



Modernisation at Statistics Netherlands

Bert Kroese

Statistics Netherlands, The Hague, Netherlands – ah.kroese@cbs.nl

Barteld Braaksma – b.braaksma@cbs.nl ; Marton Vuicsan – mhj.vuicsan@cbs.nl

Abstract

Abstract: Modernization is a strategic priority at Statistics Netherlands. The key drivers are new expectations of our users, the technology push and budget cuts. In this paper our main approach is described. In order to free enough resources for modernization a large scale Lean Six Sigma program is introduced at Statistics Netherlands, aimed at making the current production processes as effective and efficient as possible. Our modernisation aims at changing our processes and IT systems, using enterprise architecture and implementing the results of our innovation program. New (big) data sources, mixed mode techniques and automation play a key role. Agile methods, like scrum and Kanban, are widely applied. Cooperation with many national and international partners is essential. A number of examples and lessons learned will be presented. When it comes to modernization, the moment is now; for an information processing industry, the time has never been better to do amazing things.

Keywords: Modernisation, big data, Lean, Agile

1. Introduction and problem statement

He that will not apply new remedies must expect new evils; for time is the greatest innovator.
Francis Bacon

In the past decades we have been busy modernizing our processes and innovating our products using an incremental approach (1,2) and it is quite tempting to conclude that we are on the right track. However, if we delve a little deeper we can wonder if that still holds.

While this mode of operation was OK for the past ten years, major developments in the world around us indicate that our current processes, methods and tools for collecting, processing and disseminating data may rapidly become obsolete; or at least will need to be complemented by totally different and innovative alternatives. Up till now, at Statistics Netherlands our modernization efforts allowed us to keep up with the pace of change in society. But are we doing enough, now that society and technology are accelerating at a rate never seen before? How should we proceed? Which challenges can we expect? How can we collaborate successfully?

On the whole, this point in time is defined by the really stupefying number of opportunities that are waiting for us to take advantage of. However, that means that we have to accept change as a way of life and become more flexible. In this paper we will examine these questions, starting from our current and past modernization initiatives and then looking at our plans to move forward. However, we should realise we are rising to this challenge in adverse conditions: We are old institutes with staff that have been extensively trained on the current processes. Our IT is ageing and in some areas becoming legacy, while budget cuts aggravate the situation. In a time it is needed most, our capability for change is not at an ideal level to say the least...



2. Modernization at Statistics Netherlands at present

Freeing up resources for modernisation

As aptly stated in the UNECE High Level Group Vision document (8), it is paramount to free up resources, as we have no free resources available due to continuous budget cuts. Since 2014 Statistics Netherlands has embarked on a large institute wide program to implement Lean Six Sigma (L6S). To this end, a training program of 15 Black Belts, 80 Green Belts, 175 Orange Belts and 80 Champion Belts has been commissioned and is being executed in waves of approx. 4 months. The training program will run well into 2016. Basic principle is that education and practice go hand in hand. As a consequence 10-15 simultaneous L6S improvement projects are executed in every wave. Wave 4 is to end at Sept. 2015. A first evaluation of the results so far gives an encouraging picture: the return on investment per process excluding training and management costs is 1 year. The nominal cost per green belt project is around 500 hours. Although exact figures are hard to give, we estimate that - on average with a fair margin of error - 25% improvement could be possible to achieve over all processes. Improvements can be achieved in terms of turnaround time, processing time or quality. Freed resources can be partly used for modernisation and innovation, some of these resources will be used to absorb the imposed budget cuts.

Using Kanban and Scrum for better processes

Aside from the L6S process improvement cycle we also use Scrum and Kanban for our IT processes. For processes that are about administration or reactive processes we introduced Kanban in the beginning of 2014. The Kanban process uses boards (large electronic screens or paper boards around which a full team can gather) where the different stadia of the work are depicted in lanes. Work is to be advanced from lane to lane while avoiding queues and or bottlenecks. All essential information about a task is visualised on a sticker that is placed on the board in the lane it currently belongs. This way a quick impression of workload and progress, waits and problems is easy to see. The results are encouraging; the clarity and transparency helps the team and the customer to interact in a much more meaningful way where priorities and dependencies are clearly visible. Workload issues are presented in a way that enables a constructive dialogue between team and customer.

Scrum is particularly suited for development work and quite successful in that area. Scrum has a resemblance to Kanban but is project oriented. One of the main advantages is the much closer and involved contact with the customer/commissioner of the work. The scrum method gives more control over the end result and the process leading up to that without compromising quality. Project teams are operating much more efficient compared to the older methods like RUP or waterfall.

The cost reductions we are getting from Scrum and Kanban are hard to quantify but the improved customer satisfaction, improved motivation and better process flow give ample justification.

New dissemination formats

Changes in technology led to new opportunities for disseminating and presenting statistical products and services. Examples are the introduction of infographics as a tool for statistical storytelling and our Open Data site which went live in the summer of 2014. To (re)connect with the younger “always online” generation we increase our presence on social media platforms, while purposely reducing the number of paper publications. We are investing in our Twitter account, making sure we are present with relevant tweets enhancing the (breaking) news. This has resulted in 85.000 followers up till now. We are working on statistical apps to be released in the near future.

New statistical services

Prominent examples of new services include the introduction of a Centre for Policy Related Statistics, where paying customers can request tailor-made statistical information based on reuse of the in-house statistical treasure chest of microdata; and remote access and on-site facilities for accredited researchers to work directly on statistical microdata (under strict conditions).

New ways of Data collection

Cheaper data collection is one way to increase efficiency, while also contributing to the on-going demand for lower response burden. The ways to reduce data collection costs include reusing data already collected by ourselves or other statistical institutes, using data collected by other government bodies for administrative purposes by tapping off XBRL-based business administrations, introducing web-based surveys (CAWI) instead of the more expensive CAPI and CATI modes, crawling data from company websites or negotiating access to scanner data from retail chain stores to compile price information.

Modern Methodology

One of the key enablers for modernization has always been the development of new methods and algorithms, based on a sound scientific approach and usually motivated by practical problems in statistics production. More recent methodological advances include the development of small area estimators to compile detailed output, automated editing and imputation techniques to increase efficiency of the throughput process and mixed-mode strategies to make data collection cheaper. Sometimes IT tools are developed alongside theoretical methods, which greatly enhances their ability for application. We have created a method and system of automatic balancing of National Accounts. Traditionally, methodology is an area of extensive and fruitful collaboration, both with other statistical institutes and academic partners.

Enterprise architecture

An important tool to enable structured and controlled process innovation was the introduction of an enterprise architecture, with strong emphasis on a high-level business and information architecture. In particular the concept of data steady states proved very useful and intuitively appealing. Each steady state contains data and accompanying metadata in an explicitly described state of processing, with predefined quality¹. Thus, the fact that statistics production is essentially the transformation of data and/or metadata is explicitly acknowledged.

3. Trends in society

Our usual output channels are disappearing

Advertisers are struggling to reach their customers through classical media. The circulation of newspapers is shrinking. On-demand television becomes more popular, at the expense of traditional broadcasts. As we rely heavily on these traditional dissemination channels, this directly impacts on our possibilities for exposure.

The new channels will be mobile

It is expected in communication circles that mobile media are the future of communication. They will soon become the only way to reach the public. These devices however, be it smartphones, watches, tablets or Virtual Reality devices like the Oculus Rift, are fundamentally different from our traditional media. Static internet sites will have to be replaced by dedicated mobile apps that are much more versatile and interactive. This will prompt us to define a new and restyled product set.

Our output will be digested and criticised by machines

An increasing part of the output of our machines will be input of customers' machines, which will gradually learn to analyse and combine information in new and unexpected ways. Society will base more and more important decisions on machine-generated evidence. We will not be allowed our

¹ 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 or Eurostat could be a final steady state. A number of intermediate steady states can be identified between these extremes as desired.



occasional system failure. Process reliability and robustness will be a survival factor. Moreover, the machines will find inconsistencies and implausibilities across our data sets without mercy and without relent. Their big data will be the lens on our data. That implies that we have to put an effort in methods and approaches to deal with incongruities; also in relation to third-party data.

Ever cheaper storage, faster computers and ubiquitous information

Until the 1990s (or even later), statistical offices were among the largest data-processors. The landscape has now drastically changed. Data is becoming available in abundance and technological advances have created completely new industries that depend on large-scale data processing: Amazon, Facebook and Google could not exist without it. All these data can be processed in ways that we cannot imagine yet. And no doubt people will come up with statistical applications. It is unclear what the role of a statistical institute will be in that new situation, with our focus on traditional survey sampling and administrative data.

The crowd becomes an actor

According to an often quoted informal definition, “Crowdsourcing is channelling the experts’ desire to solve a problem and then freely sharing the answer with everyone.” For us as statisticians this may mean different things. For a start, we may address the crowd through apps and web-panels to collect data². This comes with obvious methodological challenges, but nevertheless offers opportunities to reach out to groups that we cannot reach otherwise. The crowd may also become a platform for disseminating statistical information, when people incorporate it in blogs and other kinds of social media. For some statistics, crowdsourcing might become an option.

4. What does this mean for us as statisticians?

We are living in exciting times (3,4,5,6). At this moment it is hard to see where we are going; but in the coming years we will most probably see a further acceleration of events. The best we can do as NSIs is to make our organisations more agile, better suited to a volatile environment. Budget cuts and the resulting urge for efficiency are an important driver. But apart from those, developments in society urge us to reconsider our outputs, methods, processes, and organisation. We have to invest in the capabilities of our staff, and empower them to implement changes in a flexible manner.

On the whole, this point in time is defined by the really stupefying number of opportunities that are waiting for us to take advantage of. We should, however, be cautious that we do not look upon them as threats for our profession. However, that means that we have to accept change and become more flexible.

Output and privacy

As NSIs we have some considerable strengths. Our independence, guaranteed by the Statistical Law, is a huge asset in making people trust our data. But independence is not enough. We have to invest more than ever in procedures that guarantee privacy and data security. And in a society awash with information, we have to stand out from other information providers. Our outputs have to be coherent. Rather than publishing isolated figures, we must aim for painting holistic and consistent pictures to explain phenomena in society. Only in this way we can support policy makers and other users in making information based decisions.

Paradigm shift in methodology

One of our better assets has always been our methodology, which helps to do our job in an efficient way while maintaining high quality standards. Now the new world forces us to rethink our processes and reconsider our treasured knowledge. The standard paradigm of survey sampling has worked quite well for the past century. We also managed to adapt it when we started to incorporate administrative

² In a sense, we are already doing this in environmental statistics by relying on volunteers for observation of birds, butterflies etcetera



sources in statistics production, because their underlying concepts and objects are roughly the same as we were already familiar with. With the arrival of big data we may have to move further away from that paradigm. Big data sources typically describe different objects than those we are really interested in, and they come with their own concepts. We need new ways to extract meaningful statistical information from big data. We must learn to apply machine learning, text mining and artificial intelligence techniques, as well as methods to combine and integrate heterogeneous and sometimes volatile sources to produce quality statistics.

Process, organisation and people

We used to build systems to last ten years or longer. It is not very likely that this will be the case in the future. Our processes will become agile, using intelligent machines to create the desired output from a variety of sources; including big data, administrative sources and a limited amount of supplementary surveys. With data masses routinely exceeding petabytes, records will be processed unseen by human eyes. Some people will supervise the processes as a whole, giving direction where human insight transcends machine knowledge. Our traditional organisation, that one of the UN/ECE High Level Group members once called “the cottage industry”, has to give way to a much more flexible setup with multidisciplinary analysis teams and modular automated production systems. The new organisation will have less people, but with different skills. Statisticians, that are trained and motivated to adapt to the new paradigm: adding value to information available in a flexible and creative way. Data scientists, who are creative, know more about IT than statisticians and more about statistics than IT people. Not only the money (shrinking because of budget cuts) will cause us to leave the paradigm of handmade statistics: it takes too long to produce relevant results in a fast paced world if you let humans do all the work.

Our predicament

Whatever we do, our customers will not accept any outage of statistical information services be it regular or on demand services. That leaves us with the problem that our on-going production is more or less sacred. Whatever we are trying to do in the meantime, “upsetting the business” is not an option.

We are old institutes with staff that have been extensively trained on the current processes. Our IT is ageing and in some areas becoming legacy. This legacy is not very agile in its structure and is becoming a bottleneck for our modernisation endeavours. Another problem is the fact that we have been investing for a long time in improvement of the same processes making the same set of products. For some, this is even being seen as modernisation. It is however clear that we have to reconsider our products; the on-going change in information needs of society fuelled by technology should be our main focus.

The upcoming change in our output will be a landslide (7): mobile and virtual reality are not things that can be done on the side. The new media are intimately connected to a drastic change in information needs. That means any developments in this area will influence the whole production chain and use a lot of resources to realise. We have to make room in our budgets to enable the new developments by making our current processes more efficient. This will prove to be difficult. We have been experiencing budget cuts since the eighties so without considerable efforts there is not much room for manoeuvring.

5. Looking forward

Taking all of the above into consideration, we see three guiding principles that are crucial for our modernization efforts in the immediate future.

Be open and agile

We must learn to be more open than ever to investigate opportunities and think out-of-the-box. That is not easy in a climate that is risk-averse and an environment governed by cost-benefit considerations. But like the UN/ECE HLG Vision states (8): ‘In most organizations there is a good supply of forward thinking people. The challenge is to unlock this potential. We should encourage an entrepreneurial attitude and look for ways to change the culture in our organizations where necessary.’ Only by creating conditions where we encourage ‘to boldly go where no man has gone before³’ we can hope for true innovation. Past Statistics Netherlands innovations like the Blaise software package, the Business Cycle Tracer and the Statline app did not originate from planned projects but from inspired individuals or small groups who found room to experiment. Openness also implies that we must be ready to work together with others to explore promising ideas.

Join forces

In the past, statistical institutes were used to operate in relative isolation. This was partly because maintaining an independent position and guaranteeing privacy and security was best served this way, and partly because the number of possible partners that could offer useful knowledge and experience was quite limited. An exception is the cooperation between statistical institutes and international statistical organizations. This cooperation has concentrated on international harmonization of statistical concepts, joint development of methodology and sharing best practices. The game is changing now. Collaboration between statistical institutes has to be reinforced, and as an industry we must seek new collaboration forms with partners, both public and private, that can provide data, knowledge and expertise. This is essential to keep up with the speed of change in society. We can approach possible partners with confidence: in return we are interesting partners because of our independent and trustworthy reputation, existing data pool and knowledgeable staff.

Organize innovation

In 2012, our office established an Innovation programme and an Innovation Lab (9). The programme aims at stimulating developing and testing innovative ideas for statistics production in a broad sense. It uses a three-stage funnel approach that gives maximum room for idea generation. This has resulted in over one hundred ideas in three years’ time, part of which have been implemented (e.g. open data, infographics, data collection with tablets). The Lab offers state-of-the-art IT facilities and a suitable environment for testing purposes. The experiences with these facilities are positive. Many innovation tracks are done in close collaboration with external partners. These include non-profit institutions like universities and government bodies and commercial enterprises like IT consultants, graphical experts and data providers. Working with external partners is not the first nature of every statistical researcher at our organisation: we need to encourage researchers to make better use of the already existing solutions and make reuse into a habit. It needs constant attention and efforts to gradually increase external orientation.

Concluding remarks

Our office is used to modernization, but it did not always go smoothly; we had our failures. We learned a few valuable lessons, sometimes ‘the hard way’.

- ✓ *Make the business owners responsible for modernization.*
- ✓ *Assure strong support by top management; otherwise don’t even try.*
- ✓ *Devote enough time to definition, selection and preparation of projects.*
- ✓ *Maintain coherence and focus by professional portfolio management.*
- ✓ *Use architecture insofar as needed for guidance, without hindering agility.*

³ Star Trek introductory text



- ✓ *Use tollgates to ensure a project delivers– or stop it if that is not the case.*

Present trends in society now force us to speed up modernization. This will be challenging, and we cannot do it alone. Focus on innovation, personal attention and consistent allocation of funds, even in times of budget cuts, are essential elements. This year we will introduce the first statistics based on big data sources like traffic detection loops and mobile phone data. We started experiments with artificial learning in coding and are building alliances to work on text mining on the internet. We still have a long way to go and international cooperation is essential.

References

1. Barteld Braaksma, Gert Buiten. Redesign of the Chain of Economic Statistics. Paper prepared for the ICES IV conference, Montreal, 2012. See [here](#).
2. Menno Cuppen, Paul van der Laan, Wim van Nunspeet. Re-engineering Dutch social surveys: From single-purpose surveys to an integrated design. *Statistical Journal of the IAOS* **29-1**, p.21-29, 2013.
3. Erik Brynjolfsson, Andrew McAfee. *The second machine age; work, progress and prosperity in a time of brilliant technologies*. Norton & Company, 2014.
4. Viktor Mayer Schönberger, Kenneth Cukier. *Big Data, a revolution that will transform how we live work and think*. Eaman Dolan/Houghton Mifflin Harcourt, 2013.
5. Ray Kurzweil. *The singularity is near: When humans transcend biology*. Penguin Books, 2006.
6. Nick Bilton. *I Live in the Future & Here's How It Works*. Crown Business, 2011.
7. Marton Vuksan. *Strategic developments, trending topics*. Preprint, 2014.
8. UN/ECE. *Strategic vision of the High-level group for strategic developments in business architecture in statistics*, 2010. See [here](#)
9. Barteld Braaksma, Nico Heerschap, Marko Roos, Marleen Verbruggen. *Innovation at Statistics Netherlands*. Paper prepared for the NTTTS 2012 conference, Brussels. See [here](#).