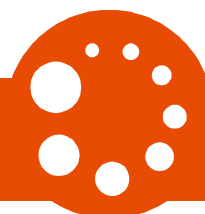




Pilot projects within the data innovation strategy

# Brief methodological evaluations in preparation for a possible transfer into production

## EXPERIMENTAL STATISTICS



Neuchâtel, 2020

**Published by:** Federal Statistical Office (FSO)  
**Information:** Jean-Pierre.Renfer@bfs.admin.ch, tel. +41 58 463 66 62  
**Editors:** Jean-Pierre Renfer, FSO  
**Topic:** 00 Statistical Basis and Overviews  
**Original text:** French  
**Translation:** FSO language services

**Layout concept:** Section DIAM  
**Downloads:** [www.statistics.ch](http://www.statistics.ch)  
**Copyright:** FSO, Neuchâtel 2020  
Reproduction with mention of source authorised  
(except for commercial purposes)

# 1 Pilot projects evaluated

The first five pilot projects from the FSO's innovation strategy 1.0 below have been evaluated in preparation for a possible transfer into production based on project reports from the end of June 2019. In summary, these are the following pilot projects:

- a) **ADELE** : Recognition of aerial images for the national land use statistics;

Complementary analysis methods used: "Various convolutional neural networks have been applied for practice on aerial images based on Keras";

- b) **ML\_SoSi** : Individual trajectories of the social security system;

Complementary analysis methods used: kmeans and optimal matching (agnes with optimal matching distance matrix from seqdist) for sequence clustering, state distribution plots (TraMineR package) and Sankey plots for visualisation;

- c) **NOGAuto** : Automation of the coding of the classifications of economic activities in the Business and Enterprise Register;

Complementary analysis methods used: for the text mining preparation phase (natural language processing, NLP); for the modelling phase, artificial neural networks and random forests;

- d) **Plausi++**: Automatic plausibility checks to verify the quality and reliability of administrative data and data collected through surveys;

Complementary analysis methods used: predictive part, gradient boosting machines and random forests. For the explanatory section, Ceteris Paribus and local interpretable model-agnostic explanations (LIME);

- e) **SAE\_JOBSTAT**: Evaluation of the potential for improvement of precision in the Job Statistics (JOBSTAT)

Complementary analysis methods used: Small area estimation (lmer) and robust model (rlm with Huber-M) by using extrapolation weights.

## 2 Evaluation

### 2.1 Evaluation criteria

The evaluation of a possible production transfer is based on individual evaluations of pilot projects. For this purpose five criteria have been selected: **documentation**, **aims**, **value added**, **quality of validation** and **compliance with the FSO's quality criteria**.

The evaluation of criteria concerning the documentation, validation quality and compliance with the FSO's quality criteria plays a key role in the global evaluation and, as a result, these criteria must be fully satisfied.

### 2.2 Status of pilot projects

The evaluation serves as a recommendation basis for the pilot project status. This falls under one of the following options:

- **Ongoing**: continuation of work according to the evaluation recommendations;
- **Finished**: completion of the project without transfer into production;
- **Stopped**: project currently interrupted (due to priorities, data, resources, competencies, infrastructure, etc.) with the option of being resumed if conditions allow for this;
- **Transferred into production**: project for a transfer into production started.

## 2.3 Evaluations and recommendations of pilot projects

- a) ADELE: The documentation summarises the process and methods used but could be further developed. There are in particular certain IT aspects which should be included in this documentation. The aims of this project have been achieved and integration into the revision of the land use statistics is anticipated. The value added for production will be evaluated in this revision project. It should be noted that validation quality criteria and harmonisation with the FSO's criteria will also be addressed in this revision project.

Recommended status

**Transferred into production:** project for transfer into production started.

- b) ML\_SoSi: The documentation does describe several aspects but could be optimised with the detailed description of certain applied methods. The project is at an advanced stage but still requires some work to reach the goals set. The project has already made it possible to add a certain added value with the initial results and their visualisation. It will be possible to evaluate the quality of validation and harmonisation with the FSO's criteria once the project is finalised.

Recommended status

**Ongoing:** Continue work according to recommendations.

- c) NOGAuto: The documentation makes it possible to have a clear overview of the issue and the planned processes but should be even more detailed. Work must be continued by generalising and trying to optimise the process. The initial results are promising in terms of the potential added value but still need to be validated and compared with the FSO's quality criteria.

Recommended status

**Ongoing:** Continue work according to recommendations.

- d) Plausi++: The detailed documentation describes the process in a clear way and provides all elements required for its reproduction. The aims have been achieved in full and the added value for production has been able to be demonstrated based on a detailed validation which notably included several methods. The follow-up to this work must focus on the process in order to guarantee quality during production in order to meet the FSO's criteria.

Recommended status

**Ongoing:** Finalise work according to recommendations.

- e) SAE\_BESTA: The documentation is quite detailed and the open points mentioned at the end of this documentation will be tackled. Work is at an advanced stage but does not yet cover all areas of analysis. It will be possible to evaluate the overall added value, the quality of the validation and harmonisation with the FSO's quality criteria once the project is finalised.

Recommended status

**Ongoing:** Continue work according to recommendations.