



Economic and Social Council

Distr.: General
31 March 2011

Original: English

Economic Commission for Europe

Conference of European Statisticians

Fifty-ninth plenary session

Geneva, 14-16 June 2011

Item 3 of the provisional agenda

Organization of data collection and sharing, and the management challenges for the implementation of Statistical Data and Metadata eXchange

Towards efficient data collection at Statistics Sweden – organisational issues

Note by Statistics Sweden

Summary

The paper gives an overview of experiences and lessons learned from the organisational model of data collection used in Statistics Sweden today and the gains expected from the ongoing projects. One particularly interesting experience is the organisational impact on how new tools and methods are implemented. An example concerning multimode data collection in business surveys is given.

I. Abstract

1. The data collection at Statistics Sweden is a process organisation where the process department owns the process and two functional departments of process users run the daily business of direct data collection from respondents. The structure of the data collection departments and units is based on the collection mode and type of respondents. One department deals with household surveys and the other with enterprise and public sector surveys. Within the department for household surveys there is an interview unit (having both a group of centrally located telephone interviewers and field interviewers placed throughout the country) and one for questionnaire surveys. Other important roles in Statistics Sweden's data collection process belong to the register units who are responsible for collecting administrative data, and the unit for cognitive methods who are experts on questionnaire design.

2. With many roles involved in the data collection, there is a need for resource pooling where some people are experts in areas used by several units. Since methodological, technical and subject matter issues need to be addressed simultaneously in the process, continuous cooperation between the different roles is also necessary. A number of common technical tools are already in use and there are ongoing projects devoted to developing a common production environment for direct data collection and making the collection process more efficient. This introduces a new way of working and it can be foreseen that the data collection environment will look very different in five years time from now. Besides improving and adapting to new collection techniques, development is needed to be able to easily access, combine and process data from different sources and different collection modes. The paper gives an overview of experiences and lessons learned from the organisational model used today and the gains expected from the ongoing projects. One particularly interesting experience is the organisational impact on how new tools and methods are implemented. An example concerning multimode data collection in business surveys is given.

II. Data collection at Statistics Sweden today

A. Background

3. Over the last decade, the data collection organisation has moved from a decentralised organisation to a centralised organisation with two data collection departments. The creation of a process organisation for the statistical production process means that there is now a matrix aspect to the organisation as well.

4. Before 2004, each business survey subject matter unit collected their own data. For household surveys, there was even then a centralised unit that carried out surveys on behalf of external customers and supported internal customers with collection activities, especially scanning. In 2002, there was a decision to centralise the data collection, and from 2004 there have been two departments of data collection at Statistics Sweden. The responsibility that these units have lies mostly in direct data collection, most collection of administrative data is carried out by register units or subject matter units.

5. After the creation of a specific department for data collection from businesses and organisation, further reorganisations have occurred. The responsibility of collecting administrative data is moving more and more to the register units, but the main reorganisation was the creation of a process organisation for the statistical production process in 2008. The process department that was created then is now responsible for

developing and maintaining methods, tools and routines, there is a process owner for the process Collect. The data collection is still carried out by the data collection departments, and the work of the process owner is carried out in close contact with the process users.

6. Within the data collection departments, there is a division into units. The department for data collection from individuals and households has two units, one unit for interview surveys and one for questionnaire surveys. The interview unit in turn consists of both a central group of Computer-assisted telephone Interviewing (CATI) interviewers and a large number of field interviewers throughout the country. The department for data collections from businesses, organisations and public sector has two units for data collection from businesses and organisations, one for data collection from public sector and one unit for large enterprise management and coordination issues.

B. Roles in the data collection process today

7. Process owner: The process owner is responsible for developing and maintaining methods, tools and routines in data collection, giving support to process users and evaluating and improving the process. Some of the most important tasks for the process owner is to:

a) Maintain the common tools in place for data collection – the web data collection tool (SIV), the interview collection tool (WinDati), the Personal Digital Assistant (PDA) data collection tool for observation in shops of prices, the scanning system, the Touchtone data entry system, the “funnel” for receiving administrative data and the two sampling tools that are presently in use.

b) Establish common routines in the data collection process, such as common default choices for data collection methods in different types of surveys, default contact strategies, routines for fining business that fail to deliver mandatory data, templates for contact letters and reminders et cetera.

c) Act as client/customer of the project to develop a new common platform for data collection (see Section III A below).

8. Process users – the data collection departments: The data collection departments can also be called process users. They are responsible for carrying out the data collection and micro editing activities in practice as efficiently as possible, and to give feedback to the process owner on possible improvements in the process. Activities to improve the process are normally carried out jointly by the process owner and the process users. To make their work as efficient and professional as possible, the data collection departments have also created expert functions that are used within the department, jointly by both data collection departments, and also by other units at Statistics Sweden. Those are for example the scanning function, a group of questionnaire makers that create the actual questionnaires, a specific unit for large enterprise management and coordination functions and a service function for businesses that feel that they are over-burdened by statistical surveys. The data collection department for enterprises and organizations also plays an important role in the work with better regulation and the reduction of the administrative burden on enterprises.

9. Subject matter units: The subject matter units are the owners of the different surveys. They buy data collection services from the data collection departments. A few units still collect their own data.

10. Data collectors at register units: The register units (for the business register, the population register and the real estate register) are responsible for collecting administrative data from other authorities and distributing them within Statistics Sweden. An important task in this work is to set up and maintain contracts with the main data providers.

11. The unit for cognitive methods: This unit provides an expert function within the process department; its staff are experts on question design, questionnaire design and cognitive methods. The function has an important role to work together with questionnaire makers at the data collection departments to create the best possible questionnaires.

C. Experiences and lessons learned from the present organisation

12. Some of the most positive effects of the centralised data collection organisation are the following:

a) The possibility to create expert functions. The creation of expert functions within the data collection organisation would probably have been much more difficult or even impossible in a decentralised organisation. In a single survey, only a small amount of resources are devoted to such actions, where in a centralised organisation it is possible to pool those resources. It is also possible for some of the staff to concentrate on these functions, such as questionnaire making, where only a centralised organisation can offer continuous work in the area to a few persons in the staff, who can thereby become experts in this area.

b) Pooling of resources. The data collection activities in a single survey are not distributed evenly throughout a calendar year. This can mean that it is hard to have enough staff in some periods, while some of the staff may have too little work in other periods. In a centralised organisation it is much easier to share resources between surveys, so that the staff can be used much more efficiently. In order to have full effects of this possibility, the production environment needs to be as standardised as possible, something that is not the case for questionnaire data collection at the moment. An ongoing project is building such a common environment, see Section III A below.

c) Learning by experience. By gathering the collection activities in a centralised organisation, it will be much easier to find good things that are done in specific surveys and learn from them, while eliminating inefficiencies in other surveys. This together with the possibility to have expert functions as described above gives good possibilities to make the collection work more efficient. This has also been shown in figures by the unit for data collection from the public sector which had a staff of 24.1 full-time equivalent employees (FTEs) at the start. Since then, work has expanded by new tasks of 6 FTEs, while the staff is now 19.4 FTEs + 1.7 FTEs purchased from the department for data collection from individuals and households (an example of pooling of resources). This means an efficiency gain of about 30 per cent in three years.

d) Implementation of new methods, tools and routines is much easier to achieve if the process users are centralised. The combined force of the process owner and the centralised data collection organisation can mean a faster implementation of newly developed tools and routines, compared to when implementation is a matter between a process owner and a large number of decentralised process users. One example here is the implementation of a new default contact strategy where the first contact with businesses only offers web data collection, and where a paper questionnaire is only sent at the reminder stage. The experience at Statistics Sweden is that collecting business data by web is more efficient, less costly and good for data quality. With the centralised data collection organisation embracing this decision, implementation has been rather fast. Today, already 32 of the 55 business surveys carried out by the data collection department have used the new strategy and there is a plan for the remaining ones. For the surveys using the new strategy, the web response rate is now 88 per cent while it is only 49 per cent for the others.

e) Addressing issues that are not survey specific. A centralised data collection organisation has the possibility to find issues that need addressing that are not survey

specific. Specific examples here are response burden issues, where having many surveys carried out within one organisational unit means that burden issues become very visible, both at the aggregate level and at the level of individual businesses.

13. While the effects mentioned here are mostly positive, there have also been some drawbacks to the centralisation of the data collection. One main negative effect is internal tension between subject matter units and data collection departments. Economic issues such as what the cost of data collection should be, and what to do with the savings that arise from a more efficient collection are and have been common, especially right after the reorganisation. There are also some issues on who should decide what, what is the division of responsibility between subject matter units and data collection departments. Size in itself can also be a problematic factor, which is seen at the interview unit where planning of resources may be difficult sometimes. This is also addressed in Section IV below.

III. Ongoing projects to make the data collection more efficient

14. There are several ongoing projects that aim, wholly or partly, to improve the data collection process. Three specific initiatives that are worth mentioning are the Triton project that aims to build a general data collection and micro editing platform, ongoing projects to make the interview collection process more efficient and a data warehousing initiative including coordination of the main registers.

A. The Triton project

15. The Triton project is an ongoing project with the goal of building a general but flexible production environment for data collection and micro editing. The aim is to cover most kinds of surveys, but in a first stage it is directed at surveys with direct data collection through questionnaires (web and paper). Even though a version of the platform is already in use, a new and significantly improved version is under development. The new version will be released at the end of June 2011. The aim is that the new platform will replace many of the old survey specific information technology (IT) systems, be usable for a majority of the surveys at Statistics Sweden, integrate the common tools already in place, introduce new common tools for additional parts of the data collection process and eliminate as much manual work as possible. Besides integrating the existing common tools such as the web collection tool and the scanning system, the platform will have three main new parts: an administration/ design tool for setting parameters for a specific survey and monitoring the survey progress, a tool for working with individual objects and a communication platform that connects all the parts of the platform.

16. The project is carried out with a team of process owner, a professional project leader, data collection staff, methodology staff, IT staff (including IT architects) and business architects. Together, the team has defined how the data collection and micro editing processes should look in the near future and what IT support is necessary for these processes. This has also led to a new desired way of working with these processes, which will hopefully lead to both a more efficient way of working and a process of good quality, leading in turn to good quality in collected data.

1. Expected gains from the project

17. Some of the more important gains expected from the project, apart from technological gains such as a better IT infrastructure that is built on modern technique and that is easy to maintain, are the following:

a) Metadata and documentation of it will have a real effect in production. The platform is built with an administration part where each survey will enter the parameters necessary to run the platform based on the survey specific conditions (when the collection is run, which tools to use, how many remainders to send, when and how those are sent, how to treat duplicates et cetera) and these parameters will have a direct effect on the actual production. For example, it will be possible to have an automatic sending of questionnaires, letters and address files to the printing and sending function at the specific date entered in the administration tool. This is a bit different to the present situation, when there is a planning process that is not well integrated with the actual production process, and also a documentation process usually carried out when production is ended, and where the decisions made et cetera are documented in retrospect. With the active handling of meta data described here, much documentation of how the process was carried out will also be available automatically.

b) More of quality assurance built into the actual production process rather than a retrospective quality control. Today, there is already much focus on quality assurance, but it is often carried out as a quality control in retrospect and in a separate tool, where there is for example a check list of things to do. In this check list, the production team leader can fill in ‘done’, ‘not done’ or ‘not relevant’. Even if this is better than nothing, there are still drawbacks to this approach. If many items are “not relevant”, there is always a risk that the whole check list will be taken lightly. And checking “done” in retrospect is not as good as having the activities that need to be carried out available all the time, in the production system itself, and to mark each activity as done as soon as it is actually carried out. The Triton platform will create just this, an area for each survey, within the administration part of the system, where activities are auto-generated based on the choices made and parameters set by the specific survey. This will also eliminate most or all of the need to mark activities as not relevant, simply because only relevant activities will be auto-generated. To have these parts of quality assurance (of course there are also other aspects of quality assurance that are not covered by these activities) as a built in feature of ongoing production will be a significant gain to the production process.

c) Easy to work in different surveys and pool resources. With the common platform, some things that differ between surveys today will be standardised. One important example is administrative codes to classify non-respondents, over coverage and such. All these kinds of codes are survey specific today, which means that staff will have to learn these codes every time they work on a survey they have not worked on before. A common interface, with standardised placement of functions on the screen, will also mean that it will be easier to jump between surveys. Sometimes there is a need to introduce extra staff to a specific survey on a very short notice, and the threshold of learning the specific IT system may be a real hindrance, this will also be eliminated with a common platform.

18. The Triton platform is up and running in an early version at the moment, but intensive work is being carried out with the aim of releasing a new version of the platform, which will be necessary to realise the gains mentioned above, by June 30 2011.

2. Organisational aspects of the Triton project

19. The project work has been facilitated by having a centralised data collection organisation. The gains of the centralised organisation described above (especially learning aspects) has been essential in the project team, where the representatives from the process users can speak for all data collection, not only a small part of the organisation. The results of the projects will also facilitate working in a centralised data collection organisation, as shown by the third point in Section III A 1.

B. Initiatives to improve the efficiency of the interview collection process

20. There are several ongoing projects that aim to improve the efficiency of the interview collection process (the projects have other goals as well, such as maintaining or increasing response rates in interview surveys). Important parts of these projects are to find more efficient contact strategies for telephone interviews (when and how often to try to get in contact with different types of respondents) and using a common set of respondents to contact for both the central CATI group and the field interviewers. Both these initiatives are dependent on having a centralised organisation for the interviewing activities, so that all interviewers, both in the central group and in the field, can be used as efficiently as possible with regards to contact strategies and the total work load.

C. Data warehousing initiatives and register coordination

21. The data warehousing and register coordination initiatives aim to make use of the advantages of a National Statistical Institute, such as access to registers and access to data from many different sources. These initiatives are still in the planning stage, but will have organisational impact if they are carried through.

IV. Future challenges in data collection

22. While the centralisation of the data collection organisation has had many positive effects, and further positive gains are expected from the ongoing Triton project, this will not mean the end of challenges within the data collection process. Some future challenges that need to be tackled are:

a) The tendency of declining response rates to surveys. It is a real challenge to motivate respondents to participate in statistical surveys and keep response rates at a reasonable level.

b) The pressure to further reduce response burden. There will probably be further pressure to reduce the administrative burden on mostly businesses, but also within the public sector. There is also a strong demand from businesses and from the government to coordinate data collection between different government authorities or agencies.

c) The need to combine administrative data and survey data. One way to collect less data and reduce burden is to re-use already collected data. To do this on a larger scale will mean at least some methodological challenges.

d) There are important issues to tackle in the interview collection, and they may be even larger when the efficiency improvements discussed in Section III B above are implemented. The challenge is large to find an optimal way to plan and pool resources and interviewers so that efficient contact strategies can be implemented in all surveys. Different strategies will be efficient for different surveys. The planning process as a whole with a large number of interviewers working on a large number of surveys is also an issue in itself.

23. The expected future challenges are difficult ones, issues that would probably be hard to solve for a single survey while the centralised organisation with its expert functions should have a better possibility to meet and handle them.

V. Conclusions

24. The data collection organisation regarding questionnaires at Statistics Sweden has gone through large changes in the last decade. In total, the reorganisations have had many

positive effects on efficiency, but also some negative effects regarding internal tensions within the organisation. An ongoing project to build a common production environment will help to further realise the possible gains of a centralised data collection organisation as well as giving new gains. However, there are still a number of future challenges for the data collection process, challenges that a centralised data collection organisation should be better equipped to meet.
