



00

Statistical basis and overviews

1790-1700

Swiss Federal Statistical Office Data Innovation Strategy

Purpose, strategic objectives
and implementation steps



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

Neuchâtel 2017

Published by: Federal Statistical Office (FSO)

Information: Bertrand Loison, FSO, tel. +41 58 463 67 70,
bertrand.loison@bfs.admin.ch

Editor: Bertrand Loison, FSO
Diego Kuonen, Statoo Consulting

Series: Swiss Statistics

Topic: 00 Statistical Basis and Overviews

Original text: English

Translation: FSO language services

Layout: DIAM Section, Prepress/Print

Front page: FSO; Concept: Netthoevel & Gaberthüel, Biel;
Photograph: © vinnstock – Fotolia.com

Copyright: FSO, Neuchâtel 2017
Reproduction with mention of source authorised
(except for commercial purposes).

Print format orders: Federal Statistical Office, CH-2010 Neuchâtel,
tel. +41 58 463 60 60, fax +41 58 463 60 61,
order@bfs.admin.ch

Price: Free of charge

Downloads: www.statistics.admin.ch (free of charge)

FSO number: 1790-1700

Table of contents

Introductory note	5
Introduction	7
Our view on data innovation	9
The current legal basis	9
Primary and secondary Data	10
The inductive-deductive reasoning cycle	11
First strategic objective	11
Second strategic objective	13
Overarching consideration	13
Realising our strategic objectives	14

Introductory note

The focus of the strategy is to augment and/or complement existing basic official statistical production at the Swiss Federal Statistical Office (FSO) in the areas where data innovation (as defined below) makes sense. The scope of this strategy, as outlined below, is limited to certain elements of statistical innovation. It does not claim to fully englobe all aspects of innovation at the FSO, nor does it aim to establish an order of prioritisation. Moreover, the strategy will be progressively revised and updated according to how the objectives and implementation steps set out below are fulfilled.

Introduction

The digital revolution is built on data and is well underway: terms such as big data, cloud, internet of things, internet of everything, fourth industrial revolution, smart cities and data economy are no longer just words on everyone's lips, but concepts that are changing the habits of consumers and businesses. This digital transformation started with the first wave of digitalisation: the technical digitalisation of converting analogue contents and services into digital ones – without fundamental changes to the underlying processes – that resulted in the (big) data revolution. However, big data are a data management infrastructure with underlying hardware, software and architecture and should not be “taken for museum purposes” only. As early as 1942 the official statistician W. Edwards Deming noted that the “ultimate purpose of taking data is to provide a basis for action or a recommendation for action”. As such, a second wave of digitalisation is needed to enable learning from big data and to generate value from this (big) data revolution for society as a whole: the FSO calls it **data innovation**. The biggest challenge in this respect is the veracity of the “data pedigree”, *i.e.* the **trustworthiness of the data**, including the reliability, capability, validity and related quality of the data, as well as the transparency of the related production process. In a world of post-truth politics and fake news, **data veracity** is more important than ever.

Official statistics has always dealt with a diversity of data sources. Since its foundation in 1860, the FSO has had to continuously assess the veracity of its data pedigree in order to deliver reliable and trusted statistical services. However, (re-)using existing data sources and/or using new data sources could provide the FSO with an innovation opportunity to augment and/or complement its official statistics and to deliver a more efficient, timely, effective and complete statistical service.

It is important to note that the data revolution is relevant not only in the world of official statistics.

At an international level, the access and use of new data sources are part of the discussion in global and regional statistical conferences. The 2030 Agenda for Sustainable Development and its sustainable development goals are game changers: data and (official) statistics are receiving much more attention at the political level. Due to such demands, the access and use of new data sources have to be discussed, tested and if possible, implemented. The Cape Town Global Action Plan for Sustainable Development Data, adopted at the 48th UN Statistical Commission in March 2017, calls for different actions related to new data sources. These include the identification

and removal of barriers to the use of new data sources, including register, administrative and other data from new sources, as well as the coordination of efforts to incorporate such data into official statistical production. The work conducted by the FSO will contribute to this action.

At a national level, different strategies aiming to enhance the access and usage of new data sources as well as register data have been drafted in recent years (e.g. the Digital Switzerland Strategy and the National Data Policy). However, these strategies are mostly technology-driven and place little emphasis on the use of such data for statistical purposes. Official statistics and in particular the FSO have encountered mounting pressure in parliament to reduce the burden on respondents by using register and administrative data, as well as other potential data sources stemming from advances in digitalisation. In this respect, the work conducted by the FSO will both help to position official statistics in the implementation of these strategic initiatives as well as contribute to the development of efficient solutions to respond to political pressure and the corresponding expectations.

Our view on data innovation

By understanding **analytics** as the science of learning from data (or of making sense of data), the FSO defines

data innovation as the application of *complementary analytics methods* (e.g. predictive analytics using approaches from advanced statistics, data science and/or machine learning) to existing (or traditional) and/or new (or non-traditional) data sources

to sustain the role of official statistics in the democratic process in Switzerland by ensuring that the information we provide remains reliable, transparent and trustworthy.

Data innovation could open up opportunities for more cost-effective interventions, as well as provide new insights that may have been overlooked by traditional approaches and analytics methods. The goal of data innovation is to enhance the quality, scope and cost-efficiency of statistical products and to reduce the response burden on households and businesses. Data innovation also has the potential to improve the timeliness and level of detail of official statistics.

The current legal basis

According to the law (Federal Statistics Act), the FSO and other official statistics bodies present representative results objectively and impartially on the status of, and changes in, important topics for data-informed decision and policy making, such as population, the economy, society, education, research, spatial planning and the environment in Switzerland. To fulfil its mandate, the FSO already uses different data sources. According to the law, the FSO initially has to check whether register or administrative data of appropriate quality are already available from the Confederation, the cantons, the communes or other entities of public law. Only if this is not the case, and if the FSO is mandated to provide statistical information according to the multi-year programme, can the FSO collect data on its own through censuses and surveys. The number and type of surveys are limited to what is strictly necessary. Data collected by the FSO have the advantage of following a clear statistical purpose and of data veracity and the related data quality being under control of the FSO.

Primary and secondary Data

In the context of data innovation, the FSO divides data sources into primary and secondary data.

Primary (“made” or “designed”) **data** have been collected – and designed – by the FSO for statistical purposes to explain and check the validity of specific existing ideas, *i.e.* through the operationalisation of theoretical concepts. Learning from such primary data is known as primary (or top-down, *i.e.* explanatory and confirmatory) analytics. The corresponding analytics’ paradigm is “deductive reasoning” that starts with an idea or theory (“idea first”). Examples of primary data are traditional data sources like censuses and surveys that have been collected by the FSO for statistical purposes.

In contrast, **secondary** (observational or “found” or “organic”) **data** have been collected – and designed – for other reasons, often without FSO supervision, and could be used to create new ideas or theories. Learning from such secondary data is known as secondary (or bottom-up, *i.e.* exploratory and predictive) analytics. The corresponding analytics’ paradigm corresponds to “inductive reasoning” that starts with data (“data first”). Examples of secondary data are non-traditional data sources such as FSO internal and external register data, administrative data, and other digital data from devices, machines, sensors, satellites, drones and social media. As secondary data sources were not designed to be used directly in official statistical production systems, they need to be made fit for purpose for statistical inference, *i.e.* so that conclusions can be drawn from them for the purpose of official statistics by deductive reasoning.

To do so, secondary data can be further classified into identifiable and non-identifiable data. **Identifiable data** can be meaningfully associated with a single unit at a given place and time, such as an individual, institution, product or geographical location (*e.g.* register data, administrative data, satellite imagery, geospatial information and product barcodes). **Non-identifiable** data cannot be made identifiable at any such level (*e.g.* Google trends data, Twitter feeds and other forms of social media). Identifiable secondary data could be made fit for purpose for statistical inference if their veracity has been successfully assessed (as is the case with the FSO’s current use of its internal register data), whereas non-identifiable secondary data are of limited use for statistical inference because it is not possible to assess their veracity.

The inductive-deductive reasoning cycle

It is important to note that the two approaches of analytics (*i.e.* inductive and deductive reasoning) are *complementary and should proceed iteratively and side by side* in order to enable continuous improvement and data-informed decision and policy making. This implies that the analytics methods currently used at the FSO will still be needed together with complementary analytics methods.

For example, the information (ideas) generated by inductive reasoning cannot explain if and why these discoveries are useful and to what extent they are valid. The confirmatory tools of deductive reasoning will be needed to confirm the discoveries and evaluate the quality of decisions based on those discoveries.

First strategic objective

The first strategic objective within the Federal Statistics' multi-annual programme for 2016–2019 is to provide basic statistical production on a permanent basis and to adapt it continuously to its users' requirements. Among several implementation steps, the development of guidelines on potential uses of new data sources was proposed.

The first strategic objective of the present data innovation strategy is directed towards creating awareness that **data innovation is a main strategic issue**.

Moreover, the first strategic objective will be progressively revised and updated depending on how the steps defined below are fulfilled.

Strategic objective 1: Develop data innovation guidelines and investigate the feasibility of the application of complementary analytics methods to existing (or traditional) and/or new (or non-traditional) data sources, along with the goal of augmenting and/or complementing any existing basic statistical production for which data innovation makes sense.

The **first step** is the application through FSO internal pilot projects of complementary analytical methods to existing (or traditional) FSO internal primary data sources and already matched identifiable secondary data sources (if such exist). It is about augmenting and/or complementing existing learning from data , *i.e.* to use such data sources in new ways to gain practical experience and make an inventory of existing challenges, resources, skills and technologies in order to perform data innovation and investigate “quick-wins”.

If feasible, a **subsequent second step** could then be to complement and/or augment existing statistical production at the FSO with data innovation generated from the application of complementary analytics methods to additional secondary data already in use at the FSO.

A **subsequent third step** could be to apply complementary analytics methods to only new – hitherto unused at the FSO – secondary data to investigate and produce new statistical information and statistics in particular statistical domains.

The preferred data source sequence for the FSO’s data innovation strategy is:

1. FSO internal primary data sources and already matched identifiable secondary data sources (if they are already used in FSO’s current statistical production);
2. additional secondary data sources already in use at the FSO;
3. new – until now unused at the FSO – secondary data sources.

The aim of this sequence is to raise efficiency, reduce costs and minimise the administrative burden on businesses and individuals. Moreover, this sequence should allow for a better understanding of methods, technologies and tools, without adding unnecessary complexities as would be the case by starting with the third step, which might raise legal, technological (IT) and related capacity problems at the same time. Furthermore, greater experience of complementary analytics methods will sharpen the skills needed to find new data sources adapted to the specific aims of official statistical production.

Second strategic objective

Data innovation requires a paradigm shift by combining deductive and inductive reasoning that will necessarily lead to a change in the production and communication of official statistics.

It is important to note that no common generic methodological approach exists for the procedures outlined above for using data innovation in official statistical production. The challenges and opportunities are specific to each data innovation application and type of data. As such, current FSO process frameworks and process models may need to be adapted and extended to enable data innovation.

This means that **communication** on these issues is key. This leads us to the second strategic objective.

Strategic objective 2: Develop and implement FSO internal and external communication measures to increase awareness of the added value of data innovation in official statistics and the related paradigm shift.

Overarching consideration

The FSO's overarching consideration when performing the data innovation steps described above must be the core values of data privacy, data protection policies, data security, data governance and public trust.

Realising our strategic objectives

To meet our strategic objectives and to enable data innovation, the following specific actions will be needed:

- Evaluate the need to adapt and extend the current conceptual FSO data quality frameworks to include data innovation (e.g. with respect to complementary analytics methods, e.g. predictive analytics);
- Evaluate the requirement for new skills among FSO staff with respect to complementary analytics methods and to the (IT) technologies and tools needed to enable data innovation. The first step mentioned above and the related pilot projects are key to establishing the requirements in terms of the skills needed for data innovation. Upon completion, this will enable the FSO, if necessary, to:
 - develop, for example, in cooperation with private companies and/or universities, a new FSO internal training curriculum aimed at the continuing education of FSO staff, and which covers the complementary analytics methods and soft skills required of a skilled workforce. These include creativity, collaboration, curiosity, problem solving, communication and storytelling;
 - define and implement the usage of an FSO internal and/or external data lab (“sandbox”) to train staff in the new skillsets needed to use complementary analytics methods;
 - define the requirements for adapting (IT) technologies, FSO internal and/or external infrastructures and tools.
- Investigate the legal issues related to data innovation, e.g. legal requirements and incentives for securing agreement on obtaining secondary data from the private sector (“data ownership”) without compromising the business interests of the data owners (see third step mentioned above);
- Discuss and exchange information with all FSO stakeholders within the FSO, the statistical systems and the research community.

The FSO regards effective and sustained multidisciplinary partnerships across the Confederation, industry, academia and the statistical community as a critical factor for assembling the full portfolio of skills and experience that will be needed to succeed with its data innovation strategy.

Orders

Tel. +41 58 463 60 60

Fax +41 58 463 60 61

order@bfs.admin.ch

Price

free of charge

Downloads

www.statistics.admin.ch (free of charge)

FSO number

1790-1700

**Statistics
counts for you.**

www.statistics-counts.ch