Revision Project of the Business Register and Business Statistics in Finland

Sami Saarikivi
Statistics Finland, Helsinki, Finland sami.saarikivi@stat.fi

Abstract

In 2009, Statistics Finland started the defining phase for a project aimed at a total revision of the Business Register and business statistics. During the defining phase, descriptions of the present status and the target status were prepared for the Business Register. From 2010 onwards, the revision progressed by sub-projects to the planning and implementation phases towards its phased introduction mainly due in 2014. Overlapping and conflicting work in different statistical systems for processing business data was found to be one of the key problems in describing the present status. In the new information system following the target status, the problem will be erased by establishing a joint production database for key business statistics, where the Business Register is at the core as the basis for data. In addition, a warehouse of data on enterprises will be set up, which will operate as the distribution channel for ready unit-specific business data. The publishing processes of data collection and statistics and their tools will be harmonised during the project. The new system enables the introduction of the actual enterprise concept at Statistics Finland. In addition to the revised technology, the information system is managed with process management tools and metadata. The benefits to be gained from the revision for business statistics include elimination of overlapping work, intensification of production processes and coherence between statistics. The revision will also lead to a reorganisation of work processes and procedures between business statistics.

Keywords: coherence, data warehouse, process management, uniform production system

1. Introduction

The task of the Business Register is to serve as a basic statistical register, data source and instrument in data collections from enterprises. Statistics, as well as information services for customers, are also produced from the Business Register. The Business Register contains data on over 500,000 enterprises and corporations. Approximately 30 staff-years are spent annually on the maintenance and development of the Register and on supporting tasks in the Business Register unit. Statistics Finland spends around 77 staff-years per year on the processing of data on enterprises. Around 300 persons inside the agency have user rights to the Business Register.

The information system of the Business Register was last reviewed between 1995 and 1998. The database structure of the solution adopted then was complicated and the system has since been developed further to meet changing demands.

Pressures to develop the Business Register arise from Statistics Finland's Operational Strategy and Strategy for Economic Statistics, and from the European Union's regulation on business registers for statistical purposes (incl. EGR - EuroGroupsRegister). In addition, the agency's development programme on data collections from enterprises has set as a target that the sampling frames of individual sets of statistics should be replaced by the Business Register. If the development needs are responded to without fundamental background work the reliability and efficiency requirements of the system will be compromised. Further development of the present system without renewal of the foundations will continue to increase the demand for labour input.

No consistent procedure has been applied in the treatment of administrative data that are entered into the different business information systems at Statistics Finland.

Diverse checks at the data entry stage take unnecessary time and the efficiency of the process needs to be improved.
Overlapping work has been carried out between business statistics in defining different unit structures (e.g. establishment and legal entity) and main classifying variables (e.g. turnover and NACE code). This is partially due to the opportunities offered by individual production databases to define your own variables and to change unit-level data in your own production system.
For instance, there are also needs in the Business Structures Statistics to renew information systems, as well as the desire to integrate the systems more closely with the Business Register. This requires simultaneous or successive reviews of the Business Register and statistics on business structures.

2. Results and objectives of defining work (2009)

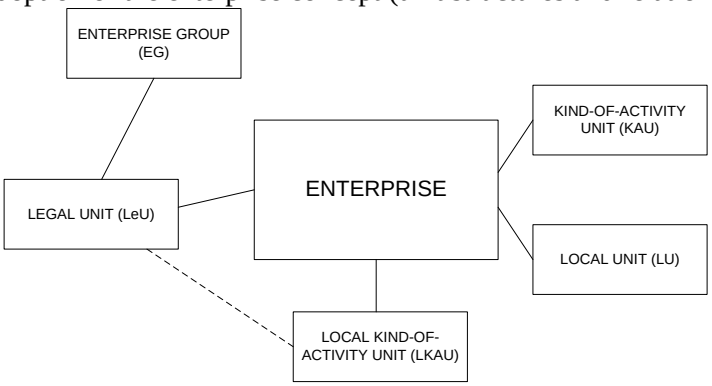
The project began in 2009 with a one-year defining phase during which descriptions were drawn of the present and the desired state of the Business Register.

Table 1. Collected results of the Business register desired state (defining phase 2009)

System part	Key development targets	Benefits to be gained
Reception of administrative data	• Utilisation of metadata • Process management • Variable editor and editing	• Intensification / Introduction of uniform practices to pre-checking of data
Direct data collections	• Uniform tools • Facilitation of responding	• Improvement of quality and intensification of activity
Databases	• Adoption of the enterprise concept • Uniform production database	• Improvement of quality • International comparability • Non-response and response burden get smaller
Application programs and processing of data	• Intensification of processing (decrease of applications) • Introduction of a process management application	• Improvement of quality and intensification of activity
Warehouse of data on enterprises	• Uniform location for data on enterprises • Basic data search • Comparison of data	• Coherence of statistics
Statistics and products	• Introduction of a unit information service (Internet) • Adoption of a chargeable target group definition service (Internet)	• Improvement of quality and intensification of activity • Improvement of customer service

As a result of the new system, Statistics Finland will adopt a new concept, enterprise unit, which will act as a central unit in the database structures (Picture 1). Implementation of the statistical enterprise unit will introduce new work phases (e.g. consolidation of turnover and maintenance of the unit structure) but will also make temporal comparability and international uniformity of the units easier.

Picture 1. Adoption of the enterprise concept (unit structures and relationships)



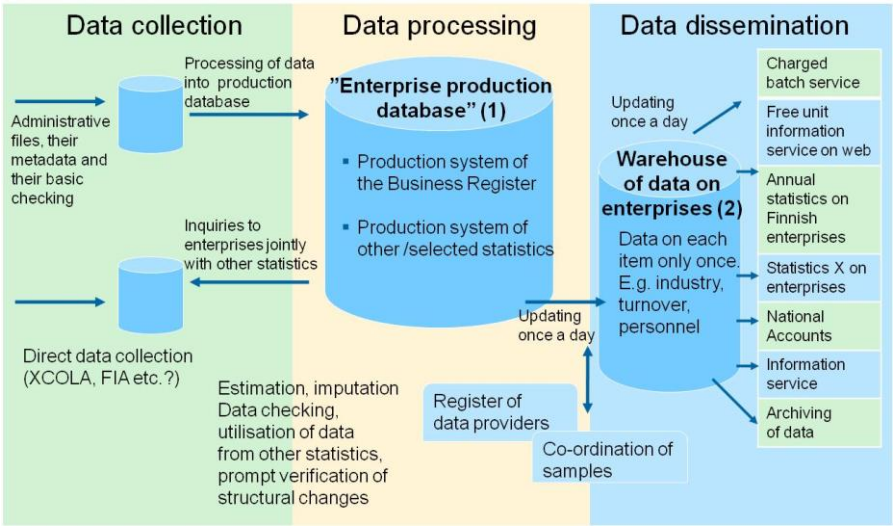
The defining work started in 2009 and the planning that followed in 2010-2014 aims at a total renewal of the Business Register and business statistics where the objective

is improvement of the coherence and functional efficiency of business statistics in the long term. The objective is to be achieved by constructing a common production database for business statistics around the Business Register and by establishing a warehouse of data on enterprises to serve as a foundation for consistent business statistics so that:

- Uniform procedures are in place for the reception and pre-processing of administrative data at Statistics Finland and these are accessible to everyone needing them.
- The Business Register forms the core of a consistent production database and statistical system for business statistics.
- Consistent business statistics are produced from the warehouse of enterprise data with uniform procedures.

Co-operation continues with the individual-based social statistics that are linked with business data.

Picture 2. The Business Register and business statistics in the consistent business data system from 2015 onwards



In accordance with the outlines drawn at the defining phase, the common production database for business statistics consists of

- the Business Register,
- financial statement statistics
- statistics on international trade in services
- commodity statistics
- regional and industrial statistics on manufacturing
- FATS
- statistical systems of the Monthly Business Indicators unit (partial link)

The selection criteria for the statistical systems was, in addition to overlapping work, usage of the same administrative data (e.g. main Tax Administration's data) as part of the production process.

The plan is to bring all main business statistics into the established enterprise data warehouse as quickly as possible after the establishment phase of the project (from 2015 onwards). When data from different statistical system are brought into the common data warehouse, consistent statistical releases can be generated from it with consistent data and consistent tools. The consistent data warehouse also makes the compilation of national accounts and further processing of business data easier (e.g. in the Research Services).

The new information system will better enable meeting the obligations imposed by the EU regulation concerning business registers. In addition, it will take into consideration

the development needs highlighted in the Strategy for Economic Statistics. Business statistics will be integrated with each other with the help of the common production database and enterprise data warehouse. The functional requirements of the charged unit-based data service of the Business Register will be safeguarded also in the future.
Further aims include improvement of the interactiveness of the Business Register, reduction of dependency on the IT department and easier upkeep.
The project will contribute to these points of emphasis in Statistics Finland's Operational Strategy:
1. We increase the comparability and combinability of statistics in order to improve their quality and our customer service.
2. Competence and services that are not expedient to sustain by ourselves are procured externally.
3. We use uniform production methods and applications.
4. We improve cost-effectiveness in economic statistics.
5. We exploit data from basic statistical registers of enterprises and individual persons widely and uniformly in statistics production.

At the project defining phase in 2009, it was estimated that the new common production database for business statistics and the production processes, including the standardised reception of administrative data, would generate a saving of 10 staff-years in labour input. The saving would come in consequence of improved efficiency once the business statistics systems have been combined.

3. Description of project work (2010-2014)

The development of Business Register and Business Statistics during 2010-2014 has been divided in modular, process adapted sub-areas. During 2010, the system for the reception of administrative data was improved and the planning of databases in terms of the production database and business data warehouse was launched. During 2011, the project continued more extensively when the planning and development of the application programmes and direct data collections required by the system started. Planning of data conversion from the old systems to the new one began during 2011. During 2012, the defining work for statistical publications and the sub-area responsible for partial integration of business trend indicators into the business statistics system started. Implementation planning has also began in addition to development work in order to be able to, for instance, arrange training well ahead of the actual introduction of the new system.

The project set has been organised into different sub-projects. The activities of the sub-projects are coordinated by the main project that is composed of the project managers of the sub-projects and the IT project managers. A fixed-term unit has also been formed for the full-time project workers for the duration of the development work. A separate steering group has been formed for the development entity. Nearly one hundred part-time and full-time employees at Statistics Finland act as members and experts in the project organisation.

The project aims at using many new and different working methods, e.g. scrum methods, that are based on short and regular meetings in order to monitor activities and recognise problems. Iterative work is also carried out between the different project phases in order to reduce the risks of incorrect specifications and resulting unnecessary planning and implementation work. We also try to include the personnel in the development work as early as possible, for instance as system testers and test users as part of the training related to implementation. In connection with implementation of the project output we also consider updating the organisational structure of Business Statistics. The renewed business statistics system also renews people's work processes and tasks will be redistributed.

4. General description of the new information system

Three main databases can be identified in the new business data system. In the first, the data collection data from administrative sources is stored in a so-called **raw data warehouse.** At this stage of the process, only technical (for instance removing duplicates and some transformations) and automatic validation have been performed on the data. The second database is the **enterprise production database** formed by the main business statistics systems that acts as the data processing place, both in terms of manual corrections of individual data and performance of mass-type and automatic batch runs. The production stage also includes selective and automatic editing and imputation stages. The **enterprise data warehouse** is the third database and operates as the distribution channel for ready unit-specific data. In the enterprise data warehouse, the data is located in a uniform place and in a uniform format in terms of classifications and unit structures. The enterprise data warehouse contains both the most recent unit-level data and unit-level data that correspond with statistical releases, and they are managed with different versions. The enterprise data warehouse is passive, so any detected errors must be revised through a separate enterprise production database. The long-term objective is to bring all of Statistics Finland's enterprise level data into the enterprise data warehouse. The data moves between the databases with the help of different ETL tools (Extract, Transform and Load) either through automatically timed or manual work phases. The following basic database principles have been considered when building the databases:

- Clarity, ease of use and understandability
- Flexibility and expandability
- Convenience of use: speed of inquiries
- Compatibility with existing data systems
- Security: access will be subject to granted user rights

The aim is to control the large information system by building a process management system that covers the entire production process. The main benefits and characteristics of the new process management application are, for instance:

- we know at what stage of the process we are
- the tasks related to the process are completed in the correct order (e.g. principle selection is only done after important data revisions)
- it can help launch automatic (e.g. mass-type data loading) and manual work phases
- system users have different tasks and roles
- process-related tasks have named persons responsible for the tasks
- task documentation is easily available
- the dates planned for the tasks are visible. If necessary, the system can send messages to persons if a critical task is behind schedule.

Another point seen as beneficial in managing the large system is maximised use of metadata. The system contains different types of metadata and it occurs throughout the production process:

- technical and material metadata (names and characteristics of variables and classifications)
- metadata related to the production applications
- process metadata in the process management application

5. Technical solutions

New and congruent tools in terms of the technical environment and application development are used in building the new information system. For the databases, Microsoft's SQL Server 2012 (OLAP with SQL Server Analysis Services and SAS Enterprise Guide as a cube browser) is used as the application platform. Reporting is

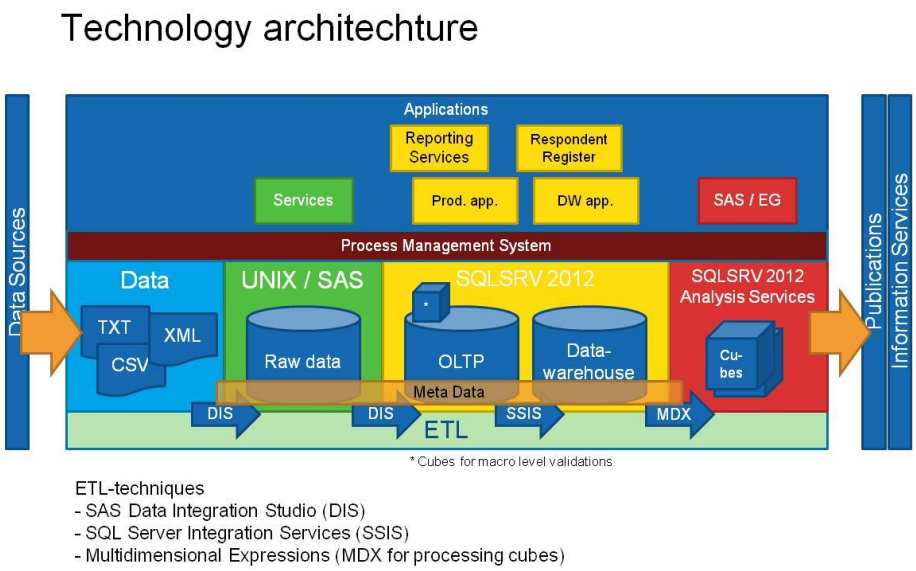
mainly carried out with SQL Server Reporting Services. The used ETL tools are SQL Server Integration Services for database to database transfers and SAS Data Integration Studio for raw data loading. The following tools have been chosen for application development solutions:

- Platform Windows w/ Microsoft .NET Framework 4.0 (programming language C#)
- Graphical User Interfaces with Windows Presentation Foundation (WPF), Microsoft Prism 4.0 as MVVM Framework
- Data Access with Microsoft Entity Framework: Code First (an ORM approach) and SQL Server Stored Procedures

The process management system is developed as an individual application at Statistics Finland and it is based on the BPMN/XPDL standard. The system requires a diagram that defines the flow and the tasks to execute and you can execute, for instance, Microsoft .NET assemblies, SAS procedures, database procedures, Windows executables and Unix scripts.

The metadata system is also Statistics Finland's internal system. There are different tools (variable editor, classification editor and process editor) related to editing metadata.

Picture 3.



6. Conclusions

There are several reasons for the renewal of the Business Register and the statistical production system. Diverse national regulations and statutes, and Statistics Finland's own Operational Strategy demand renewal of the production system. Technically the usable age of the Business Register is also coming to an end. In addition, divergent procedures and overlapping work have caused inefficiency in different business statistics. In order to solve problems and ensure a more efficient production process, the key business statistics form a uniform production system with the Business Register at its core. In terms of data collection, uniform procedures are used both in receiving administrative data and in direct data collection. Thanks to the uniform production database, the revising of business data is centralised and interdependence is taken into account. One data warehouse helps produce more coherent statistics and the business data is found in a uniform place with uniform classification variables and unit structures. The system is managed with process management tools and with maximal utilisation of metadata. The system is also built using up-to-date technology.

At this stage of the project, key sub-projects are being finalised and gradual implementation of the system is being prepared.