Proceedings of MSP2009
Modernisation of Statistics Production 2009

Redesigning a statistical institute; The Dutch case

Barteld Braaksma[1][2]

Statistics Netherlands is facing a number of major challenges. On the one hand, budget cuts and the IT maintenance burden call for increasing efficiency. On the other hand, the statistical landscape is changing rapidly. This applies both to the input side, where attention shifts from primary data collection to the use of administrative registers; and to the output side, where users ask for increasing flexibility while maintaining high quality standards. In order to keep in control of these competing challenges an ambitious redesign program, the Masterplan, was started in 2005.
The general ideas of the Masterplan are represented in a comprehensive enterprise architecture. Some key elements are the identification of steady states in statistical processes, consisting of data sets with guaranteed quality to promote re-use of data and supported by a Data Service Centre; a new organizational unit providing generic data collection services; and introduction of a standardized statistical toolbox, supporting the core production processes. Thus, the traditional stovepipe-based approach will be gradually abandoned in favour of a flexible process– and service-oriented approach.
Over the past years, substantial resources have been allocated to prepare the new approach and supporting IT solutions. We have now reached the stage where the general ideas of the Masterplan are to be implemented on a larger scale. Over the next few years, several hundreds of statistical processes will have to be redesigned accordingly. Central redesign teams have been established to accompany the redesign projects, supporting them in carrying out their business analyses and in designing software solutions compatible with Masterplan principles.
This contribution aims at giving a broad overview of the Masterplan from a business perspective, covering both the ideas and their implementation. Not only successes will be presented, but also the false steps that were –almost unavoidably– taken and that have led to rethink the initial approach. The approach will be illustrated by some concrete examples.

Key words: Architecture, economic statistics, efficiency, generic services, rule-driven processing, service-oriented approach

Acknowledgements: In line with the general redesign philosophy, the author has freely re-used material available within Statistics Netherlands. Sometimes this material was purely plagiarized, sometimes adapted and sometimes merely used for inspiration. The author wishes to thank those who provided the original material. Furthermore, the author wishes to thank Geert Bruinooge, Carina Fransen, José Geurts, Hank Hermans and Adrie Ykema for helpful comments and suggestions.

1. Introduction

Statistics Netherlands is in the middle of radical changes. A number of very heterogeneous driving forces pose major challenges that can only be met when the way the institute operates is thoroughly reconsidered. Van der Veen (2007) gives an overview of the present situation and the

[1] Deputy Head of the Department for Process Development and Quality Control at Statistics Netherlands, b.braaksma@cbs.nl
[2] The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands

challenges ahead in a broad context. In particular, efficiency and quality of key statistics must be improved, while the administrative burden is significantly lowered at the same time.
Statistics Netherlands is facing a number of major challenges. On the one hand, budget cuts and the IT maintenance burden call for increasing efficiency. On the other hand, the statistical landscape is changing rapidly. This applies both to the input side, where attention shifts from primary data collection to the use of administrative registers; and to the output side, where users ask for increasing flexibility while maintaining high quality standards.
Below we first give a brief overview of the general setting. We describe the driving forces behind the redesign and the approach of the Master Plan. Next we turn to the 'facilitators' realised at a central level: enterprise architecture, generic services and the statistical toolbox. Then we turn to the redesigning process itself with extra attention for the important example of the chain of economic statistics. We conclude with some remarks on program organization and a number of lessons learned.

2. Driving forces behind the redesign

The redesign is driven by a number of forces that act on Statistics Netherlands, for example general goals of the Dutch government. Below we discuss these driving forces.

2.1. Increasing efficiency

Statistics Netherlands is – like all other Dutch government institutions – confronted with ongoing budget reductions due to the intention of central government to increase efficiency of government agencies in general and to reduce the number of government employees. In the next years, reductions of at least 15–20% have to be realized for our institute. Hence, the redesign must yield substantial efficiency gains.
A specific aspect concerns the reduction of IT cost. The current application landscape of Statistics Netherlands is considered less than transparent, entangled and (too) expensive. Over the years, many different dedicated 'tailor-made' systems have been developed and implemented by our IT division. The number of systems amounts to several hundreds, including subsystems several thousands. All of these systems need to be maintained, and this costs a lot of effort from both IT and statistical staff. The maintenance burden has grown to such proportions that necessary innovations suffer. Hence, an important objective of the redesign program is to clean up the IT landscape and reduce the maintenance cost. New and more efficient ways for dealing with systems development have to be developed.

2.2. Reduction of administrative burden

Another stimulus for the redesign derives from the national pressure to reduce the administrative burden on the business sector. Although statistical reporting is only a small part of the total administrative burden, the private sector and politicians insist that this burden should be reduced. Statistics Netherlands fully supports this endeavour. The pressure to minimize the reporting burden is illustrated by a recent proposal[3], endorsed by a majority of the Dutch Parliament, which requires that Statistics Netherlands stops all data collection from small and medium-sized enterprises. Although the Dutch government immediately replied that European obligations make it impossible to carry out the proposal, it indicated that major efforts must be made to reduce statistical reporting.

[3] Motion of Aptroot and Van As, members of the House of Representatives of the Dutch Parliament, 2005–2006, 29 515, nr. 111

2.3. Quality issues

The quality of statistics needs to be improved, where quality is to be understood in a broad sense. For example,

- coherence between subsequent (monthly/quarterly/annual) estimates for the same variables;
- improved timeliness of publications;
- consistency between data on the same population;
- reproducibility of production processes;
- transparency of compilation methods,

are all aspects of quality in this sense. In particular for key economic indicators it is clear that steps must be taken to improve the current situation. Criteria need to be developed that facilitate guarding these quality aspects.

A specific quality aspect concerns flexibility and robustness. This applies both to the output and input side of the statistical process. On the output side, the increased complexity of society and the rapid changes it is going through imply an increasing demand for reliable statistical information. The focus is shifting towards thematically presented information providing more insight into how developments and sectors are related. Statistics Netherlands wants to speed up its flexible response to the changing needs for information. This requires combining data from various sources. On the input side, the increasing dependency on external data suppliers like register holders calls for quick and flexible responses to complications with external data, and at the same time the production process should be robust enough to cope with environmental fluctuations.

3. The Masterplan

In order to stay in control of the competing challenges mentioned above, an ambitious modernization program, the Masterplan 'Rekenen op Statistiek'[4], was started in 2005. The original focus of the Masterplan was on reducing IT cost. Although this goal is still considered important, recently some of the other driving forces have come more to the front. In particular, the needs to increase general efficiency and to reduce administrative burden are now primary concerns. Accordingly, the focus of Masterplan activities shifts somewhat from time to time. Below we present some important elements of the Masterplan. Ypma and Zeelenberg (2007) describe the general approach in more detail.

One of the first steps in the Masterplan was the development of a comprehensive high-level enterprise architecture for Statistics Netherlands. It is service-oriented and serves as a reference framework for the future organisation of the statistical processes. All statistical processes will gradually be adapted or redesigned in such a way that they adhere to the architecture. The HEcS+ program described below is only one, though important, case.

In addition, projects were started aiming at developing a number of generic services to be used throughout the institute. First and foremost among these are the Data Collection Service, a newly created business unit that will take care of all statistical data collection within the institute; and the Data Service Centre, that will take care of storage and retrieval of important data. A first instalment of a toolbox to support statistical data processing has been decided upon and steps are being taken to further complete this toolbox.

[4] Counting on Statistics

With the construction of these prerequisites for the redesign under way, attention is shifting towards carrying out the actual redesign projects for the many production processes that are in place. Responsibility for these local redesigns essentially lies with local management, but resources for development, support and advice are also supplied at central level and distributed according to centralized prioritisation procedures.
An important condition on the side while carrying out the Masterplan is that business continuity must be guaranteed at all times, in particular the quality and timely delivery of those statistics that are legally compulsory must not be compromised in any way.

4. The enterprise architecture

The general ideas of the Masterplan are represented in a comprehensive enterprise architecture, consisting of an overall business architecture and a software architecture. It serves as a reference framework for the future organisation of the statistical processes and supporting IT systems. Most importantly, the new business architecture definitely leaves the old 'stovepipe' approach, where each statistical process has its own relatively autonomous systems and procedures, for a more harmonised approach that unifies design and production concepts. The most striking feature of the new software architecture is that the traditional approach to develop new systems and applications in-house is abandoned in favour of an approach where COTS (commercial off-the-shelf) software is considered as a first choice.
Work on improving and extending the architecture is still under way. For example, the original version of the architecture excluded the use of open source software. Since the open source model is now considered to have some clear advantages and also because the Dutch government has a policy of actively promoting the use of open source software, the architecture was adapted to enable the use of open source software. Another specific output category concerns access to anonymized micro-data for selected research partners, both on-site and remotely.

4.1. The IAF framework

The enterprise architecture is developed using the Integrated Architecture Framework (IAF) approach. IAF is service-oriented, which makes the approach flexible and future-proof. It gives a concise, complete and practical overview of all relevant aspects. It is useful for the development of systems, but its main function is to show the relations and interdependencies of the organisation and IT in all its major elements. For many architectural efforts only part of it is of paramount importance. Still, the overall framework provides the right scope and positioning of the aspects and areas involved and the way they need to be included. The framework is applicable at enterprise level of architecture, as well as at domain and project level of architecture. It describes a basic design of the statistical process at an abstract level in terms of activities, functions and objects using a uniform terminology. Key elements of the architecture are expressed through a set of general principles that incorporate the goals of the Masterplan. Below we describe the key elements from a business perspective in somewhat more detail.

4.2. Output orientation

Output requirements are leading in the (re)design of statistical processes. This implies that output needs to be specified at an early stage of redesign. Legal obligations, in most cases concerning European regulations, are to be met at all time. For other less formal output requirements it needs to be (re)assessed whether merit and costs match, observing the consequences of the driving forces described above.

An important new element in the output strategy of Statistics Netherlands is that the institute aims at responding more rapidly and flexibly to new or incidental demand for statistical information from society, in particular strategic relations. In many cases, information demand can be satisfied by re-using data that are already in-house, having been collected for another purpose. In other cases, specific contract research may be an option. This strategy is supported explicitly by the enterprise architecture, for example by the introduction of the notion of steady states (see below).

4.3. Chain orientation

The statistical production process is modeled as a value chain[5] of coherent subprocesses operating between steady states. Every sub-process adds value to the data being processed. The chain for the statistical process consists of two elements: business activities and information products. Information products form the input and output of the business activities. At a high level of abstraction, the information products represent the objects and the business activities the processes that can be distinguished within the statistical process. If the processes are ordered in a logical manner, it is possible to represent the process chain in the form of the diagram in Figure 1. The diamond shapes represent the business activities (processes), whilst the rectangular shapes and file icons represent the information products (objects).

The schematic diagram of the process is static and does not do justice to the dynamic nature of the statistical process. That dynamism consists of repeatedly going through parts of the chain, depending upon the possibility of reusing processes and information products. Changes at the front end of the chain may have an effect upon the subsequent logical stages in the chain, but will not have a (direct) effect upon the things that preceded them in the chain. The onward dependency upon the chain is strongly determined by the extent to which re-use takes place at the start of the chain.

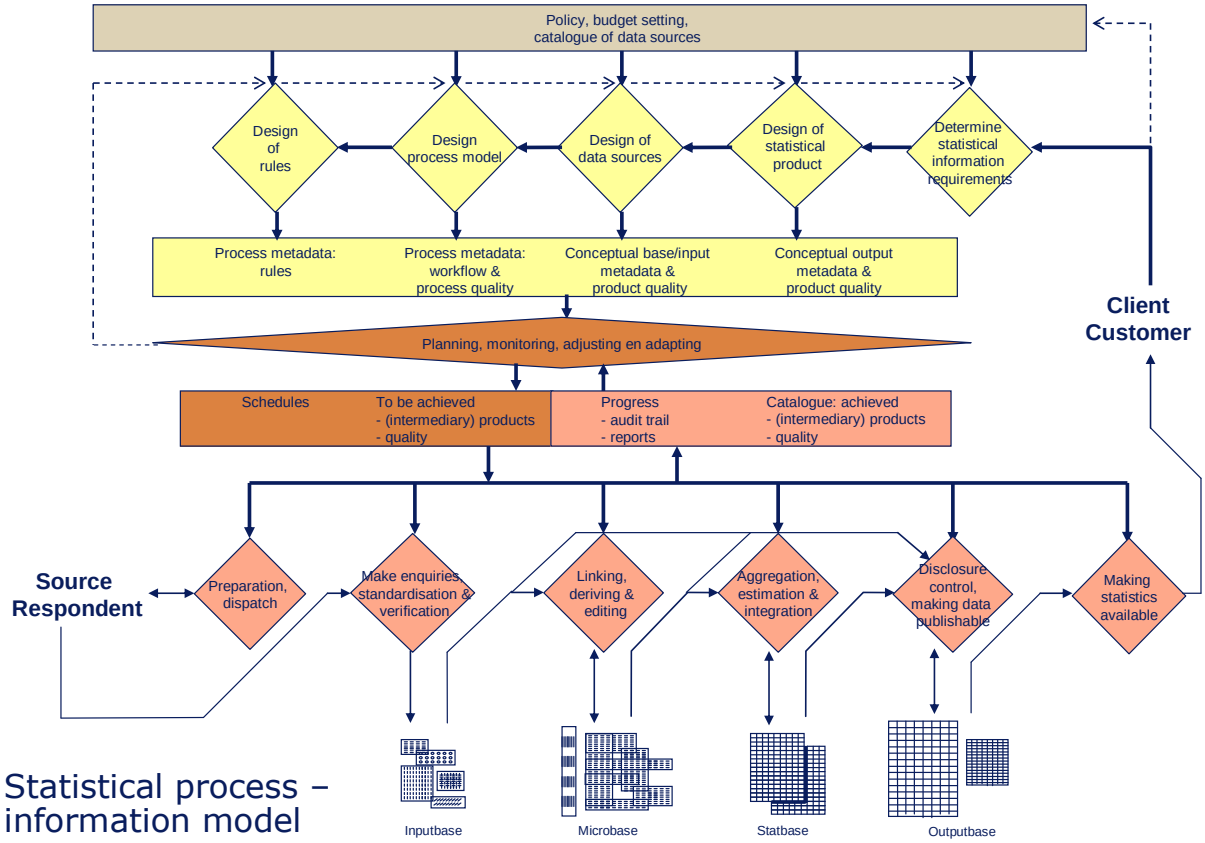
The chain orientation applies both to production and design processes. In a design process, the necessary data sources and required end-products are defined and specified and the production model of the activities to be carried out and the rules governing the production method are laid down. A production process involves working through a chain of activities in a logical sequence from data collection right through to publication. The process is carried out under the direction of the entire chain, in which the cycle plan, monitor and adjust/adapt are included.

4.4. Chain management

It is imperative that production progress can be planned and controlled across the whole chain against quality and performance expectations. This implies that statistical processes should generate high quality process data in the form of audit trails, performance indicators, process logging and/or progress reports. When anomalies occur, necessary measures have to be taken through a system of chain management. For example, when the response rate to a survey is likely not to meet the predefined standards, the data collection process (or unit) must signal this to the client process (or unit) in time such that adequate measures can be taken. For example, either the collection efforts could be increased or the lower response rate can be accepted and dealt with by statistical techniques.

[5] Or rather, it is not a chain but a network of processes. Many cross-relations exist between processes in the chain.

Figure 1. *Schematic representation of the statistical value chain, including the design process chain and chain management*



Chain management plays an increasingly important role, the more data collection shifts from primary (surveys) to secondary (registers) and as the integration of processes and statistics progresses. Re-use and mutual dependencies are increasing, as a result of which the statistical process is becoming ever more complex and harmonisation ever more difficult. As external dependencies increase, internal chain management will be influenced to an even greater extent by external factors.

4.5. Steady States

The architecture identifies a number of steady states for the chain of statistical processes. Each steady state contains data and accompanying metadata in an explicitly described state of processing[6] with predefined quality. For example, raw data obtained from a sample survey or from an administrative register are considered as a first steady state; output disseminated to the Statistics Netherlands website, press releases, Eurostat or other channels is considered as a final steady state. A number of additional steady states are identified between these extremes. Mutual use and re-use of source data and intermediary results is facilitated through interaction of the

[6] It is important to realize that steady states are a conceptual notion. It is tempting to visualize them as databases, and often this will indeed be the case, but it is not necessarily so.

steady states. At the same time, they guard the coherence of the whole statistical system. Furthermore, steady states play an important role in chain management.
Within the enterprise architecture, at first four types of steady states were identified: the *input base* for source data, the *micro base* for statistical micro-data, the *stat base* for aggregate statistical information and the *output base* for publishable data. Soon it became apparent that at least two further steady states should be identified: the *pre-input base* for untreated source data (in particular registrations) and the *post-output base* for data that have actually been published. The reason behind this lies in contacts with the outer world. For internal use, data are stored in a format that is standardized as much as possible (preferably open standards). However, external suppliers and external customers are not obliged to conform to these standards. In order not to burden the statistical production with a great diversity of formats and technical data models, the pre-input base and the post-output base were added to the architecture.
Steady states also play a role in version management. There is a trade-off between meeting agreed quality standards and timeliness of statistical output. This conflict can be solved by producing several quality versions of the 'same' data set. The solution has a drawback, however: too many quality versions confuse the users. So, the number of different quality versions should be limited and designed in advance. A steady state may thus contain different versions of the 'same' data set when they have a clear role in the statistical process, for example preliminary next to definitive estimates.
The idea of explicitly distinguishing steady states seems to be new for a statistical enterprise architecture. Thus the fact that statistics production is essentially all about transforming data and/or metadata, according to well-defined procedures, is explicitly acknowledged.

4.6. Rule-driven processing.

One of the architectural principles reads

> *'The formulation and maintenance of rules is a statistical activity that is carried out separately from the systems that ensure the automatic generation of data.'*

Although this principle may seem a natural requirement, it is not. In many legacy systems the rules are hard-coded in such a way that it is difficult to apply even the slightest modifications. For example, legal and administrative rules applying to wages, salaries and social benefits are often changed. This implies that source data for income statistics are changing as often and thus, the processes and supporting systems relying on these sources are also subject to frequent changes. In the traditional approach this leads to cumbersome change procedures. Change requests first have to be specified by subject matter specialists, then implemented by IT staff, and the system changed accordingly has to be tested by both IT staff and subject matter specialists.
The rule-driven approach simplifies the situation by allowing subject matter specialists to implement changes all by themselves without need to rely on IT staff. Rules for processing must be documented explicitly and stored externally to the IT applications, which then act simply as rules engines to execute the rules. This should allow for relatively quick changes to the rules, if necessary. Moreover, this approach is expected to contribute significantly to the Masterplan goal of IT cost reduction.

4.7. Domain architectures

The enterprise architecture described above serves as a blueprint for the whole institute and thus lacks sufficient detail for more specific domains. For example, the data collection domain is only

described sketchily in the enterprise architecture. When preparations for a generic data collection service were started, the first project (early 2006) was to elaborate a more detailed architecture for the data collection domain, within the context of the IAF framework. This domain architecture has helped to determine the scope of data collection and also to establish a common language to unite all local data collection dialects.
Other specific domain architectures relate to the output domain, the Data Service Centre, the statistical throughput process and the chain management process. A specific statistical example is the domain architecture for the chain of economic (GDP-related) statistics, which is described below. The latter example can as well be considered a high-level project architecture. Similar high-level project architectures have also been established for other large redesign programs like the redesign program for Consumer Price Indices and the redesign program for Demographic Statistics.

4.8. IT architecture

Since this paper focuses on the business side of the Masterplan, we will only briefly touch upon the IT part of the enterprise architecture. The IT architecture applies both to software solutions and the underlying infrastructure. Like the business architecture, the IT architecture is formulated in the context of the IAF framework. Of course, business and IT architectures are fully aligned.
Again, important elements are formulated through key principles, see Project Group Architecture (2007). For example, both solutions and supporting infrastructure are required to be scalable. The IT realization of business services should follow a service-oriented approach that facilitates re-use of components, both on the software level and below. Another important principle is that when new solutions are needed, first re-use of already existing solutions should be considered and then solutions derived from COTS (commercial off-the-shelf) packages.
Especially the latter principle constitutes a major departure from the approach in the past, where custom-built applications were usually developed in-house by own IT staff. This approach is now abandoned because it is not cost-effective and, moreover, it is not feasible to keep up with the rapid succession of technological innovations in the IT field. In the new situation buying goes before making. Major parts to be made will be put to the market to be built according to specifications by an outside partner.

4.9. The methodology series

In parallel to the development of the enterprise architecture, a series of preferred statistical methods is being documented. This series consists of methods and techniques that are commonly regarded as sound and sufficiently applicable in common statistical production. This will lead to a fairly complete documentation of all methods that are in use at Statistics Netherlands.
All methods are described according to a standard format that includes suggestions for use, as well as limitations and prerequisites. The descriptions are not meant to give a thorough scientific exposition, but they give enough information to a developer of statistical processes. They are reviewed internally and then published on a dedicated intranet site. In future, the methodology series will be made publicly available on the Statistics Netherlands internet site.
Reason for development is twofold. The first goal is to document the way that statistics are being produced in a transparent way for the users. The second goal is to gradually drive statistical production processes to use methods that are included in the methodology series. This is particularly relevant for designing or redesigning processes. The second goal is much more ambitious and will have be to be pursued with care. It requires some allowance for exceptions that undoubtedly occur and also calls for proper maintenance of the methodology series: on the

one hand, methods may become obsolete and on the other hand, new methods may need to be included.

5. The data collection service

In line with the enterprise architecture a single business unit is being created for data collection. This unit will take care of all statistical data collection within the institute, including related logistics and administrative issues, such as reminding late respondents and data entry. All modes of data collection will be addressed: including web-based surveys, field interviews, and paper questionnaires. A special unit will be assigned to register handling. Other statistical units will thus be able to concentrate on their core business of data processing.

5.1. Approach

The data collection unit is being positioned as a generic service. The service interacts with their (subject-matter oriented) clients, data processing units, by means of so-called *data collection orders*. These orders are standardised forms that for example specify the size of a survey, collection modes, non-response rates allowed, and the time frame within data collection should take place. The task of actually formalizing data collection orders is carried out by an intake organisation inside the data collection unit, based on agreements discussed with their clients. Within the specifications agreed, the data collection service is free to organise its own ways of working. This applies for example to organising various kinds of staff resources in an efficient way.

The data collected by the service is delivered to the respective clients in a raw state (without further processing). In this sense the data collection service is a 'dumb' service: it simply collects what it is being told to collect, without knowledge of the content. This implies that in case it appears necessary to re-contact data providers with questions on what has been reported –for example, when possible errors are identified in editing or analysing survey data– the subject matter unit concerned should do this.

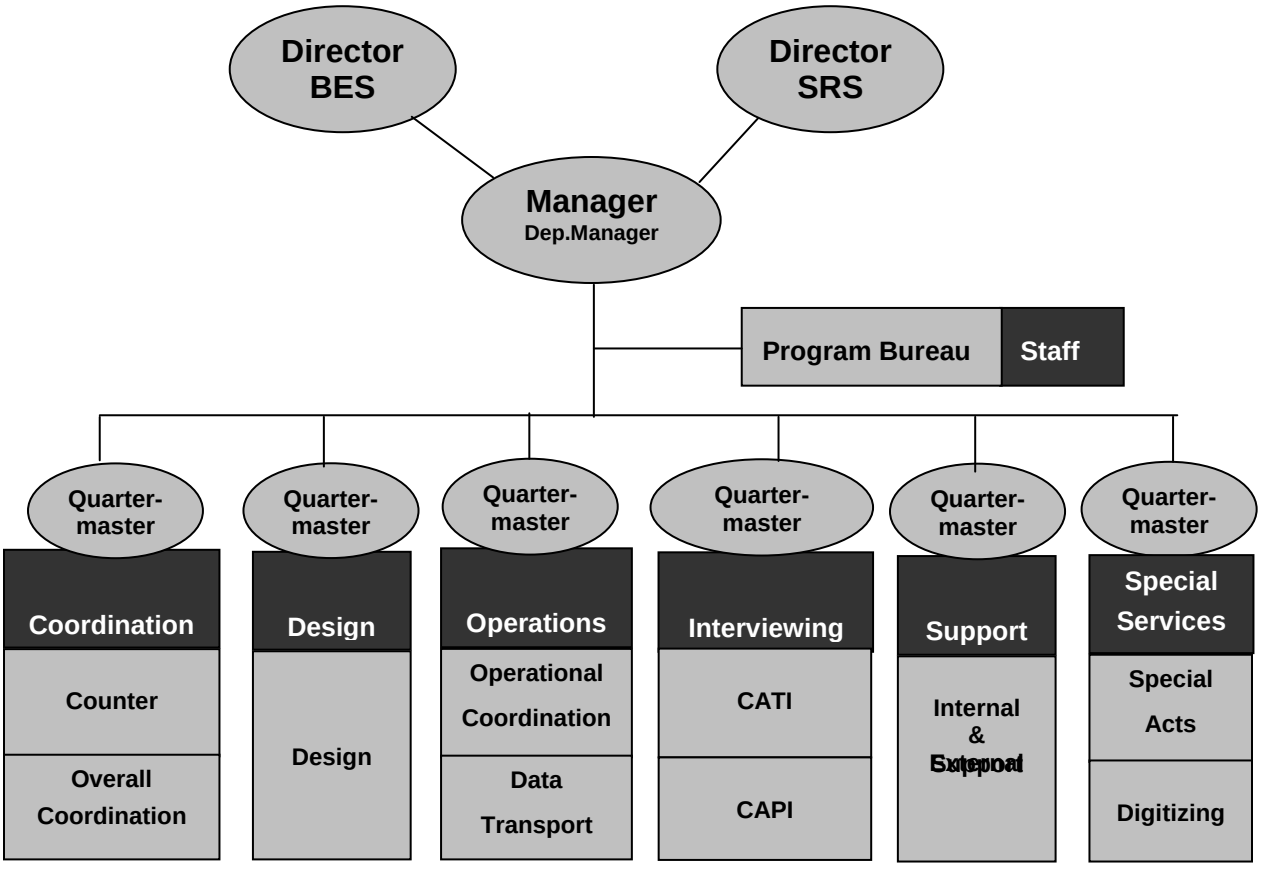
5.2. Organisation

In the past, subject matter-oriented business units usually took care of their own data collection. In a major reorganisation of Statistics Netherlands in 1999–2000, this traditional stovepipe-oriented approach has already been partly abandoned and the organisation of statistics production has already become more process-oriented. Logistic and administrative processes were generally carried out by dedicated units, but still managed on a decentralised level.

As a first step of centralisation, all local units that were identified as having data collection as their core business were brought under a single management. As a second step, a new organisational structure was created and staffed with people from the centralised local units. This new organisation will first operate on an pilot basis. After some time, when the new situation is evaluated positively, it will be approved formally. This formalisation will probably coincide with a major reorganisation of the entire institute that is envisaged for the near future. The new unit is expected to yield large efficiency gains[7].

[7] The efficiency target is to reduce the current data collection formation of 336 fte (including field interviewers) with 70 fte. Outside the direct scope of the new unit, an additional 27 fte effciency gain should result from the new data collection strategy

Figure 2. *Pilot organization of the Data collection service*



5.3. Data collection tools

Data collection is a separate discipline and requires its own set of tools. Coming from a decentralised situation, it is natural that the legacy toolset is somewhat heterogeneous. For example, in order to manage contacts with source data providers, three main CRM (customer relations management) systems were used, all of them tailor-made and developed in-house. It was decided that these systems should be replaced by a single CRM application to be based on COTS software. Mid 2008 the development of a Siebel solution has started. It is planned to be completed by 2010 in two stages.

The full transition to the new CRM application will take 2–5 years. An important difficulty that has to be overcome is that all three legacy systems contain additional functionality, like process management facilities. This kind of functionality is outside the scope of a pure CRM application and will therefore not be included in the new Siebel system. To make sure that ongoing production processes are not disrupted, additional measures must be taken to ensure continuity. It even turned out to be necessary to develop a legacy system further, in order to guarantee production of critical statistical output.

In order to support the core primary data collection processes, it has been decided to standardise on the survey tool Blaise, developed in-house. This should cover the whole spectrum of

questionnaire design, building and maintenance. This means for example that new facilities for E(lectronic)–surveys and XBRL-based reporting should only be developed for this tool.

6. The Data Service Centre DSC

In order to facilitate the exchange of statistical data and metadata between internal users, a dedicated Data Service Centre (DSC) is being created. Like the Data Collection Service described above, it is being positioned as a generic service. The main functions the DSC provides are storage and retrieval of statistical data and associated metadata. This refers in particular to data and metadata associated with the steady states that have been identified in the context of the business architecture. Standardised tools for access of the DSC are being developed.

A leading principle of the DSC is that data should be collected, not delivered. That is to say, sets of statistical data (so-called steady states) are made available to internal users in a standard format and accessible via a catalogue. An authorized user who requires a DSC data set can retrieve this set and is responsible for transforming it into a suitable format and further processing himself. Making data available to *external* users is part of the dissemination strategy of Statistics Netherlands and is not a primary DSC responsibility.

6.1. Approach

The Data Service Center (DSC) collects the definitive statistical data. Note that DSC's data storage not only concerns the published data, but also the "frozen" intermediary datasets (both microdata and aggregates) that have led to the published data. From the moment these data and associated metadata are being stored in the DSC, they are accessible to any authorized user inside Statistics Netherlands. To be precise, all internal data exchange will be channelled through the DSC service.

Needless to say, all data that are ultimately stored in the DSC are first handled by statistics compilers during their normal production process. During the processing period, the data are still in a dynamic state which is not yet suitable for DSC storage. Local databases will store the data at that time.

The DSC has two main functions. First, data can be shared throughout Statistics Netherlands, which enables statistics compilers to re-use data produced by others for their own production processes. Second, it functions as a permanent archive of data that can be used to reproduce statistical results. In order to perform these functions, the DSC offers a number of elementary services to manage, query, search and access its content. In addition, it provides some supporting services that are needed to e.g. secure the statistical data and its descriptions against misuse. Like the Data Collection Service, the DSC is a 'dumb' service in a certain sense. It has no subject matter knowledge of its content. Ownership of its content remains with the party that originally provided the data –unless otherwise decided– and only the owner can decide about authorisation rights.

6.2. Metadata

According to one of the guiding architectural principles there are no statistical data without metadata. This principle is enforced by the DSC: it simply refuses data that does not comply.

Originally, a Metadata Service Center (MDSC) was planned alongside the DSC as a generic service in its own right, although it is of course closely linked to the DSC. For example, populations and classifications applicable can be defined independently of specific datasets. Later

on it was decided to integrate DSC and MDSC into a single service[8]. The main function of the MDSC is now to serve as a catalogue for the DSC, in order to ensure that all statistical data are described appropriately. For all data offered for storage in the DSC, associated meta descriptions are loaded into the catalogue. Thus any internal stakeholders may utilize the catalogue to search for data that may be relevant for their own statistical process, but are produced elsewhere in the organization.
Ultimately the catalogue should handle various kinds of metadata. During development of the DSC most attention has been paid to conceptual metadata, that is metadata relating the statistical data to real-world concepts and allowing the user to understand and interpret the statistical data. This kind of metadata is considered the most important from a user perspective. A model was developed for conceptual metadata that deals with both micro-data and aggregate data. This model is explicitly linked to the enterprise architecture. A variant of the same model, dealing with data in a dynamic state, is under development and will be implemented eventually in the Digros tool.
Besides conceptual metadata, DSC distinguishes structural metadata and managerial metadata. Structural metadata are meant to describe the structure and the presentation of statistical datasets from a user perspective. This may involve the heading of a dataset, the choice of the variable names, the syntax of variables, keywords, and so on. The purpose of managerial metadata to administrate the owner of the dataset and its users or subscribers. This type of metadata is important both to manage the authorization and authentication process in view of securing datasets and from a chain management point of view.

6.3. IT support

The creation of a DSC is a non-trivial task. It touches upon many aspects of statistics production and affects many parties in the institute who all have their own interests and goals. This non-triviality is reflected in the development of IT-tools supporting DSC. From 2007, steps were taken to realize a full DSC system based on COTS components to be supplied by external partners. In 2008, however, when the first version of the system was tested it appeared that the approach was too ambitious. The approach was downscaled and it was decided to expand DSC functionality gradually using a step-by-step approach. Only after a step has been tested thoroughly and put in production, the next step is taken up.
From 1 January 2009 a first version of DSC has become operational, including a meta system based on Documentum. This first version has limited functionality and for example, only supports micro-data. Support for aggregate data is scheduled for early 2010. The road to be taken further depends on experiences and perceived needs. Many questions relating to the future service level of DSC have to be answered along the way, based on efficiency and usability considerations and practical restrictions. For example, will DSC in future only provide retrieval services for ‘full’ data sets or will it also offer a selection service? Should DSC offer a linking service for data pertaining to the same population but coming from different data sets? Is a notification service useful, announcing the arrival of new data to subscribers? Is it a DSC task to achieve a certain level of coordination (see Willeboordse, Struijs and Renssen, 2005) between data sets?
One of the main practical restrictions concerns the amount of storage needed by DSC. Statistics Netherlands has over the course of 2008 and 2009 outsourced its IT infrastructure, including virtualisation of workstations. This means also that all data are stored in a data warehouse

[8] A bit confusing, the resulting combination of DSC and MDSC is also called DSC.

maintained by an external party. Rough estimates indicate that future capacity of the data warehouse will not suffice unless measures are taken. Some of these measures inevitably relate to the amount of data to be stored by DSC. For example, what data are kept and for how long, and whether it is allowed to compress data, which reduces storage but increases access times.

7. Statistical data processing

Broadly speaking, the domain of statistical data processing contains all activities (including monitoring) that are required to convert the collected statistical data from the pre-input base into either a publishable file of micro-data or a publishable table of aggregate data in the post-output base. A distinction can be made between processes that intrinsically improve the data and processes that do not. A process of the former type typically consists of collecting data from a certain steady state (in the input base, micro base or stat base), checking the quality (optional), statistically improving the data, checking the quality again and, if approved, returning the data to a different steady state (micro base, stat base or output base). This constitutes the statistically relevant core of data processing that will be elaborated upon in this chapter.

Data processing between the pre-input base and the input base consists merely of processes that do not intrinsically improve the data, such as executing technical transformations of formats and deleting unwanted data. Similarly, data processing between the output base and the post-output base consists of customer-specific processes that do not intrinsically improve the data either. Examples include technical transformations of formats, making selections tailored to an (external) customer, and recoding. Writing press releases is another typical activity that could be placed between the output base and the post-output base.

7.1. Basic process steps and the toolbox

The statistically relevant core of data processing occurs between the input base and the output base. Usually a wide variety of activities must be carried out in order to obtain useful statistics from source data. These activities typically depend on the nature of the data and the goals of the compilation process at hand. Nevertheless, some general recurring elements can be identified.

The underlying idea for standardization of data processing is that most of the processes can be built up from a list of basic process steps or rather elementary business function, and that each business function can be implemented using a limited number of standard tools. The selection of standard tools depends on the statistical methods that are required for data processing. This yields an important connection to the methodology series (see above). A first rough draft of a list of elementary business functions was drawn up by Renssen, Wings and Paulussen (2008). In a research project in 2009, this list has been completed and refined further.

The question remains which tools are preferred for a given business function. Obviously, the preference depends on the business context in which the tools are to be used, such as education level and experience of staff members, size of datasets to be processed, processing frequency, and complexity of the methodological solution. The choice will also depend on total cost of ownership, technical considerations like the facilities offered by a tool to link to other systems, and architectural requirements like abilities to allow for rule-driven processing and to separate design from implementation.

All tools currently in use plus a number of additional tools were rated by Renssen, Wings and Paulussen (2008) according to several criteria. Two of these were critical: the tool *must* be able to handle metadata and the tool *must* be able to separate design and implementation. Mainly based on the critical criteria, a first instalment of the toolbox was proposed, consisting of 18 preferred tools (see Box 1). In Box 1 both general and specific tools are included. General tools are

required for virtually all forms of processing, as distinguished from specific tools that cover niches such as statistical disclosure control, matrix algebra and seasonal adjustment. Note that this toolbox is meant to provide the statisticians with tools that enable them –to a large extent– to specify and implement their own processes, such that involvement of IT staff can be minimized. The IT department still keeps its own set of tools for software development and related tasks, but these tools are –as before– not available to non-IT people.

Box 1. First instalment of the toolbox

General tools: Clementine, Manipula, RuleBurst, SPSS

Specific tools: μ- and τ-Argus, Bascula, Blaise, BizTalk, Classification server, Documentum, JC, PES, Quat, SharePoint, Siebel, Slice, Characterization module, X12-Arima

The criterion that eligible tools must be able to separate design and implementation refers to the architectural principle on rule-driven processing. In order to accommodate the rule-driven approach, three new tools were selected: Clementine for statistical ETL (extraction, transformation and loading of data), BizTalk for BPM (Business Process Management) and RuleBurst for BRM (Business Rules Management). In fact, other tools in the toolbox can also be considered ruled-driven. For example Blaise comes with its own scripting language.
Note that commonly used tools like Access and Excel are absent from the toolbox, because they do not support the critical criteria.
The toolbox approach and its first instalment have been approved at the level of the Board of Directors and was accepted surprisingly quickly by statisticians and developers. In order to manage tactical and operational decisions on the toolbox, a Configuration and Change Management Board has been set up consisting of the heads of R&D departments and the business and IT lead architects.

7.2. Blind spots

For several reasons, the first instalment of the toolbox is not complete. The set does not cover all business functions and methods and thus contains blind spots[9]. Furthermore methodological innovation, e.g. new estimation techniques, may require additional tools. Additional projects are needed, not just to fine-tune and secure the future set of tools, but also to establish for each preferred tool the relation to a business function and one or more statistical methods.
An important reason for the blind spots is that, until recently, mostly tailor-made IT systems were developed to support a specific statistical process. This applies in particular to those areas where relatively much specific subject matter knowledge needed to be incorporated. Examples are data editing and estimation, where applicable rules are usually very specific. In many cases, these rules were hard-coded and thus 'hidden' in the IT systems. This leads to systems that are not transparent and also difficult to maintain. A second reason is that in a few cases existing tools are considered technically obsolete and need to be replaced by more up-to-date alternatives. A third reason is that some tools are used outside the scope for which they were initially intended. Blaise is a good example: due to its powerful possibilities and adaptability, it is sometimes used as a 'Swiss army knife' far outside its intended scope.

[9] On the other hand, duplication may also exist among the 18 preferred tools. But this is subtle: some tools might be considered duplicate from a functional point of view, but different (and possibly complementary) as regards performance, manageability, adaptability, usability, etc.

It is possible that further investigation yields that other tools currently in use cover some of the blind spots, but probably some blind spots are genuine and should be covered by new solutions. Hence, it is worthwhile to explore whether COTS solutions exist, or whether development of new generic solutions for these blind spots is feasible. These solutions are then to be added to the set of standard tools. In particular, new solutions chosen should be in line with the rule-driven approach of the enterprise architecture

Statistics Netherlands intends to put significant resources into further elaboration of the toolbox, based upon the approach sketched by Renssen, Wings and Paulussen (2008). Already some areas are defined where common tools should be developed. For example, MacroView is developed to become a generic tool for top-down analysis and editing. This tool is scheduled to be applied in the HEcS+ program (see below) and the redesign program for transport and traffic statistics. Further examples include a generic tool for interactive (manual) data editing; a generic database approach (Digros) for dynamic data (i.e. data being processed before they reach DSC); and a tool for gluing together different process steps. All these tools are developed in a Statistical Informatics program consisting of a number of relatively small projects.

8. Redesign of individual processes

So far we have mostly discussed generic elements of the redesign, like the architecture, generic services and the toolbox. But the ultimate goal of the Masterplan is to redesign all statistical processes according to the architectural principles, which implies adoption (where applicable) of the toolbox and the generic services. Redesign of all hundreds of statistical processes is a huge task and can only be accomplished step by step.

8.1. Project architectures

As mentioned above, each redesign project is required to adhere explicitly to the enterprise architecture. To accomplish this the redesign project must develop a project architecture consisting of a business architecture document (BAD) and a software architecture document (SAD). Before the IT implementation part of a redesign project starts, the architectural documents need to be finalised and approved by a formal review procedure.

The BAD contains both a description of the output and input data, including accompanying metadata and expected quality, and a sufficiently detailed process design including a description of the processing rules and the intermediate steady states used. The BAD should also describe which generic services are to be used and how.

The SAD is based on the BAD and contains a description of the software tools to be used, data storage requirements and their interactions (also with generic services). One of the main goals of the SAD is to decide to what extent it is feasible to develop a system using standard software from the toolbox. This is the preferable option. In some cases it may be decided, however, that development of a custom-built system is still the best option. Finally a hybrid approach may be advised, consisting of a combination of standard tools and custom-built components. Strict rules apply in order to guarantee that the system to be built complies with architectural principles and does not become a new 'black box'.

Currently the feasibility of introducing a MAD[10] (methodological advisory document) is being considered. The MAD should contain an outline of envisaged statistical methods which is

[10] We are aware that the English words *MAD*, *BAD*, and *SAD* do not elicit very nice associations, yet decided to leave the acronyms like that, maybe due to a slightly deviant sense of humour; at least they are easy to remember…

sufficient to judge the feasibility of the whole redesign approach. The statistical methods should preferably originate from the methodology series (see above).

8.2. Redesign approach

There is no centralised approach for the redesign of the business process other than the general architecture. In principle each redesign project can follow its own path as long as it adheres to the enterprise architecture and obeys a number of agreed guidelines:

- use the Prince-2 method for project management;
- apply the RUP method for software development;
- define a clear and convincing business case;
- appoint a business executive at the level of at least head of department.

Although there is no centralised approach for redesigns of business projects, each redesign project can call for support on three levels.

First, within each of the three statistical divisions a development department exists that includes project managers, business analysts and tool experts. These are available to help staff redesign projects. Each of these categories has specialised knowledge that is usually not available within the individual subject matter departments. For example, tool experts (a newly created discipline) may assist in implementing solutions using the statistical toolbox. Allocation of these resources is done within each division based on their own priorities.

Second, on a central level business and IT architects are available for advice and support, organised in central redesign teams. They have a deep knowledge of the enterprise architecture. Using central redesign teams is not obligatory, but virtually guarantees that tests for compliance with architecture are passed. Allocation of these resources is decided on a central level, based on prioritisation and compliance with a number of intake criteria, including for example existence of an explicit project plan and sufficient commitment from the subject matter department.

Third, also on a central level, a limited number of software developers are still available for development of tailor-made systems in cases where solutions consisting entirely of standard tools do not suffice. Allocation of these resources is also decided on a central level, based on prioritisation within pre-allocated budgets per statistical division.

Although a generic approach to transform statistical processes is not explicitly aimed at, some standards are gradually emerging. This derives partly form central stimuli and partly from local initiatives. Each project must comply with the same mandatory elements and the central redesign teams use a common approach (the software developers already have one based on RUP). In addition, specialists in development departments are organising themselves in interest groups to exchange experiences, share best practices and develop common approaches.

One of the key elements that is recurring in successful projects is that relatively small, dedicated and motivated interdisciplinary teams yield the best results. Clear commitment from the responsible management is also a key success factor.

8.3. Example: energy statistics

A number of projects have already been carried out along the above lines. One exemplary project concerns the energy statistics. This is a rich and complicated set of mutually related statistics, serving many different internal and external users, that are based on a large variety of different sources. Over the years a complicated set of processes and supporting systems had developed, which were badly documented and contained a lot of less-than-transparent manual operations.

When it was decided (early 2007) to start a redesign project for energy statistics, the first few months were devoted to a thorough analysis of the situation, including definition of the project

vision and ambition, and design of the outline of the new process and identification of steady states according to the business architecture. Then a project initiation document (PID, a Prince-2 artefact) was written describing planning, resources and deliverables of the project.
It was decided to leave methodology unchanged at first. Focus should be on disentangling processes and supporting them with new systems and tools, while satisfying important additional requirements such as improved flexibility[11] and better consistency of output figures. Methodology will be reconsidered at a later stage, when the advantages of the well-structured new situation can be exploited. It was also decided to focus on core statistics first.
A system has recently been completed (it is now being tested) using the standard tools Blaise and Clementine; and the database tool Digros. Moreover, limited use of Excel, Access and VBA code is made. Blaise is used as a temporary solution for manual editing, until a generic solution becomes available. Similarly, Excel is used as a temporary solution for generating output tables, also by lack of a dedicated generic solution for this purpose; and the graphical user interface has been implemented in Access ADP. The project has been carried out with only limited IT resources[12], in line with the new strategy of Statistics Netherlands to rely less on IT staff. Additional functionality will gradually be added to the system.
The project was initially scheduled to be finished by end 2008, but that has proven to be too ambitious due to several reasons.

- the complexity of the project was greater than anticipated. This is a side effect of the lack of transparency of the old situation, which inhibits good complexity estimates. In particular, it has proven very difficult to make the (often intuitive) knowledge of subject matter specialists explicit.
- it was a pioneering project for the new redesign approach, which means that in many cases it was not clear how to proceed. Sometimes even approaches had to be developed from scratch, like testing the tool-based system;[13]
- the project is carried out with an interdisciplinary team under single project management, instead of the traditional setting where separate business and IT projects come each with their own management. Although the interdisciplinary approach has clear advantages, it has also taken a significant amount of time to get used to it.
- it proved difficult to obtain and secure project resources.

Several measures were taken to deal with the setbacks the project encountered. For example, the project was broken down in three clusters in order to cope with the complexity, the planning was adapted and in many cases creative solutions or *work-arounds* were developed. All in all, the project is considered a success and an example for other redesign projects. Project results, methods and approaches can be re-used.

9. Redesign of the Chain of Economic Statistics

In the fall of 2006, preparations have started for drafting the HEcS+ redesign program. The HEcS+[14] program aims at redesigning the whole chain of economic statistics from an integrated

[11] e.g., it should be easy to incorporate new energy sources such as bio-fuel or to compile new publications
[12] So far, some 55% of resources is supplied by the development department, 20% by the subject matter department, 20% by the IT division (including 5% hired externally) and 5% by the Methods division according to the project administration (of course, these percentages are approximate and may change slightly over time)
[13] The standard testing procedures and testing environments at Statistics Netherlands are designed for testing applications built by IT staff using development tools like Visual Studio
[14] The acronym HEcS stands for 'Herontwerp Keten Economische Statistieken' or 'Redesign of the Chain of Economic Statistics'. It is a pun on the Dutch word 'Heks', meaning 'Witch', and thus hints at the touch of magic

perspective. It focuses on the core of economic statistics relating to economic growth, in particular structural business statistics and related short-term turnover statistics. The system of National Accounts also belongs to the scope. The program is responsible for both R&D and implementation of the results. A lot of R&D work is carried out, supported by practical experiments and feasibility studies. The outlines of the new system for economic statistics have been developed using a pressure cooker approach involving all key stakeholders.
The HEcS+ program has three main goals, in line with the general Masterplan goals: increasing efficiency[15], reducing administrative burden and improving quality. Reduction of GDP (gross domestic product) revisions is the main quality issue driving the HEcS+ program. An analysis carried out in 2005 on annual and monthly/quarterly turnover estimates from primary statistics yielded differences that were considered too large, and which had an adverse impact on the ensuing quarterly (including flash) and annual GDP estimates. The analysis yielded a number of recommendations to improve the situation. It turned out, however, that pursuing these recommendations was often not feasible in the then-present situation due to essential limitations in the processes, methods and systems being used. Also important is the (political) urge to reduce the administrative burden on enterprises. That means that the use of registers (mostly fiscal) will have to increase.
Migration will be a complex issue, since discontinuities must be kept to a minimum. Moreover a benchmark revision of National Accounts is planned for 2013 , in order to accommodate changes in SNA/ESA and a time schedule for introducing the new NACE classification has been agreed upon internationally. An open question remains the structure of the new HEcS+ organisation. Both keeping the current division structure and creating a new dedicated HEcS+ division are options. All these aspects will have to be synchronized one way or another with the HEcS+ redesign.
Below we describe some of the key elements of the HEcS+ program.

9.1. Top-down approach and chain orientation

Economic statistics are traditionally compiled ‘bottom-up’ with considerable detail. The statistics considered as the most important, however, are aggregated indicators like the growth of GDP, commonly referred to as ‘the economic growth’, and retail trade turnover.
The focus of the HEcS+ approach is on compiling such indicators ’top-down’ as accurately as possible, and with as little revisions as possible. Specific results by line of industry and other refinements are considered less important and may be of lesser quality. In all cases, however, the approach should be such that mandatory output[16] can be compiled with sufficient quality.
This approach applies for example to annual structural business statistics processing, where traditional unit-by-unit interactive editing is abandoned. Instead units will only be subject to editing after analysis at meso-aggregate level, using the new MacroView tool, yields implausible results.
One of the main conclusions from the 2005 analysis was that communication across the whole chain of economic statistics was less than perfect. The various chain links tend to optimize their local processes without being aware of the needs further on in the chain. This leads to quality loss

---

that is needed to carry out all that needs to be done. The ’+’ was added later on in order to stress that the program should not only cover R&D but also implementation of its results
[15] The efficiency target is to reduce the current HEcS+-related formation of 325 fte with 65 fte
[16] For example, reesulting from European legislation or agreements with important national customers

and inefficiencies. Hence, it was decided to start a redesign of the whole chain of economic statistics and organize it as a coherent program.
Where possible, improvements in the current production processes were already introduced. One of the most important measures taken was introduction of working groups by line of industry, consisting of experts across the whole production chain. These working groups are meant in the first place to discuss and where possible solve production issues, but they also serve as quality assurance for the HEcS+ redesign projects.

9.2. Direct estimation of turnover from VAT data

Turnover statistics, both annual and infra-annual (monthly/quarterly), will be compiled as much as possible on the basis of Value Added Tax (VAT) data. Moreover, also the production processes will be similar. The reasons for this are twofold. On the one hand, according to the new Dutch Statistical Law that entered into force in 2003, Statistics Netherlands is only allowed to collect data when no adequate administrative sources are available. On the other hand, it is expected that use of a single source for both annual and infra-annual turnover statistics contributes significantly to reduction of GDP revisions. In line with the top-down approach mentioned above, emphasis in the compilation process is on obtaining high-quality results in particular on macro levels.
To improve usability of VAT data, the way statistical units are constructed in our General Business Register is adapted by taking into account the structures used for fiscal purposes. Previously more than half of the VAT data could not be used for statistical purposes due to problems of matching fiscal (VAT) units to statistical units in the General Business Register. This matching problem occurs in particular for larger units and is distributed unevenly across specific lines of industry and thus introduces bias, which is highly undesirable when VAT data are used as the basis for compilation of statistics. The price to be paid for the improved usability of tax data is less homogeneity in economic activity of statistical units, but the trade-off is considered worthwhile.
Introduction of the first stage of VAT-based turnover statistics, relating to monthly retail trade turnover statistics, was scheduled for early 2009. This introduction had to be postponed, however, because Dutch government decided to change tax declaration rules as a financial crisis countermeasure[17]. At the time of writing it is still being investigated how to handle this government decision. It clearly poses new methodological challenges and may require continued surveying in some cases. Apart from this setback, preparations for compilation of quarterly and annual turnover figures continue as planned. Full-scale introduction of VAT-based estimation is scheduled for 2010.

9.3. Tailor-made approach for large and complex units

In the traditional approach, large and complex units are basically treated the same as any other units. Such units are by their nature part of almost any business survey carried out by Statistics Netherlands. This causes undesirable duplication in questionnaires and results in inconsistencies for individual units, which pose major problems at the level of national accounts. The problem has worsened significantly in recent years, for example due to globalisation issues. Many of the larger Dutch companies operate on a multinational level with complicated cross-border relations.

[17] To be precise, enterprises who previously had to declare VAT turnover on a monthly frequency may now choose to declare on a quarterly frequency. It is as yet unclear how many and which enterprises will make this choice

In order to deal with these issues, the HEcS+ program aims at developing a tailor-made approach for the top three hundred large and complex enterprise groups[18]. For each unit at hand, data collection will be streamlined and data checking, editing and imputation will take place in an integrated environment. Consistent results on these units are delivered to individual statistical processes for joining with results for smaller units and further processing in order to produce specific output. Additional editing is not allowed, however.
From pilot studies it appeared that the approach works. Moreover, the enterprises involved appreciate the fact that Statistics Netherlands tries to avoid duplication in questionnaires and unnecessary questions and thus contributes to reduction of administrative burden. The approach is not expected to yield direct efficiency gains, however. Successful implementation is rather labour-intensive and will require well-qualified staff.

9.4. Redesign of National Accounts

A full redesign of the system of National Accounts takes place as part of the HEcS+ program jointly with the redesign of primary business statistics. Important elements in the National Accounts redesign are a rationalization of the National Accounts compilation process, separation of the error detection (using a new tool, the *Dashboard*) and balancing processes and the introduction of sophisticated automatic balancing techniques where possible. Moreover the level of detail for compiling annual supply-and-use tables has decreased significantly. The National Accounts redesign has both efficiency and quality goals[19]. Efficiency goals partly rely on measures taken elsewhere. For example, when the tailor-made approach for large and complex units is in place, it is expected that the job of National Accountants to check and balance their sources becomes significantly less labour-intensive.

9.5. The HEcS+ architecture and the Expert Team

A central role in the HEcS+ program is played by the HEcS+ architecture. This is a domain architecture that refines the enterprise architecture (business part) to a level of detail sufficient to sketch a clear perspective of the chain of economic processes. Moreover, the HEcS+ architecture also serves to organize the HEcS+ program, to determine the scope and interactions of the individual projects and to provide a starting point for the migration path towards the new situation, including organizational changes. The HEcS+ program is organised in several projects that each construct their own project architectures (both business and IT parts) within the framework given by the HEcS+ architecture and the enterprise architecture. All in all, the HEcS+ architecture has proven a successful tool with a number of unexpected applications.
While pursuing the HEcS+ program, it appeared that not only process-related issues had to be solved (which is an architectural activity) but also content-related issues. In order to deal with these –sometimes long-standing– content-related issues a team was established consisting of senior experts. This expert team is to decide on such issues, which in some cases calls for a Solomon's judgment. For example, the treatment of continuity issues (e.g. population mutations) and the composition of the Top-300 enterprise group are issues for the expert team.

[18] The number of kind-of-activity units involved is yet an order of magnitude bigger
[19] Since National Accounts do not collect data themselves, reduction of administrative burden is not a goal

10. Program organization

A large program (or rather, collection of programs) like the Masterplan requires special attention to program organization aspects like governance, resources and communication. Below we discuss these aspects of program organisation.

10.1. Governance structure

Due to the sheer size of the Masterplan it has proven difficult to establish a governance structure that is both efficient and not too bureaucratic. Over the years, gradually improved insight and changes of focus and perspectives have led to various changes in the governance structure. Below we describe the current situation (autumn 2009) which was established after a re-assessment of the situation by an external partner.

Seven programs are identified at central level: Data Collection, Standard Tools (the Toolbox), HEcS+, three redesign programs for each of the statistical subject matter divisions, and a combined program for the IT services division and central staff services. The last one of these programs has not yet been clearly defined, the other programs will be discussed below.

Formal responsibility for the Masterplan rests (as before) with the deputy Director General, who is advised by two boards. A Central Portfolio Board, whose members are directors, advises on strategic choices and allocation of central resources, and monitors progress of central projects. A Central Architectural Board deals with strategic architectural issues. A Program Bureau supports the deputy Director General and the two boards in operational and administrative duties. When necessary, issues are escalated to the full Board of Directors or the Executive Board.

For the Data Collection Service and the DSC, implementation managers have been appointed who are responsible for implementing these generic services. A specific IT program 'generic services' was, until summer 2009, responsible for development of tools for the generic services Data Collection and DSC and for completion of the statistical toolbox. The IT program was thus placed formally outside the IT division, which was (and still is) working on its reorganisation. In July 2009 it was decided to delegate further development of Data Collection tools to the Data Collection implementation manager. Responsibility for the statistical toolbox sub-program, including DSC tools. will remain on a central level.

Responsibility to redesign individual processes lies with local management and this is now formally organised in a redesign program for each statistical subject matter division. This means that the directors of the three statistical divisions BES (Business Statistics), SRS (Social and Spatial Statistics) and MSP (Macro-economic Statistics and Dissemination) are accountable for their own processes. All local redesign projects have their own project boards, usually chaired by the head of the receiving business unit. Progress is monitored locally but significant delays and other problems also have to be reported centrally. The scarce central resources are allocated centrally by the Central Portfolio Board.

The HEcS+ program for the redesign of the chain of economic statistics is, unlike other redesigns, managed by a Program board chaired directly by the deputy Director General because it involves two divisions, BES and MSP. A dedicated HEcS+ Program manager and a Program Bureau support the deputy Director General and the HEcS+ Program board. In addition, projects within the HEcS+ program have their own project boards that operate under a specific project mandate granted by the Program board. Project resources are supplied by the Program board and managed by the project boards.

10.2. Resources

In order to realize the Masterplan goals, the Ministry of Economic Affairs has provided an additional budget of € 24.6 million on top of the normal budget. The largest part (approximately 90%) of this budget is allocated to the IT program 'generic services', for example to finance systems development by external partners. In a pre-selection process, seven possible partners were selected. One of them has been selected as managerial partner, also providing strategic support. The other six partners could apply for the different projects like the creation of a CRM solution. Further external partners are sought for possible outsourcing of development and maintenance. In addition to this out-of-pocket budget, internal resources amounting to 120 fte (of which 50 fte internal IT staff) are allocated to the IT program 'generic services' uses some, distributed over the years 2006–2010 with a peak in 2007–2008.

For the HEcS+ program, internal resources amounting to 230 fte are allocated, of which 50 fte internal IT staff. These resources are distributed over the years 2007–2011 with a peak in 2008–2010. No external partners are used, apart from a limited number of project managers. For other redesign projects the allocation of resources is decided separately according to standard internal budgeting procedures. No estimates for the number of fte's involved, but this will certainly amount to hundreds of fte's over the next years. The remaining 10% of the additional budget (€ 2.3 million) is allocated to support specific aspects of the redesign projects, for example for hiring tool experts (a new discipline) for a limited time.

10.3. Communication

It is important to communicate about the Masterplan both internally and externally. There are important differences between both types of communication, however.

External communication with stakeholders such as strategic relations is mainly a responsibility of line management. These stakeholders are interested mostly in changes in statistical output due to the redesigns, for example when moving to new methodologies or new data sources like registrations. They are informed through regular communication channels like the various advisory councils and also, when considered useful, through dedicated meetings and/or written communications such as newsletters. Apart from the Central Statistics Committee (the supervisory board for Statistics Netherlands) in general no direct communication about the introduction of generic services takes place.

The explicit goals for internal communication are to generate support and to overcome resistance. This applies in particular to those elements of the Masterplan that directly affect daily work of employees. Both the HEcS+ program and the program to create a Data Collection service have special ongoing projects on communication. These projects produce on a regular basis publications on program progress like newsletters, FAQs and they maintain communication sites on the internal web, and they coordinate oral presentations of plans and results, meetings etc. The IT program 'generic services' had no special communication project since its results reach most employees indirectly.

11. Lessons learned

In a program as big as the Masterplan, almost inevitably false steps are taken and the approach must be reconsidered from time to time. Below we present some of the lessons we learned so far.

11.1. People

Ultimately all redesign projects and the ensuing new production processes must be run by people. In many cases people are highly motivated and clearly enthusiastic for redesigns. They are aware

of the need for change and contribute actively. In other cases, however, people may be reluctant to support redesign projects, because it involves extra work on top of regular production tasks, because changes will happen in the working manners they feel comfortable with, or because they feel downright threatened to loose their jobs; which in some cases is inevitably true. Support may also suffer from delays or inherent uncertainties in redesigns. Generating support is a management task that calls for open communication and dialogue, while taking existing ill feelings and concerns seriously. Sometimes it is necessary to explain over and over again the reasons behind a redesign.

One of the main concerns is furthermore that the 'right' people are often not sufficiently available. This applies to lack of specialized developers like tool experts, architects and business analysts; lack of skilled IT personnel for development and testing; and lack of production staff with knowledge of the new tools. In particular, there is a lack of people with overview like senior subject matter specialists, experienced methodologists and project managers. This lack of qualified people hampers progress. In some cases, steps could be taken to fill the gaps. For example tool experts for new tools like Clementine and RuleBurst were selected from existing staff and then trained 'on the job'. Experienced project managers, developers and system testers were hired on temporary contracts. In other cases solutions are less clear-cut, for example in the case of senior subject matter specialists there is serious competition for resources between ongoing production and redesign tasks.

11.2. Architecture and redesign projects

The fact that the results of the Masterplan are so far only in part available adversely affects support. It is difficult to explain that a redesign project should comply with the enterprise architecture when the toolbox is not complete, or when the generic services are not yet fully operational. A related problem is that standard tools sometimes offer less (specific) functionality than the custom-made systems people are used to.

Besides this, there is a natural reserve to start using the generic services and the toolbox. First, given the complexity of the IT-landscape it is difficult to replace only parts of the local systems and still obtain efficient processes. Second, there is a discrepancy between the short term targets of the local processes and the long term corporate challenge to increase overall efficiency of production processes. Third, different local processes have different needs and expectations with respect to each service, which may not entirely match the scope and quality of the implemented version.

It is essential that a redesign project is aware of the consequences of its design choices from the outset. A recent redesign of Income Statistics was based on a full income tax register that has become available recently[20]. A business architecture document (BAD) and a software architecture document (SAD) were written, and only when the SAD was reviewed by the lead IT architect it appeared that the approach proposed was not feasible due to IT constraints. Storage of the full dataset would amount to 4.9 Terabyte and processing time would take far too long. This example shows that the mere fact that a full register is available should not automatically result in full processing. The incident has led to an urge to start a redesign with a methodological advisory document (MAD) that outlines key methodological choices, such that it is possible to assess the consequences early on.

Central (corporate) and decentral (local) goals for redesigns differ. For example, the primary concern of local subject matter units is to produce statistical output as agreed. Operational

[20] In the past, a relatively small sample of selected variables from this register was used for statistics production

management is often reluctant to change proven processes because this entails a risk for production continuity and because this draws away resources from production tasks. Strong commitment from higher management is needed to convince operational managers to cooperate. When carrying out a redesign project the natural tendency of local units is to optimize their own situation. This means that incentives must be created to make it more attractive to satisfy corporate goals, for example by granting or discontinuing IT budgets depending on (non)compliance for certain corporate goals. For example, special budgets were made available for 'early adopters' to use the DSC.

The idea of chain management proves difficult to implement. Although there is enthusiasm to start using *quality* and *process indicators* associated to steady states, there is reluctance to introduce chain management in a systematic way. On the positive side, more or less informal approaches to chain management, are introduced, resulting in better communication between different actors within the chain. For example in the case of the chain of economic statistics working groups by line of industry are established (see above) and also regular meetings of the heads of units to discuss chain issues are now organized.

11.3. Tooling and systems development

Construction of a well-tuned toolbox is a difficult task. At first, the aim was to select a very limited set of tools (Clementine, RuleBurst and BizTalk) that should suffice for all data processing work. Although this approach was opposed by many the path was persevered for some time, which has led to a number of non-performing applications and thus waste of resources. After some time this was considered a dead end and a more extensive set of 18 tools was selected (see above). Support for this approach is considerable, but on the other hand there is a tendency to submit many additional 'private pet tools' for inclusion in the toolset. Attention must be paid to prevent the toolbox from growing too large.

The initial goal to stop completely with developing custom-built systems has proven too ambitious. In some cases, dedicated software or a hybrid solution is still to be preferred to rule-driven processing. This may be for performance reasons, or because non-standard functionality is required that is not available in the toolset or COTS solutions. Development of a custom-built or hybrid system is now decided on the basis of a software architecture and requires explicit management approval.

A specific problem, as yet unsolved, concerns maintenance and support responsibilities. This used to be straightforward. All systems were developed under responsibility of the IT division, which was also responsible for maintenance and (technical) support. In the new situation, other parties may also be involved in systems development, in particular when rule-driven systems are concerned. An even more complex situation arises for hybrid systems, partly consisting of custom-built components developed by the IT division and partly consisting of rule-driven components developed locally. Who is then responsible for correct operation, maintenance and support of the full hybrid system?

Another only partially solved problem occurs when developing and testing a new system. A well-established approach is in place for custom-built systems. These are developed and submitted to technical tests in a separate development environment accessible to IT staff only, then transferred to a production environment accessible to users for acceptance tests and finally production. This approach can not simply be copied to rule-driven and hybrid systems, but the proposed alternatives turn out to be difficult to implement.

11.4. Outsourcing

Application development using external partners is a radical change from the past practice of developing applications in-house with IT personnel that understands statistics and is used to rather loose and 'dynamic' specifications that are easily modified. External partners require more strict specifications and, given their own business goals, are less willing to accommodate change requests (and if they do, additional costs are likely[21]). Moreover, the deliverables must be tested thoroughly. These are lessons that we have learned 'the hard way'.

Another lesson learned concerns the European rules for the awarding of contracts. Statistics Netherlands had little previous experience with these rules and violation may lead to substantial fines. The European awarding rules turn out to be very bureaucratic and cause additional delays amounting to several months, which makes planning well ahead necessary. Furthermore they require detailed specifications in advance that also cannot be simple altered or amended after a contract is awarded (a 'level playing field' condition), which makes it very difficult to operate flexibly.

11.5. Governance

Prioritisation between development and regular production and how to organise this prioritisation process proves to be difficult. Many changes in the overall governance structure were driven by a need to get a better grip on planning and control of redesign activities at corporate level. The current approach is to allocate resources based on central assessment of project proposals and corresponding business cases. Once a project is granted a budget, the responsible project board should watch over adequate execution within budget constraints.

Local managers like to be in control of their own situation. It is important to acknowledge this fact and give them responsibility for redesigning their own processes, including resources needed to carry out the task successfully. On the other hand, it must be made clear what room is available for manoeuvring within corporate frames like the enterprise architecture and the toolbox.

It is important that a receiving party for Masterplan products is identified in time in order to set targets. In some cases, a specially appointed change or implementation manager is needed who can help with the migration strategy, with the (re)organization of the receiving unit(s) and with creating support within the organization. For example, development of the domain architecture for data collection was started early 2006, before the generic data collection service had been established as the receiving organization and an implementation manager was appointed. Due to lack of a receiving party the project was allowed to run too long and the level of detail became too high. The goals of the implementation manager turned out not to be fully met[22], as observed in an external audit for this project in 2008. Building upon these experiences, one of the first steps taken when establishing a specific architecture for the output domain end 2008 was to identify its future owner. Furthermore, additional stakeholders were identified and involved in the creation process.

A final issue relating to governance is commitment. It is crucial that in particular the top level management shows clear commitment and communicates the vision, strategy and goals pursued with confidence. It should also be ensured that they are consistent over (a long) time, in order to avoid the risk that people tend to wait for the next turn.

---

[21] In fact, modifying specifications when developing in-house also causes additional costs, but these are less visible

[22] Nevertheless, the data collection domain architecture has proven to be helpful in developing the generic data collection service and also to further develop the enterprise architecture. For example, the notion of pre-input base was developed here and later added to the enterprise architecture.

12. Concluding remarks

We realise that the redesign program as described above is very ambitious and is far-reaching in its consequences. There is, however, a high pressure to achieve results. If the Masterplan does not deliver its promises, Statistics Netherlands will run into serious problems.

We have now entered a crucial phase. Some key generic elements of the Masterplan, like the enterprise architecture and the data collection service, are already in place and are acknowledged as successes. On the other hand, elements like the DSC and the statistical toolbox are still in their infancy and need to be developed further. The actual redesigns of statistical processes according to the Masterplan principles have started, exemplified by the cases of energy statistics and the HEcS+ program described above. There lies a huge task ahead for the whole organisation to carry through the program, where difficult choices yet need to be made and priorities set. The success of the program is without doubt crucial for Statistics Netherlands.

References

Bredero, R., Dekker W., Huigen, R., and Renssen R. (2009). Statistics Netherlands Architecture; Context of Change. Statistics Netherlands discussion paper 090187.

Huigen, R., Bredero, R., Dekker W. and Renssen R. (2009). Statistics Netherlands Architecture; Business and Information model. Statistics Netherlands discussion paper 09018.

Project Group Architecture (2007). Statistics Netherlands Architecture, Reference Architecture Information Systems Conceptual, version 2.0. Internal publication

Renssen, R. and Van Delden, A. (2008). Standardisation of design and production of statistics; A service oriented approach at Statistics Netherlands. Paper presented at the IAOS congres 'Reshaping Official Statistics', October 2008.

Renssen, R., Wings, J. and Paulussen, R. (2008). Processes, methods and tools. Statictics Netherlands internal report

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, February 2007. Statistics Netherlands, Voorburg.

Van Velzen, J.H. (2005). The embedding of a uniform statistical process. Paper presented at the UnStatCom/ECE Conference of European Statisticians, May 2005.

Willeboordse, A., Struijs, P. and Renssen, R. (2005), On the role of Metadata in Achieving better Coherence of Official Statistics, Statistics Netherlands.

Ypma, W.F.H. and Zeelenberg, C. (2007). Counting on Statistics; Statistics Netherlands' Modernization Program. Paper prepared for the Seminar on Increasing the Efficiency and Productivity of Statistical Offices at the plenary session of the UnStatCom/ECE Conference of European Statisticians, June 2007.