The SSP Lab: The French NSI (very young) experience

Dr. Elise Coudin
INSEE, FRANCE

Abstract : INSEE has just created a new lab unit to promote innovation in terms of new data sources, new statistical methods and new ways of working inside the French official statistics system, *i.e.* the NSI and the statistical services of governmental departments.

Inserted in the Directorate of Methodology, Statistical Coordination and International Relations, the SSP Lab is a data science resource and networking center for applied research and experimental development using new sources of data, new technologies, and new methods in relation to the production of official statistics. Sponsored by departments in charge of these productions, the SSP Lab conducts experimental projects in partnership with these departments. It is in charge of monitoring and disseminating innovative statistical methods . It also leads networks within the French official statistics system on innovative topics in close relationship with the IT team, and works in collaboration with different external partners (research centers, peers, ...).

The project was born from the strategic reflections on the medium-term of INSEE (INSEE 2025), during which one working group recommended to set up a unit dedicated to research and development, and another one proposed to create a research watch and animation unit.

This intervention would aim at presenting this new unit composed of eight permanent data scientists, its mode of operation, and at outlining some of its current experiments, among which a project on the automatic identification of the employers of Census interviewed persons in firm registers. This experimental project gave the opportunity to organise an open two-day hackathon inside the French official statistics system and its close partners, during which around sixty persons in a dozen of teams faced a fuzzy identification problem, proposed solutions using search, indexation, text mining or webscraping techniques. Another experiment deals with anomalies detection and data edition in social administrative data, another one faces log data handling for analysing student exercice resolution strategies. Each time, the team is mixed between members of the SSP Lab, and members of departments in charge of statistical production.