Redesign of the Chain of Economic Statistics in the Netherlands

Barteld Braaksma, Gert Buiten

1Statistics Netherlands, Postbus 24500, NL-2490 HA Den Haag, The Netherlands

Abstract

In 2006, Statistics Netherlands started the ambitious HEcS+ redesign program for the chain of statistics relating to economic growth. It has three main goals: increasing efficiency, reducing administrative burden and improving quality.

Major innovations in the HEcS+ program include an integrated approach for large enterprise groups, use of VAT data for turnover estimates, sophisticated new tools for analysis of intermediate results and automated balancing methods for National accounts. Important parts of the HEcS+ program have been implemented in 2010 and 2011; the remainder will follow in 2012.

The HEcS+ program has benefited greatly from a comprehensive architecture. This architecture simplified integral management of the development program and scoping of individual projects. Moreover, a new and efficient organisational structure for economic statistics has been created along the lines of the HEcS+ architecture.

This contribution aims at giving a broad overview of the HEcS+ program and its architecture, covering both the ideas and their implementation. Not only the successes, but also the false steps that were -almost inevitably- taken will be presented.

Key Words: architecture, efficiency, administrative burden, quality, economic growth

1. Introduction

In the autumn of 2006, preparations started for drafting the HEcS+ redesign program. By the end of 2011, the program had largely been completed. Major parts of the program are already implemented. Full implementation is expected to be realized by the end of 2012. This paper focuses on the important role played by architecture throughout the whole, rather extensive, redesign program.

The structure of the paper is as follows. Chapter 2 describes the main goals, scope and key ingredients of the HecS+ program. Chapter 3 highlights some important features of Statistics Netherlands’ enterprise architecture. Next, Chapter 4 shows the various uses for the domain architecture developed for the HEcS+ program. Chapter 5, finally, presents some lessons learned.
2. The HEcS+ program

The HEcS+ program aimed at redesigning the whole chain of economic statistics from an integrated perspective (Zwijnenburg, 2012). It focuses on the core of 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.

In this chapter we briefly discuss the goals of the program and its key elements, including their current (mid 2012) status.

2.1 Goals of the program
The HEcS+ program had three main goals: increase efficiency, reduce administrative burden and improve quality (Van der Veen, 2007; Braaksma, 2009).

Reduction of GDP\(^2\) revisions was the main quality issue driving the HEcS+ program. An analysis carried out in 2005 on the reasons for GDP revisions yielded differences in annual and infra-annual (monthly/quarterly) turnover estimates from primary statistics that were considered too large, and which had an adverse impact on the ensuing quarterly (including flash) and annual GDP estimates.

The efficiency target was to reduce the HEcS+-related staff with some 20%, from 325 fte to 260 fte. These reductions affected all organisational units involved.

Also important was the ongoing political urge to reduce the administrative burden on enterprises. That meant that the use of registers (mostly fiscal) would have to increase. Moreover, according to the new Dutch Statistical Law that entered into force in 2003, Statistics Netherlands is only allowed to collect data with surveys when no adequate administrative sources are available.

2.2 Chain orientation
One of the main conclusions from the 2005 analysis on GDP revisions was that communication across the whole chain of economic statistics was less than perfect. The various chain links tended to optimize their local processes without being aware of the needs and possibilities in other parts of the chain. This led to quality loss and inefficiencies. Hence, it was decided to organize the redesign of the whole chain of economic statistics as a coherent program.

Below we present the main processes from the HEcS+ program. Between parentheses are the (Dutch) acronyms for these processes.

2.3 Direct estimation of turnover from VAT data (DRT)
Turnover statistics, both annual and infra-annual (monthly/quarterly), are compiled as much as possible on the basis of Value Added Tax (VAT) data. The reasons for this are

---

1 The acronym HEcS stands for ‘Heronde Keten Economische Statistieken’ which means ‘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 that was needed to carry out all that needed 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

2 Gross Domestic Product, from which economic growth figures are derived
twofold. Apart from the contribution to reducing response burden, it is expected that use of a single source and processing method for all turnover statistics contributes significantly to reduction of GDP revisions.

The new approach has been implemented mid 2011 for the first two lines of industry (wholesale trade and car trade), while others gradually follow. Full-scale introduction of VAT-based estimation should be finalised during 2012.

2.4 Adaptation of the General Business Register (EHB)
To improve usability of VAT data, the way statistical units are constructed in our General Business Register had to be reconsidered. Dutch law and the taxation system make it often attractive for companies (especially the larger ones) to set up complicated legal structures. 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. Hence, the construction algorithm for statistical units was adapted to take fiscal structures explicitly into account. This also meant that the register of fiscal units, maintained by the Tax Office, had to be included as an additional source for constructing the General Business Register.

2.5 Tailor-made approach for large and complex enterprise groups (ConGO)
In the traditional approach, large and complex enterprise groups were basically treated the same as any other units. Enterprises in these groups are by their nature part of almost any business survey carried out by Statistics Netherlands. This not only causes undesirable duplication in questionnaires but, more importantly, also results in inconsistencies in the data from different surveys for individual units. This poses 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.

In order to deal with these issues, the HECs+ program has developed a tailor-made approach for the top three hundred large and complex enterprise groups\(^3\) (Vennix, 2012). For each of these units, data collection is streamlined and data checking, editing and imputation takes 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 of the top units is not allowed at this stage anymore.

For some 1500 enterprise groups just below the top three hundred a less labour intensive survey approach is still being used. They are considered too important to just rely on VAT data, but not important enough to justify the tailor-made approach.

The new approach was implemented end 2010 and is gradually being improved. Apart from the quality effects, the enterprises involved appreciate the tailor-made approach because it contributes to reduction of the perceived administrative burden.

2.6 New approach for annual business statistics (NOPS)
The introduction of direct turnover estimates from VAT data on the one hand and the tailor made approach for large and complex units on the other hand had a profound

\(^3\) The number of kind-of-activity units involved is yet an order of magnitude bigger
impact on the compilation processes for annual business statistics. Both new approaches had to be integrated and combined with the survey approach for the 1500 enterprise groups described above and the sample survey for the 60 thousand smaller enterprises.

At the same time, it was decided to shift the focus in the compilation process from manual micro-editing to top-down analysis of meso-aggregates (Aelen and Smit, 2009; Hacking and Ossen, 2011). A number of automated algorithms for micro-editing were introduced. Only when the meso-analysis shows implausible results, the underlying units are manually examined for anomalies, and possibly edited. A new tool, MacroView, was introduced to support this analysis. It turned out that some existing systems could be reused, which reduced the development costs.

The new approach was first used in 2011 for the compilation of the structural business statistics for reference year 2009, though without the VAT estimates (for which the processes and systems were not yet ready). The VAT turnover estimates are now gradually incorporated.

2.7 Redesign of National Accounts (Integratie)
A full redesign of the production system for the Dutch National Accounts (including sector accounts and government finance statistics) was carried out as part of the HEcS+ program. Important elements in the National Accounts redesign were a rationalisation 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 (Bikker et al., 2010).

The National Accounts redesign had both efficiency and quality goals. Efficiency goals partly rely on measures taken elsewhere. For example, introduction of the tailor-made approach for large and complex enterprise groups made the job of National Accountants to check and balance their sources significantly less labour-intensive.

The new system of National Accounts has now (mid 2012) become fully operational with finishing the Sector accounts part. The Supply and Use tables part (resulting in GDP estimates) had been completed a year ago.

3. The enterprise architecture of Statistics Netherlands
Starting in 2005, Statistics Netherlands has developed a comprehensive enterprise architecture (Bredero et al., 2009; Huigen et al., 2010), consisting of a business architecture part and an IT architecture part. It serves as a reference framework for the organisation of the statistical processes and their supporting IT systems.

The new business architecture abandons the old ‘stovepipe’ approach, where each statistical process has its own relatively autonomous systems and procedures. Below we describe some key elements.

---

4 Actually, the redesign of National Accounts started in 2005 as an independent program before being incorporated in the HEcS+ program

5 Since National Accounts do not collect data themselves, reduction of administrative burden is not a goal
3.1 Value chain
Statistics production is modelled as a value chain\(^6\) of coherent sub processes operating between *steady states*. Every sub-process adds value to the data being processed. The value chain consists of two elements: business activities and information products. 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.

![Diagram of the statistical value chain](image)

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

If the processes are ordered in a logical manner, the process chain can be represented in the form of the diagram in Figure 1. The diamond shapes represent the business activities, whilst the rectangular shapes and file icons represent the information products.

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 re-using 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 what preceded them in the chain. The onward dependency upon the chain is strongly determined by the extent to which re-use takes place.

The chain orientation applies both to production and design processes. In the design process, the necessary data sources and required end-products are 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

---

\(^6\) Or rather, it is not a chain but a network of processes. Many cross-relations exist between processes in the chain.
activities in a logical sequence from data collection right through to publication. The process is carried out under chain management procedures, in which the cycle plan, monitor and adjust/adapt are included.

3.2 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 with predefined quality. For example, raw data obtained from a sample survey or from an administrative register can be considered as a first steady state; output disseminated to the Statistics Netherlands website, press releases, or Eurostat 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 steady states.

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.

3.3 Chain management
It is imperative that production progress can be planned and controlled across the whole chain against quality and performance expectations. When anomalies occur, countermeasures have to be taken through a system of chain management. For example, when the response rate to a survey does not meet the predefined threshold, the data collection process (or unit) must signal this in time to the client process (or unit), such that adequate measures can be taken. Then, either the collection efforts could be increased or the lower response rate could be accepted and dealt with by statistical techniques.

Chain management plays an increasingly important role as data collection shifts from primary (surveys) to secondary (registers). Re-use and mutual dependencies are increasing, as a result of which the statistical process is becoming ever more complex.

3.4 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. 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 (and cheaper) changes to the rules.

7 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.
3.5 Domain architectures
The enterprise architecture serves as a blueprint for the whole institute and thus lacks sufficient detail for more specific domains. Hence, a number of domain architectures have been developed. For example, the data collection domain is only described sketchily in the enterprise architecture. When preparations for a generic data collection service started, the first project (early 2006) was to elaborate a more detailed architecture for the data collection domain. 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.

A specific statistical example is the HecS+ domain architecture, which is described below. This can also be considered as high-level project architecture. Similar high-level project architectures have been established for other large redesign programs like the redesign program for Consumer Price Indices and the redesign program for Demographic Statistics. The use of several general steady states assures data interchange between the domains.

3.6 IT architecture
Since this paper focuses on the business part of the architecture, 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. A key principle is, for example, that both software solutions and supporting infrastructure are 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.

4. The HEcS+ architecture and its use
Right at the start of the HEcS program, an outline of the overall HEcS architecture was developed. This architecture is an elaboration of the generic Enterprise Architecture of Statistics Netherlands described in chapter 2. The HEcS+ architecture has played a crucial role in the whole program, both in expected and unexpected ways.

The HEcS+ architecture partly inherits elements directly from the enterprise architecture (Renssen and Van Delden, 2008). Rule-driven processing is, for example, heavily used in the new production system of the Dutch National Accounts, where three different systems for automated corrections and balancing have been introduced, and also in automated editing and the new MacroView tool for meso-analysis. On the other hand, it adapts or interprets enterprise architecture elements specifically for the HEcS+ situation.

4.1 The HecS+ steady states
It soon appeared that the concept of steady state would play a key role in the HEcS+ architecture (Braaksma and Van Delden, 2009). At first, three phases were distinguished:

- The Source phase. Automated checks, edits and standardisation within a single source. This yields a set of microdata within each individual source.
- The Combination phase. Editing, imputation, aggregation and combination of data from various sources. This leads to a set of aggregate data (in some cases still microdata) based on all available sources.
• The Completion phase. Confrontation with other data sets and model-based estimation, including treatments like seasonal adjustment and statistical disclosure control. This leads to a set of data ready for publication, usually at macro– or meso level (but exceptions are possible).

Next, these phases were broken down into eleven separate steady states, each with ‘guarantees’ specifying the value added to this steady state compared to the previous one. These eleven phases are used extensively throughout different aspects of the HEcS+ program, as shown below. We give a few examples.

Steady state #3 ‘Standardised concepts’ (which connects the Source phase to the Combination phase) guarantees

√ Source metadata are transformed to conceptual standard metadata: standard units, variables, classifications and periods;
√ Corresponding data are transformed accordingly.

Thus, before reaching this steady state, data and metadata are processed as much as possible in their ‘original’ formats. Obviously, as soon as (meta)data from various sources are being combined it helps when they are standardised.

Steady state #7 ‘Consistent results’ (in the middle of the Completion phase) guarantees

√ Figures for “The Netherlands” are complete and satisfy predefined identities;
√ Plausibility is guaranteed for (relations between) these key figures;
√ An indication of their reliability can be given.

In this steady state, important aggregates like GDP and total retail trade turnover are finalised. More detailed figures might still need additional processing.

Different versions of a steady state should be distinguished. For example, when producing quarterly GDP estimates each quarter a new version of associated steady states is produced. Moreover, different versions may also arise for the same reference period. The flash GDP estimate (produced six weeks after end of quarter) is not the same as the regular GDP estimate (produced three months after end of quarter) or the revised estimates after benchmarking quarterly figures to annual GDP. Different versions of steady states also arise when an error, detected later in the process chain, makes it necessary to rerun some processes.

4.2 Program planning

The core of the HEcS+ architecture was drawn up in the first six months of the program. It has since served as a framework for organising and managing the program activities. The overall HEcS+ program plan, which was maintained throughout the program (and regularly updated), used the architecture to split the program into several logical and manageable projects and other activities.

During the program a number of possible scenarios were explored, either by feasibility studies or research projects. Some scenarios were further pursued and incorporated in the core HEcS+ development program, while others were abandoned. The overall HEcS+ architecture has helped to keep track of all these scenarios and was able to smoothly adjust to the choices made. This shows both its flexibility and robustness.
4.3 HEcS+ projects

The core HEcS+ program was broken down into a number of development projects, along the lines of the main processes described in sections 1.3–1.7 above. For each of these projects the deliverables, resources and timelines were specified. Projects could then run relatively independently within their mandate.

One of the standard deliverables for each development project was a Business Architecture Document (BAD). The BAD describes the business part of the project architecture, in particular steady states and connecting processes, in sufficient detail. Each BAD should be fully aligned to both the enterprise architecture and the overall HEcS+ architecture. Formal reviews were conducted to confirm alignment. Even more important was that each BAD was written with active involvement from business architects. Analogously, a SAD (Software Architecture Document) was required. Later on, a MAD (Methodology Advisory Document) was introduced in order to document the methods used. All of these documents served as inputs for specifying the IT systems to be built.

The use of steady states guaranteed that interfaces between the projects were made explicit and also that no gaps would emerge. A supporting project ‘Coherence in Processes’ (SiP) was established to help in fine-tuning the interfaces, overseeing interdependencies and taking care of program-wide elements like preparing the introduction of chain management.

Apart from the SiP project, a number of other supporting projects were created. As one of the first activities, the statistical outputs of the HEcS+ program (target variables and target subpopulations to be described) were documented, in line with a general architectural principle – start each (re)design with output requirements. Another supporting project concerned processes and methodology around the new meso-analysis approach mentioned in section 1.6.

4.4 Organisational changes

Statistics Netherlands has two offices, The Hague and Heerlen, which are some 250 km apart (a three hour train journey). Traditionally, business statistics were organised along lines of industry with construction, transport and trade being processed in The Hague, while manufacturing and commercial services were processed in Heerlen. Furthermore, different directorates were responsible for primary statistics on the one hand and National Accounts on the other hand. This division of labour caused a number of inefficiencies. Communication between the involved organisational units was often cumbersome, resulting in inconsistencies in methods and tools with sometimes contradicting results as a consequence.

Hence, it was decided end 2009 to carry out a reorganisation. All processes belonging to the HEcS+ program, from the General Business Register to National Accounts, were brought together in a new directorate dedicated to the chain of economic statistics. At the same time, the DRT and NOPS processes were concentrated in The Hague, while the ConGO process was concentrated in Heerlen. This means that the new organisation essentially followed the framework given by the HEcS+ architecture (organisation

---

8 The EHB and Integratie processes were already concentrated in Heerlen and The Hague, respectively, and remained there
boundaries are effectively defined by steady states); an unforeseen application of the architecture.

4.5 Chain management
While the main role of the HEcS+ architecture was at design time, it still plays a part at run time. This is most obvious in chain management, where the concept of steady state is rather important (Van Delden et al., 2009).

A steady state consists of a data set including associated metadata describing important characteristics, such as desired quality. At run time it is not always possible to achieve the quality specified at design time. For example, non-response in a sample survey may cause higher variance than desired, or time is running out before all editing/analysis work can be completed. By defining desired quality explicitly, it becomes easier to compare realised quality with desired quality. If results do not meet the preset requirements, it is possible either to take some action or to accept the lower quality.

Moreover, steady states facilitate more transparent and structured communication between the several actors in the process chain. Steady states are both an output of a process phase and an input to the next phase. These phases are often carried out by different people and/or systems, which may even belong to different organisational units. Steady states allow for negotiating between actors such that needs and possibilities can be reconciled, throughout the whole chain.

HEcS+ chain management has been implemented by appointing a dedicated chain manager, reporting directly to the director responsible for the chain of economic statistics and supported by a team of representatives from all organisational units involved. A number of tools have been created, including a Chain management manual, a Chain management game and a ‘HEcS+ Underground Map’ showing the intricate routings and dependencies in the HEcS+ process chain.

5. Concluding remarks
With a large program like HEcS+ it is inevitable that not everything works out as planned. Below we present some of the lessons learned.

The architectural approach was still new when the HEcS+ program started. Its power and limitations were not well established, and not everybody was convinced of the merits of an architectural approach.

At first, the HEcS+ architecture project was given instructions to start from a ‘green field’ approach, abstracting from any existing statistics or processes. This turned out to be unrealistic and also unnecessary. In the end, a number of important elements (processes, systems) from the existing approach were kept.

After the initial phase of constructing steady states, a lot of effort was put into elaborating the architecture to a much more detailed level. This was less efficient with hindsight, because many content-related issues were not yet resolved. Only when these issues were sufficiently clarified (which in some cases took a while) it was possible to develop the architecture. To a large extent this was done through the BADs delivered by the development projects, or by the SiP project.
A specific example of the above is the development of chain management. A lot of effort was put into theoretical work trying to define different approaches and supporting instruments (such as quality reports), but the organisation was simply not yet ready to decide on the desired forms of chain management.

Overall though, the most important conclusion is that an architectural approach helps to manage a large redesign program in many different ways. In particular, the concept of steady states has proven to be very powerful.

**Figure 2:** The HEcS+ process chain (or rather network) in the form of an Underground map. Colours indicate the main processes.
Acknowledgements

The authors would like to thank Boris Lorenc for organizing and chairing the session ‘Integrated systems and architectures for production of economic statistics’ at the ICES IV conference, and for inviting us to present this paper there. We thank Kees Vennix and Jorrit Zwijnenburg for their many useful comments on a draft version of this paper.

References


