

# **Data modelling – merging supervision, statistics and monetary analysis data into a joint data set for financial enterprises**

Olav Stensrud<sup>1</sup>

## **1. Introduction**

This paper is based on experiences gathered in creating optimal data sets and structures by data modelling procedures in Norway. Our main vision is to create well structured data sets serving multiple purposes: financial statistics, supervision, analysis and research of the various financial sectors, on a micro and macroeconomic level.

Cooperation between government agencies is vital in small countries such as Norway due to quite limited resources and few available experts. The legal framework in Norway also requires coordination of data specifications and data set sharing between government agencies. Besides this there are ongoing projects led by the Ministry of Industry to ensure simplification and reduction of the data reporting burden of private and public enterprises.

The first efforts on data structuring for financial enterprises started in 1975 with joint collection<sup>2</sup> and compilation of 38 different forms for statistics and supervision purposes for the financial sector – mainly banks.

In 1986 this financial statistics collection and computing system of balance sheets and profit and loss accounts, interest rates, capital adequacy etc. was restructured into six matrix based data sets which were stored in a shared data base<sup>3</sup> accessible by all three government agencies. In this new computing system electronic submission of data was also implemented and was later (in 1996–98) strengthened with web-based, automated reporting and check-routines handled by the reporting entities themselves. The data model, data sets and computer system – ORBOF – still comprise the main source of information for statistics, supervision and analysis in Norway in 2008. A revision of the data model was made in 2007 and a change of technical environment is planned from 2009.

In the following I will describe some efforts and challenges we have met and also give some input on future development ensuring high quality data sets serving statistics, supervision and analysis.

## **2. The role of Statistics Norway as collector, adviser and data provider of financial statistics for end users – data modelling**

Statistics Norway (SSB) has defined duties serving other government agencies and reporting entities as official statistics provider and intermediary to the public, to national analysts and

---

<sup>1</sup> Senior Adviser at Statistics Norway.

<sup>2</sup> Statistics Norway, The Financial Supervision Authority of Norway and The Central Bank of Norway developed this form and computer based system together, but the data structure was initially very raw with only a few cross-checking routines to establish a moderate level of consistency between the datasets.

<sup>3</sup> The technology used in the data base and computing system is Adabas and Natural from Software AG.

as reporter of data to international organizations. The statistics functions of SSB are legally based on the Statistics Act, which defines the major tasks. Statistics Norway is the main, domestic government agency for statistics production, but its duties also encompass advisory services to every data collection system suitable for official statistics purposes – independent of who collects and computes the data. The advisory role gives Statistics Norway wide authority in all major domestic data collection and compilation systems.

Thus, for financial statistics purposes Statistics Norway has been the key government agency from 1986 to 1996 and again from 2007 performing data structuring and modelling, data collection and compilation and acting as data base host for Statistics Norway, The Central Bank of Norway and The Financial Supervision Authority of Norway. The financial statistics computing system consists of accounting based data sets organized in a shared data base serving financial statistics, financial sector supervision, monetary policy and financial stability analyses.

Serving these four “functions” is a difficult task with few data matrixes / data sets, as the requirements of the cooperating government agencies are different. There is always a danger of collecting too many details and then also making too complex and large data structures when serving many data requirements. However, Statistics Norway’s main goal has been to harmonize the data requirements into a common structure based on a comprehensive data matrix. The matrix is built from core balance sheet data and profit and loss data on which attributes and variables of statistics information of institutional sector, industry, country and other attributes such as currency, risk, interest rates, etc. are added. Relations between the different data, variables and attributes are defined clearly and monitored continuously.

The comprehensive data matrix model was developed in 1986. The model has been extended according to changes in data requirements stemming from changes in international and national standards and needs, but its main features remain the same. In 2007 the inclusion of IAS/IFRS changes was not very complex as it necessitated inclusion of only a few new variables and attributes in the model. However, the challenge of every large change is to handle and link time-series data, and IFRS was no exception at this point.

A redesign of the data model, data matrixes plus the technical system is planned from 2009.

### **3. Implementation of standards for statistics, accounting, supervision into the data model and data sets**

#### **3.1 Overview**

Traditional financial statistics may often be built on standards and methods used for national accounts, standards for institutional sectors, balance of payment standards etc. In Norway we also have implemented official accounting standards and data specifications in the definitions and the data structure, to ensure consistency and high, stable quality between published official financial statements and the financial statistics data sets.

We believe that precise definitions and defined links between the actual, official financial statements of accounting data and the financial statistics data set are crucial for both reporting entities, authorities and end users of data; as a single, reconciled data set should be the basis for statistics, supervision and analysis.

However, there is a change of focus from national accounts based data towards new data better fitted for financial stability analysis and supervision, where new attributes and variables are implemented. The change is strengthened by mergers and acquisitions among financial enterprises on a cross border and cross sector base which implies a need for reporting entities based on different consolidation principles. These changes have to be implemented

in existing and new data sets, both on the variable/attribute side and on the reporting entity side. New reporting entities covering the legal entity then have to be implemented alongside the financial sector based reporting unit. New variables and attributes defining risk, weighting, estimation of value, multiple sectors (direct versus ultimate risk counterparties), maturity etc. have also to be considered in the data modelling activities and in a change of data structure as well.

When restructuring data sets and redesigning data models, there are several choices to be made:

### **3.2 Full integration of all variables in a single, comprehensive data matrix**

If this is the choice, then in principle all data – entities, variables and attributes – are collected and stored once in a detailed full information data matrix. The matrix does not use marginal distributions or aggregates, and all information can be extracted directly on a detailed level – and might be transformed to the different uses by extraction by “rule based” computer programs.

**Pro:** All data are defined and reported in a detailed way, and double and multiple reporting and storage is avoided. This may also ensure full consistency in the data set and will throw light on missing variables and attributes and slumps in data quality when reports are sent in from reporting entities and automated quality checks are made.

**Contra:** Large, detailed data sets may be difficult to collect, compute, store, correct and extract in a data base. The base might be too large and too costly to maintain. Reporters and users might lose their overview of the data – it might simply be too large. The data base may have many “missing observations”, which may be difficult to handle technically and practically.

### **3.3 Partial integration of variables in a few data matrixes**

If this is chosen separate data sets or smaller matrixes directly suited for statistics, supervision and analysis might be constructed. Data should be collected and stored as few times as possible, suited for the purpose at hand. In such a structure it is important to define a core data set that is the basis for reconciliation of the different matrixes / data sets.

**Pro:** Data defined and reported in a more aggregated way than the single data matrix method might be more easily handled, collected, stored, collected and extracted. Acceptable consistency in the data set may be achieved by implementation of automated check routines. Data sets can be made smaller than the single matrix method, positive for costs for the authorities.

**Contra:** There may be multiple reporting and storage of data, and it may be difficult to establish full consistency and transparency of data in and between the different data sets. The danger of inconsistency may affect the end users of statistics, supervision and analysis if actions are based on partially reconciled data in different parts of the data base. Errors and omissions in the data might be concealed by aggregated data. Cross checking routines might be mistaken for data quality check routines. However, for the reporting entities the preparation of different forms and aggregates that need to be reconciled might easily be at least as resource consuming as the comprehensive data matrix method. The reason for this is that what often matters most for the reporting entities, accounting routines, reconciliation and computer systems preparing data sets for the authorities, are the total number of variables and attributes.

### **3.4 Recommendation**

Based on our experience in Norway a single, comprehensive data matrix method is the best data model solution, when data quality and avoidance of multiple storage and collection of data is concerned. Also when it comes to flexibility of analysis and research activities, a detailed matrix should be the first choice. However, in practice the level of detail which will be the product of the number of variables and attributes, might lead to a huge data set that is difficult to handle. The choice is then often the partial integration of data requirements into a few data matrixes. Then these data sets/matrixes should still consist of quite detailed number of variables and attributes to ensure reporters focus on data quality of the data input into the reporting financial statistics system. Its also essential to establish the core data set giving the key to relevant reconciliation among the different data sets. A core data set should be based on balance sheet data and profit and loss accounts data preferably strongly linked to the official financial statement of each reporting financial enterprise.

## **4. Reporting burden and data specifications**

Enterprise reporting burden has been a key issue in Norway for at least ten years. As previously mentioned, Norwegian legislation (the Act on business reporting obligations and the Statistics Act) requires government agencies to cooperate and harmonize data specifications, definitions and share data collected from financial and non-financial enterprises. The legislation not only encourages but requires harmonization and cooperation.

Another important issue relevant for the data specification process is that all domestic public authorities have to have in mind that each reporting entity meets the sum of all data requests from each government agency and that the harmonization of data to a few consistent data sets without too many different reports is important to reduce the reporting burden. To structure, redefine and cut off irrelevant marginal reports and aggregates and coordinate and share data between government agencies is crucial in this work.

Lately there is also an efficiency improvement project lead by The Ministry of Industry where all official data reports to government agencies (including tax authorities) are scrutinized and actions of harmonizing data are to be planned within 2009. The professional and industrial associations of the financial corporations take part in this project. The main objectives in the efficiency improvement project are to further develop web-based data collection and response systems, reduce rules and costs of reporting and map and reduce the authorities' reporting burden on financial and non-financial enterprises.

The professional and industrial associations of the financial corporations also initiate input and actions to harmonize and simplify the reporting of financial data to the authorities. For this purpose Statistics Norway has established a forum for financial enterprises, their professional associations and computer service companies to meet, discuss, prepare and implement changes in the financial reporting system. This forum meets several times each year and working groups are also established solving specific issues. Lately the future IFRS changes are on the agenda of this group, and changes are planned implemented in 2009.

## **5. International organizations, standards and reporting requirements**

Most international organizations (OECD, IMF, ECB, Eurostat, BIS) use the official national accounts statistics standards and related standards as a base for reporting requirements. Many international organizations also share and exchange data between themselves.

However, when it comes to accounting standards and data requirements for financial stability purposes and data specifications in country reports to different international organizations,

there are a large number of specifications and additional data, consolidation and estimation activities to be made on the national level. Data requirements may at times seem to be determined from theoretical needs and models, but general theoretical needs are often hard to meet in several countries.

Sometimes standards may seem paradox. For instance the IFRS-standards might express transparency in official financial statements for the end user, but also encourage flexibility of the data presentation of the financial statements. However flexibility might hamper the comparison of financial statements between financial enterprises without a thorough study of statements and annotations. Flexibility might then mean a more complex and resource consuming analysis of financial statements.

Unfortunately, different reports to international organizations still use different definitions, and different data specifications and templates. This means that the national statistics agencies both have to collect additional data from the reporting units and make several estimations on the data beyond the available details on the national level. Additional data specifications and different definitions increase the reporting burden on the reporting enterprises and is also a threat to data quality when the reporting unit cannot understand the usefulness of the additional, slightly different specifications.

In the case where a national statistics agency has to estimate data for non-identified observations in the available data sets, based on key figures that often may be debatable, we may question the quality of the data sent to international organizations.

In my view, efforts should be made on further harmonization of definitions, data sets and data exchange between international organizations to ensure highly reliable data at an acceptable cost to reporters, national authorities and international organizations.

## **6. Recommendations**

To sum up:

- The creation of a comprehensive data matrix based data set within a consistent data model is crucial for data quality, data collection, compilation and use of data.
- The data set should cover data requirements for statistics, supervision, analysis and research purposes.
- Data requirements and specifications should be harmonized more thoroughly:
  - National authorities and reporting entities should develop data sets in cooperation.
  - Statistics authorities should take part in the development of standards for accounting etc. where data also is relevant for and will be used for financial statistics purposes.
  - International organisations should work together to harmonize definitions, data sets and templates.

All participants in the standardization and harmonization process must focus on simplification and consistency of the data. Data sets should be quite few and with relevant detail ensuring good quality. Data should be easily reconcilable and handled within a well structured data model.

# Data model – ORBOF – revised for 2009

