

# **Electronic Co-operation**

## **Towards a service oriented architecture in Statistics Norway**

By Hanne Mette Janson and Rune Gløersen, Statistics Norway

### **Introduction**

The public sector in Norway has arranged a large work program to improve the value of its electronic services. The idea is to make the virtual room a wider and more efficient arena for exchanging information, for citizens, enterprises and government departments. This imposes higher requirements on the standardisation of exchange formats, both for data and metadata. Statistics Norway wishes to become a driving force by arranging such a coordinated infrastructure with its closest business partners. The Electronic Co-operation project has been established, with funding from the Ministry of Finance, in order to carry out this work. The aim of the project is to develop synergetic solutions for

1. User-friendly reporting to the public sector, through Internet solutions (primary data capture) and direct reporting from administrative systems for enterprises (secondary data capture)
2. Re-use and exchange of information between government departments
3. Coordinated dissemination of all official statistics in Norway.

In order to increase the efficiency and improve the cooperation between government departments regarding the capture, use and dissemination of government information it will be necessary for IT systems, initially built for production, to become service oriented. This implies that they must be adjusted for the exchange of data and information about data (metadata) with other users and other systems and that this must happen through the exploitation of standards and technologies established for this purpose.

In 2005 we will prioritise the building of an IT platform based upon principals for service oriented architecture, so that in the next round we can build new solutions and services in a uniform and well thought through manner.

In the following, we outline our priorities for the work in 2005.

### **Project portfolio 2005**

#### **Data collection**

The primary goal is to complete the extensive development tasks in reporting from municipalities to the state (Kostra) and from industry to Statistics Norway (Idun) as described below. At the same time we are moving towards the establishment of re-useable and stable production solutions for both electronic questionnaires (e-questionnaires) and direct reporting from the respondents administrative systems. The cost of maintaining two almost identical data capture platforms like Kostra and Idun is incommensurately high so in 2005 we will begin investigation of the possibility of combining these two solutions in accordance with service integrated principles.

#### *Idun/Altinn*

The work in 2005 consists of two parts. Firstly, we will finalize the development of e-questionnaires for the 25 remaining annual statistical inquiries. All of these should be ready in the first quarter of 2005. Then we have reached our goal of offering e-questionnaires for every survey in Statistics Norway.

Secondly, an integration between Statistics Norway's Idun solution and reporting data from private industry to governmental institutions (AltInn) must be made, so that respondents in the industrial sector can log on to Altinn and from there have access to e-questionnaires in Idun. A separate project

plan for this integration has been made. According to this plan it should be possible to reach Idun via Altinn (version 3.1) by December 2005.

In 2005 we also plan to implement our intention of providing tailor made statistics as feedback to our respondents. This work, which was one of the central goals of Idun from the very beginning, was given lower priority because of the breadth of work within Altinn. The project will firstly consist of technical development work and secondly of an investigation among selected respondents as to what type of information they would be interested to receive. We expect that providing respondents with relevant statistics after they have submitted their response through our e-questionnaire will increase the percent that report by web and reduce the perceived respondent burden.

In this way, we establish good and generic solutions for use of electronic questionnaires and direct reporting from respondents administrative systems.

#### *Kostra*

We will concentrate our efforts to find a more dynamic platform for questionnaire generation in Kostra, based on the Idun-platform. This approach will rest upon the new metadatabase model for our common Questionnaire Repository. The main part of this modelling will take place in 2006.

## **Dissemination**

### *Catalogue of services for automatic data exchange*

We wish to move over to a development in which both external and internal end-users have a more integrated relation to Statistics Norway's data sources, and can therefore make use of Statistics Norway's data directly in their own applications. The services shall be adjusted so that end-users can tailor-make the amount and structure of the data they wish to extract. At the moment, such access requires interactive participation from the end-user. Both internal and external users have requested automation of such extraction routines.

- We wish to have a particular focus on external users, and have amongst other activities entered into a dialogue with the Norwegian Pollution Control Authority (SFT) regarding the establishment of such an automatic extraction service. SFT wishes to present numbers from Statistics Norway on their web pages without having to store the data in their own system. The Norwegian Bank has requested a similar service. A secondary effect of such an automatic extraction is that copies of data to a lesser extent need to be stored by users. This can make Statistics Norway's role as a master system for many datasets clearer.
- With regard to internal users, we shall develop basic mechanisms that make it easier to re-use data from StatBank Norway and other data sources, directly in other applications. At the same time, publishing work for the subject matter divisions will be made easier since data only need to be published once.
- In accordance with the intention of the Norwegian Statistics Committee, a proposal has been made to create a national network for dissemination of official statistics through [www.ssb.no](http://www.ssb.no). In this model there is one very important principle i.e. that each producer of official statistics has the responsibility for maintaining their own systems, data and metadata but that the portal offers a common set of services to the participating departments.

### *Automatic extraction from StatBank Norway (Data shooting)*

Based upon the extraction mechanisms discussed over, a structure will be built over StatBank Norway that sends data according to a previously agreed upon specification. In this way the receiver does not need to check for new relevant data in StatBank Norway. StatBank Norway becomes the active partner in the data exchange. We imagine building several transport channels for this:

- completely automatic, over e-mail and http post
- half automatic via a warning over e-mail, http post, or RSS (Rich Site Summary) that data now lie in a particular table ready to be downloaded at the receivers convenience.

### *Integration of statistical production systems for publication*

The project will enable an automated transfer process for finalized data from statistical production systems to StatBank Norway. This involves a streamlined integration with all production/publication tools in use in Statistics Norway (SAS, Fame, Excel, Oracle, Super Cross etc).

## **Metadata systems**

### *Common metadatabase for Kostra/Idun*

Integration of metadata for e-questionnaire design (web, xml4dr and other formats) is a prioritised area within our metadata strategy. As mentioned earlier, this is also an important precondition for integration of questionnaire production in Kostra/Idun. The final goal of this project is to develop an optimal data model for these two data capture solutions. The main criterion will be to distinguish between layout and business rules in accordance with the guidelines in our IT-strategy. The metadatabase must have an interface with metadatasystems both inside and outside Statistics Norway. This could be Altinn and/or our internal master systems for documentation of variables (Vardok) and classifications (Stabas). Users must be involved in the development of a stable and user-friendly interface for creation and editing of the contents of the metadatabase.

As mentioned, data modelling activities take place in 2006. The work in 2005 has two goals:

- Short-term adjustments in questionnaire production for Idun: small improvements of Idun metadatabase
- Prepare a user requirement specification for a combined Kostra/Idun e-questionnaire design tool

### *Catalogue of services for metadata exchange*

Statistics Norway's metadata strategy, approved by the board of directors on the 13<sup>th</sup> of January 2005, purposes a series of implementation projects that shall contribute to a more general metadata system for our production of statistics. Central to such a harmonised system are our three authoritative metadata sources: Stabas, Vardok and our documentation system for files (Datadok). By cultivating these as our sources for classifications, definitions of variables and file descriptions respectively we can achieve a consistency and efficiency in the storage and re-use of our metadata in the entire production cycle. In order to reach this goal, our sources must be supplemented with routines/services and entry points that give both our external and our internal users easy access to metainformation. Such services should also make it easier to attach the right metadata throughout our statistics production. Our ambition is to base these exchange routines on service integrated principles, and through this, achieve a set of open and generic routines independent of the master system's platform.

On top of these services, we also wish to build a user-friendly common entryway with general access routines. The steering committee for metadata projects has not yet decided the priority this activity will have in 2005.

### *Metadata framework*

In order to build more services and exploit new possibilities within electronic co-operation, it is necessary to achieve as format independent data communication as possible. W3C has taken the initiative to develop a structure for this called RDF (Resource Description Framework). This framework is based on an XML notation where definitions of formats and standards are released from more proprietary data models and applications. In this way a significant increase can be achieved in the re-use of metadata stored in different systems and models.

Progress in this project will be steered largely by external actors that we cooperate with. It is possible that the SDMX (Statistical Data and Metadata Exchange) initiative will be one of our prioritised areas within this work. We await possible participation in OECD projects within SDMX. The project will therefore go more as a background process in 2005 and be activated in step with emerging partners.

## **Annex**

### **Service Oriented Architecture**

We draw heavily upon a technological development that makes it possible to implement services/operational processes more directly, and to a larger extent independent of underlying database tools, programming languages and network platforms. Together these ways of thinking and communication solutions are called integrated services and are primarily directed towards efficient exchange of large amounts of data/metadata across different system platforms in a network. Such an infrastructure is supported by a collection of re-useable (generic) components, which can be set together in a flexible way to provide new services. Technically these mechanisms are based mainly upon web standards such as SOAP (Service Oriented Application Protocol) and XML (Extensible Markup Language). Other middle-ware technology can be used internally. Several departments that Statistics Norway co-operates with have begun to use such data and metadata exchange solutions.

By taking this development into use, we can achieve more long-term system solutions that are robust against changes in the underlying infrastructure. At the same time we will be better prepared to make electronic services connected to Statistics Norway's data sources directly available for external users. We must prepare for a broad increase in competence in order to make the best use of this development, both conceptually and technically. Amongst other considerations, will exposing more of our internal services/processes to external users, require increased foresight and consideration of the vulnerability this could potentially inflict upon our data sources. Through this increased competence we will develop innovative creators and coordinators in a comprehensive IT-environment. The Department of IT and Data Capture has a special responsibility for solutions related to data capture, metadata exchange and dissemination. We begin in 2005 to exploit service integrated principles across three system platforms in order to develop services that promote electronic co-operation.