Statistisk sentralbyrå
Statistics Norway

Establishing an Innovation lab for new data sources and techniques at Statistics Norway – experiences so far

Anders Holmberg, *Statistics Norway*

Introduction

In recent years many NSOs have started exploratory work trying to utilize new types of data, technology, methods and analytical tools. In the attempt to organize these initiatives many NSOs has established "Big data centers", "innovation labs", "Data science labs" or similar. The level of ambition, scope, magnitude and organizational form varies between different initiatives, but they have in common to try to put together a multidisciplinary arena for statistics innovation across the organization.

The establishment of Statistics Norway's Innovation Center (SNIC) was decided by the board of directors at the end of 2017, and the first half of 2018 has been much about building the fundamentals with infrastructure, people, projects and establishing a management structure. This is still an ongoing process.

In this paper I want to share some thoughts on reasons to organize innovation activities in general, and further some thoughts on the establishment of the innovation center. I will explain the background for this initiative, what we consider as key preconditions and features in the innovation center, and how to place the innovation center in the rest of the organization. The latter is experimental, since many features are new and crosses the current practices of line management.

Background and motives for an innovation laboratory

The past decade Statistics Norway, just as many other NSOs and international statistical bodies, has been busy with modernisation and transformation of its statistical production process (see e.g. [1], [2], [3]). One may debate the speed of success in this endeavour, but there has been some progress in terms of standardisation, process quality and importantly the organisational maturity in understanding the needs for including new methodology, technology and new unconventional data sources in statistics production. At Statistics Norway, this is predominantly illustrated by the launch of an organisational strategy that, among other themes such as competence building, includes a major modernisation program.

One of the first and very important outputs from this modernisation program was a high-level target for Statistics Norway's future statistics production environment. Besides, describing the sought ideal situation for internal processes, the target picture also encompasses interaction points with the outside world, and as a support feature to test and feed new ideas and functionality, an innovation environment or innovation center. Here, it is important to bear in mind that Statistics Norway's organisation does not have a formal unit that solely is dedicated to applied research and development for improving its business processes. These activities have been done decentralised, in patches through project funding. To guarantee technical resources and assistance for such projects, they have been managed in so called portfolio projects. Other expertise such as methodology have not been allocated through such channels. Hence, the creation of a multidisciplinary innovation center, manned and run by internal staff is something new and one of the pieces for realising the modernisation target to increase Statistics Norway's R&D capability for making statistics.

Parallel to setting the scene for the modernisation program, Statistics Norway has begun to tackle the appearance of new disruptive technologies and new (sometimes big) data sources. This has been done on a limited scale. It has been done aside from production development and away from the formalisation inside the organisation wide modernisation program. There are several good reasons for this approach.

New data sources, with greater dimensions than previously experienced, and with questionable consistency poses new challenges for official statistics. It enforces a general rethinking that involves tools, software, methodologies and organisational structure. This implies that, before there is a reasonable certainty of a viable business case to pursue, considerable explorative investigations must be done. An externally funded modernisation program cannot take (many) such risks. It must start from what are known gaps and plan accordingly to fill those gaps with what would yield the best return of investment. This was one of the reasons why it was decided to put investigations about new methodology and new unconventional data sources outside the umbrella of the modernisation funding. Keeping such exploratory undertakings with fuzzy goals small and inside an innovation environment, is a way to control risk in initial development stages and at the same time (in Statistics Norway's case) making useful organisational experiences.

To promote innovation activities a decision was taken to connect the exploration of unconventional data sources with starting the SNIC. The intention is that SNIC should be an environment that, without major project development, can explore, experiment and try solutions that may improve statistics production. The activities / projects should be flexible in content, but targeted towards SSB's strategic goals. Each activity must have limited time frames and thus limited scope. After each activity phase, an evaluation is carried out and the results tuned and communicated. The SNIC activities/projects are small and should therefore follow good resource management in the sense that if they fail, they fail fast.

Financial transaction data: The business case for starting joint development of methodology, technology and multiple subject matter domains.

The concept of creating SNIC coincides with a statistical product demand to enable use of new methodology and new sources with financial transaction data. This opportunity rose partly by chance and partly from a project that investigated alternative statistical designs for a Household Budget Survey (HBS). In addition, an earmarked small funding was made available from a supplementary government budget round. Together, these circumstances paved the way to begin exploring the benefits of financial transaction data. The now existing embryo of SNIC orbits these data sources, and the activities are focussed on how to utilize them for statistical purposes.

To join multiple disciplines and get traction in innovation with new data sources, it is important to have a case where the potential business benefits clearly can be sensed by everyone. Moreover, when the solutions are not straight forward, or obviously classical, and when a prerequisite for making successful progress, is a mix of new methodology, technology testing and subject domain demands; then the conditions originally intended for SNIC apply.

A difference between Statistics Norway's approach compared to innovation activities and development around Big data sources set-up at other statistical organisations (e.g. Statistics Netherlands, Statistics Sweden and the ONS) is that it is less top-down. While it is common to start with a management decision about organisation / start-up, budget etc. and then chose cases. SNIC is in a sense built from a case(-s) and then possibly expanded.

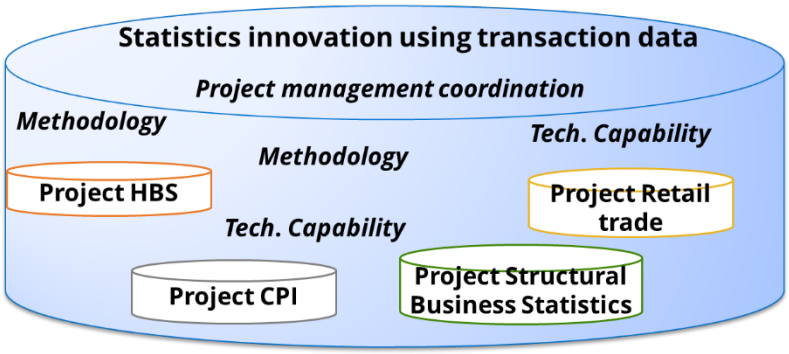

Figure 1 Innovation Projects/activities centered around financial transaction data

Since several different statistical outputs are concerned with the potential of financial transaction data, there is a common part and a subject domain part. This is illustrated by Figure 1. The statistical domains, HBS, Consumer Price Indices (CPI), Retail trade and Structural business statistics share the same technical infrastructure and certain general competence such as methodology and IT. But because of varying end objectives or demand, they also run different subprojects where they do tailored activities and analysis. A common project management coordination assures that resources and results are shared and distributed across the participating projects.

One advantage of this set-up is that, instead of each subject matter unit creating and doing their own innovation activity for financial transaction data (or for that matter methodology or IT experts trying 'cool stuff' without clear ties to a statistical need) it facilitates and ensures a common direction and resource sharing. A further spin-off from cooperating like this on a multifaceted innovation theme (in this case a set of unconventional data sources which not yet are determined fit for statistics production) is that it lowers barriers between staff working in different units, and it broaden staffs' insights of other statistical areas.

Since SNIC has the marks of an organisational experiment its start-up activities with financial transaction data have a relatively low profile. Much of the work so far are focussed on retrieving, safeguarding and understanding data sources. These tasks are mutual and cross topic for the projects in Figure 1. Presently, only the tasks related to the HBS project can be said to have started domain specific studies.

However, the organisational set-up also has human resources aspects. Mixing disciplines across unit and department boundaries with a relatively light touch governance creates some foreseeable issues. Two obvious issues are; a) time allocation for key staff members between innovation activities in SNIC and prioritized production and modernisation program work and b) assuring that staff doing a splendid 'volunteer' work in SNIC, outside their normal line management duties, get proper recognition. In this build-up face, these issues are dealt with by defining a set of principles for the innovation environment and by anchoring the content to a steering group consisting of members from the board of directors.

Guiding principles for Statistics Norway's Innovation Center

Steering, prioritization and content principles

Administratively SNIC is placed at the Department for Digitalisation and Shared Business Services (DSBS). The practical reason for this is that the department consist of units responsible for Methodology, IT and Data collection. All of them are central for building capacity for the use unconventional data sources as well as any general development in production processes and infrastructure. However, SNIC's steering group who is responsible for content and resources allocation is manned by the directors from all subject matter departments and the Research department. All departments at Statistics Norway have participated in the practical project work.

As mentioned previously, the activities in SNIC are aimed at improving statistics by using unconventional data sources. The ultimate purpose and the innovation objective for a project must be one (or several) of the following; i) to replace an existing data collection or statistics production solution; ii) to increase quality in an existing statistical product or production process and iii) create new statistics within a subject matter domain. Given that this the purpose requirement is fulfilled, the steering group prioritizes between suggested projects based on organisational needs and existing capacity.

The projects/activities are limited in volume and have test characteristic. When they are evaluated the conclusions are either, a) to stop the activity altogether, b) to put on hold, c) to continue and do another phase or d) to increase the efforts by pushing for implementation or a bigger development project. The last alternative will happen outside SNIC.

Staffing principles

SNIC is a resource platform and a knowledge environment which is manned by methods, IT and subject competence. Methodology and IT skills do not mean personnel only from DSBS. Expertise in these areas also exist elsewhere and an important aspect is that it is knowledge, expertise and interest that guides who contributes, not which department / unit staff belong to.

Another major principle is that SNIC, besides the administration, should not have permanent staff attached. The environment shall be organized in a way that facilitates rotation of resources and where people from different parts of the organization participate in shorter or longer periods of interest, desire and needs (related to the relevant competence or project). The motivation is to promote the flexibility, resourcefulness and avoid the dual risk of being self-serving, or alternatively so task-oriented that it becomes pedestrian. Another motive is to dare to give various staff the opportunity to learn, to create networks that could continue to function after a project in SNIC is finished, and spread experiences further out in the organisation.

The right competence is crucial. SNIC must be a place where relevant extramural research can be brought in, to assist, inspire, speed-up and correct when needed. SNIC shall be a good surface to benchmark and collaborate with other agencies, in Norway and internationally, and an environment where it is easy to set up various types of internships to get relevant project assistance. Statistics Norway's developed partnerships for statistics with Big data will be channeled through SNIC.

Communication and principles for competence dissemination

A purpose of SNIC is to make SSB acquire new and modern skills and to build and spread knowledge and experience further in the organisation. The staff rotation and interdepartmental network building is one way, but it must also be complemented by active communication. All results (including failures) from the innovation activities should be documented and regular seminars about techniques, tools, methods and experiences from working large amounts of data is held. In this way, knowledge will be spread across the organization. This is about topics such as artificial intelligence, analysis programs and programming languages, as well as infrastructure requirements on data storage and processing that has been tested.

Information and data management principles

Since the innovation center presently focusses on methods and techniques for unconventional data sources, and since the handling of these data is apart from regular statistics production. It is necessary to set up SNIC's information and data management principles. The innovation process of new data sources which are guided by these principles are depicted in Figure 2.

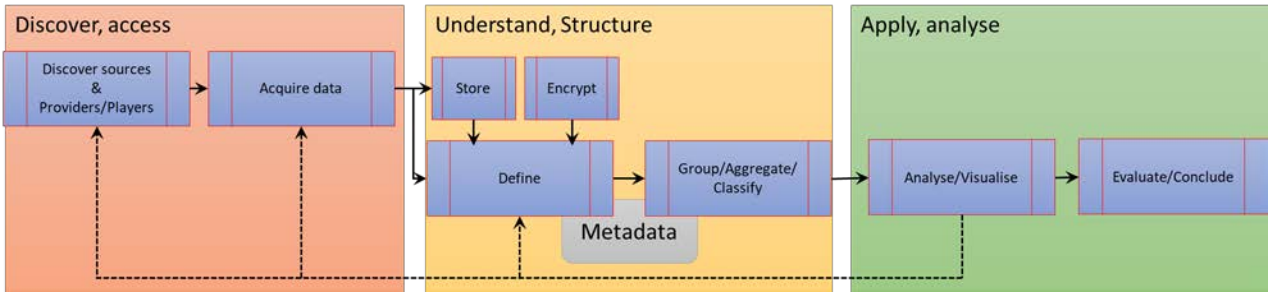

Figure 2. The value chain for working with unconventional data in SNIC

It would be to go far in this paper to discuss every process step and their principles in this value chain, but there are factors worth mentioning that I believe increases effectiveness throughout the organisation.

- Multidisciplinary teams with staff from different departments that work together would contribute to standardisation by adopting and following the same set of principles in the SNIC. E.g. with respect to routines for acquiring and quality assuring data, analysing tools, suitable methodology that fit the problem and so forth.
- Routines for data security, accessibility and uniform methods to process and store data and metadata effectively will be spread through SNIC experiences. Colleagues who participate will gain competences outside their normal production environment.
- Outcomes from innovation projects are scrutinized and evaluated from a methodological standpoint to the emphasize quality dimensions of statistics and value added to Statistics Norway's core functions. This increases general statistics production competence and cultivates awareness of common business solutions. A good example of this from the current SNIC activities are the collaboration and centralized dialogue with external data owners.
- For the three main phases in Figure 2 above, the SNIC principles build an infrastructure and a know-how bank. This enables Statistics Norway to make quicker assessments of future data sources and better choices between alternative statistical designs.

Extending the scope by covering new grounds

During its build-up the first eight months, SNIC has applied innovation activities on financial transaction data. The work with different financial transaction data sources have been done simultaneously with developing SNIC's guiding principles, the staff working arrangements, technical infrastructures etc. In its backlog, Statistics Norway also has had other data sources and problems that is waiting to potentially enter as an innovation activity. Figure 3 depicts a potential expansion with new activities involving other unconventional data sources. This happens only if there are enough resources and strong enough internal demand.

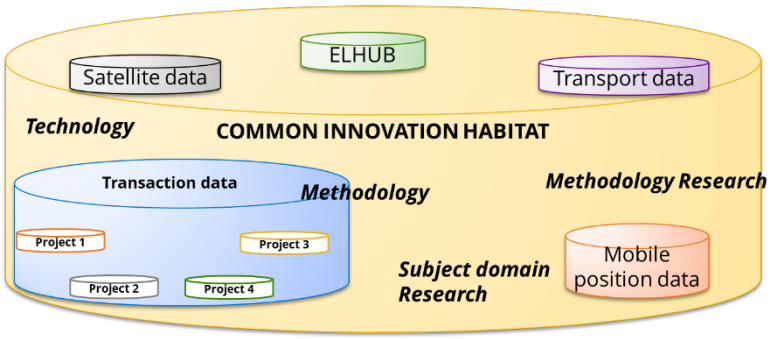

Figure 3. A potential expansion of the Innovation Center's scope would have activities involving more unconventional data sources than only transaction data.

In collaboration with external partners there are plans to start investigating sensor data for transportation purposes, mobile phone position data with a whole set of applications, and smart meters measuring electricity (Elhub). Presently the Elhub data is closest to start in SNIC activity, but that can change quickly. SNIC activities depends entirely on ongoing prioritizations and freed staff resources. There are little or no room to initiate new activities from the center itself. This might be viewed as a drawback, but it is not. It was never the intention that SNIC should have a research agenda of its own, as for example in a permanent R&D department with fix staff would. SNIC is meant to be something smaller and more agile in terms of choosing activities. The advantage is that SNIC activities will not be caught in the competence web of permanent staff, the drawback is vulnerability in terms proper manning and possibly a lack of continuity.

Summarizing the experiences so far

It remains to see whether this format to do innovation activities can deliver. It has been set up in a small scale in order not to compete with other major undertakings. It is also an experiment in the sense that staffing, organisation, funding and some of the planning is unorthodox and crosses traditional line management routines. To some extent, it is based on volunteer work and enthusiasm. This may work well while the size of the center is small and bilateral managerial agreements are solved through practical agreements and task swapping.

However, in 2018 SNIC was backed by extra funding and strategic prioritization from the Methodology unit and DSBS. The funding for 2019 is not so straight forward and the model for SNIC requires every participating member to individually set aside resources. By having business cases with a clear end-goal this becomes a matter of managerial long-term priority. No solution for using (big) financial transaction

data will come out by 2019. Perhaps it is a dead end and the outcome will be experiences and competence only? Does this suffice? Competence building is a strategic argument that would work in favour of SNIC and it appears unlikely for a single unit at Statistics Norway to establish competence building opportunities for new methodology and technology that combines team work with relevant strategic themes.

Yet another factor to reconciliate is the interaction of SNIC outputs with other development work. Enterprise, Business and Methodology architectures are under development and innovative discoveries from SNIC may both contribute to new and disrupt earlier solutions [4]. Thus far, there have been continuous meetings with the architecture group to discuss ways forward for the value chain for financial transaction data and similar future sources. Still, this does not guarantee that the 'not-invented-here-syndrome' turns up when results from SNIC are evaluated and taken further. This is something Statistics Norway must watch out for and find ways to handle.

Finally, the experiences until now have been characterized by the mindset of explorers and a positive interest from the rest of the organisation. The multidisciplinary team working in SNIC has begun to deliver concrete findings by acquiring a type of data that Statistics Norway never has worked with before. The first seminar which will show preliminary empirical outputs is due in August 2018. It remains to see if the progress is fast enough, and if the principles for SNIC will work in practice over long term and if the environment grows in scope. A nice proof that it is moving in the right direction is that the innovation work more often is brought up as a vital tool for Statistics Norway's capability building, and as an opportunity for staff to increase individual competence but in group.

References

[1] S. Penneck, (2009). "The ONS Statistical Modernisation Programme: What went right? What did we learn?" in Modernisation of Statistics Production, Conference at Statistics Sweden, Stockholm November 2–4, 2009.

[2] Pink, B., Borowik J. & Lee G. (2009). The case for an international statistical innovation program – Transformingnational and international statistics systems. Statistical Journal of the International Association for Official Statistics, Vol. 26, 125–133.

[3] Götzfried A. (2015). Modernising Official Statistics: A Complex Challenge. In: Lausen B., Krolak-Schwerdt S., Böhmer M. (eds) Data Science, Learning by Latent Structures, and Knowledge Discovery. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg DOI: https://doi.org/10.1007/978-3-662-44983-7_1

[4] Eltinge J.L., Biemer P.P., and Holmberg A., (2013). A Potential Framework for Integration of Architecture and Methodology to Improve Statistical Production Systems, *Journal of Official Statistics* 29, 113-124. DOI: 10.2478/jos-2013-0007