# Modernizing statistical production at Statistics Norway

Trygve Falch, Statistics Norway, trygve.falch@ssb.no

Rune Gløersen, Statistics Norway, rune.gloersen@ssb.no

Jon Folkedal, Statistics Norway, jon.folkedal@ssb.no

**Abstract**

NSIs are facing increasing demands to produce statistics with higher quality and faster production time, while reducing cost and staff. New challenges and expectations from stakeholders, future role of statistical organizations, skills, methods, ways of working and producing statistics are drivers for technological change.

To meet these challenges and expectations Statistics Norway are modernizing the production of statistics, with a digitalization effort to adopt new technology, utilizing new competence and employing new methods for carrying out our mission.

As part of the modernization activities, there have been a coordinated and focused effort to base the modernization on the ModernStats standards, including, GSBPM, GSDEM, GSIM and CSPA. Operationalizing these standards in a systematic way have always been a challenge. However, using Open Source tools, modern system development techniques and best practices, we’re closer to being able to put the standards as well as Open Source components into production.

Statistics Norway have actively participated in developing GSIM and is now implementing GSIM in a new production wide distributed datastore (Linked Data Store). Statistics Norway is actively participating in developing the next version of CSPA.

The use of standards will help us in establishing a modernized platform for developing and operating services for statistical production, aimed to increase control and reuse of data and metadata, implement reusable services and automate production processes.

Statistics Norway intend to establish the new statistical production platform using hyperscale cloud services, to deliver

* efficient and secure development, management and operation of components and services
* fast, frequent and reliable products
* cost-effective capacity utilization
* flexibility in choice of technology (cost effectiveness to change and replace technology)

The use of Open Source components that are fuelled by the UNECE ModernStats standards, can also be the basis for more collaboration between Statistical Organizations.

***Keywords:*** *standards, cloud, data storage, open source, modernization*

**1. Introduction**

Today's statistical production is characterized by many parallel production pipelines, where data flows heterogeneously, with a large degree of manual processing, tailor-made data and applications. Existing practice where data and applications are linked to statistical products is a challenge, as the same data and functionality occur in several variants in several places in different production chains. It reduces the quality of both data and production, leads to bottlenecks and involves a significant consolidation challenge, especially in the development and production of composite, transverse statistics as well as for the development of new products wholly or partly based on existing data.

Statistics Norway operates and maintains about 300 different systems and applications within the statistics production. Large parts of this portfolio are outdated and consist of fragmented and deficiently documented tailor-made solutions. The portfolio is based on obsolete architecture and outdated technological solutions. This poses a significant technological and security risks and leads to accelerating management costs and significant investment needs when external changes imposed on us must be implemented in several separate systems rather than once in a common, standardized production system. This technical vulnerability is high and poses a risk that Statistics Norway will not be able to streamline and develop the business so that requirements for reduced costs, lower response burden and the need for more statistics and a better basis for decision-making in society can be met.The conclusion is that the existing platform does not meet future needs or have future carrying capacity.

An important task is directed at the need to better utilize the business expertise in Statistics Norway, through better utilization of technology one would be able to reduce manual work processes and streamline routine production processes in order to increase the capability to do more analyzes and better utilization of existing data and information.

Statistics Norway considered it necessary to carry out a comprehensive modernization of the business to meet future challenges. The development work focused on five areas of action, where various alternatives where investigated and prioritized:

* Establish a modern IT platform based on hyperscale cloud technologies
* Better information management and security
* Process improvement, product development and self-service
* Transversal management and change capability

A major challenge for realizing the measures for more automation and self-service in business operations lies in the need to acquire and develop better IT solutions in the statistics production. Statistics Norway's assessment is that the business cannot face upcoming challenges without a significant increase in the cost level, if extensive investment in new technology is not implemented.

The modernization of statistics production creates external benefits by contributing to more up-to-date and relevant statistics and analyzes, eased and improved reuse of data, and further reducing the response burden by utilizing more of existing data. The internal benefits are reflected in the form of more efficient and scalable production, improved information security, better utilization of resources and reduced staff and digital vulnerability. The effects are permanent and give Statistics Norway an ability to increase production and at the same time reduce future cost levels. Statistics Norway is implementing measures in the following main areas:

* Extensive modernization and upgrading of Statistics Norway's IT solutions.
* Method, organizational and competence development to use new data sources and streamline work processes.
* Better solutions for standardized storage and documentation of data that will provide better overview, availability and reuse of data as a basis for more analyzes both internally in Statistics Norway, for external users and for research purposes.
* Better information management should lay the foundation

## 2. Vision and strategy

The target picture visualizes Statistics Norway's future statistics and is a conceptual vision for future statistics production. It shows how modernization supports strategic goals and how it will contribute to implementing the strategy for modernization.



Figure 1. Target picture for future statistical production.

The vision describes several dimensions and levels that will contribute to value creation in Statistics Norway.

The development of Statistics Norway in the 2020s must be based on the new Statistics Act. The ambitions, frameworks and roles described in the proposition is a natural starting point for assessing Statistics Norway's strategy. There will be a need to set clear targets for the 2020s that can also provide guidelines for the development work. The existing target for technological development of Statistics Norway's production processes will be an important element in this process. Much will still be relevant and important, while technological development and a new statutory law may indicate that we elaborate and further develop other parts of the target picture. For example, it may be relevant to consider how the ambitions in a new statutory law may make it necessary to set clearer targets for user value for the statistics.

The technology allows, among other things, registry data to become increasingly available in near real-time. Along with the removal of manual processes, this is expected to provide increased automation and improved timeliness. Another area where additional benefits can be obtained is the transition from manual correction and editing to more algorithm-based processing where possible.

**Automation** of production processes can be an important tool free up capacity and increase speed to modernize and develop Statistics Norway's statistics, analyses and other services. A possible strength with a product-focused organization may lie in the **product** team developing expertise in its statistical area - from understanding the capabilities and challenges of the raw data at one end of the **value chain** to understanding the **users** and their needs in the other. The fact that the team is in control of "its" value chain facilitates efficient development with little need for waiting time and coordination with transversal functions. On the other hand, there can be economies of scale in that functions such as data reception, information management and environment for analyses serve as common platforms that are utilized by the product teams.



Figure 2. Strategic themes.

Statistics Norway is working on the professionalisation of software development throughout the organisation. This applies to areas such as version control, documentation, code sharing, processes for sharing **competence**, quality reviews of code and other activities that characterize development organizations. The need for programming expertise is great, and Statistics Norway's ability to provide employees with the right development opportunities, recruit and retain the necessary expertise is important for future development.

Important prerequisites for success with product teams will be that interdisciplinary expertise works continuously in the team (**partnership**) and that the teams have clearly defined, and measurable goals related to Statistics Norway's main priorities.

## 3. Open standards and open source

Open standards are a prerequisite for meeting national architectural principles of interaction, but also for co-operation at national and international level.

The use of open standards ensures that you have a community that can be used to exchange experiences and expertise, which can contribute to more efficient, robust and flexible solutions that can be shared and reused.

Interoperability and collaboration across businesses and national borders requires an effective flow of information, including through technical interoperability ensuring that technical standards facilitate well-defined interfaces, transfer protocols and formats.

We develop models on several levels with different purposes. Conceptual models, where the purpose is to communicate internally and externally, and to highlight connections and reduce complexity. Conceptual models describe what and not how. Logical models that break down and detail conceptual with the purpose of communicating with architects, suppliers etc. and to construct and choose solutions and describe them (work processes, information, functionality, sharing services, architecture selection, service-oriented mindset). We also prepare physical models, where the purpose is to describe reality as it is and solutions as they should be implemented and applied.

Our models and standards are based on international technology standards and on international standards / models in the statistical field:

* Generic Statistical Information Model (GSIM) for information model
* Generic Statistical Business Process Model (GSBPM) for business processes
* Common Statistical Production Architecture (CSPA) for services and data / service flow

The use of open source code gives freedom to choose products, and technology that suit us. It also gives freedom to choose suppliers, but also expertise in the market. Sharing code nationally and internationally can also contribute to improving proprietary components.

## 4. Business case for modernizing statistical production

The pace of technological change and change in acquisition of data from owners and providers, requires our organization to be more responsive to these changes and opportunities. Especially the pace of technological change challenges our ability to provide timely statistics, but also to keep our private datacenters safe and capable to meet our needs. With accelerating technological changes, there is also an substantial change in the type of security threats we are faced with. This requires an increased capability of change to our on-premise datacenters, including keeping up with demands for new competencies for managing and securing our datacenter. We will always try to balance security with increased speed and increased agility when it comes to producing timely statistics, but a lot of resources ultimately have to be prioritized into costly maintenance and upgrade of our on-premise datacenters.

Our approach is to utilize modern, scalable cloud technologies, that does not require large up-front investments, and which will help us to focus on what is our core mission as a statistical office.

In the planning of the modernization several approaches were considered. We did consider building a generic business platform for our statistical production as a business functionality approach. However, project complexity, the length of time and high project cost were considered to give high project risks. An alternative was to do a business case approach looking at one or a smaller group of statistical products, still targeting at developing a generic statistical production platform. This was considered to increase speed and complexity as well as lowering the project costs and earlier break-even point. As we are establishing the generic technical platform, this approach gives the flexibility to start implementing business functionality for other statistical product alongside the chosen business case.

We have started a project, where we aim to establish a base platform which will enable us to develop and deploy new applications fast, and to create a first version of an analytical platform with big data capabilities, combined with extensive metadata support. As a first use case for this platform we’re re-doing most of our tax, income and structure statistics by modernizing both the production processes and technical services for the business functions.

### 4.1 Business case: Modernizing statistical products

In the period 2018 to 2021, the Tax Administration will carry out major restructuring of its work processes and information models. This will affect Statistics Norway's use of administrative data from the Tax Administration, and it is necessary for Statistics Norway to initiate work on adapting the collection and management of data in order to safeguard the continuous production of several statistics. The change will also facilitate modernizing work processes internally in Statistics Norway, and the project is proposed to be implemented mainly as part of the modernization activities.

The business case and following project – called SIRIUS - will deal with the changes in statistics production that come because of the Tax Administration's development of a new, digitized tax return, the SIRIUS project. For Statistics Norway, the SIRIUS project entails changes in

* the information base for the statistics that are based on tax and accounting data for the business sector and person
* the way in which tax and accounting data comes in to Statistics Norway
* the work processes for data collection, preparation, editing and analysis of the statistical products

The 28 statistical products that are affected are identified, and the project will develop a new, comprehensive statistics production platform that enables Statistics Norway to produce these statistics with the right quality and improved timeliness.

At the same time the Tax Administration also modernizes the National Population Register and the information flows that Statistics Norway receive from the National Population Register. The project will also handle the changes imposed by the modernized Population Register, limited to what is required by the selected statistical products

The affected statistics are representative for register-based statistics, and statistics that combine data from samples and registers, and in addition, the data processed by the statistics concerned is widely used elsewhere in Statistics Norway. For these internal secondary users of data, the project should help ensure that data can be delivered and reused in an efficient manner. The realization of a new statistical production platform for the statistics concerned is therefore an appropriate first part of the modernization of statistics production in Statistics Norway, and in accordance with the mandate for the concept phase, shall be based on the principles that follow from the vision for the modernization work.

In order to be able to modernize the statistical products related to Tax Administration Statistics Norway needs to establish a technical platform that meet the requirements.

### 4.2 Business case: The cloud service model

In planning the establishment of a Modern IT platform, the choice of concepts and technologies depends on the decision on the delivery model for the IT platform. The delivery model defines how IT functionality (such as networks, processing, storage, services, and applications) is delivered and made available for use in the development, operation, and management of the new technical platform.

When considering different solutions for modernizing Statistics Norway it was recognized that the existing platform does not have future carrying capacity. In order to fulfill the business needs regarding ability to automate manual work, utilize new expertise and new methods and use more data and information a cloud native platform was considered as the best approach.

Based on the vision and strategy and business requirements, this means that the IT platform will facilitate:

* Efficient and secure development, management and operation of components and services for modernized statistics production.
* Fast, frequent and reliable deliveries (production of altered or new IT functionality).
* Cost-effective capacity utilization of technical infrastructure (also for unknown changes in capacity requirements, such as innovation activities and use of big data).
* Technology choice flexibility (cost effective to change and replacement of technology components).

A Cloud service model is the delivery of IT resources dynamically as self-service services. The cloud service provider and consumer are the two key players in the model (cf. Cloud Computing Reference Architecture (NIST) and The Cloud Ecosystem Reference Model (Open Group)).

Key features of the cloud service model:

* The IT functionality is delivered when you need it and can be scaled up and down as needed
* IT functionality can be acquired quickly, and often through self-service
* The IT functionality is available through standard online mechanisms
* The use of IT functionality is documented (measured, controlled and reported) and is readily available to both the provider and user of the functionality
* The investment cost is low as you pay for what you need when you need it, instead of investing in hardware and software to handle peaks
* IT functionality can be offered from various locations, so data storage and processing need to be controlled through agreements
* The responsibility between the provider and the consumer of IT functionality is well-defined and contract-regulated
* Technology dependency and flexibility between teams is achieved through vertical integration via cloud services

The recommendation was to realize the modern IT-platform on a hyperscale cloud platform provided by a hyperscale cloud vendor.

A hyperscale cloud platform is considered to provide the best cost-effectiveness in terms of establishment, innovation, accessibility, scalability, energy efficiency, pricing models and operating costs. It is also our understanding that the risk of security incidents and breaches by public cloud suppliers from a technical perspective is similar or lower than that of private cloud providers in and outside Norway. We think that the risk of misleading data is more dependent on vendor choices and what security measures are being implemented than if the infrastructure services are available from a public or private cloud. Regarding compliance with legislation, general cloud providers operating in Europe are subject to the same regulations as private cloud providers. The European legislation applies to all suppliers operating in the EU / EEA area.

In the length of the recommendation Statistics Norway carried out a risk and vulnerability analysis. It is also a prerequisite for successful implementation that the further process of choice and acquisition of cloud solution is rooted in Statistics Norway.

## 6. Risk and vulnerability analysis

The risk analysis has not revealed any regulatory constraints that prevent the transition to the use of public cloud-based services. However, archiving of data archives must still take place in Norway in accordance with the existing archive law, but NOU 2019:9 (Official Norwegian Report) delivered on April 2, 2019 states the following:

“The Committee proposes to repeal the current prohibition on the transfer of digital documentation abroad, with certain geographical limitations, requirements for risk assessments and opportunity for revisions”. Proposal for a new law on archives must be consulted and can be expected to be presented to Stortinget in the spring of 2020, with effect from 2021.

Through the analysis, several areas that can represent a risk were identified, including:

* Own ability to manage a cloud service: How to manage our own security related tasks
* Loss of Governance: How to maintain overall control even if some liability is left to the cloud supplier in areas that may affect security.
* Lock-in: How to avoid being locked to a specific cloud supplier for technical or other reasons
* Lack of isolation: How to isolate data processing from other actors using the same platform
* Rights risk: How to ensure that the cloud supplier's security management and processes comply with formal security requirements / frameworks and certification requirements
* Compromising the administration interface: How to avoid being attacked through the supplier's internet-based administration interface
* Data protection: How to ensure that the supplier safeguards necessary security against data loss
* Uncertain or incomplete data deletion: How to ensure complete and timely deletion of data when necessary
* Damage caused by malicious insides: How to prevent and avoid damage caused by malicious insides

The risk and vulnerability analysis have furthermore revealed that several measures will have to be implemented prior to any general cloud services being used, including the following:

* Statistics Norway must design and implement enough technical measures that support security based on the sensitivity of the data to be lifted into the cloud, including encryption, differentiated access management, “zero trust” architecture and monitoring.[[1]](#footnote-2)
* Processes and routines must be established in the organization that support the technical measures. Other necessary organizational measures must also be established, including measures related to the ordering role. There must also be training in new processes and routines and possibly also other attitudinal measures.
* Furthermore, the Safety Manual must be evaluated and updated in accordance with uncovered change needs.

In parallel with the work on the risk analysis, the projects is establishing and safeguarding the necessary information security, where best practice has been laid. The analysis forms a set of general requirements for safety in technical solutions and organizational ability.

While several risk minimization measures are being implemented, it is important to point out that security measures will never eliminate the security risk. The measures could reduce the risk, but there will always be a residual risk. The management of Statistics Norway will be responsible for this residual risk. It should also be mentioned that there is also security risk associated with the current solution in Statistics Norway. It is therefore important that Statistics Norway continuously renews its IT infrastructure in line with new opportunities that the technology provides and implements security measures that correspond to the risk picture of the data in production.

In the further work, we will clarify the need for decisions related to risk elements and vulnerabilities, propose technical and organizational risk mitigation measures and make recommendations.

## 7. References

Fang Liu, Jin Tong, Jian Mao, Robert B. Bohn, John V. Messina, Mark L. Badger, Dawn M. Leaf (8. September 2011), NIST Cloud Computing Reference Architecture. Special Publication (NIST SP) - 500-292. Available at: https://www.nist.gov/publications/nist-cloud-computing-reference-architecture.

NOU 2019:9 (Official Norwegian Report): Fra Kalveskinn til datasjø – Ny lov om samfunnsdokumentasjon og arkiver. (In norwegian only)

OpenGroup, The. The Cloud Ecosystem Reference Model. Available at: http://www.opengroup.org/cloud/cloud\_ecosystem\_rm/index.htm

Pratt, Mary K. (2018), What is Zero Trust? A model for more effective security. Available at: https://www.csoonline.com/article/3247848/what-is-zero-trust-a-model-for-more-effective-security.html (Accessed 16 May 2019)

1. What is Zero Trust? A model for more effective security.

https://www.csoonline.com/article/3247848/what-is-zero-trust-a-model-for-more-effective-security.html [↑](#footnote-ref-2)