UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
ModernStats World Workshop 2022
27-29 June 2022, Belgrade, Serbia

An Updated Process Model For A Modernised Production Process

Abstract
Many factors drive the need to further modernise 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, we see increasing user needs regarding both detail and coherence. One of the main ingredients of a modernised statistical production process is the cohesive design of whole statistical areas together rather than designing separate surveys. The modernised process is driven by design, data and metadata, and is sometimes referred to as "Statistical production 4.0".

Can our traditional process models such as the Generic Statistical Business Process Model (GSBPM) describe the modernised production process? Statistics Sweden has developed an updated process model since the model we use will benefit from some updates. The main focal points in the update are design of statistical areas, harvest and use of different types of data sources, re-use of data across surveys and areas, and a data driven view of the production process including clear handover points between process steps.

This paper will describe the updated process model, and its connection to the GSBPM.

Keywords: The statistical production process, modernisation, GSBPM

Background

Statistics Sweden's Process Model (in the following referred to as the Swedish 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. 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 process. The Swedish Process Model is similar to, but not exactly the same as the GSBPM. 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 in how to carry out the statistical production process from start to finish.

Modernisation

The production of statistics is changing. This is nothing new, changes to how we produce statistics have happened before, but it can be argued that the changes now are fundamental due to the digitalisation of society in general and the so called data revolution with data being produced in larger amounts than ever before and in a rapidly growing manner, while traditional direct data collection from both individuals and businesses is getting more difficult and expensive due to falling response rates and lower willingness to participate in surveys. So the statistical production is moving more and more to a "direct collection as a las resort" strategy, with registers and new types of data sources becoming the primary data sources. Of course, registers have been used for a long time, but digitalisation creates new registers that can be used for statistics. Furthermore, what will characterise the production of future statistics is the combination of different types of data sources.

Other changes to the statistical production process regard user needs. There are increasing user needs of flexible and timely data, but also more coherent statistics. One way to respond to these needs are redesigning statistical programs and whole statistical areas from the perspective of fulfilling user needs in the best possible way using different data sources to produce coherent statistics. It can also mean collecting data for several statistical programs together, for example using data in the administrative systems at businesses. This could mean moving to a collection by system and periodicity rather than keeping the same collection surveys as before. This would mean a more flexible approach where several data sources are combined in different ways to create the statistics that fulfil the user needs.

A data perspective on the production process

In the modernised statistical production process, data will even more than before be a key asset. The flexible approach describes requires full control over data; which data comes into the statistical office from which source and when, which statistics use which data sources and so on. The target for Statistics Sweden is to gain that control through metadata management and clear handover points between process steps and between statistical programs. The term handover points is similar to the concept of "steady states" but since those states are in reality not as steady as the term suggests, it was decided to use another term. The handover points outlined are the following:

- HP 0: raw data,
- HP 1: transformed raw data,
- HP 2: treated observation register,
- HP 3: final observation register,
- HP 4: statistics
- HP 5: published statistics and data.

Data in the handover points must be described with metadata and documented including a quality assessment. Some statistical programs collect data and generate as their output treated observation registers. Those are used and combined by other statistical programs to create final observation registers and statistics. There might be a need for further treatment before dissemination due to the result of disclosure control mechanisms, hence the last handover point. Collection (or rather harvest) of data and use of data will be supported by additional metadata on technical contracts of the data exchange, which means an automatic control over what data is used where will be gained.

Adapting the process model

Can the existing process model accommodate the changes due to ongoing modernisation? In some ways it probably could, but some of the main features of the modernised process would not be clearly visible. The Swedish Process Model and the GSBPM are very much a description of a linear process that is run through by each statistical program. It does not describe complex statistical areas and exchange and use of different data sources as well. Compared to the GSBPM the adjustments that would be useful are for example:

- Incorporation of a coordinated design of whole statistical areas consisting of several statistical programs. This can be an overarching process, but the results of such a design approach should influence the Specify needs phase.
- More focus should be put on the harvest of new data sources and flexible use of different kinds of sources.
- More focus on setting up metadata and rules for automated data handling.
- Ending the main process phases when data reach the handover points. This means an updated division between the Collect and Process phases as well as between the Process and Analyse phases.
- The modernised production process is mostly supported by a common IT platform. Therefore, it would be preferrable to remove the building of the IT platform from the statistical production process and instead let it be a part of the system development process. Instead, a larger focus should be put on the configuration (through parameters or code) within the platform. That configuration should be part of the design phase. This also means an updated division between the Design and Build phases.

Architectural support of the updated process model

Since some of the keywords of the updated way to produce statistics are flexibility end diversity, additional demands will also be put on the production platform. New data sources will appear while others change in their coverage and content and yet others become obsolete. The production platform must stand on a common solid ground yet must be able to support flexibility. Key features will be the possibility to add new statistical program and re-design existing programs in parallel with the ongoing production, so that future production rounds can be planned while not disturbing existing ones. Data exchange must be easily set up through new or adjusted contracts. To achieve this, common standardised metadata is a necessity, not only content metadata but just as important context metadata.

Challenges when updating the process model

If we introduce an updated process model, there will be a bigger distance between the national model and the GSBPM even if mapping is possible. This means it could be discussed when and how changes should be implemented. An alternate way would be to incorporate the changes as descriptions connected to the existing GSBPM model and strive for an international update of GSBPM.

- The SPS that is built on the national model needs to support both the modernised and the more traditional way of working. This could make it more difficult to keep it updated because also the more traditional process must have relevant support.
- Acceptance and useability in the organisation can be a challenge, since most of the staff are used to working in the traditional way, and the new process might initially be seen as strange or difficult to implement. On the other hand, implementation of a new way of working could be helped by having a new model as guidance.