Counting on Statistics: Statistics Netherlands'Modernization Program

Winfried Ypma en Kees Zeelenberg

Discussion paper (09032)

Statistics Netherlands

The Hague/Heerlen, 2009

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2007–2008	= 2007 to 2008 inclusive
2007/2008	= average of 2007 up to and including 2008
2007/'08	= crop year, financial year, school year etc. beginning in 2007 and ending in 2008
2005/'06–2007/'08	= crop year, financial year, etc. 2005/'06 to 2007/'08 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

6008309032 X-10

Counting on Statistics: Statistics Netherlands' Modernization Program

Winfried Ypma and Kees Zeelenberg[1]

Summary: Counting on Statistics, the modernization program of Statistics Netherlands has as aims improvement of quality, more use of administrative data, higher efficiency, and reduction of the number of ICT applications. It consists of 5 subprograms: business architecture, methodological framework, generic processes, modernization of the main economic statistics, and modernization of other statistics

Keywords: Statistics Netherlands, modernization, business architecture

1. Introduction

From 1999 on, Statistics Netherlands (SN) has been confronted with budget cuts, which has led to a major reorganization in 2000. The traditional stove-pipe organization with small units each responsible for the complete process of a small part of the statistical output, has been replaced with a more process oriented organization in 3 large divisions for business statistics, social and regional statistics and macro-economic statistics (see Willeboordse, 2000, for a more detailed description). This reorganization left the statistical methodology basically more or less unchanged, although in particular in business statistics a more efficient approach to data collection and processing was chosen. At the same time, SN was also confronted with external pressures and demands for reduction of the response burden, which has led to a clause in the 2004 Act for SN which requires it to use all available administrative date before any surveys may be held. From 2004 there are again budget cuts of in total 10 percent over period till 2012, which are to be met by efficiency gains without reducing the statistical output. More recently, it was recognized that the quality of statistics was not always as good as it should be, which led to a greater focus on methodology and quality. Finally, stakeholder surveys showed that SN should improve the flexibility of its output, i.e. major users want a more timely reaction to needs for new output. Discussions on these external and internal factors (Van der Veen, 2006) have led to the Modernization Program (MP) *Counting on Statistics* for

---

[1] Deputy Director for Statistical Modernization, respectively Director of Methodology and Quality

Paper prepared for the Seminar on Increasing the Efficiency and Productivity of Statistical Offices at the plenary session of the Conference of European Statisticians, 11 June 2007

SN's statistical processes. The next sections focus on the modernization program and its components. Section 2 gives an overall view of the program. Section 3 describes the business architecture for the statistical production process and the architectural rules that govern the modernization. Section 4 describes the methodological framework within which the modernization takes place. Section 5 describes the part of the program that will create generic statistical tools and services. Section 6 describes the redesign of the main economic statistics, which is a major part of the modernization program.

2. Overall view

The main objectives of the Modernization Program are

1. improvement of quality, in particular with respect to coherence, flexibility, consistency and reproducibility
2. more use of administrative data
3. higher efficiency
4. reduction of the number of ICT applications

The Modernization Program consists of 5 subprograms:

1. Business architecture
2. Methodological framework
3. Generic processes for data collection, data storage, metadata, rule-based processing, and output
4. Modernization of the main economic statistics (short-term business statistics, annual business statistics, National Accounts)
5. Modernization of other statistics

Work on the first 4 subprograms is now under way. We will describe them in the following sections after an overall view in the present section.

The Modernization Program is partly a centralized development program and partly a decentralized development program. The centralized part consists of the first 3 subprograms. The first two subprograms set the standards and frameworks and the third one provides the generic tools for the actual redesign of the statistical processes, which is carried out in the 5^{th} subprogram. The redesign of main economic statistics is a separate subprogram because of their important role as the provider of the main economic indicators such as the growth rate of BBP. The programs 3 and 4 fall directly under the Board of Directors, and a member of the Executive Board and the deputy director for statistical modernization are responsible for the operation of the program.

3. Business architecture

The program started in 2005 with the development of a business and information architecture for the statistical production process. The architecture consists of an abstract representation of the process and of general principles for the activities within that process. Figure 1 shows the main lines of the statistical process and its information model.

The main architectural principles are

1. *re-use* of data and meta-data
2. distinction between the *design* of a statistical process and the actual carrying out of the processes
3. the design is modeled as a *value chain* of activities governed by rules that are explicitly documented
4. design is directed by *output requirements* within *frameworks* for methodology, data collection, finance and organization
5. no production without *metadata*, consisting of the process model and its design
6. metadata are *standardized* with respect to units, concepts, classifications, quality aspects, and process definitions
7. in line with the architecture, each process has 4 databases: *inputbase* for the input data, *microbase* for the statistical microdata, both raw and processed, *statbase* for statistical outputs and *outputbase* for publishable data.

Early 2006, the business and information architecture has been completed (Huigen, 2006); thereafter it has been complemented with architectures for information systems and for ICT (Brederode and Dekker, 2006a and 2006b; Windmeijer, 2006). The maintenance of the architecture now rests with an Architectural Review Board (ARB), consisting of specialists from the methodology department and the IT department, and an Architectural Steering Committee (ASC), consisting of managers from the methodology and the IT departments and representatives from the statistics divisions.

The architecture is imposed on every development project by means of a project start architecture (PSA) document, which describes how the process that the project is to develop will fit within the architecture.

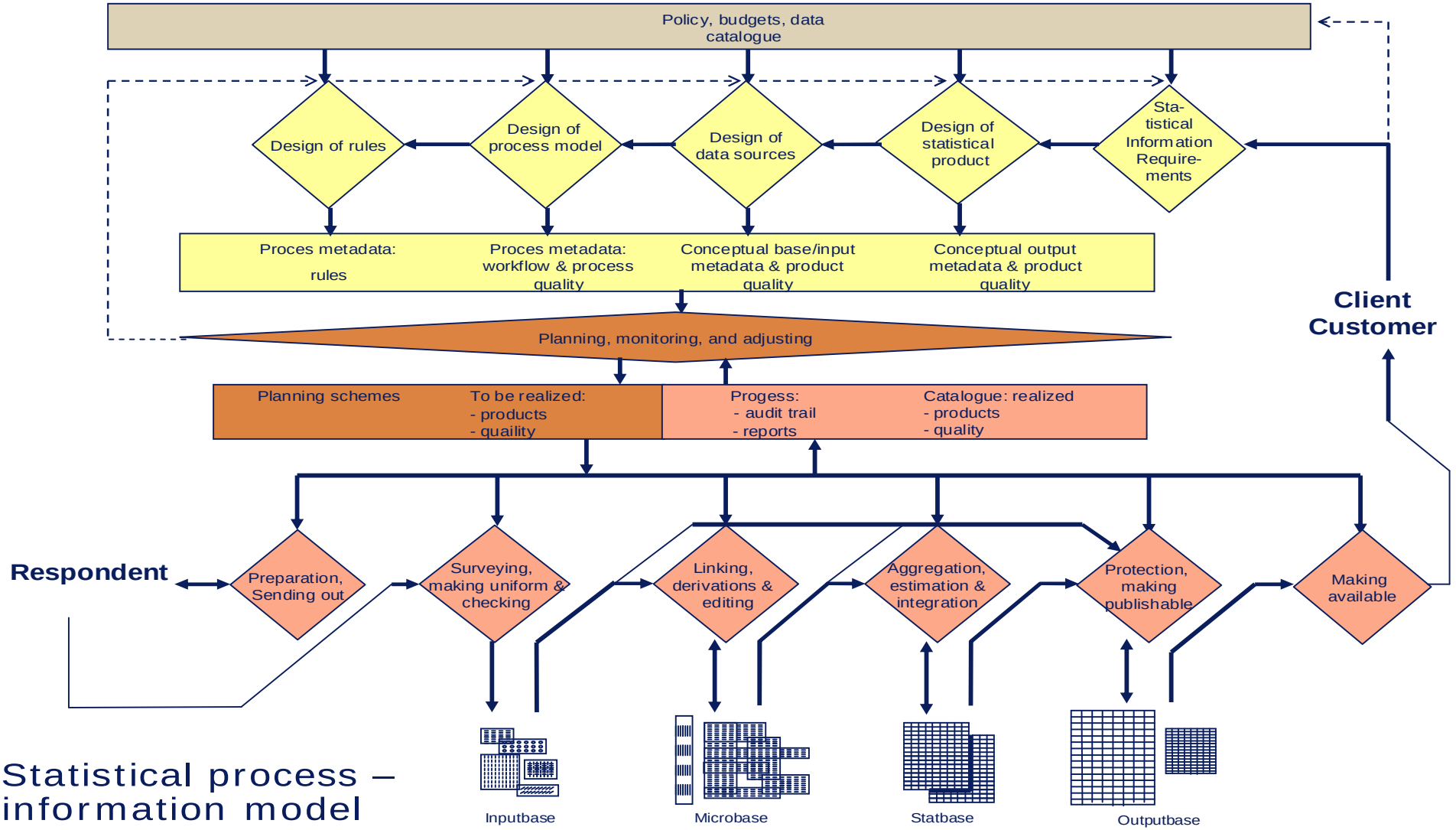

Figure 1: Business architecture and information model of the statistical production process

4. Methodological framework

The methodological framework is the set of validated statistical methods that may be used in the statistical processes. Methods outside the framework may not be used. The main purposes of the framework are standardization of methods, improvement of quality, transparency and IT-standardization (by linking methods to tools). The framework has of 3 elements:

1. The *methodology series* consists of descriptions of the methods and criteria for their application. It is particularly relevant for designing or redesigning processes. The series starts with a general perspective on statistical processes along the lines of Willeboordse (1997). Thereafter the whole body of statistical methodology will be described in a series of papers, which are not meant to give thorough scientific expositions but will give enough information to a developer of statistical processes. This will lead to a fairly complete documentation of all methods that are in use at SN.
2. The r*ules for process development* describe how processes are to be developed taking into account available methods, budgets and quality norms. They are clearly related to methods for project development, such as Prince-2, and methods for software development, such as RUP. The rules will be based on Total Survey Design (TSD) principles. Amongst others they will include procedures for implementing new processes as replacements of old ones, e.g. guidelines for analysis of breaks in statistical output series and measurement of quality of administrative data, and for change management.
3. *Quality control* for processes, mainly by means of statistical audits. The aim is to review the key statistical processes once in every 5 years. The focus will be on chains that lead from input data to output data, e.g. the chain leading from the inputs for the Labor Force Survey (LFS) and the Wage and Employment Survey to the Labor Accounts, which are part of National Accounts.

5. Generic processes

The main drive toward generic processes came from the observation that the software of SN can be described as a bowl of spaghetti. The number of applications is very high, maintenance cost is high and growing, documentation is poor, and there is so-called grey (unauthorized) automation.

As solution we have adopted the *service approach* directed towards re-use of applications: less applications means lower maintenance costs but also lower development costs. This approach also has its advantages on a business level. The service approach starts with a good architecture of the complete SN process; see section 3. The main services will be:

Data collection

All data collection will be delivered as one service, with a subdivision into a number of channels using control functions to steer individual channels and to facilitate switching between channels leading to flexible mixed mode data collection. Amongst others, it should provide CAPI, CATI and web channels.

Data service centre (DSC)

This is the central structure of the overall process. We construct 'resting points' between processes of the value chain. They range from an *input base* (resulting directly from data collection) to an *output base* (giving outside access to output). Main services to be delivered are data storage, accepting data and handing out data. These services will be provided through metadata requests. This is the most challenging part since user friendliness, easy access and performance are important. Also, one physical means of storage will probably not do.

Meta

The meta service is closely linked to DSC. It should handle conceptual meta, qualitative meta and process meta. It will consist of metaservers on variables, classifications, units and populations as well as a catalogue describing the content of DSC.

Output

The output service is in part a special case of DSC. It takes care of the output database and the connection between the last stage within the chain and the output database.

Chain control

The architecture shows a distinct activity that gives directions to the production process within the value chain. As such it seems to have been underrated within Statistics Netherlands. This neglect has caused quality problems in the past.

Rule based tools

This is a strange element in the list. It stems form the principle that business rules and software should be separated. Also Statistics Netherlands wants to minimalize the ICT input for minor changes in its processes. That can be achieved if statisticians maintain their own business rules. That means less SQL and less VB and more input from the statistician. Of course, then appropriate testing and acceptation environments remain essential as well as ways to maintain versions of rule sets. Statistics Netherlands is looking into tools like Clementine and Ruleburst.

All this is done within one program *generic services*. What this program does not aim at is building generic data processing solutions. In our view that is one bridge too far. The program just offers efficient tools for data processing.

The ICT realization of these business services is done by implementing a service structure on the software level and below as well, leading to a Service Oriented Architecture on several levels.

In this process buying goes before making. Major parts to be made will be put to the market to be built on our specifications by an outside partner. For managing this program also an outside partner has been contracted.

6. Modernization of main economic statistics

Creating services is one thing, making actual production processes go over to use them is another. Modernization at Statistics Netherlands therefore has at least two dimensions. The first is where the services are developed; the other is where the transition is made from the old process to the new service based one.

One process or better one chain of processes that will make that transition is the chain of economic statistics. Modernization is not only and not even in the first place led by architectural considerations. More important is the need to improve quality. At present, the differences between the several estimates, such as short-term statistics, production statistics and national-accounts estimates are too large. Also important is the (political) urge to reduce the administrative burden. That means that the use of registers (mostly fiscal) will have to increase. A second reason for modernization is to increase efficiency.

To reach these goals a new chain of processes is set up following the architecture. To begin with, a more detailed architecture for this chain is developed. One of the issues to be dealt with is the tuning of the monthly, the quarterly, the provisional yearly and the definite yearly growth figures. The next step is to decide upon the necessary output. An ongoing discussion is the one about what output should be produced above the output we are obliged to make by (international) law. The next step is to identify the possible administrative data sources that can be used to create the output we need. After that we can look for the most effective and efficient methodology.

We are looking at the chain starting directly after data collection (itself a service separated from the rest; see section 5) through data processing on micro and on macro level towards national accounts. The approach is essentially top down. Important issues are the use of administrative data instead of direct surveys and a separate, intensive handling of the larger and more complex enterprises (‘Topxxx’) and the means to manage and control the complete chain of production.

Migration will be a complex issue, since discontinuities must be kept to a minimum, and a big bang transition is not attractive. A revision of the national accounts is already planned for 2013, the NACE classification is set up for 2011, and so we will be looking into ways to synchronize these with the major redesign.

Both programs, generic services and the modernization of economic statistics, will develop new means for the statisticians. It may also lead to changes in the organization.

For the change in SN to be successful, it is necessary that there is a receiving party for the programs. So, a change manager is needed who can help with the migration strategy, with the (re)organization of the receiving departments and with creating support within the organization.

The program on generic services will last until 2009, the modernization of economic statistics program lasts year longer. The complete migration of SN's production process will take several more years.

References

Brederode, R., and W. Dekker, 2006a, *Conceptuele IT-architectuur* (Conceptual IT architecture). Internal report, Statistics Netherlands, Voorburg / Heerlen.

Brederode, R., and W. Dekker, 2006b, *Logische IT-architectuur* (Logical IT architecture). Internal report, Statistics Netherlands, Voorburg / Heerlen.

CBS (Statistics Netherlands), 2002, *Statistics that count*. Strategic Plan, Statistics Netherlands, Voorburg / Heerlen.

Huigen, R., 2006, *Statistics Netherlands Architecture: Business and Information model*. Discussion Paper, Statistics Netherlands, Voorburg / Heerlen.

Van der Veen, G., 2007, *Changing Statistics Netherlands: driving forces for changing Dutch statistics*. Paper presented at the Seminar on the Evolution of National Statistical Systems, New York, 23 February 2007. Statistics Netherlands, Voorburg.

Willeboordse, A. (ed), 1997, *Handbook on the design and implementation of business surveys*. Eurostat, Luxembourg.

Willeboordse, A., 2000, Towards a new Statistics Netherlands: blueprint for a process-oriented organization structure. *Netherlands Official Statistics* 2000-1, 46-9.

Windmeijer, D., 2006, *Logische informatiearchitectuur* (Logical information architecture). Internal report, Statistics Netherlands, Voorburg / Heerlen.