Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

FSO's Data Innovation Strategy

Prof. Dr. Bertrand Loison, Vice-Director

Swiss Federal Statistical Office

UN Global Working Group on Big Data for Official Statistics
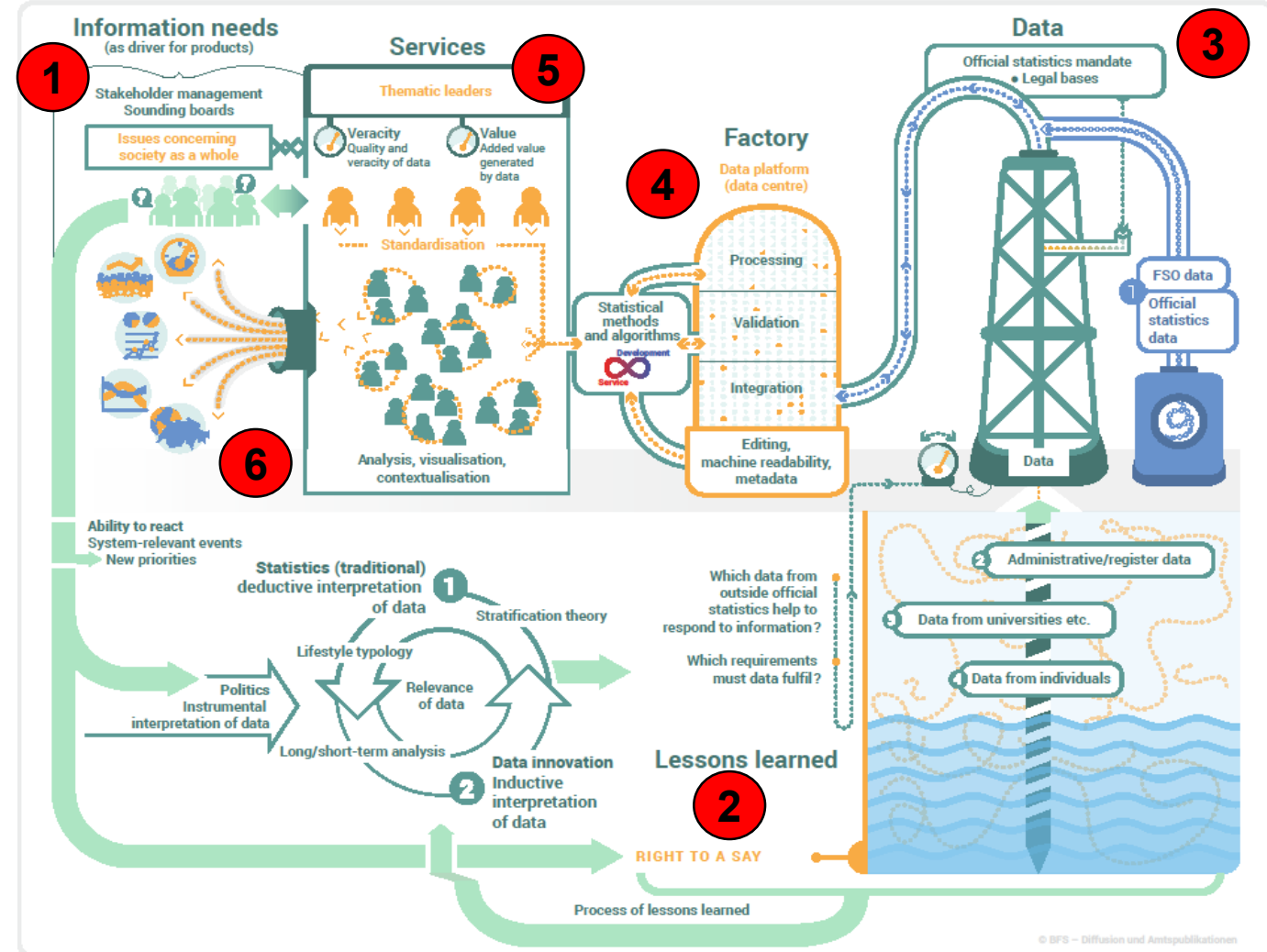
Kigali, Rwanda, 29 April 2019

Agenda

1. **The FSO's Value Chain**

2. Preparing the FSO's Data Innovation Strategy

3. The FSO's Data Innovation Strategy

4. The implementation of the FSO's Data Innovation Strategy

5. Pilot projects

6. Conclusion

An output-driven value chain for official statistics

1 Information needs

2 Lessons learned

3 Data

4 Factory

5 Services

6 Output

Agenda

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

Model: Cornerstones of an Data Innovation Strategy

CURRENT STATUS

OBJECTIVE/ASPIRATION

HOW ARE WE CLOSING THE GAP?

CURRENT STATUS

OBJECTIVE/ASPIRATION

HOW ARE WE CLOSING THE GAP?

CURRENT STATUS

OBJECTIVE/ASPIRATION

HOW ARE WE CLOSING THE GAP?

CURRENT STATUS

OBJECTIVE/ASPIRATION

HOW ARE WE CLOSING THE GAP?

Strategy   People

Data   Tools

Source : BAL, IMD, Lausanne, 2018

Agenda

1. The FSO's Value Chain

2. **Preparing the FSO's Data Innovation Strategy**

       - FSO's Cornerstones of an Data Innovation Strategy

       - FSO's Data Innovation Strategy: Defense vs Offense

       - FSO's Five Steps of an Analytics Roadmap

       - FSO's Data Category Map

3. The FSO's Data Innovation Strategy

4. The implementation of the FSO's Data Innovation Strategy

5. Pilot projects

6. Conclusion

Model: Data Innovation Strategy: Defense vs Offense

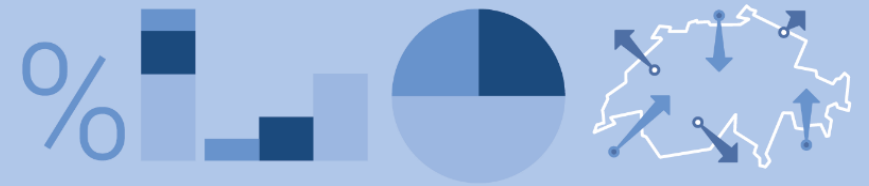|	⤴ DEFENSE	⤢ OFFENSE
KEY OBJECTIVES		
CORE ACTIVITIES		
DATA-MANAGEMENT ORIENTATION		
ENABLING ARCHITECTURE		

Source : adapted from "what's your data strategy?" by leandro dallemule and thomas h. davenport, may–june 2017, BAL, IMD, Lausanne, 2018

Agenda

1. The FSO's Value Chain
2. **Preparing the FSO's Data Innovation Strategy**
   - FSO's Cornerstones of an Data Innovation Strategy
   - FSO's Data Innovation Strategy: Defense vs Offense
   - **FSO's Five Steps of an Analytics Roadmap**
   - FSO's Data Category Map
3. The FSO's Data Innovation Strategy
4. The implementation of the FSO's Data Innovation Strategy
5. Pilot projects
6. Conclusion

Model: Five Steps of an Analytics Roadmap

Identify the Data needed?

Determine the Analytics methods

PROBLEM

TOOLS

PLAN

DATA

PEOPLE

Identify the Problem

Locate Data Sources

Set up Feedback Loops

Feedback Loops

Source : BAL, IMD, Lausanne, 2018
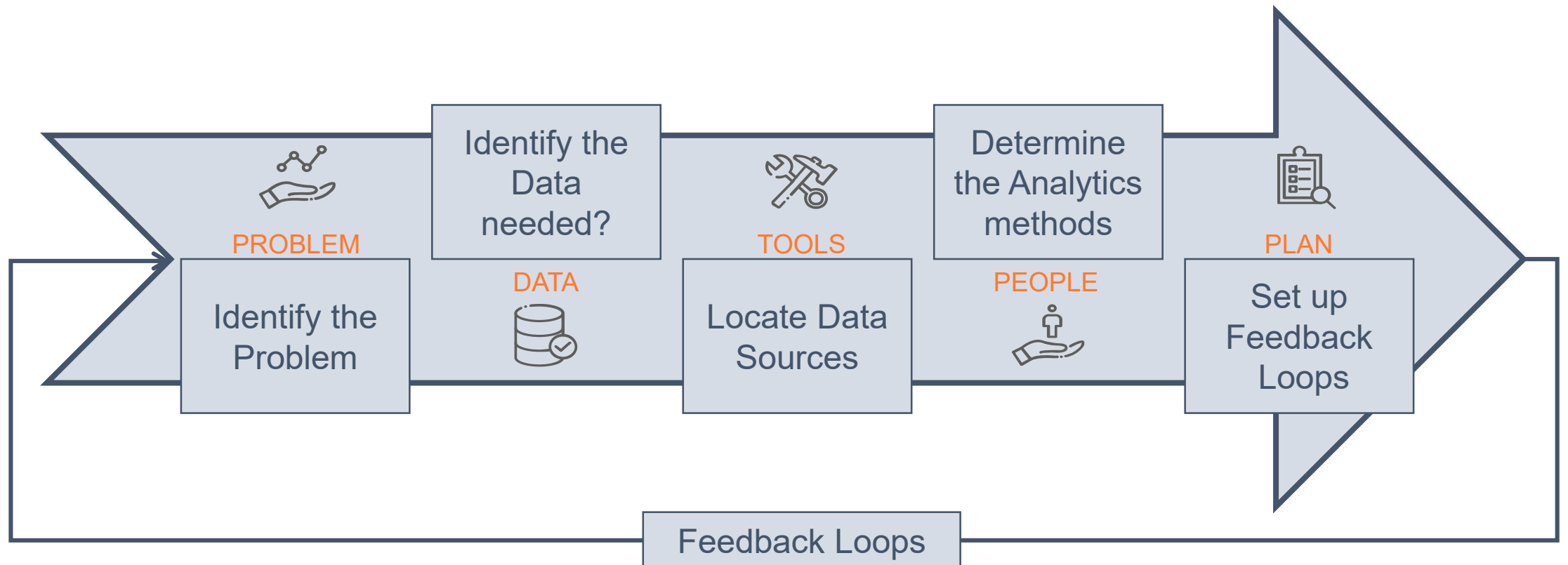
Agenda

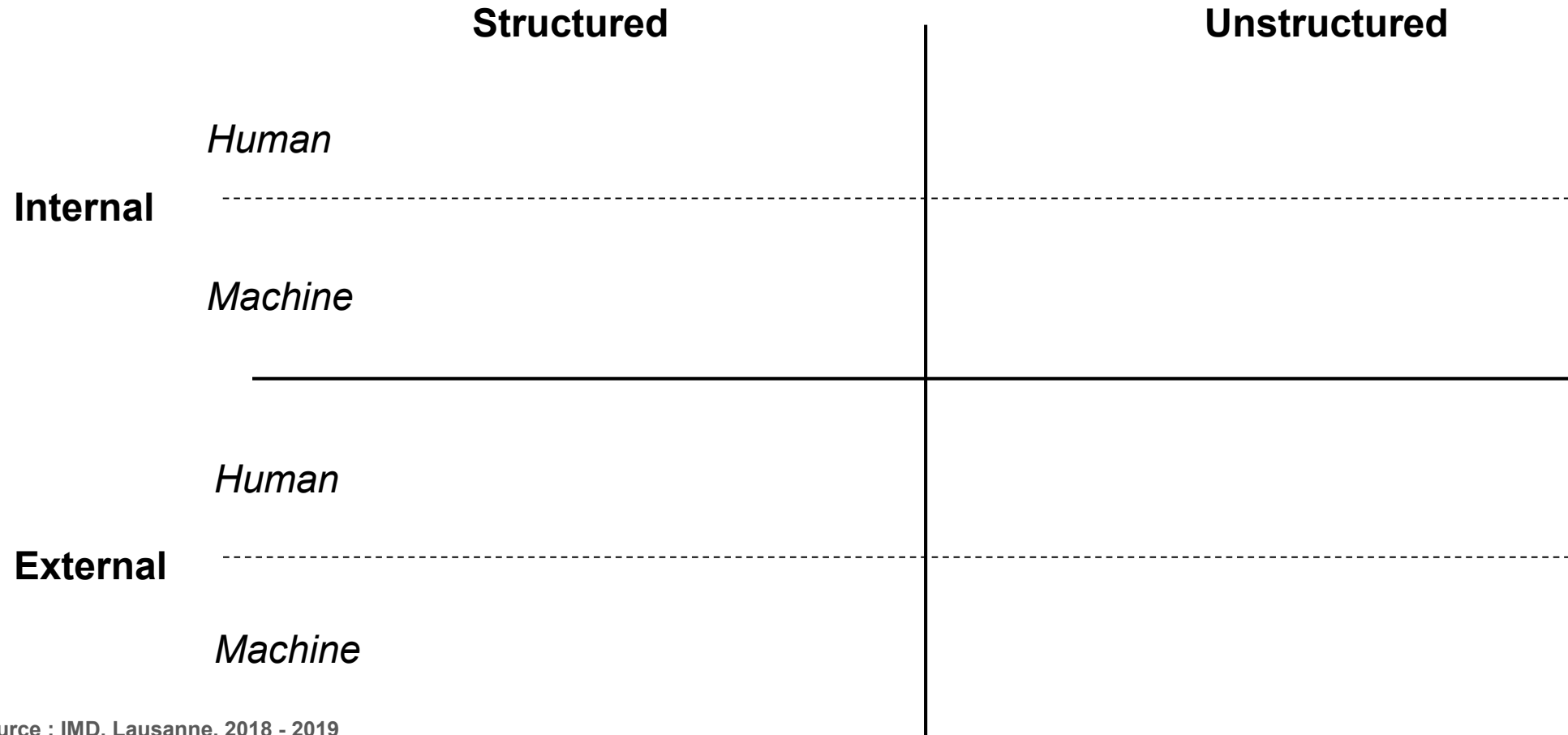1. The FSO's Value Chain

2. **Preparing the FSO's Data Innovation Strategy**

   - FSO's Cornerstones of an Data Innovation Strategy

   - FSO's Data Innovation Strategy: Defense vs Offense

   - FSO's Five Steps of an Analytics Roadmap

   - **FSO's Data Category Map**

3. The FSO's Data Innovation Strategy

4. The implementation of the FSO's Data Innovation Strategy

5. Pilot projects

6. Conclusion

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

Model: Data Category Map

	Structured	**Unstructured**

Human

Internal

Machine

Human

External

Machine

Source : IMD, Lausanne, 2018 - 2019

Agenda

What is Data Innovation ?

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.

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

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

Swiss Confederation

Neuchâtel 2017

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

Strategic objectives

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.

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.

How did FSO define primary and secondary Data ? - #1

Primary data ("made" or "designed") 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 analytics (or top-down, i.e. explanatory and confirmatory).

The corresponding analytics' paradigm is "deductive reasoning" that starts with an idea or theory ("idea first").

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

How did FSO define primary and secondary Data ? - #2

Secondary data (observational or "found" or "organic") 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 analytics (or bottom-up, i.e. exploratory and predictive).

Corresponding analytics' paradigm corresponds to "inductive reasoning" that starts with data ("data first").

How did FSO define primary and secondary Data ? - #3

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. 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.

Why is the inductive-deductive reasoning cycle so crucial ? - #1

Theory

Data
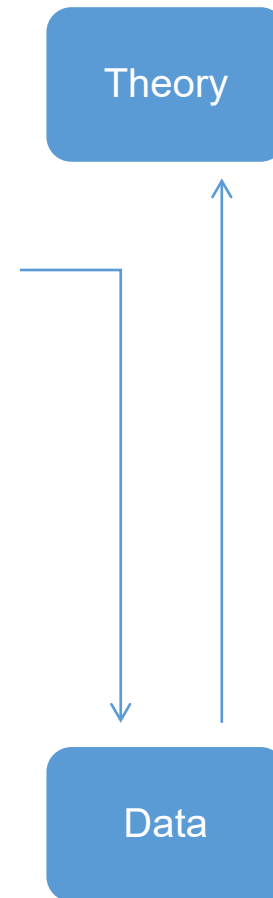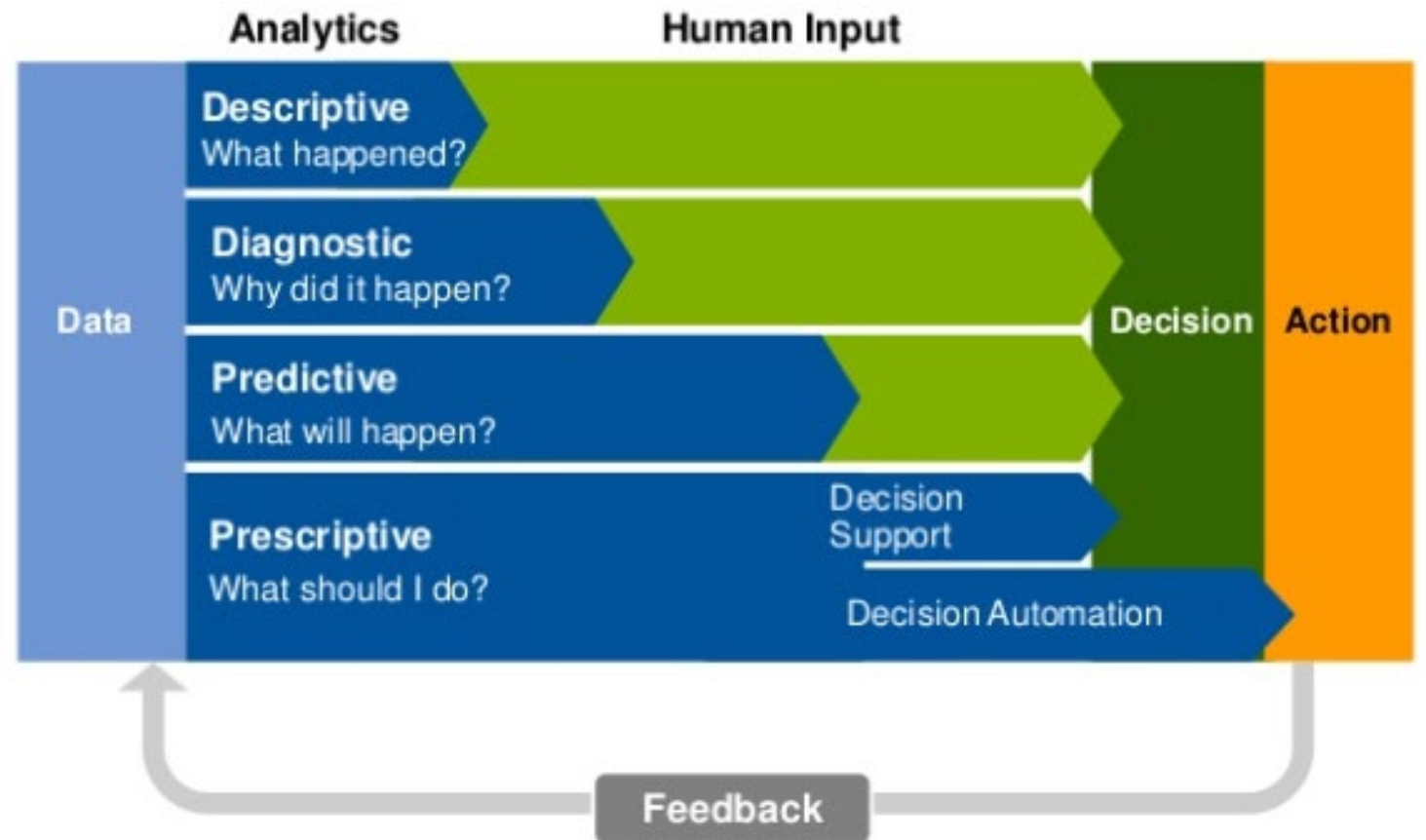
Why is the inductive-deductive reasoning cycle so crucial ? - #2

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.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

Why is the inductive-deductive reasoning cycle so crucial ? - #3

Questions that analytics tries to answer are strongly related to the correct application of the inductive-deductive reasoning cycle…

… which means having a thorough understanding of the algorithms and statistical methods and models you use

… and not just to use new data sources !

Source : **https://www.gartner.com/en/newsroom/press-releases/2014-10-21-gartner-says-advanced-analytics-is-a-top-business-priority**, **2014**
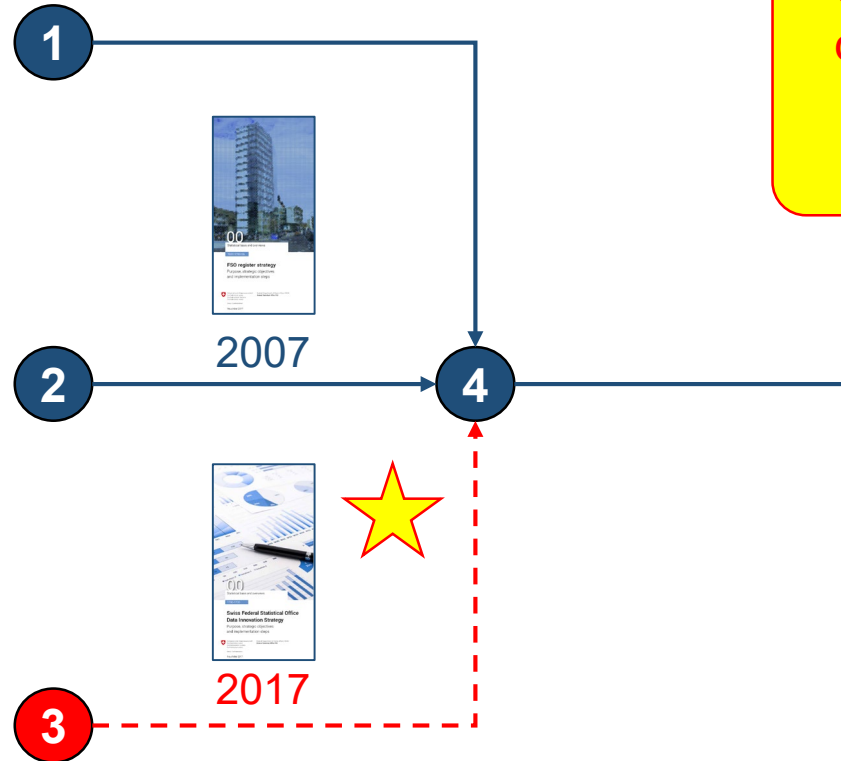
Agenda

1. The FSO's Value Chain

3. Preparing the FSO's Data Innovation Strategy

4. The FSO's Data Innovation strategy

5. **The implementation of the FSO's Data Innovation Strategy**
   - **- Data collection (input)**
   - **- Data processing (throughput)**
   - **- Data dissemination (output)**

6. Pilot projects

7. Conclusion

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

Data collection (input)

Surveys

1

2007

Registers and
administrative data

2 ——————— **4** ——————→

New data sources
(big data)

2017

3

The FSO has been in the digital era for over 10 years. New data sources (big data) are simply the next step… but a major one !

Data collection (input)

According to the Federal Statistics Act, the institutions of federal statistics must limit the administrative burden on data suppliers (individuals, enterprises and authorities).

To this end, since 2007 the FSO has implemented a strategy to use administrative data that already exists within the public administration wherever possible instead of conducting surveys (on paper, online or by telephone).

The data innovation strategy aims to apply complementary analysis methods (e.g. predictive analysis by advanced statistical techniques, data science and/or automatic learning) to the FSO's primary sources and identifiable secondary sources already crossed (if they are already used in the FSO's current statistical production, to additional secondary data already used at the FSO and to new secondary sources (not used at the FSO so far).
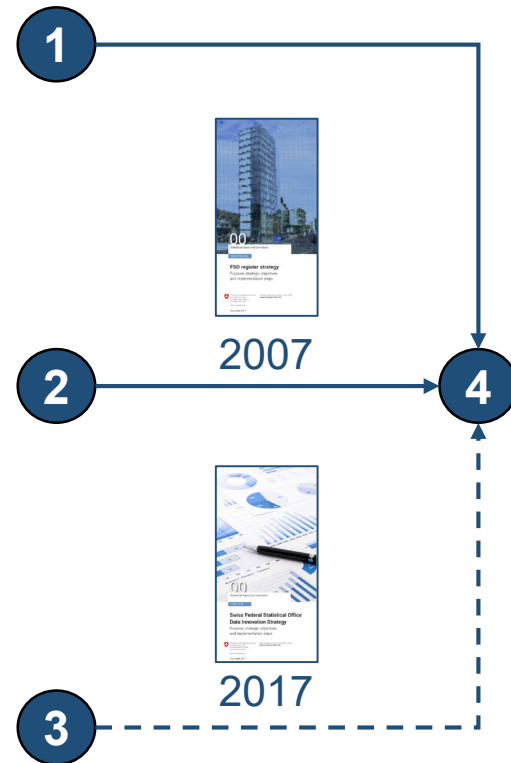
Data processing (throughput)

Surveys

①

Registers and administrative data

②

2007

New data sources (big data)

③

2017

④

Between 2008 and 2010, the FSO developed a corporate Statistical Information System (SIS ver. 1.0) to fit the challenges of register based statistics. A new SIS release (SIS ver. 2.0) is currently being developed to fit the needs of new data sources.

2010
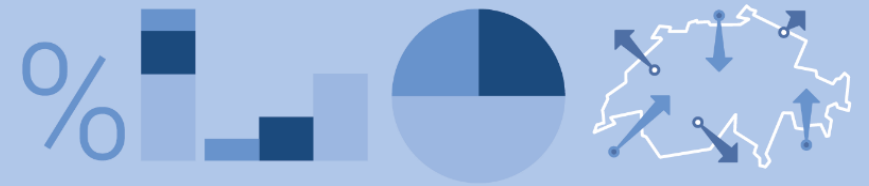2017

⑤

Statistical Information Systems (SIS)
+

⑥

Complementary analysis methods

The FSO has adopted a new data innovation strategy. This constitutes the first response by official statistics to the challenges posed by the digital revolution.
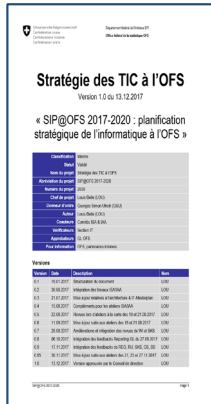
2017

Data processing (throughput)

The aim of the SIS ver. 2.0 program is to set up a statistical information system (SIS) which takes into account the constraints of integrated statistics, data protection and traceability.
This system must be in line with the FSO's added value creation process and must make available to all of the FSO's productive organisational units, all data providers and all kinds of data sources the functions that they need.
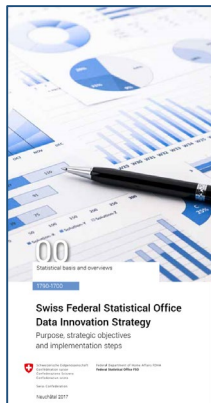
The data innovation strategy aims to apply complementary analytics methods (e.g. predictive analytics using approaches from advanced statistics, data science and/or machine learning) to apply complementary analysis methods (e.g. predictive analysis by advanced statistical techniques, data science and/or automatic learning) to the FSO's primary sources and identifiable secondary sources already matched (if they are already used in the FSO's current statistical production, to additional secondary data already used at the FSO and to new secondary sources (not used at the FSO so far).
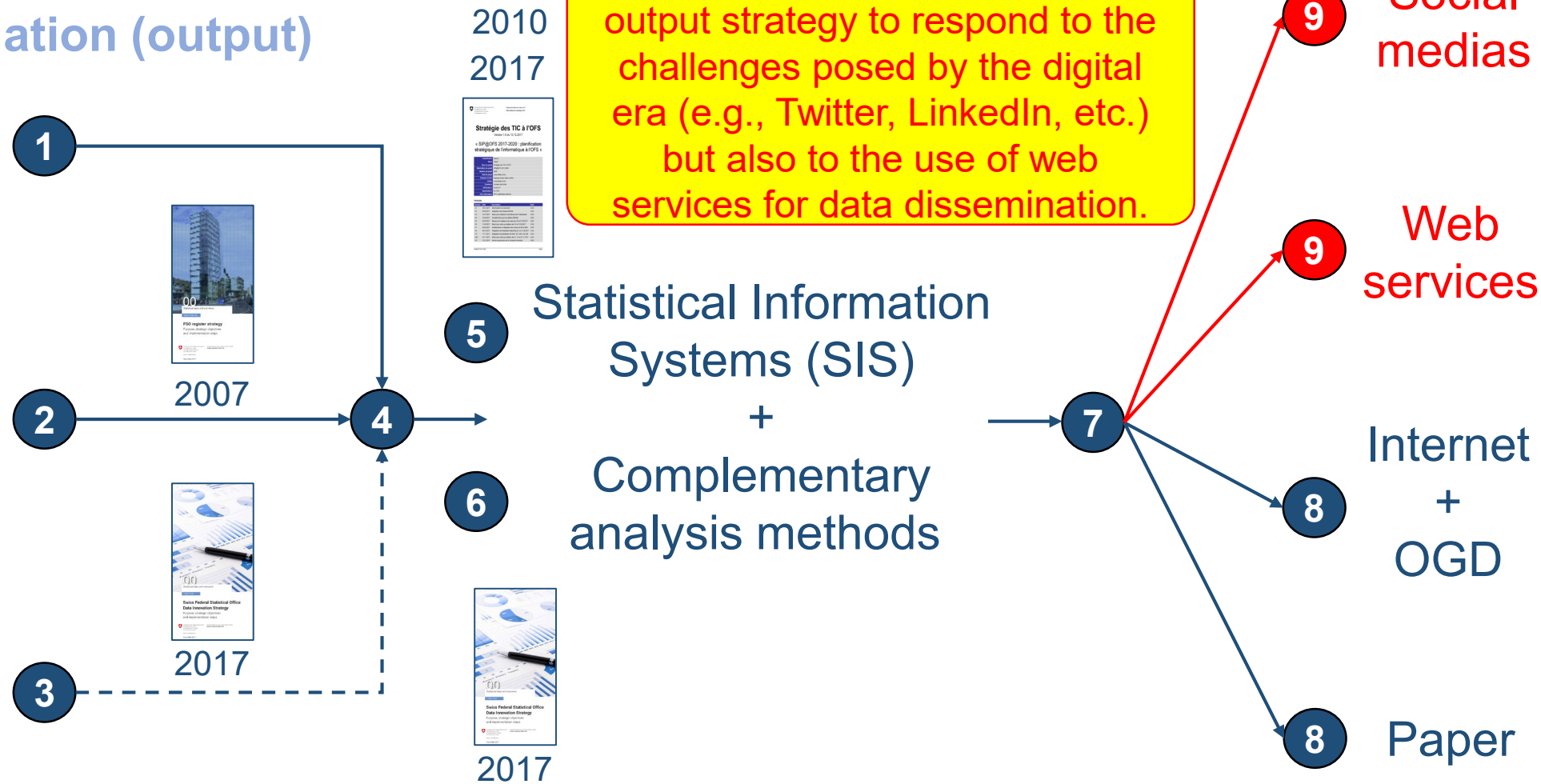
Data dissemination (output)

Surveys

Registers and administrative data

New data sources (big data)

2010
2017

In 2017, the FSO renewed its output strategy to respond to the challenges posed by the digital era (e.g., Twitter, LinkedIn, etc.) but also to the use of web services for data dissemination.

Stratégie des TIC à l'OFS

« SIP@OFS 2017-2020 : planification stratégique de l'informatique à l'OFS »

FSO register strategy

2007

Swiss Federal Statistical Office
Data Innovation Strategy

2017

Swiss Federal Statistical Office
Data Innovation Strategy

2017

1
2
3
4
5
6
7
8
8
9
9

Statistical Information Systems (SIS)
+
Complementary analysis methods

Social medias

Web services

Internet + OGD

Paper

Agenda

1. The FSO's Value Chain

3. Preparing the FSO's Data Innovation Strategy

4. The FSO's Data Innovation Strategy

5. The implementation of the FSO's Data Innovation Strategy

6. **Pilot projects**

7. Conclusion

Pilot project 1 : Business Registers

Coding of economic activities of enterprises
One of the FSO's key tasks is the correct coding of the economic activities of enterprises. This project strives to automate the coding of the economic activities of enterprises using machine learning methods applied to data already available within the FSO (data from surveys, descriptions in the commercial register, key words, explanatory notes for classifications etc.) with a view to supporting the production units.

Pilot project 2 : Territory

Automate the visual interpretation of aerial images
The FSO's land use statistics form an invaluable tool for the long-term observation of the territory. This project involves learning and mastering the use of artificial intelligence (IA) technologies to eventually automate (even partially) the visual interpretation of aerial images in order to detect and classify changes.

Pilot project 3 : Business Surveys

Small area estimation method for the Job Statistics
The project evaluates the potential of the small area estimation method for the Job Statistics. The aim is to produce reliable estimates of the total number of jobs and FTEs for cantons, major towns and NOGA levels that were not anticipated in the sample plan. The aim of the present project is to evaluate the small area estimation method's potential to respond to these needs without increasing sample sizes.

Pilot project 4 : Social Security

Grouping of typical prospective trajectory patterns
The grouping of typical prospective trajectory patterns concerning the receipt of benefits in the social security system and employment and the estimation of group affiliation through the use of individual variables and retrospective trajectory data applying a machine learning approach.

Pilot project 5 : Education

Extend and accelerate plausibility checks using machine learning algorithms

This project aims to extend and accelerate plausibility checks in the FSO using machine learning algorithms, while at the same time improving data quality.

Statistical offices carry out plausibility tests to check the quality and reliability of administrative data and survey data. Data that are either clearly incorrect or seem at least questionable are sent back to data suppliers with a correction request or comment. Until now, such plausibility tests have mainly been carried out at two different levels: either through manual checks or automated processes using threshold values and logic tests.

Machine learning could help to ensure faster and more accurate checks. This approach would rely on an algorithm using historical data at first.

# FSO's Experimental statistics

Experimental statistics are produced using new methods and/or new data sources and are therefore in line with the ==FSO's data innovation strategy== and the Confederation's multi-annual programme for federal statistics. This site contains descriptions of the (pilot) projects currently being developed.

By publishing them we can involve users and partners at an early stage for both the development and consolidation of projects.

The aim of these statistical projects is to better meet users' needs in terms of efficiency, quality and speed. However, these statistics still have the potential to evolve, especially regarding their methodology, which is still being assessed. For this reason they are clearly marked as experimental and carry a logo that can easily be recognised.

EXPERIMENTAL ST
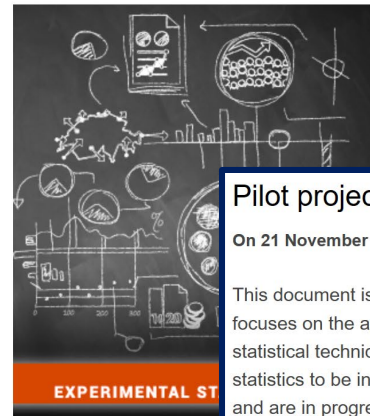
Published statistics

EXPERIMENTAL STATISTICS

Small area estimation (communes) of economic ac[...] rate in the structural survey

The structural population survey provides important information on the population, including information ab[...] work. The whole purpose of Small Area Estimation is t[...] the boundaries imposed by standard methods.

The study showed that it is possible to obtain reliable estimates for both annual economic activity rates for communes that had a sample of at least 100 people.

Pilot projects within the data innovation strategy

On 21 November 2017, the FSO published its ==data innovation strategy== 🔗.

This document is the FSO's first response to the wider subject of digitalisation. More specifically, it focuses on the application of complementary analysis methods (e.g. predictive analysis using advanced statistical techniques, data science and machine learning) that enable the current production of official statistics to be increased or completed. Five pilot projects have been chosen to implement this strategy and are in progress. Each project is described below.

EXPERIMENTAL STATISTICS
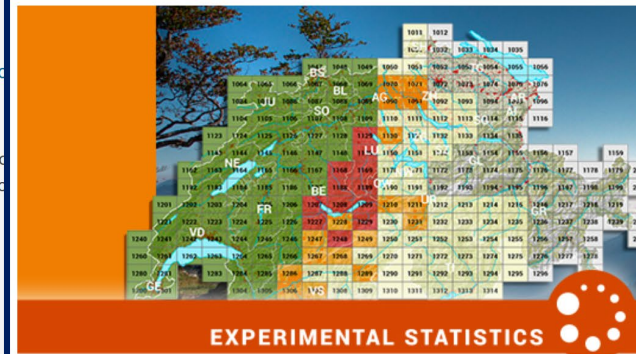
Project "Area Statistics Deep Learning" (ADELE)

The FSO's land use statistics are an invaluable tool for long-term land observation. Thisproject involves learning and mastering the use of artificial intelligence (AI) technologiesto eventually automate (even partially) the visual interpretation of aerial images in order todetect and classify changes.

EXPERIMENTAL STATISTICS

Project "Automation of NOGA coding" (NOGAuto)

Automation of the coding of the economic activity of enterprises using Machine Learningmethods applied to data already available within the FSO (data from surveys, descriptionsin the commercial register, keywords, explanatory notes for classifications etc.) to supportcoding.

Agenda

1. The FSO's Value Chain
3. Preparing the FSO's Data Innovation Strategy
4. The FSO's Data Innovation Strategy
5. The implementation of the FSO's Data Innovation Strategy
6. Pilot projects
7. **Conclusion**

Conclusion - #1

1. 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).

2. Evaluate the requirement for new skills in FSO staff with respect to complementary analytics methods and to the (IT) technologies and tools needed to enable data innovation (capability building).

3. 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).

4. Discuss and exchange information with all FSO stakeholders within the FSO, the statistical systems and the research community.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Ufficio federale di statistica UST
Uffizi federal da statistica UST

Conclusion - #2

Work in progress !

5. The FSO is currently defining version 2.0 of its data innovation strategy for the years 2020 - 2023. The creation of an official Data Innovation Lab and a Competence Center for Data Science will (probably) be at the heart of the new strategy.

Strategic objective 1: To create a central, cross-departmental and independent "Data Innovation Lab" (DIL) with a focus on data innovation services, i.e. the application of complementary analytics methods to data (and not to the type of data source and/or technology), for the FSO (I), the whole Swiss public statistics system (II) and the whole federal administration (III).

Questions & Answers