and experiences have been shared.

The new editing process is based on several different components that need to be designed and developed as a whole to achieve the intended effect in terms of quality and resource consumption. This means that at least expertise from data collection, methodology, and subject matter needs to be brought together, and the work should be carried out in close cooperation within production teams. During the implementation of the new editing process, it has been evident that the work could be conducted efficiently where there has been, or rapidly established, cooperation focusing on the contributions of the different parts to the whole.

It was evident very early in the implementation work that it is complicated to formulate the quality requirements of the statistics sufficiently clearly and to design a well-balanced editing

through the quality components. Continued work is needed in this area on a broad front to strengthen the ability to operationalize the quality concept. It is important to note that this capability is not limited to the implementation and development of the new editing process but is necessary for the ongoing transformation of how we capture and process data in a broad sense.

With that said, and although there is more to be done to reduce the editing, the achieved result is unparalleled at Statistics Sweden in modern times. This result involves both significantly reduced editing in some products, some of which originally had quite extensive editing, and that results have been achieved across a wide range of products. Therefore, large parts of the organization have been affected, either directly through implementation work or indirectly by other work needing to be prioritized.