

The system of collecting administrative data and how it responds to the guidelines of the code of practice and the peer review

Sinikka Laurila*, Janne Eskelinen and Johanna Sisto
Statistics Finland, Helsinki, Finland

Abstract. Recently, Statistics Finland developed a centralised system for collecting and receiving administrative data. That work was well in line with Statistics Finland's strategic plans to harmonise processes, develop generic solutions and centralise data collection activities and know-how.

The work was organised as a project to take place in the period from 2013 to 2015. Statistics Finland now has a new system that utilises process metadata to control and inspect the received administrative data. Project results also included the implementation of a generic tool for technical validation and distribution of administrative data.

The European Statistics Code of Practice gives guidelines for statistical offices about using administrative data and registers in statistical production. This presentation evaluates how these guidelines are met in the new information system that receives and validates administrative data at Statistics Finland. A brief review is also presented on how it corresponds to the Peer Review (2013 to 2015) recommendations today.

Keywords: Administrative data, information system receiving administrative data, data collection process

1. Background

1.1. Using administrative data in statistical production

Statistics Finland has a long history of using administrative data and registers when producing statistics. One of the main successes in collecting and using administrative data is the population census data, which have been combined from various official registers since 1990. Today, most of all the data that are used in statistical production at Statistics Finland come from administrative data or registers and about 150 administrative data or registers are collected and used when producing statistics.

1.2. ICT strategy implementation

Statistics Finland updated its current ICT strategy

(2015 to 2019) during 2014. The mission of ICT activities is to develop and generate efficient ICT solutions for statistics production. In addition, the target is to offer a reliable and easy-to-use interface to Statistics Finland's services for data suppliers and statistics users.

As a result of the 2008 strategy guidelines, Statistics Finland decided to centralise the collecting of administrative data from various official data providers and started a project with the main task to build an information system and a technical solution for centralised data collecting.

1.3. 'Havas' project

The Havas (Hallinnollisten aineistojen vas taanottojärjestelmä in English The system for receiving administrative data)-project started in April 2013 and lasted two years ending in March 2015. The main goals of the project were:

- To find and record all the administrative data that are collected and used by Statistics Finland.

*Corresponding author: Sinikka Laurila, Statistics Finland, Helsinki, Finland. Tel.: +358 29551 3573; E-mail: sinikka.laurila@stat.fi.

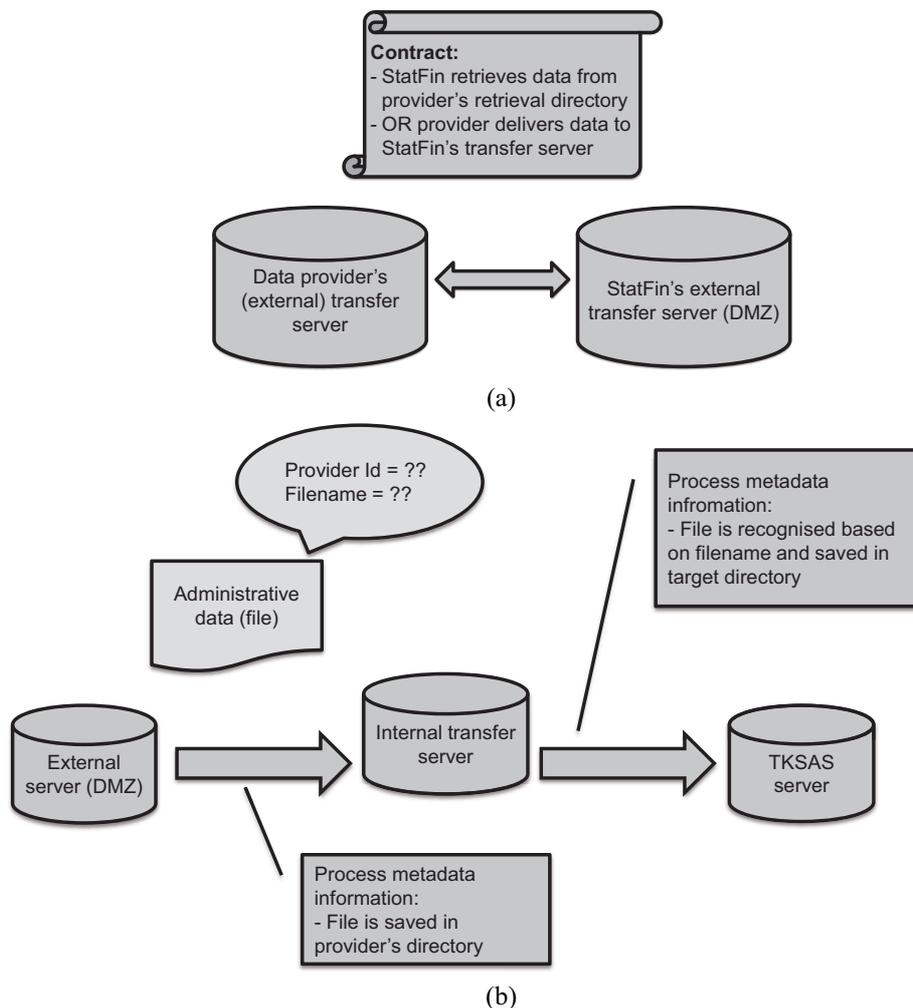


Fig. 1. (a) Data transfer between transfer servers; (b) Servers and process metadata information.

- To define and build a centralised IT system for receiving administrative data by line transfer.
- To define and build a SAS system for pre-controlling the data received.
- To describe all the data and variables in a centralised metadata system. The descriptions were to be used in pre-controlling the data before it is handed over to statistical production.
- To organise a team to whom the centralised administrative data collecting was handed over after the project.

The Havas team was formed and it started its work alongside the project in autumn 2014. One of the most important tasks of the team is the collaboration with the official organisations that are data providers. The goal of the collaboration is for Statistics Finland to have data contracts with all the data providers by the end of

2017. This is also mentioned as a task which the ESS (European Statistical System) Peer Review appointed to Statistics Finland in 2014. Today, in 2016, nearly 70 per cent of the administrative data is collected and received using the centralised information system.

2. Administrative data collection process

The centralised administrative data collection system is built on the following ideas:

- Process metadata information controls data transfer and directs arriving administrative data files into the right target directory.
- Data files arriving in flat file format are transformed into SAS file format using the generic ap-

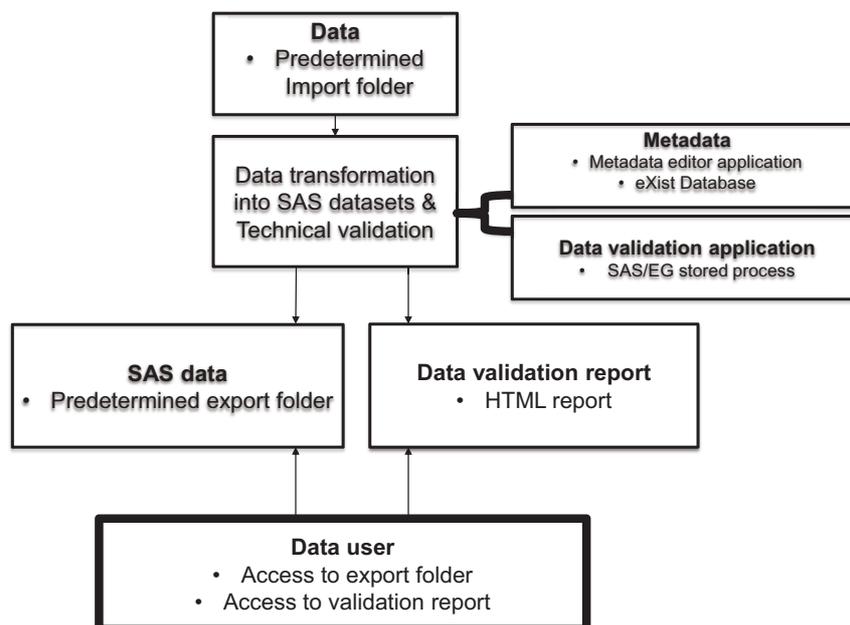


Fig. 2. Technical validation, data transformation and routing data to its predetermined target folder.

plication and metadata information (file descriptions).

- Data files are validated technically utilising metadata information (file descriptions).
- Accepted data files are saved in customer directories and delivered to users.

2.1. Data transfer

Contracts with data providers determine how administrative data go between external transfer servers of the data provider organisation and Statistics Finland. Basically, there are two different possibilities (see Fig. 1(a)).

- Either Statistics Finland retrieves the data from the external transfer server of the data provider organisation, or
- The data provider organisation delivers the data to Statistics Finland's external transfer server.

Once the data arrive at Statistics Finland's external transfer server (DMZ-area) they are transferred to the internal transfer server (see Fig. 1(b)). At this point, the Havas team receives an automatic email stating that the data have arrived in the directory dedicated to the data provider.

In the next phase, process metadata information saved in the eXist database directs the data files into the right target folder on the TKSAS server based on

the filename and provider ID of the data. In order for this phase to end successfully, it is essential that data providers follow the contracts and name the data files as agreed. Otherwise, the file cannot be directed to its target folder. If that happens, the Havas team receives an automatic error message about the orphan file.

2.2. Technical validation and data transformation

Once the administrative data have been saved in the predetermined import folder, a Havas team member transforms the data into a SAS dataset and validates the data technically. This step utilises metadata – i.e. file descriptions that have been saved in the eXist database using the metadata editor application – and the data validation application (see Fig. 2).

The transformed and validated administrative data are saved in a predetermined export folder as SAS datasets. As an output from the technical validation phase, the Havas team receives an email including a data validation report in HTML format. If the validation report shows that the data meet expectations, the Havas team sends the validation report to the end user(s) and informs them that the data are now available in the export folder.

The centralised administrative data collection process needs standardising of certain phases in order to work properly.

- Standardised process metadata.

- Standard naming of the files.
- Standardised data descriptions (data and variable metadata).

In practice, by means of this information, it is possible to automatise the entire data collection process, which includes automatic technical validation, data transformation into a SAS dataset and routing data to its predetermined target folder.

3. Benefits of centralised administrative data collection

Successful centralised administrative data collection comes with many benefits compared to a situation where many separate statistical production units are in charge of their administrative data collection. By means of the centralised data collection and consistent data collection procedures we are able to achieve many benefits, such as.

3.1. *More efficient and systematic administrative data receiving processes*

This means the standardisation of our data receiving processes that enables process automatisation. This also means standardising our operating environment including a consistent folder structure for the data received.

3.2. *Better transparency of the utilised administrative data*

In the future, it is possible to establish a much better overall picture of the administrative data utilised at Statistics Finland. In other words, we can increase our knowledge of collecting and receiving administrative data including the information on what, when and what sort of data we are receiving. Earlier it was difficult to form a general view because all the know-how was fragmented.

3.3. *Rationalised data requisitions/requests and advanced data usage*

In the future, one team is responsible for all new administrative data requisitions, hence we can avoid overlapping of requisitions. A common objective is that the joint use of administrative data increases and that the statistical production units are henceforth more aware of the existence and possibilities of the administrative data gathered.

4. Code of practice and the guidelines of the peer review 2013 to 2015

4.1. *Code of practice*

Statistics Finland has made the quality commitment to implement the European Code of Practice when producing statistics. The Code of Practice was revised in 2011 and it contains 12 principles including institutional environment, statistical processes and statistical output to guide the production of statistics.

Principle 8 Appropriate Statistical Procedures.

Appropriate statistical procedures, implemented from data collection to data validation, underpin quality statistics.

As mentioned earlier, Statistics Finland has a long history of using administrative registers and data in producing statistics. In the Code of Practice (CoP), Principle 8 focuses on data collection practices. In the Quality Assurance Framework (QAF) indicators 8.1, 8.7, 8.8 and 8.9 of Principle 8 consist of information and recommendations on collecting administrative data.

Indicators 8.1 and 8.7

When European Statistics are based on administrative data, the definitions and concepts used for administrative purposes are a good approximation to those required for statistical purposes.

Statistical authorities are involved in the design of administrative data in order to make administrative data more suitable for statistical purposes.

The administrative data that are collected by official organisations are not primarily meant to be used to produce statistics. However, the concepts and definitions that are used in these data collections are for the most part good approximations for statistical purposes as well. It is important that the statisticians, who are using the data, know the definitions of the variables and the collection process well enough to guarantee that the data are suitable for statistical production. This means constant and frequent collaboration with data owners. This also verifies that the interests of statisticians are taken care of when changes are made to the administrative data collection.

Indicator 8.8

Agreements are made with owners of administrative data which set out their shared commitment to the use of these data for statistical purposes.

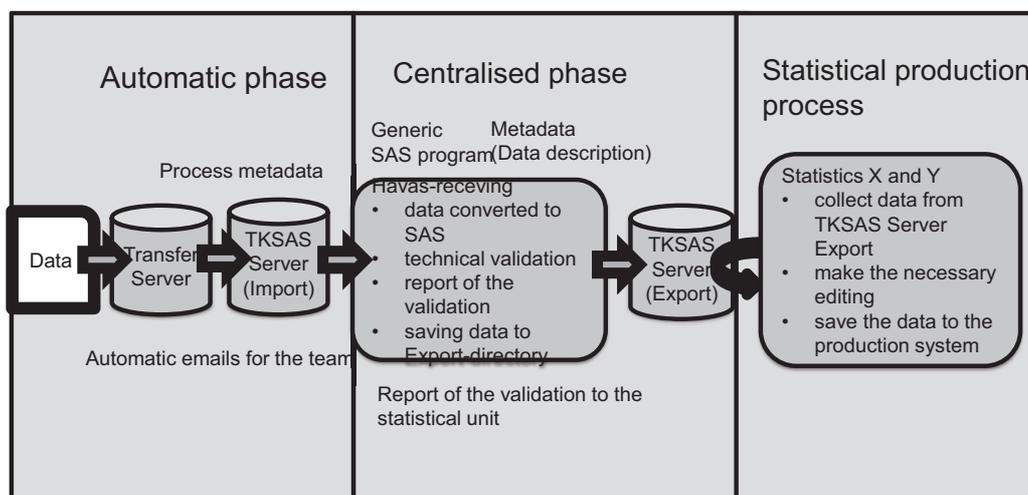


Fig. 3. Administrative data receiving system.

One of the main tasks for the team was to guarantee that there were agreements made between the official data providers and Statistics Finland. Whenever this is not the case, the agreement negotiations are started.

There are two different types of agreements that are made with data providers, *skeleton agreements* and *data acquisition agreements*.

Skeleton agreements are made with important data providers. In addition to the data acquisition, the agreements determine regular collaboration between the organisations. This collaboration usually involves annually arranged meetings between both parties. If the other party is a large data provider the meetings are held at the highest level and the general directors of both organisations take part in them. In these annual meetings important issues concerning changes or future plans in administrative data are discussed. The *skeleton agreements* are always signed by the general directors of both organisations.

Data acquisition agreements are made with smaller organizations. The agreements consist of detailed information about the data collection, such as:

- How the data are transferred.
- In what format they are sent.
- When and how often the data are either sent or retrieved.
- The contact persons for the data transfer process in both organisations.
- The content experts for the data in both organisations.

In *skeleton agreements* the data acquisition agreements are enclosures and they contain the above-

mentioned detailed information about the data transfer. The data acquisition agreements are always updated whenever any changes in data occur.

Indicator 8.9

Statistical authorities co-operate with owners of administrative data in assuring data quality.

At Statistics Finland, this indicator is not applied yet as it should be. Today the administrative data is checked technically in the data receiving process. Further checks are made by statisticians in the production process. More collaboration between the data provider, statisticians and data receiving unit (Havas-team) would be desirable. This could be arranged for instance by small projects involving administrative data quality measurements. In future these data quality evaluations should be made regularly within given intervals. This will be one of the future challenges of producing statistics.

4.2. Peer review guidelines 2013 to 2015

The Peer Review measured both the realisation of the guidelines of the CoP at Statistics Finland and gave recommendations for producing statistics. In all, the Peer Review stated 21 recommendations consisting the whole range of activities and processes in producing statistics. Below are mentioned and commented those recommendations which contain activities mainly in collecting administrative data for statistical processes.

Recommendation 8. Finnish authorities should consider the feasibility of reinforcing the legal framework to take account of the requirements of official statistics.

In particular, official statistics should be considered as one of the purposes of the main administrative registers. Reasonable advance notice should be given in the case of proposed system changes, and Statistics Finland should be consulted before changes are finalised (European Statistics Code of Practice, indicators 2.2, 8.7, 9.3, 9.4, 9.5, and 10.3).

Situation in 2016:

In Finland, the possible use of administrative data for statistical purposes is taken into account in the national legislation. According to the Statistics Act, statistical authorities must use administrative data instead of using a direct data collection. The owners of the administrative data are obliged to supply the data to Statistics Finland. When changing the legislation, there is a procedure in place according to which a statement is asked from the essential interest groups.

Improvement Action 8.

Amendments to the statistics act according to administrative registers

Statistics Finland has agreed to perform a feasibility analysis with the Ministry of Finance to identify the amendments to the Statistics Act. In this connection, it is possible to reinforce the purpose of statistical use when founding or changing administrative registers. The amendment of the European statistical act also supports this aim.

Timeline:

2015: Find out needs for change.

2016–2017: Amendments to the Statistics Act. Timing depends on the schedules of the Ministry of Finance and the Parliament

Recommendation 7. The data-sharing agreements between Statistics Finland and administrative data owners should, as far as practicable, take account of the changes that could affect official statistics (European Statistics Code of Practice, Principle 2, indicators 8.8, 8.9, and 10.3).

Situation in 2016:

At the end of 2013, Statistics Finland started a project to renew the contracts with other producers of statistics and suppliers of administrative data. There are about 50 owners of administrative data. New model contracts have been prepared and accepted by the Director General. The contracts will be signed at the Director General level.

According to the contracts, the suppliers of administrative data must inform Statistics Finland well in advance about planned changes in the content or mode

of the data delivery. In the annexes related to specific data, exact dates for data deliveries are set. The deadlines for possible changes in the following year are also agreed upon.

Improvement Action 7.

Renewing the contracts with the suppliers of administrative data

Statistics Finland will renew the contracts with the suppliers of administrative data starting from the most important data suppliers and oldest agreements.

Timeline:

The aim is to renew five agreements per year.

The renewing started at the beginning of 2014.

The most important agreements will be renewed by the end of 2016.

All agreements will be renewed by the end of 2017.

5. Next steps for further development

Centralising administrative data collection is still ongoing development at Statistics Finland. At present, the Havas team handles the data receiving process, technical validation and distribution of about 140 administrative datasets (separate files), which represents about 65 per cent of all administrative data (files) suitable for centralised data collection.

5.1. Increasing the level of automation

There are some important development steps to be taken before the remaining datasets can be included in the centralised data collection. One of these steps is to increase the level of automation. In practice, this means that, e.g. the Havas application would work without the user selecting the right metadata folder, file description, target folder and other required parameters. This development is crucial especially for the time critical data that Statistics Finland receives weekly.

5.2. Widening the range of file formats

Another step is to develop the centralised data collection system so that it becomes possible to widen the range of file formats included in the system from flat files and SAS files to XML files – and maybe also to Excel files. This is important, because some of administrative data providers want to send the data in XML format in the near future.

5.3. Developing a user interface for controlling the process

There is also a clear need to develop the Havas team's user interface that could be used to control and manage the whole process of centralised administrative data collection. The user interface would offer tools to update and maintain process metadata, contracts and contact information, for example.

At present, there are neither enough up-to-date information nor a suitable system available to follow the process of administrative data collection systematically. However, by far the bigger problem is the availability and ease of use of information, not so much that it would not exist. The target is that information systems would enable printing, e.g. weekly or monthly reports on exact timetables, arriving and received/accepted datasets and timeliness or costs of administrative data. This would mean a considerable improvement in the Havas team's possibilities to plan and develop its work. A better overall picture would also benefit the data users.

6. Conclusion

As stated earlier, today about 70 percent of the administrative data is collected by the central data receiving system at Statistics Finland. The central system has led to wider knowledge and more users of the administrative data for different statistical purposes. There are however many challenges already in the near future which reflect data collection.

1. *The changes in the administrative data* have to be known in time, so that the negotiations with data providers can minimize the effect of the changes. This is why the collaboration with the data providers is crucially important.
2. Data today is collected in various ways, for instance as *xml-message files*. The death certificate data is received today at Statistics Finland as xml-messages instead of mass data of a certain time. The same development is widening to the other data sources, for instance population data or traffic data.
3. *New data sources* are tested today. Consumer price index is calculated using big data from product or service prices.

The central administrative data collecting system will be used also in future. However, the changes in data collection and new data modes, make it necessary to develop the system more diverse and flexible, so that it can answer to the new challenges.

References

- [1] Statistics Finland ICT strategy, 2008 and 2015.
- [2] European Statistics Code of Practice, 2011.
- [3] Quality Assurance Framework of the European Statistical System (ESS QAF).
- [4] Statistics Finland ESS Peer Review (2013–2015) recommendations for 2016 to 2019.