Proceedings of MSP2009
Modernisation of Statistics Production 2009

Standardisation for improvements in Statistics Norway

Rune Gløersen[1] and Hans Viggo Sæbø[2]

Statistics Norway has been running a programme on standardisation of processes, systems and methods (abbreviated FOSS). FOSS can be regarded as a business reengineering project based on our quality work, strategies and international trends.

The main objective of FOSS has been to improve quality and efficiency in production of statistics. Standardisation will also reduce risks and facilitate internal rotation. FOSS also covers quality assurance and human resources and organisational development supporting standardisation.

Previously planned development activities have been reviewed and fitted into this framework. As a basis for the standardisation work we have used a detailed process description or business process model or value chain. All projects are related to this model. The most important projects are related to the development of corporate framework systems.

In order to meet the challenges of intensified development, and the continuous need to provide effective IT operations and maintenance, IT in Statistics Norway has been reorganised and a new governance model is under implementation.

1. Towards a coherent statistics production system

Statistical institutes' long history in IT driven statistics production, combined with the independence of the statistical domains, resulted in the traditional stove piped production systems. For many reasons, most of the statistical institutes are currently working on a complete restructuring of the statistics production. Often, the developments start as local initiatives. Even though such bottom-up initiatives seem to be promising for a wider corporate usage, there is a need for a top-down approach to develop a coherent statistics production system.

The work in FOSS has shown that some important preconditions must be in place in order to meet the objectives of creating a more coherent statistics production system. We need

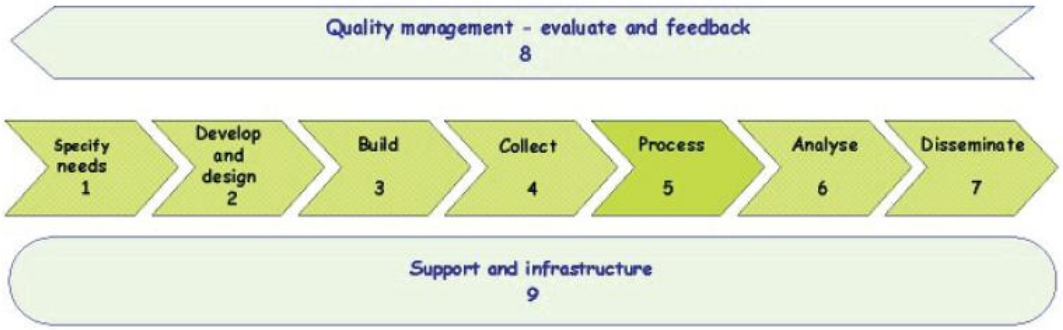
- a well defined business process model to describe how we perform our work
- an information model defining data and metadata involved in all processes
- an IT system supporting the processes, based upon standard interfaces defined by the process and metadata models, modularised to be flexible to technological change and change in user needs

Our Business Process Model has been developed and accepted (Statistics Norway 2008). The BPM is an adapted version of the BPM used in the statistical institutes of New Zealand and Sweden. From the high level shown in figure 1, another two levels have been described. Several

[1] Director of IT and statistical methods, e-mail rgl@ssb.no
[2] Director of Planning and finance, e-mail hvs@ssb.no

projects have been launched in order to build more generic systems, according to the outlined processes.

Figure 1. Statistics Norway's business process model

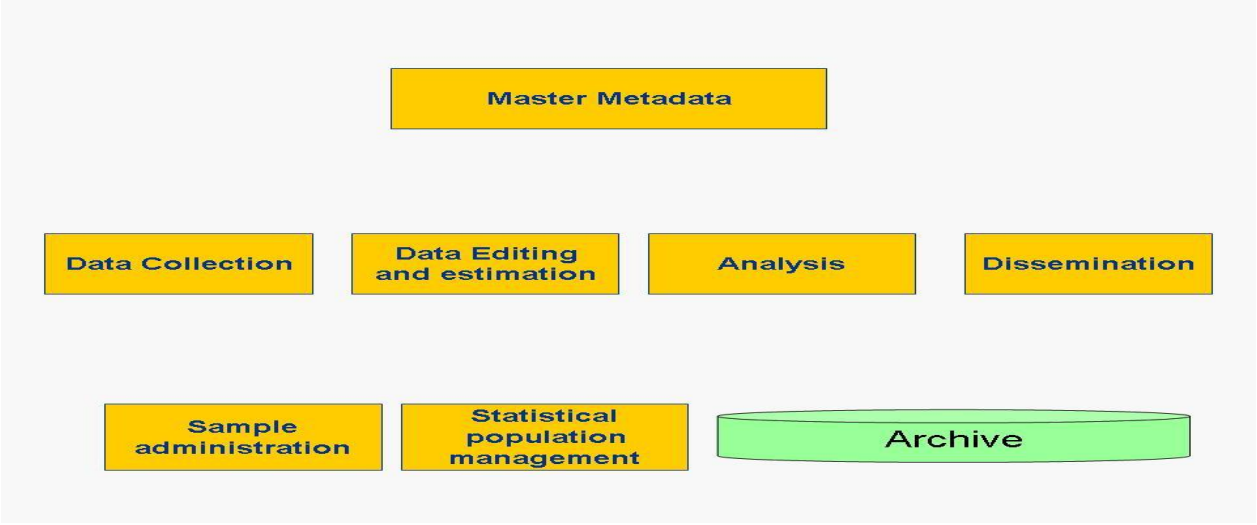


Important projects that are related to the development of corporate framework systems include:

- Check-list and indicators for data collection
- System for interview surveys
- Coordination of samples for business surveys
- A framework system for data editing and estimation
- A coherent metadata systems environment
- Development of a geo database for ground property maps
- Improving data archiving and secondary use of data for researchers
- Extensive developments in dissemination and redesign of website
- Overview of guidelines and handbooks
- Quality assessment covering all statistical products and processes

Infrastructure comprises metadata and administration of samples and stastistical populations. We will in the following briefly describe our developments and objectives grouped according to figure 2, showing families of applications serving most of the business processes used for the production of primary statistics.

Figure 2. Issues of standardisation in Statistics Norway



1.1. Metadata

The key factor in order to build a system that is coherent based upon interaction between system modules is that the modules must be metadata driven. Statistics Norway developed a metadata strategy in 2005 (Sæbø 2005). One important issue to begin with was to establish a common understanding by defining key concepts linked to metadata. In addition, the roles and responsibilities for the core metadata products and activities were put in place. The strategy advocates a stepwise development, building on already existing metadata sources. The standardisation of metadata will establish the foundation for interaction between the different sources, and common interfaces for access to the metadata should be developed.

Currently, Statistics Norway has established the following master systems for managing metadata

- The product register, defining the statistical products
- The "About the statistics" content management system, describing the statistical products
- The variable database defining and describing statistical variables
- The classification database managing statistical classifications and standards
- The input data definition database, describing data on arrival at Statistics Norway
- The data documentation system describing data in the data archive

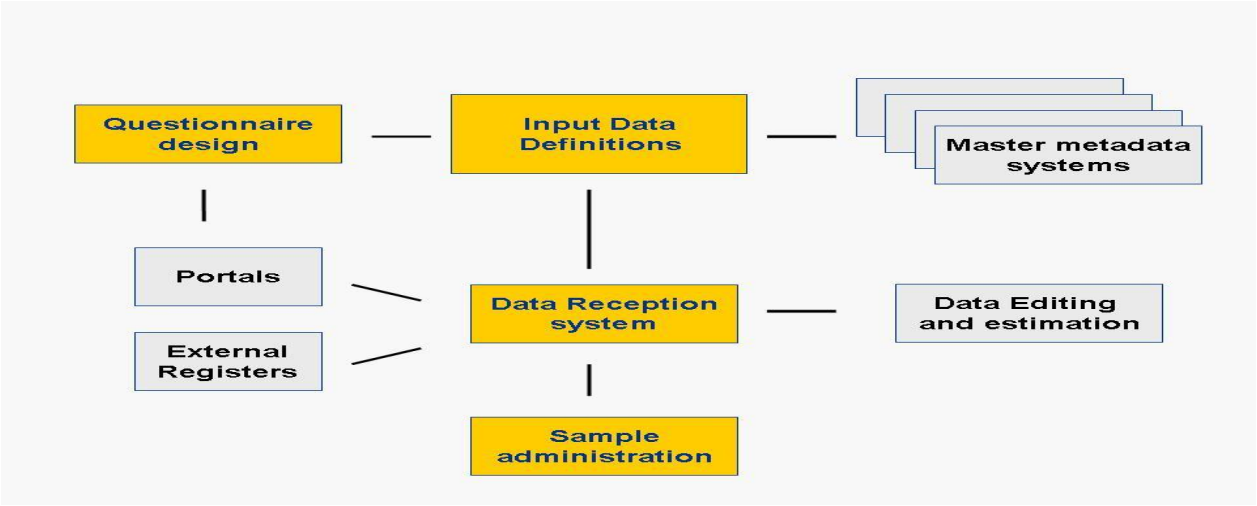
All these sources will be interconnected via their attributes, and a common user-interface will provide seamless access to the metadata. A number of the connections and interfaces have been implemented, but further developments are ongoing or planned. See also Hustoft and Linnerud (2006 and 2008).

The main challenge is to establish the necessary interfaces between the master metadata systems, and the systems used throughout the statistics production. Some interfaces are already in place. However, the interfaces must cover all exchange of information needed, and they must provide two-way communication, for instance to allow for updates of metadata when appropriate. This is necessary in order to meet the overall objective of updating a single piece of information once, and reuse it wherever and whenever it is needed.

1.2. Data collection

The main elements of our data collection are briefly shown in figure 3..

Figure 3. Main elements of data collection in Statistics Norway



Electronic data collection is currently managed through three main channels. These are the Kostra portal for municipalities, the Idun portal for businesses and the Altinn portal which is the common initiative by the Norwegian government institutions. Our aim is to consolidate all these channels into one, by using the Altinn portal only. However, for various reasons we need to keep a possibility for direct reporting to Statistics Norway for a while.

All the necessary metadata needed to provide sufficient reporting facilities to respondents are created and maintained internally, but they are exchanged with external portals like Altinn to serve as the basis for generating questionnaires and to define data flows through the Altinn portal. Relevant metadata added for use in external portals, like when creating questionnaires in the Altinn portal, are imported into our input data definitions database. The reason for this is to keep control and consistence of the metadata for further reuse in the internal production process. The Input definition database is closely connected to the Variable definition database and to the Classification database.

We are aiming for best possible integration between the Input data definition database and the questionnaire design tools used. We still use tools specific to the different input channels. However, this is about to change. We are aiming at one tool in addition to the Altinn questionnaire design environment. This tool will also take up the necessary functions to provide a end-user interface to the Input definition database.

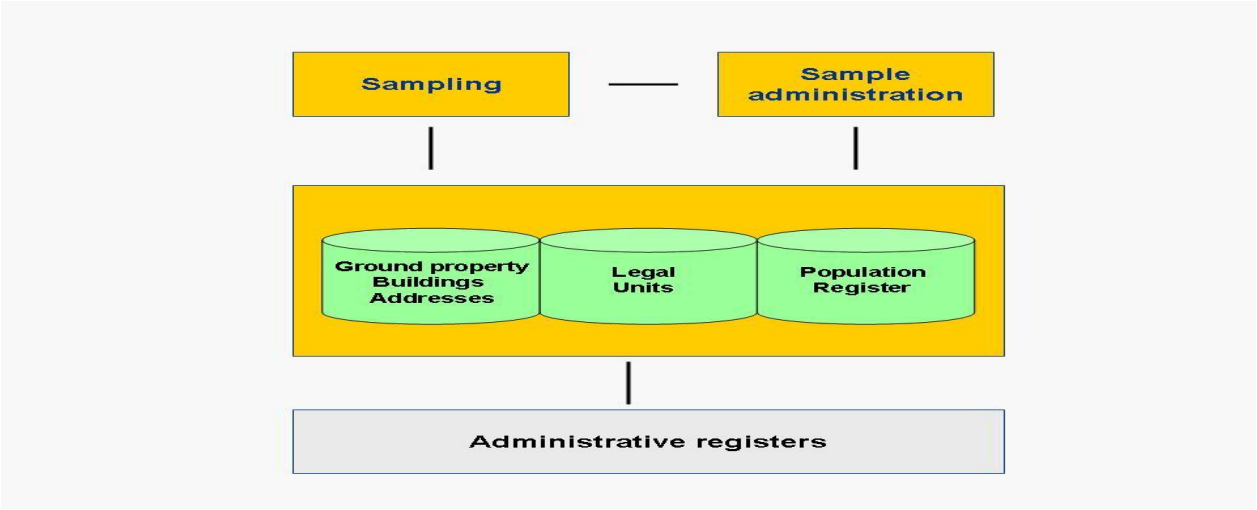
Besides the Input data definitions database, the Central data reception system forms the other cornerstone of the consolidated data collection system. All data received will be made available for further internal processing from this system. This means that data collected from electronic questionnaires or using file extracts from internal systems in businesses data from paper questionnaires, after having gone through scanning, and data from registers will be processed and further distributed from the Central data reception system.

A complete new system for managing field interview work is under development and will be released next year.

1.3. Statistical population management

A simplified view of statistical population management is shown in figure 4.

Figure 4. Statistical population management in Statistics Norway



The main challenge is to maintain the quality of the statistical populations themselves. That is to keep them aligned with the changes in the administrative registers, but to maintain the changes according to the statistical needs. Recently, efforts have been put into the management of the register of Ground property, buildings and addresses including dwellings. The developments include management of map information and linkage with geo-referenced information.

The complete coverage of the administrative registers gives Statistics Norway a unique advantage in terms of controlling the population. Given this advantage, we can still improve our systems for sampling and sample administration, both in terms of standardisation and consolidation, and in terms of functionality.

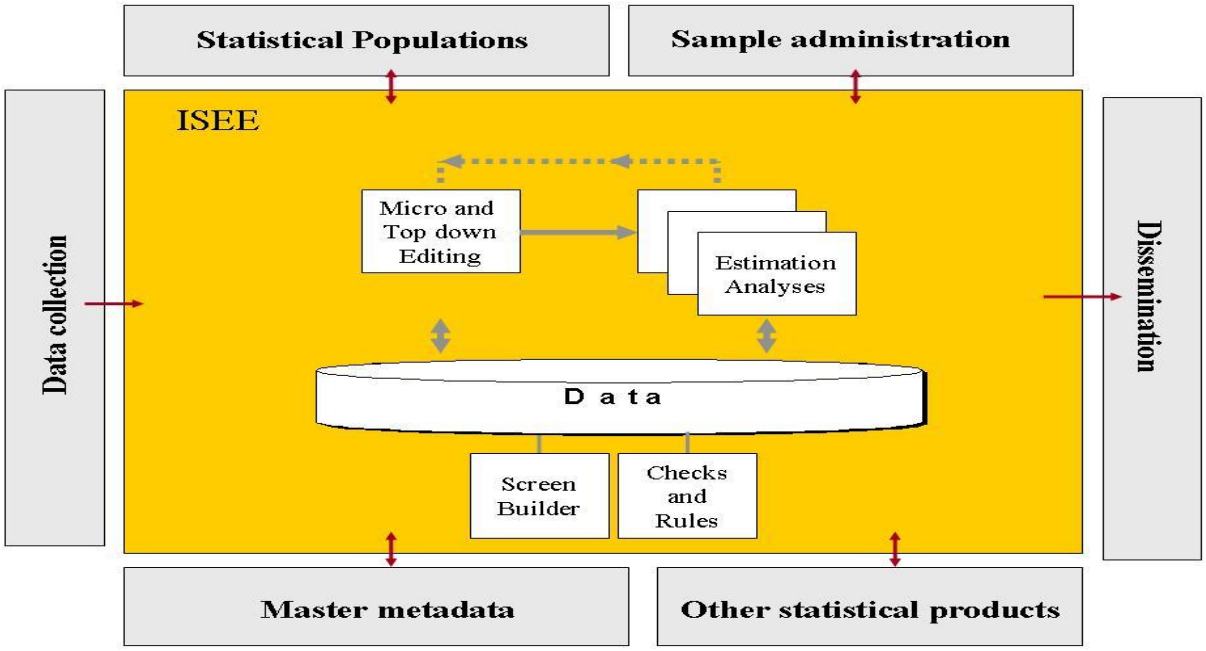
The sample administration system has been extended to serve both the Legal unit register and the Population register, i.e. samples of both persons and businesses, and will be integrated with the Central data reception system. The system is within certain limits flexible to specific needs. The term Sample administration is somewhat misleading, because the system should be used to administer respondents in a wider perspective. For instance, when collecting data from all municipalities, all necessary information on the respondents side should be kept in this system. In such a case the system can be said to correspond to and be similar to a Customer Relation Management system. Therefore, the system will also be used when, for various reasons, we need additional information from the respondents.

The sampling system covers all statistical populations, and is currently extended with functionality to manage the response burden on businesses by monitoring the burden on each business and by functionality to spread the burden using for instance twins in the population.

1.4. Data editing and estimation

The framework system for data editing and estimation is briefly outlined in figure 5.

Figure 5. System for data editing and estimation



The Integrated system for editing and estimation (ISEE) represents one of the most promising approaches to becoming a widely used framework system within the statistics production in Statistics Norway. It is basically designed to help the implementation of the so called top-down approach to data editing. It is also designed to help the statisticians to become self reliant in implementing new statistics and maintaining the existing statistics in production. For the moment, its main usage is for business surveys.

The system consists of two main building blocks which can be used together or separately

- The *Dynarev* which is a system for micro editing that uses the input data definitions to allow end users to build screens and checks for editing.
- Connected applications integrated with Dynarev. For the moment we offer two applications, *Pris* for estimation of price-indices and *Struktur* for estimation of totals and variance. The connected applications integrate in their respective ways statistical macro editing with the traditional micro editing in Dynarev. In connection, Dynarev is used to look up units based on indications of errors recognised elsewhere.

Further applications are planned for integration within ISEE. We are currently investigating whether and how ISEE can be extended for register-based statistics (Zhang 2009). We also plan to integrate a generalized editor to browse and edit datasets, to be able to cope with large datasets and not only limit changes to be made unit by unit.

The ISEE is integrated with the statistical population registers, the sample administration system, in addition to building directly upon the Input data definition database. Furthermore, ISEE links to common data sources internally and provides a common interface for uploading final results and tables.

ISEE is currently used for approximately 60 statistical products.

1.5. Analysis

Except for those using the macro applications in ISEE, no initiatives have yet been planned to provide a standardised environment for analysis and secondary use of data. Statistics Norway uses SAS as a tool for tabulation and analysis, but except for the harmonisation that follows the use of a common tool, almost no coordination has been done to harmonise the use of tools and work processes in this area. Within each statistical domain, there exists a library of SAS programs used for different types of analysis. Whenever a new task occurs, SAS programs are reused or slightly altered to meet the requirements.

When we have finalized some of the ongoing projects within the other fields of standardisation, a project will be established to investigate possibilities for harmonisation in the field of analysis and secondary use. Most likely, we will investigate a possibility for building domain specific and also cross-sectional data warehouses. This means that we will provide a framework allowing for implementation of a small number of different warehouses established, maintained and integrated with common sources of micro data and metadata based upon common concepts, definitions and rules.

1.6. Dissemination

The website of Statistics Norway, www.ssb.no, is currently undergoing a complete modernisation. The information architecture has been remodelled, and the website will be modularized and layered to allow for multiple usage, i.e. serving both persons and other web-sites, blogs etc. in a more adaptable way. The statistical databank will be the master source for all tabular data presented on the website. This means that table snap-shots should be derived in a common way from the statistical databank, and not produced separately as fixed table extracts. To meet this requirement, the existing statistical databank must also be provided with a layer of services providing both additional functionality, and standardised access to the data from end users and end user applications.

The statistical databank will also be the source for exchange of multi dimensional data using the SDMX format.

Due to the other framework applications initiatives, the internal process of providing statistical results and analyses for dissemination can be more automated. Today, uploading is regarded as rather cumbersome. This is not necessarily due to the upload procedure itself, but the need to tailor interconnections with the huge number of statistics production systems, i.e. the need to train a relatively large part of the staff on how to upload tables for dissemination.

1.7. Data archiving

Only a part of the total amount of data stored in Statistics Norway is archived in the Data documentation system. It is mainly the final, cleaned micro data files that are stored and

documented. In recent years, also raw data is stored. Data in databases, or data kept in formats specific to certain tools are not kept in the archive, and are therefore subject to ad hoc means of documentation. The main objective of the data archive is to store the source micro data in a technologically independent format for secondary and historical reuse. The conditions for archiving are described in a user guide.

One of our objectives is to control and document more of the data kept for statistics production. We also want to establish better links between the metadata and the data, i.e. provide easier and more flexible access to micro data. This will serve researchers in a better way.

We are investigating upcoming standards for structuring and documenting archived data. The DDI (Data Documentation Initiative) v3 seems very promising in this aspect, especially when the standard has taken up means to support data lifecycle management.

1.8. IT architecture principles

The outlined coherent and modular statistics production system requires equal flexibility and capability of the supporting IT system. For the moment, the objectives of the Service Oriented Architecture are driving the developments within IT. Statistical institutes, processing vast amounts of data, must implement the SOA principles with particular attention to the specific needs of the business.

Statistics Norway has agreed upon the following principles for the IT architecture:

- IT-solutions must be built upon standard methods, a standard infrastructure and be in accordance with Statistics Norway's business architecture
- IT-solutions must support Statistics Norway's business processes i.e. IT-business alignment
- Statistics Norway shall use open standards in IT and information systems.
- Our IT-solutions must be platform independent and component based, shared components must be used wherever possible
- It must be possible to create new IT-solutions by integrating existing and new functionality
- Our services must have clearly defined, technology-independent interfaces
- Distinguish between user interface, business logic and data management (layered approach)
- End-user systems must have uniform user interfaces
- Store once, reuse may times (avoid double storage)
- Data and metadata must be uniquely identifiable across systems

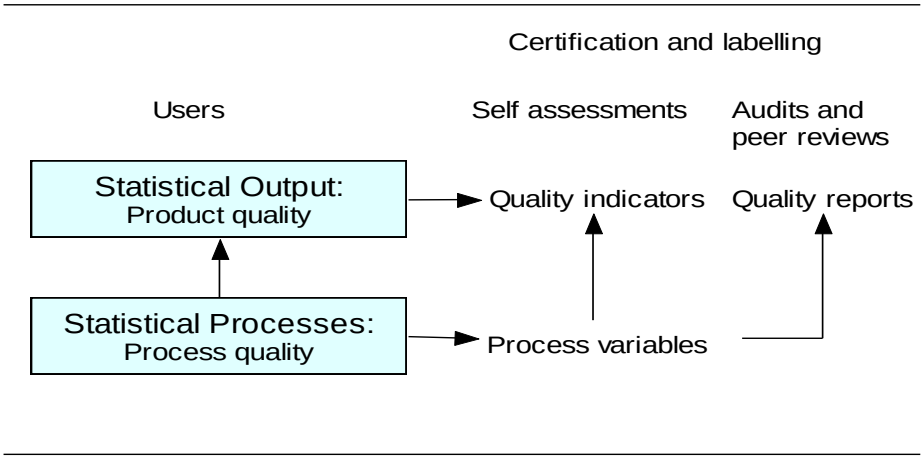
The architecture principles will be taken care of by a specific group of systems architects, each of them paying respect to a defined part of the statistics production system.

2. Improving process quality

Quality work in statistics has gradually developed from assessing and reporting on output quality to a process-oriented approach, following the whole statistical business process. The development has been facilitated by some basic and common principles of quality management. To meet the objectives outlined in the FOSS programme is a precondition to improve the over all quality of the work processes.

Sæbø (2007) has discussed measurement of process variables within the framework of quality assessment, an important technique for continuous improvement. Figure 6 illustrates the main links between process variables and other data quality assessment methods. Eurostat (2007) provides an overview of such methods.

Figure 6. Data quality assessment methods



What has often been lacking is a systematic approach to identifying and measuring such variables. Work on establishing quality assessment systems including self-assessments and audits, is important to facilitate this. As a part of the FOSS programme, Statistics Norway carried out a quality assessment by using the DESAP self-assessment questionnaire, see Næs (2008) and also Eurostat (2003). All of our approximately 200 statistics were covered in 2008. Even if this has been a self assessment, people responsible for the different processes have been supported by central quality or methodology staff in filling out the questionnaires. Main areas of improvement identified comprise data collection and editing. In both these areas more study of process variables are recommended. These areas are also emphasised in our FOSS programme.

Quality self-assessment is well suited to give a snap-shot of the quality in different areas at a certain moment, which can be summarised to a general view of the statistical production. However, it is costly and will therefore be conducted infrequently. But, in a very heterogeneous environment, quality self-assessments are perhaps the only systematic approach to follow. It is challenging to run quality audits in an extremely stove-piped production environment. You will spend a lot of resources in examining the different production systems themselves, and there is a great risk of bias in the judgements to be made. Therefore, implementation of a quality management system and standardisation of processes and systems are closely related.

Systematic monitoring of the process variables that have a high influence on resulting product quality (such as accuracy, timeliness and comparability of statistics) is necessary to improve the processes. Examples of process variables that have been used by statistical institutes are measurements of interviewer performance, non response of different types, costs and use of time for different processes. Monitoring of the necessary variables will be implemented in the

framework systems, providing quality information for the statistics production supported by the system, systematically and comparable between products.

3. Governance

As an outcome of the reorganisation of the subject matter departments in 2008, and also with respect to the ongoing work on standardisation of work processes, Statistics Norway is about to implement necessary roles and procedures to establish a sufficient and effective central service provider organisation. The basic principles have been defined within an order-delivery model, where the newly centralised Department for IT and statistical methods is responsible for deliveries to the four subject matter departments and the departments for data collection, human resources and communication, planning and finance, and research.

The two main building blocks in this model are

- To establish Service Level Agreements (SLA) with all other departments
- To establish Portfolio Management for all large projects

The objective of these measurements is to be able to control the resources, and especially to improve the ability to prioritise the use of scarce resources like IT developers and statistical methodologists, to promote development and standardisation.

3.1. Development of Service Level Agreements

This year we will finalize the work with SLAs covering all IT services. This has been prioritised because of the centralisation of personnel. The SLAs are structured in three levels,

1. The formal agreement between parties, quite similar in all agreements
2. The service library with descriptions of the services provided
3. The agreed details within each service, like a list of products, expected periods of peak work or planned readiness for specific products or tasks etc.

Important tasks when establishing SLAs are

- To establish a complete inventory of all existing and supported products
- To designate the ownership of these products
- To designate the responsible service provider for these products

Next year we will also include statistical methods services in the SLA. It is our assumption that work carried out on a regular basis within the department for IT and statistical methods, whether it is IT or statistical methods, should be described as services and be documented and agreed upon in the SLAs. The SLAs will distinguish between work to be carried out within the normal application lifecycle management, and large or complex developments that need high level attention and prioritisation and therefore will be lifted to the portfolio management regime.

3.2. Portfolio management

All large development projects are prioritised and followed up within a system of portfolio management, described by Sæbø (2009). To ensure standardised solutions is one of the objectives of portfolio management in Statistics Norway.

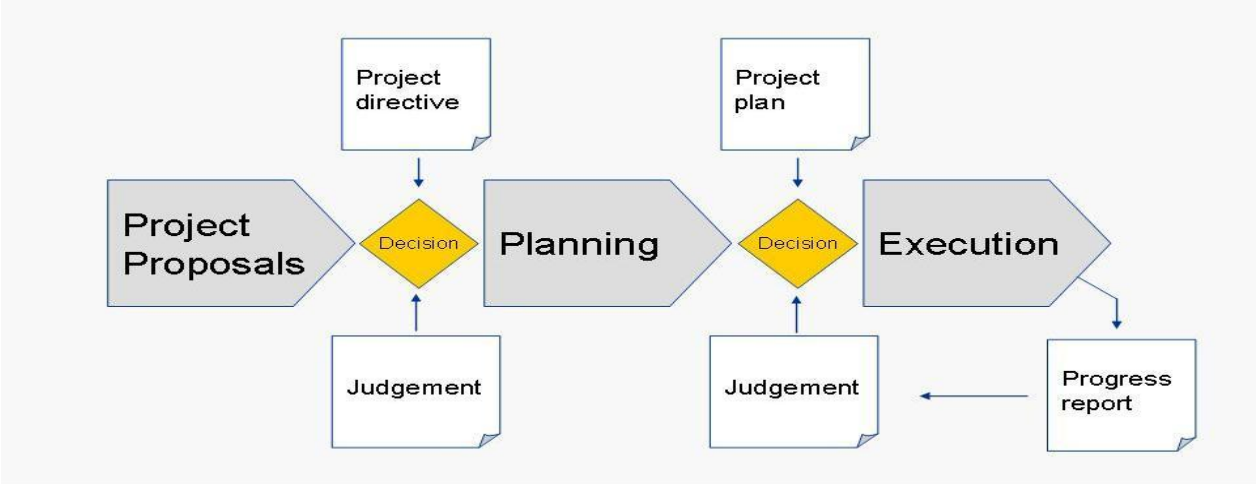
The purpose of portfolio management is to improve quality and efficiency by better prioritisation of projects based on systematic information. In addition to the normal information on projects included in project mandates and plans, a set of key parameters that can be measured or given values (points) as a basis for prioritisation has been developed. In addition to these parameters with summary points, final prioritisation has to be based on judgements including other information such as the purpose and cost of projects, available resources and competence, dependencies between projects and the totality of the project portfolio.

Portfolio management includes the following tasks

- To decide upon project proposals by judging them according to the overall priorities of the organisation
- To ensure that the project plans include sufficient quality in the estimates of the resources needed
- To follow up the progress of the projects in execution, to be able to balance the portfolio during prioritisation and incorporate new projects for execution

The phases of portfolio management are shown in figure 7, following the life cycle of the projects.

Figure 7. Phases of portfolio management



The parameters used for prioritisation of project proposals are:

Parameters which are given values on a scale
Coherence with strategies and plans
Contribution to standardisation of processes
Laws and EU-regulations
Reduced risk in Statistics Norway
Quality (relevance, accuracy, timeliness and punctuality, coherence and comparability, accessibility and clarity)
Response burden
Benefit/gain

Other parameters to be taken into account
Purpose
Costs
Competency requirements
Complexity
Links with other projects, totality of the project portfolio

Judgements have to be used for final prioritisation carried out by the top management, but these judgements are be based on systematic information about the projects. Projects are queued in front of each decision point. No projects are brought to the next stage unless there are sufficient resources available for planning or execution.

Portfolio management in Statistics Norway is partly based on experiences from the FOSS programme. From June 2009 the portfolio of projects in the FOSS programme were handed over to the common portfolio management and the FOSS programme as such was completed.

4. Conclusion

Standardisation and integration is the key to both quality and efficiency improvements in a statistical institute. This is the case even if quality and efficiency may be regarded as opposites in some cases. However, without efficiency gains there will be no resources left for new developments and improved quality for keeping statistics in accordance with new demands. In Statistics Norway common efforts for improvement have comprised both work on quality, standardisation and improved governance by reorganisations and new management systems and tools such as portfolio management. This approach is in accordance with visions expressed by many statistical agencies. The outcome of the improvements will also strengthen the capabilities for external cooperation, for instance within the public sector according to the long term objectives of e-government and within the statistical society to support the objectives of the European Statistical System (ESS).

References

Eurostat (2007): *Handbook on Data Quality - Assessment Methods and Tools.*
Published on http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/quality_reporting

Eurostat (2003): *Checklist for Survey Managers (DESAP).*
Published on http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/quality_reporting

Hustoft, Anne Gro and Linnerud, Jenny (2008): *Managing metadata in Statistics Norway*.
Documents 2008/6
Published on: http://www.ssb.no/english/subjects/00/90/doc_200806_en/doc_200806_en.pdf

Hustoft, Anne Gro and Linnerud, Jenny (2006): *Statistical Metadata in Statistics Norway*.
Documents 2006/5.
Published on: http://www.ssb.no/english/subjects/00/doc_200605_en/doc_200605_en.pdf

Næs, Peder (2008): *Kvalitetsgjennomgang, DESAP*.
Interne dokumenter 2008/33 (Norwegian only)

Statistics Norway (2008): *Statistics Norway's Business Process Model.* Documents 2008/17
Published on http://www.ssb.no/english/subjects/00/90/doc_200817_en/doc_200817_en.pdf

Sæbø, Hans Viggo (2009): *Quality and standardisation of processes.*
Paper presented at ISI 2009.
Published on: http://www.statssa.gov.za/isi2009/ScientificProgramme/IPMS/0886.pdf

Sæbø, Hans Viggo (2007): *Measurement of process variables within the framework of quality assessment.* Paper for the ISI Conference 2007, Lisboa, 22-29 August 2007

Sæbø, Hans Viggo (2005): *Metadata strategy in Statistics Norway.*
Eurostat Metadata Working Group meeting june 2005
Published on: http://www.ssb.no/english/subjects/00/doc_200508_en/doc_200508_en.pdf

Zhang, Li-Chun (2009): *Ideas on editing of statistical registers.* Paper for the Work Session on Statistical Data Editing, Neuchâtel, 5-7 October 2009