Data scientists vs. Statisticians: Lessons learned after two years of practical experience

Prof. Dr. Bertrand Loison, Swiss Federal Statistical Office, Vice-Director

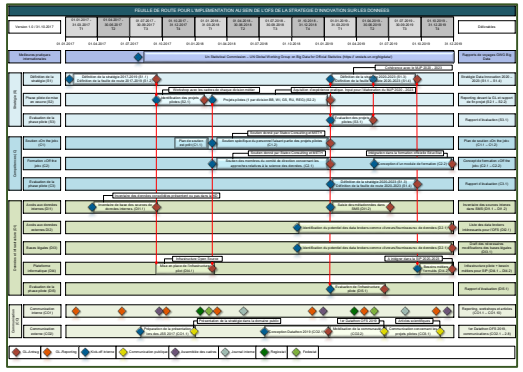
5th International Conference on Big Data for Official Statistics

Session “Data Science and Capacity Development in Official Statistics”

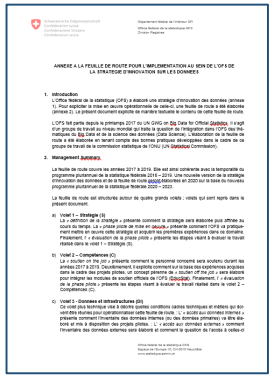
Kigali, Rwanda, 3 Mai 2019

Two years of practical experiences (2017 – 2019)

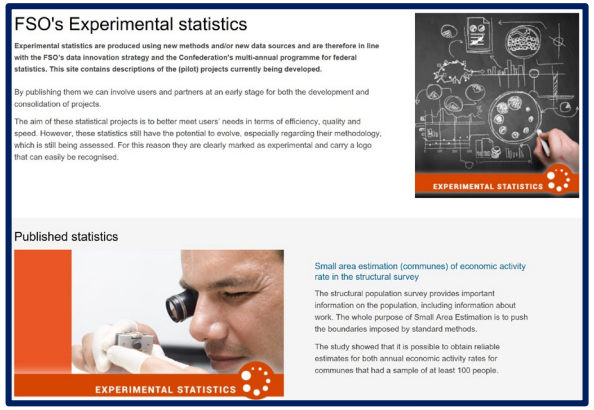
+



+



=

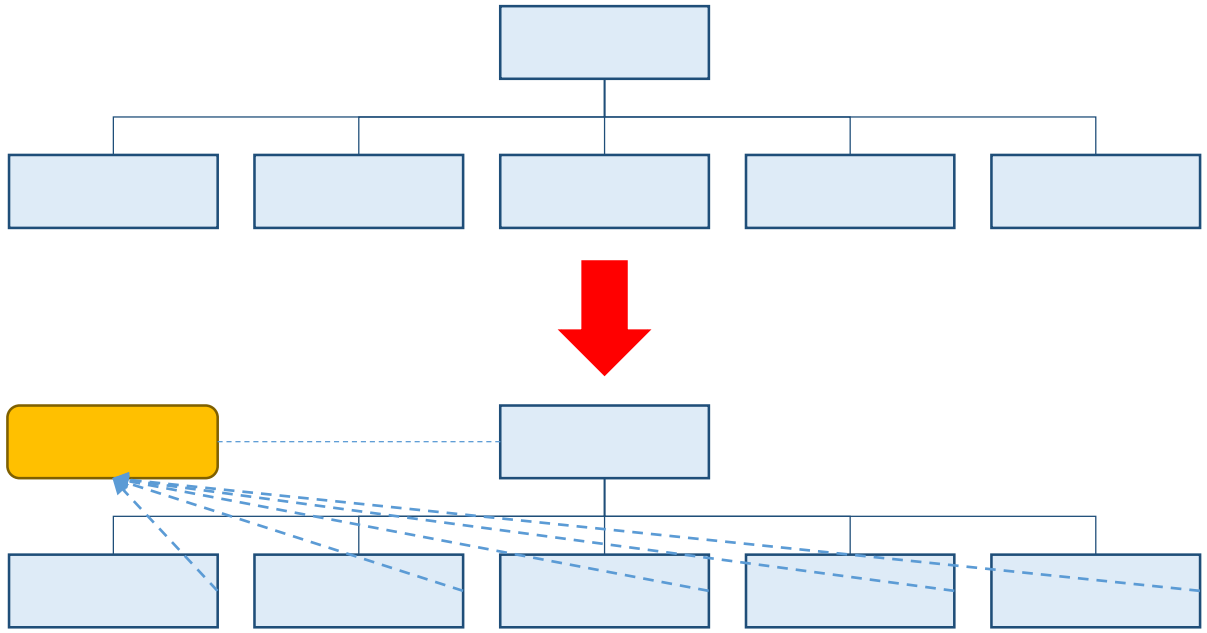


Sources:
Data Innovation Strategy 1.0 : https://www.bfs.admin.ch/bfs/en/home/news/whats-new.gnpdetail.2017-0673.html
Pilot Projects : https://www.experimental.bfs.admin.ch/en/

Agenda

1. **What did we do ?**
2. What are our experiences ?
3. What will be the next major step ?

We have created an **ad-hoc** agile organization **inside** FSO - #1

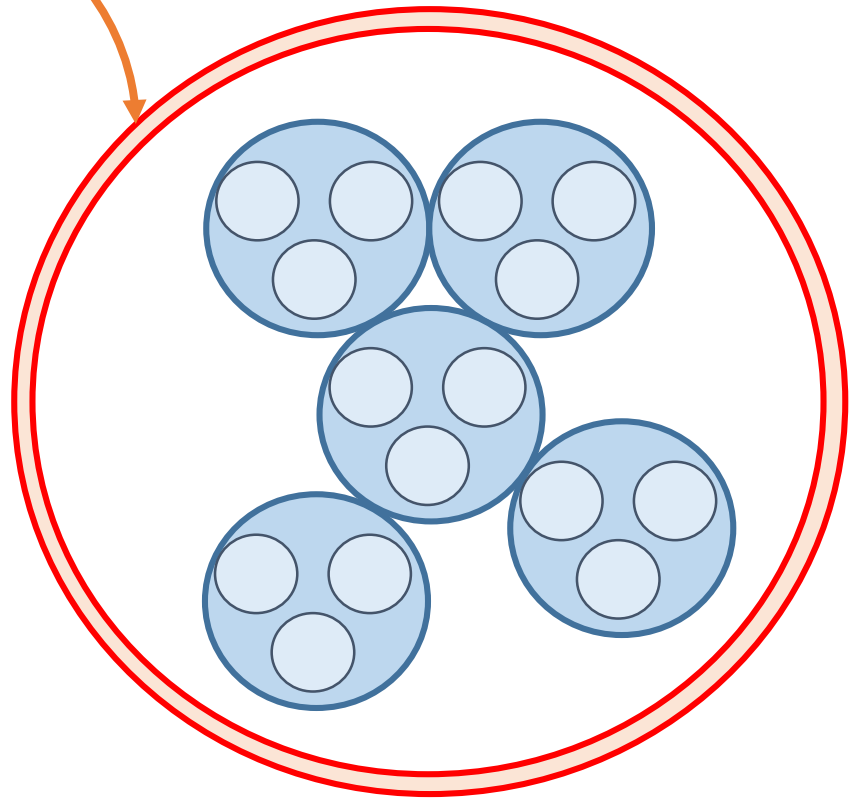


26 people (Statisticians, Methodologists, IT Specialists) aged between 26 and 58 years old !

Rules

1. Ad-hoc organization is under the final responsibility of a board member.
2. Members of the ad-hoc organization have been stayed subordinated in their line organization !
3. Five pilot projects are managed by the ad-hoc organization. The final goal is to go into production at the end of 2019.

1. We have created an ad-hoc agile organization inside FSO - #2



- Classical hierarchy that needed to be forgotten…
- High degree of autonomy was and is still required…
- Choosing the right people at the beginning was not easy…

2. We have complemented their skills with off- and on-the-job teaching - #1

Off-the-job (9 x 1 day = 9 days dispatched over 3 months)

- R for Data Science
- Python for Data Science
- Selected topics in data preparation/preprocessing
- Dimensionality reduction (*e.g.* PCA)
- Clustering (*e.g.* k-means)
- General introduction to supervised methods
- Regression (*e.g.* linear, logistic)
- Classification (*e.g.* k-NN, support vector machines)
- Decision trees
- Bayesian learning
- Neural networks
- Deep neural networks (*e.g.* CNN, RNN and LSTM)
- Model evaluation
- Feature selection

2. We have complemented their skills with off- and on-the-job teaching - #2

On-the-job (over 2 years)

Scientific advisors

- Professor for Data Science (University of Geneva) and the FSO's deputy head of methods have accompanied/supervised the five pilot projects and validate the results from the scientific point of view.

Team of "Data Scientists"

- Key idea is to share/transfer the skills and the knowledge inside the FSO Organization.
- Trying to avoid an organizational split between "traditional statisticians" and "modern statisticians or data scientists".

Agenda

1. What did we do ?

2. **What are our experiences ?**

3. What will be the next major step ?

What are our experiences ? - #1

Hard skills

- Math & statistical skill, Technical skill and subject matter can be learned. These skills are not always easy to acquire but it seems not to be the hard part of the game!

➢ Up to 2020 we will integrate a standard curriculum "Data Scientist" in our internal teaching program. We will collaborate more with Universities to define the exact content.

Soft Skills

- Soft skills (*e.g.* collaboration, creativity, problem solving, communication, story telling) need a change in the mindset of all employees.

➢ These skills are not so easy to teach… they primary need to be practiced on a daily basis

➢ Up to 2020 we will open a new mandatory curriculum "Agile Organization".

What are our experiences ? - #2

Academic curricula

- Swiss Universities offer standards curricula in data science. The content of these curricula often does not match with the FSO's expectations.

➢ FSO does not really need "Supermen or Superwomen" that can cover all the production process on his own.

National Statistical Institute as employer

- Recruiting Data Scientist for a NSI likes FSO is not easy. Salaries are not so attractive and the tasks that we can offer to a young data scientist do not offer the expected variety and freedom that they are looking for.

➢ FSO wants to empower its employees and not to replace them with Data Scientist!

Agenda

1. What did we do ?

2. What are our experiences ?

3. **What will be the next major step ?**

What will be the next major step ?

Work in progress !

Version 2.0 of our Data Innovation Strategy

The FSO is currently defining a 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