Phoenix: Redesign of the data collection process and systems
Marc Houben (Statistics Netherlands)

mhf.houben@cbs.nl

Abstract

Within Statistics Netherlands we are in the process of redesigning the whole data collection process and supporting IT-systems. It is a major effort from 2015-2019. It consists of developing processes, IT-systems, implementation within a new organizational structure and transitioning surveys to the new situation. Already 8 surveys made the transition.
Phoenix: Redesign of the data collection process and systems

Marc Houben (Phoenix Business architect and Chief product owner)¹

¹ Statistics Netherlands, Department of Data Collection, Heerlen, The Netherlands, mhf.houben@cbs.nl.

Abstract. Within Statistics Netherlands we are in the process of redesigning the whole data collection process and supporting IT-systems. It is a major effort from 2015-2019. It consists of developing processes, IT-systems, implementation within a new organizational structure and transitioning surveys to the new situation. Already 8 surveys made the transition.

Keywords. Data collection, Modernization, Process redesign, IT-system development, architectural driven, Survey transition, Organizational change

1 Statistics Netherlands and Department of Data Collection

1.1 Statistics Netherlands

Statistics Netherlands (SN. In Dutch: CBS) was founded on 9 January 1899.

As of 3 January 2004, SN has the status of an autonomous administrative body. This means that SN performs public service tasks, but operates independently and not under the direct authority of a Dutch ministry. The Minister of Economic Affairs is politically responsible for relevant legislation, budget and conditions. SN is financed from the state budget.

The mission of SN is to publish reliable and coherent statistical information which responds to the needs of Dutch society. The responsibility of SN is twofold: firstly, to compile (official) national statistics and secondly to compile European (community) statistics.

Statistics Netherlands has two branches: in The Hague and Heerlen.

The Department of Data Collection is one of the departments within Statistics Netherlands. The organization chart is shown below.
1.2 Department of Data Collection

The department of Data collection (DVZ) came into existence in 2011. Since then (most) data collection activities are centralized within this department. Before, each statistical department conducted its own data collection. The statistical processes were stove pipes. Nowadays the Data Collection department is responsible for the data collection and delivers the collected datasets to the two statistical departments.

Some characteristics of the data collection:
- 330 FTE
- 150 surveys\(^2\) resulting in 800 occurrences
- 2,000,000 sample units (ranging from 10 - 160,000)
- > 70% mixed mode surveys
- 14,000,000 contact attempts
- 1,500,000 cawi questionnaires
- 200,000 visits for capi (200-250 staff)
- 360,000 phone numbers used for cati (50-90 staff)
- 80,000 calls and 40,000 mails handled by contact centre inbound
- 4,000,000 datasets collected

\(^1\) Non-official diagram
\(^2\) 30 on persons and households and 120 on business
2 Data Collection – the change

2.1 Challenges

The Data Collection department still reflects more or less the stove pipes it inherited from the decentralized data collection processes. Processes for business surveys are apart from the processes for person surveys. This is definitely also true for the IT-systems. Too much manual activities are necessary to conduct the process and too much issues/problems arise during conducting surveys. Furthermore DVZ faces different challenges:
- Budget is decreasing;
- Processes are not flexible enough; adaptation to new technologies is cumbersome;
- DVZ wants to compete with other data collection organizations but are too expensive; and time to market is too long;
- Response rates are decreasing;
- DVZ reputation is at stake;
- IT-systems are not flexible enough; conducting changes is cumbersome;
- IT-systems are end of life (legacy).

3 Non-official diagram
2.2 Attempts to change

Between 2011 and 2015 different attempts were made to get ride off the stove pipe processes and systems. E.g. in 2014/2015 SN looked on the market for a new tool that would support the larger part of the business process. At the end we concluded that there was not really a tool that met our requirements. The process within a statistical bureau (at least SN) is quiet complex.

So we decided to start a new project (actually a programme with different projects) where we focussed on a green field approach to design new business processes and build new IT-systems for the data collection process. This is called the Phoenix programme.

3 Phoenix

3.1 The goals

- The data collection process should be a future-proof, modern, efficient and flexible process; for primary and secondary data collection;
- Be able to respond to external developments (technological and in society);
- Be able to cope with new survey communication strategies;
- Shorter time to market;
- Optimal use of internet mode, selective use of other modes.

The Phoenix programme is responsible for:

- Renewal of business processes;
- Development of IT-systems;
- Transition of surveys from the old current situation to the ‘Phoenix’ situation;
- Connect the data processing systems to the data collection process;
- Implementation of the new processes and systems in the organisation.

3.2 Business architecture – the concept

The change is architectural driven. To cope with the complexity and size of the data collection process the process was divided in conceptual domains. Each domain has its own responsibilities for executing part of the data collection process. The domains are loosely coupled. This is configured by defining information streams between the domains.

\[^{4} \text{The ‘Process’ activity in the GSBPM}\]
These domains are:
- Survey design, with subdomains:
  o Questionnaire design
  o Design of letters and other communication means
  o Design of the output of the data collection process
  o Sample design
  o Design of the survey strategy
  o Design of the control and monitoring needs
- Planning & monitoring (planning, monitoring and response analysis of all of the surveys)
- Sample management (responsible for creating the survey sample)
- Survey management (managing the logistical process on survey level)
- Case management (managing the logistical process on case / unit level)
- The different collection/communication channels (internet, capi, cati, paper, e-mail, etc.)
- Contact management (where all contacts with the collection units are managed: one view of all contacts with a respondent/business)
- Unit and agreement management (where agreements with businesses / persons are stored)
- Data access (where the output data of the data collection process is stored and made available to the next step in the statistical process)
- Secondary data collection management and channel (responsible for receiving data from secondary data sources / suppliers).
3.3 IT architecture - some aspects

From an IT perspective the domains are also loosely coupled. Each domain will have its own system(s) and database(s). The domains/systems communicate through messages. It is a webservice based architecture.

User interfaces are browser based.

Systems can be built by SN or can be bought as a commercial off-the-shelf system (COTS). Principles with COTS systems are: they will not cross the boundary of a domain and will be used as is (only configuration; no additions).

**Fig 3: Domain model of the data collection process**

---

[Diagram showing the domain model of the data collection process with labels and arrows indicating flow and interactions between domains such as Survey design, Planning & monitoring, Data access, Secondary data collection management, Sample management, Contact management, etc.]

---

6
3.4 How we organize it

About 60 people are engaged in the Phoenix programme. There are 3 software development teams (IT-architect, information analyst, developers, testers). There is 1 team responsible for the integration test and end user test. Business architecture and process design is done by a group of business architects and analysts. There is a project responsible for the transition of surveys and there is a project responsible for the implementation in the business and engaging the business change.

For the development a scrum agile way of working is used. Each domain has a product owner and there is one chief (overall) product owner. Each domain has its own key users and stakeholders.

Development is green field.

There is a total budget of little less than € 42 million. Each domain has a part of the budget allocated.

3.5 Where are we now?

We started with developing systems for the domains Survey Management, Case Management, Unit & agreement management, Contact management and the Cawi internet channel (based on Blaise 5). When the core functionality of these domains was ready, this gave us the opportunity to actually transit some surveys from the legacy to the new ‘Phoenix way of working’ 5. Because one of the most important Phoenix principles is ‘Design driven’ we needed to develop a temporary solution for the Survey Design domain. And for some functionality we made temporarily links with the legacy (e.g. to the cati and capi channel).

After this we continued with extending primarily the Case Management and Contact Management domains, and developed the Paper Channel. At this moment about 8 surveys transited from the legacy to Phoenix.

At this moment we are developing functionality for the Data Access domain, the Design domain concerning the survey strategy, and for the E-mail channel, and Upload channel. Also for Paper questionnaires a – rather simple 6 – solution is made. We expect to release large parts before Q4 2017.

---

5 Social and business surveys. E.g. the Health Survey and
6 Because the use of paper questionnaires is to be minimized
In this way more and more processes and systems are developed and more and more types of surveys can transit to the new situation.

Gradually also the organizational change is developed. Ideas are formed and detailed of how the new organization should be look like.

3.6 Our challenges and our celebrations

We experienced that working under architecture worked very well. The domain architecture and principles we devised at the early beginning is still standing with minimal adjustments. Developing greenfield is a great advantage. Also the scrum way of working works very well. We needed about 1 year of development to create the first release and then surveys already could be transited to the new situation. This gave an enormous boost to the DVZ organization and the Phoenix programme: it really worked!

Scrum has however also the disadvantage that things change and you need to adapt already realised processes and systems. This needs extra effort; migration and release management is increasingly a more important issue.

From a business perspective it is a challenge to involve key users from the business. Key users have the old situation as a framework. It is not easy to see the new situation and to think from a new perspective. The whole is organically growing.

3.7 Where to go from now?

We will develop new processes and systems through 2018 en 2019. Biggest chunks are the remains of the Design domain and the capi and the cati channel.

Always when there is new functionality available, the type of surveys which need that functionality can and will transit to the new situation. When the first legacy system can be shut off we will reach a major milestone to celebrate!

Also 2018 will be the year that a new organizational structure will come into being. We will still have surveys within the old situation and more and more surveys in the new situation. The new organizational structure should cope with both; we will not build a new organizational structure next to the existing (there’s no green field here).
4 Presentation at the conference

At the presentation at the conference I will use a top down approach to give more insight in the Phoenix programme and I will pay more attention to some aspects of the development of processes and systems and the principles and policies we use.