

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)

**ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Luxembourg, 9-11 April 2008)

Topic 2 (iii) Metadata and the statistical cycle and Implementation

CASE STUDY: STATISTICS NORWAY

Submitted by Norway¹

| | |
|--|---|
| Revision History..... | 2 |
| Organization Details..... | 2 |
| 1. INTRODUCTION..... | 3 |
| Metadata strategy | 3 |
| Current situation..... | 3 |
| 2. STATISTICAL METADATA SYSTEMS AND THE STATISTICAL CYCLE..... | 4 |
| 2.1 Statistical business process | 4 |
| 2.2 Current system(s) | 4 |
| 2.3 Costs and Benefits..... | 5 |
| 2.4 Implementation strategy | 6 |
| 3. STATISTICAL METADATA IN EACH PHASE OF THE STATISTICAL BUSINESS PROCESS..... | 6 |
| 3.1 Metadata Classification | 6 |
| 3.2 Metadata used/created at each phase..... | 6 |
| 3.3 Metadata relevant to other business processes | 7 |
| 4. SYSTEMS AND DESIGN ISSUES..... | 7 |
| 4.1 IT Architecture | 7 |
| 4.2 Metadata Management Tools | 7 |
| 4.3 Standards and formats | 7 |
| 4.4 Version control and revisions..... | 7 |
| 4.5 Outsourcing versus in-house development | 8 |
| 4.6 Additional materials | 8 |
| 5. ORGANIZATIONAL AND WORKPLACE CULTURE ISSUES | 8 |
| 5.1 Overview of roles and responsibilities | 8 |
| 5.2 Metadata management team..... | 8 |
| 5.3 Training and knowledge management | 9 |
| 5.4 Partnerships and cooperation | 9 |
| 5.5 Other issues | 9 |
| 6. LESSONS LEARNED | 9 |

¹ Prepared by Anne Gro Hustoft (agt@ssb.no) and Jenny Linnerud (jal@ssb.no).

METIS COMMON METADATA FRAMEWORK (CMF)

PART C CASE STUDY

NORWAY / STATISTICS NORWAY

| Revision History | | |
|------------------|--------------------|-----------|
| Date | Section(s) updated | Comment |
| 29/02/2008 | | Version 1 |

Organization Details

Organization Name Statistics Norway

Number of staff 1000 employees. Approximately 550 in Oslo and 450 in Kongsvinger.

Contact person (for Metadata) Jenny Linnerud
Senior advisor
jal@ssb.no
+47 21 09 45 22

Organization structure

(see appendix 1)

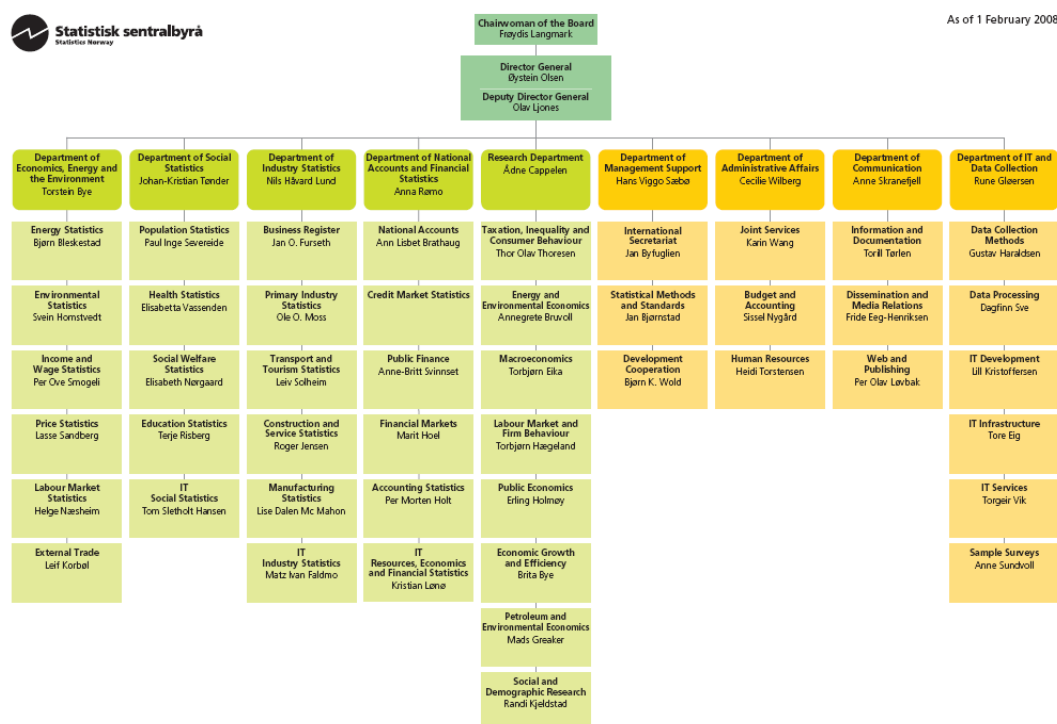


Fig.1. Organization structure (see appendix 1)

1. INTRODUCTION

Metadata strategy

Statistics Norway has in the course of time developed many different metadata systems. This led to the same information being stored several times in several places making the availability of updated and consistent information difficult. In recent years, there has been a strong focus on the need to link existing systems and a requirement that new metadata systems should not be built in isolation. To facilitate this, Statistics Norway developed a metadata strategy, which was approved early in 2005. The strategy focuses on establishing a conceptual framework, clear roles and responsibilities, and a stepwise development involving integration and linkage of systems.

Current situation

The objective of Statistics Norway's current work on metadata is to develop an integrated metadata system that will contribute to effective statistics production and dissemination, in addition to improved quality of statistics. Different metadata systems are being linked together making the metadata more easily accessible for all users. Metadata should be updated only in one place.

Metadata projects completed after 2005 include:

1. Documentation of key metadata concepts
2. Metadata portal on the Internet
3. Variables documentation system
4. About the statistics as a content management system instead of just a document.

Current metadata projects include:

1. Improved editing tool for our classification database
2. Analysis of the end-to-end creation and re-use of metadata in one production cycle for one statistic.
3. Continuing development of a master metadatabase for questionnaires.
4. Continuing development of a service library for master metadata systems.
5. Continuing development of a metadata intranet portal.
6. Improved plan system for projects, products and processes.
7. Improved access to micro-data for researchers (About the data collections).

2. STATISTICAL METADATA SYSTEMS AND THE STATISTICAL CYCLE

2.1 Statistical business process

We are currently discussing a statistical life cycle model based on the work done by Statistics New Zealand and by Statistics Sweden. We have also attended one Statistical Open Source workshop hosted by Statistics Netherlands in 2007 on this topic.

The phases we are currently discussing are as follows: 1. Need, 2. Plan and Design, 3. Develop, 4. Collect, 5. Process, 6. Analyse and 7. Disseminate.

Comparison with the CMF statistical survey life cycle model: CMF 3. Build is currently part of our 3. Develop while CMF 8. Archive and CMF 9. Evaluate are currently sub-processes for each of our processes.

2.2 Current system(s)

Datadok - File descriptions (implemented)

We document all permanent archive data files in our file documentation database Datadok. The database was built in 1998 but wasn't mandatory until 2002.

Vardok - Variables documentation system (implemented)

The overall purpose of the variables documentation system is to document variables in a central location, accessible by all, and to function as a tool for harmonising names and definitions.

There is a two way link between Vardok and Datadok (file descriptions database), a one-way link from Vardok to Stabas (standard classifications database), a two way link between Vardok and StatBank (dissemination database), a two way link between Vardok and Metadb (system for documentation of event history data) and a one way link from About the statistics, About the data collections and the statistical metadata portal to Vardok, via web services.

2006 was the last year in the development phase for the Vardok-project.

Stabas - Standard classifications database (implemented)

The overall aim of Stabas is:

- To make work with and the use of standards simpler and more efficient
- To ensure systematic use of standards across different statistical areas

One main task is to make approved versions of the central statistical classifications available in a database system where they can be taken out at different aggregation levels, together with texts in different languages and relevant documentation, and where the classifications can be exported to other IT tools.

2004 was the last year in the development phase for the Stabas-project. However, we are now planning an improved editor tool.

Service library for metadata systems (under development)

The purpose of this project is to

- Create a library of services for the master systems Vardok, Datadok, Metadb and Stabas.
- Define a framework for the description and formulation of SSB's metadata based on international metadata models (e.g. Neuchâtel) and standards (e.g. ISO/IEC 11179).
- Investigate how RDF (Resource Description Framework) can be integrated into SSB's data communication.

The project began in 2005 and will end in 2008.

Metadata portal (implemented on Internet, under development on intranet)

The overall purpose of the metadata web page is to make Statistics Norway's metadata systems more accessible and easier to use. Both internal and external users will get easier access to the metadata by displaying the contents of these systems in a common web page. The project began in 2005 and will end in 2008.

Metadb - metadatabase for event history data (implemented)

Metadata for FD-Trygd (Social security database) and NUDB (Norwegian national Education Database).

FD-Trygd: details on demography, social conditions, social security, employment, search for employment, government employees, income and wealth. Data from 1992 to the present. Continuous regulatory and technical changes.

NUDB : All individually based statistics on education from completed lower secondary education to tertiary education from 1970 to the present.

Systems for questionnaires and rules and planning information for statistical products are all under development

Systems exist but these will be replaced.

About the data collections (implemented)

Researchers frequently use data collections from Statistics Norway for their research. However, the process from finding out what you need, to actually getting the data, may be long and troublesome, especially for inexperienced researchers. Statistics Norway has therefore (with support from the Research Council of Norway) developed a website to make information about this process more easily available. Among other things, this page provides the users with documentation of several data collections. Each data collection has a general description e.g. of data quality, and it also contains a list of relevant variables, including variable documentation from Vardok. A new system is being scoped, hopefully with even more automatic solutions.

About the statistics (implemented)

About the statistics is metadata that describes each statistics that is published by Statistics Norway. It contains administrative information, information about statistics production, variables, concepts, sources of errors and uncertainty, comparability, coherence and availability. About the statistics now uses a CMS (Content Management system)-platform. CMS makes it possible to link About the statistics to Vardok and Stabas.

StatBank - dissemination database (implemented)

StatBank Norway is a service where you may select scope and content of each table, and then may export the result in various formats to your own PC. This system contains both metadata and data unlike all the other systems described above.

2.3 Costs and Benefits

Two examples of costs:

Vardok:

A total of 12690 man-hours have been used in development with ca. 70% of resources from IT. A total of 476 man-hours from standards were used in 2007 for continued

harmonisation of names and definitions, and training of personnel in the six new divisions. 294 IT man-hours were used in 2007 for maintenance and minor changes to the system.

Metadata portal (man-hours):

| | 2005 | 2006 | 2007 | 2008 | Total |
|------------------|------|-------|-------|-------|-------|
| Senior adviser | 200 | 300 | 325 | 250 | 1 075 |
| System architect | 200 | 200 | 210 | 150 | 760 |
| IT developer | 150 | 1 310 | 975 | 300 | 2 735 |
| Web designer | - | 390 | 370 | 300 | 1 060 |
| Total | 550 | 2 200 | 1 880 | 1 000 | 5 630 |

2008 are planned man-hours. The rest are used man-hours.

Benefits:

Statistics Norway's Strategies 2007 emphasise systematic quality control of products and processes. Statistics Norway's IT-strategies 2007 emphasise that

- metadata systems contribute to simplifying, improving and re-use of work processes
- data that are disseminated and exchanged must in addition to an agreed structure have sufficient metadata to give them meaning
- use of metadata systems are a pre-condition for the development of efficient data capture solutions according to Statistics Norway's data capture strategy.

2.4 Implementation strategy

All our metadata projects are based on a step-wise approach.

3. STATISTICAL METADATA IN EACH PHASE OF THE STATISTICAL BUSINESS PROCESS

3.1 Metadata Class- ification

The only metadata classification we have needed in-house so far is the distinction between conceptual and contextual metadata. This probably reflects the fact that we are only just beginning to try to integrate the metadata in the production lifecycle.

3.2 Metadata used/create d at each phase

Need - Plan system

Plan and Design - Check what is already available in all metadata systems. Ideally update all these but in practise this happens as the final step in the cycle, if time permits.

Develop - Update the metadatabase for questionnaires and rules. Service library for metadata systems.

Collect - Metadatabase for questionnaires and rules. Metadb.

Process - Service library for metadata systems. Metadatabase for rules. Stabas.

Analyse - We use commercial software such as SAS. Metadata support could be better.

Disseminate - Metadata portal, About the statistics, About the data collections, StatBank. Service library for metadata systems (Stabas, Vardok, Datadok, Metadb). Archival of flat files in Datadok.

3.3

Metadata relevant to other business processes

Our current plan system is linked to management and budgeting.

4. SYSTEMS AND DESIGN ISSUES

4.1 IT Architecture

Statistics Norway's technical solutions shall be built mainly upon the principles of service-oriented architecture. Guidelines on this are presented in *Norway's eGovernment plan*. All solutions for external users and most solutions for internal users shall:

- Have support for *open standards*.
- Be *platform independent*.
- Be *component based*.
- Have support for the packing in of data and functions in the form of *services* (web services).

These are central principals in service-oriented architecture. By applying these principles, applications and services can reuse existing functionality/components completely independent of the system they were developed in. In addition, by use of this technique, we can extend the lifetime of older applications, which have important functionality we wish to expose, just by creating a service layer on top of these. This increases the possibilities for collaboration between old and new applications in a completely new way, which gives benefits in the form of shorter development time, increased reuse and more consistent systems. This also enables us to replace systems behind the scenes, because communication with these is not directly exposed to the users.

System architects

System architects are introduced for each of the following areas in the top-level information architecture: data collection, metadata and dissemination. The mandate for this role will be made and will support the system architect's responsibility to ensure that IT development projects are in line with the IT strategy.

4.2 Metadata Management Tools

All our master metadata systems have Oracle databases. Links are hard-links between these databases. Our web applications obtain metadata from our service library for metadata systems.

4.3 Standards and formats

Our classifications system is an implementation of Neuchâtel Terminology Model Part 1 Classifications v2.0.

Our variables system is a partial implementation of Neuchâtel Terminology Model Part 2 Variables. The extent to which we follow ISO/IEC 11179, is best described by the comparison in the appendix of this Neuchâtel document.

We are considering using DDI in connection with micro-data for researchers.

We have used definitions of key metadata concepts from SDMX MCV where possible.

4.4 Version control and

Metadata have valid to, valid from and last updated.

revisions

Program code for web-services is checked in and out of subversion.

4.5 Outsourcing versus in-house development

1. Improved editing tool for our classification database - previous editor was developed out-of-house but the new one will be developed in-house.
2. Analysis of the end-to-end creation and re-use of metadata in one production cycle for one type of statistics - no system development required in analysis.
3. Continuing development of a master metadatabase for questionnaires - in-house development.
4. Continuing development of our service library for master metadata systems - in-house development.
5. Continuing development of our metadata intranet portal - in-house development.
6. Improved plan system for projects, products and processes - in-house development

4.6 Additional materials

Some data models and XML-schema can be made available on request.

5. ORGANIZATIONAL AND WORKPLACE CULTURE ISSUES

5.1 Overview of roles and responsibilities

Subject matter statistician, survey manager, metadata manager, senior advisers in standards, IT developers, system architects and web designers are all important roles in metadata/statistical lifecycle management in Statistics Norway.

5.2 Metadata management team

Our core team for development of metadata systems consists of one senior adviser in Division for Statistical Methods and Standards in the Department of Management Support, and one system architect for metadata systems and two programmers in the Division for IT Development in the Department of IT and Data Collection. When necessary we draw on expertise in data capture, web and web-services from the same IT division and on resources from other IT divisions.

Metadata system maintenance is carried out by two or three people spread across the Division for IT Development, the Division for IT services and the Division for IT infrastructure all in the Department of IT and Data Collection. Maintenance of the system contents, i.e. the metadata, is carried out by all statistical divisions with support from the senior adviser in the core metadata team.

Details per system are as follows:

Our systems for classifications and variables and our metadata portal are owned by the Division for Statistical Methods and Standards in the Department of Management Support. In-house development (2 developers and two customers) has been carried out by the Division for IT Development in the Department of IT and Data Collection. Maintenance of the systems is carried out by the same divisions (1 developer and 1 system owner) with the addition of two people from the Division for IT Services also in the Department of IT and Data Collection in the case of classifications. Maintenance of the metadata in the systems is the responsibility of all 20 statistical divisions. One person in each division is identified as coordinator for variables for the division.

Our master metadatabase for questionnaires and for business rules is being developed by 7 developers from the Division for IT Development in the Department of IT and Data Collection, the Division for IT in the Department of Economics, Energy and the Environment and in the Division for IT in the Department of Industry Statistics. This metadatabase will replace two previous metadatabases maintained by the same divisions.

Our service library for master metadata systems is being developed and maintained by 7 developers from the Division for IT Development in the Department of IT and Data Collection and the Division for IT in the Department of Social Statistics.

Our plan system for products, projects and processes is now owned by the Department of Management Support. Development of the previous system was outsourced by the previous owner (Department of Administrative Affairs) but the new system will be developed by 2 developers from the Division for IT Development in the Department of IT and Data Collection in cooperation with the new system owner.

Our system for file descriptions was developed and is being maintained by the Division for IT Development in the Department of IT and Data Collection. Maintenance of the metadata in the system is the responsibility of all statistical divisions.

Our system for event history metadata was developed by the Division for IT in the Department of Social Statistics. Maintenance of the metadata in the systems is the responsibility of two statistical divisions (Division for Social Welfare Statistics and the Division for Education Statistics).

5.3 Training and knowledge management

All new employees participate in an eight day course designed to give them an overview of all areas in Statistics Norway. 2 hours of this course is devoted to metadata in theory and in practice. Most new employees attend this course within 3-6 months of starting work in Statistics Norway.

All employees can attend the metadata forum, held approximately twice a year, where relevant problems and projects are presented and discussed.

The system architect for metadata systems has the responsibility to ensure that new development is in line with the metadata strategy and the IT strategy. This is possible through an approval process for system development documents.

A senior adviser from the Division for Statistical Methods and Standards has regular meetings with all division leaders to discuss metadata and gives metadata presentations at all three levels of the organisation (see attachment).

5.4 Partnerships and cooperation

Statistics Norway participates in relevant international meetings e.g. METIS and informal groups e.g. Neuchâtel (classifications and variables). In addition, there is a strong tradition of Scandinavian collaboration (e.g. dissemination database StatBank) and collaboration within the Statistical Open Source group (e.g. architecture).

5.5 Other issues

So long as metadata is not an integral part of the statistical production cycle it will be prioritised lower than the publication of statistics. There is then only a short window of opportunity to create/update metadata before the start of the next cycle.

6. LESSONS LEARNED

- 6.1** Top management support is essential.
- 6.2** Make a metadata strategy. It is important that we can refer to formal documents like the metadata- and IT-strategy (which has been approved by the board of directors) in our metadata work. In the same way it is useful that the list of key metadata terms promoted for use within the statistical office has an official "stamp".
- 6.3** Use step-wise development of metadata systems with active user involvement and regular delivery of functionality.
- 6.4** Ensure continuous follow-up of progress and quality with direct feedback to users and regular reports to middle and top management. One of the biggest challenges in management of metadata is allocating the necessary resources. Releasing good quality statistics within the planned time schedule is the primary task for the subject matter divisions and documentation will often have a lower priority. It is therefore crucial that the management stresses the importance of documentation and increases the status for this kind of work.

- 6.5** Harmonising variables between subject matter divisions is also a considerable challenge and an important tool to improve the quality of metadata. Several subject matter divisions may use the same variable names, but define them differently. In some cases this is necessary because of laws and regulations, but this is not always the case. We have meetings where contact persons from divisions using variables with similar names come together and discuss the definitions, e.g. if a division could change the wording of their definition to such an extent that other divisions might use it as well, which would allow us to reduce the number of definitions to one instead of e.g. three. This is a time consuming work which we have started, but which will require a lot more of resources, both to monitor where harmonisation is needed and to do the job.
- 6.6** The possibility to release metadata on the Internet makes it easier to motivate subject matter divisions to document metadata and improve metadata quality.
- 6.7** We think that to really make metadata work a natural part of everyday life in the subject matter divisions, we have to include the metadata systems in the production cycle. Then we can establish routines where the handling of metadata is included in all relevant production steps. So far the metadata work in Statistics Norway has been focused on implementing metadata systems and filling them with relevant documentation. This year we will start investigating the role of metadata (systems) in the production cycle.

*** END ***