The French SSP Lab: bringing it to life

DGINS : 10th to 11th of October 2018

Insee

Mesurer pour comprendre

What are the challenges for Official Statistics?

-Context of **"data revolution"/digitalization** in all its forms

New data and new technologies offer **new potential opportunities** for Official Statistics

but NSIs also face an evolving and more **competitive environment**, with new producers of data and statistics

The **challenge is** to **adapt** skills, production methods, approaches, in order to remain **meaningful and relevant (trust)**

- **Eurostat priority in the 2020 Vision**

- **Innovative entities are setting up in within NSIs**

and the specific context in France?

- **Positive regulatory context: the French Digital Law (October 2016)** allowing access to private data for the production of statistics **[link]**

- **Some experiments undertaken:** retail scan data for CPI at Insee (production in 2020); webscraping of job offers, satellite data for agriculture, health data in ONAs

- **Creation of a unit in 2012 aiming at identifying innovative statistical methods useful for the production of statistics** (with first one, then two datascientists experimenting with Machine Learning, new data, new IT tools and infrastructure)

- **Strategic plan (INSEE 2025) included setting up a new "lab" dedicated to applied R&D: the SSP Lab, for "Official Statistical Service Lab"** (Final deployment roadmap for the lab presented to the INSEE executive committee in October 2017)

and the specific context in France? (2)

- **Creation of:**

  - **the SSP Lab within the Directorate of Methodology** in May 2018, based on the previous unit created in 2012

  - **the EASI Unit (Entreprise Architecture, Security & Innovation)** within the IT Directorate in September 2018

- **Top young datascientists and IT developers** trained at our dedicated schools (ENSAE and ENSAI) chanelled into the French statistical system

- Choice of a "**ramified**" vision of innovation aiming at developing synergies with business units

How does the SSP Lab function in practice ?

- A new unit (8 member team): datascientists, econometricians, IT specialist with complementary skills and experience (*seniors/juniors*)

- Governance: INSEE executive committee + chiefs of Ministerial Statistical Services review the Lab experimental outputs, set overall directions without impeding creativity and reactivity. (Investment decisions made *ad hoc*)

- A **datascience resource and networking center**
  that promotes innovation in production of official statistics within the French statistical system (SSP)
  by conducting applied research and experimental developments in collaboration with units in charge of production (experimental projects) and with EASI IT unit

- **Fields of action**: new sources/types of data, new datascience methods, new tools, new angles of study

- Participating in **networks** around innovative topics, including with European peers (Eurostat, ESSnet Big Data), academic and business partnerships (e.g. Orange Sense Lab, IPP-Paris School of Economics)

- **Disseminating** knowledge and techniques at different levels: training, technical documents, workshops, seminars, hackathons, Big Data newsletters

Activities: experimentation

What is an experimental project?

- one that is proposed and sponsored by a business unit (INSEE or Ministerial Statistical Service)
- must last around 6 months (or longer if broken-up into different stages and deliverables)
- an exploratory approach, on a "small" scale
- a sort of lab testing with different kinds of deliverables
- outcomes determining any eventual scaling-up

How do we carry out an experimental project?

- by putting together a team of Lab and Business Unit (sponsor) specialists with the required skills (datascience, IT, etc.)
- by ensuring contractual commitment to the project (several flexible models)
- by promoting ‘agile’ work practices, with regular outputs
- by disseminating results and sharing experience and feedback

Activities: main ongoing projects

New data

-Student digital evaluation log files (new information to assess student response strategies), together with the ONA from the Ministry of Education

-Mobile phone data (residential population, population present at a given place in time, social segregation), associated with Orange Labs, ESSnet Big data

-Satellite data and city heat islands/SDG indicators (open public space), together with the ONA from the Ministry of sustainable development, ESSnet Big data

New methods

-Automatic detection of employer in Census/classification issues (statistical analysis for textual data, machine learning, webscraping). Improving a inefficient automatic procedure, together with INSEE Social Studies Directorate

-Detecting wages/paid hours anomalies within the mass of employer payroll declarations (Machine Learning), together with INSEE Social Studies Directorate

-Career and wage predictions by Machine Learning used as a basis for pension microsimulation, together with IPP-Paris School of Economics

Activities: dissemination and training

Technological watch and training

-Machine learning

-Statistical analysis for textual data

-Python for datascience

Dissemination and collaborative work practices

-IT innovation platform with container orchestration

-Hackathons, collaborative workshops (on webscraping)

-Reading groups: Machine learning and econometrics

-A dedicated Intranet page to experimental projects/ (eventually experimental results on the Internet?)

-Yammer, blog, github & gitlab

-Big Data newsletter, Big Data seminars

- The SSP Lab still in its early months
- Positively welcomed throughout the Official Statistics service
- Promising in terms of acquiring and sharing datascience knowledge and experience
- Mobilising people within our organisation
- Helping to inspire the next generation of our statisticians

Keep in touch

insee.fr

Insee

Mesurer pour comprendre