The Innovation Lab

**Keywords:** Innovation, BIG-DATA, Machine Learning, Integration, Research.

# Introduction

The evolution of the demand for statistical information, puts the National Statistical Institute (NSI) facing an important challenge: to improve its ability to innovate in processes and products in order to respond more effectively to new and growing needs.

The Laboratory for Innovation is one of the infrastructures adopted to respond to this new challenge with the aim of facilitating the development of innovation and reinforcing the role of research as a founding value and a tool for strategic growth of the Institute and staff. The Innovation lab is an “environment” to enforce the research role and to develop innovative ideas. It is a place to test new solutions, new processes and new products.

# Design

The laboratory was designed to provide useful infrastructure to colleagues to spend time on research, to test innovative ideas in a dedicated space. It enables the institute to invest in innovative projects to improve statistical information, processes and products and strengthen internal and external relations through partnerships with universities, institutions and public and private research institutions.

It is an open space suitable to encourage the development of innovative projects and to stimulate the collaborative experience between different competences such as data science, statisticians, somputer science and so on. All with the aim of making a contribution to the institute in terms of innovation and stimulate the publication of new experimental statistics.

## Innovation Program

An innovation program has been drawn up which defines the priority areas considered to be innovative and of greater interest for the activities finalized for statistical production. The elements identified are innovative as they offer new methodological, technological and process opportunities in relation to the specific phases (acquisition, processing and dissemination of data) of the statistical processes. The four elements that drive the program are the following

* Big Data and new data collection methods: Internet of things, mobile devices, sensors etc.).
* Improvement of statistical processes: reuse of sources and adoption of standards and guidelines provided by international initiatives (for example, UNECE GSBPM - Generic statistics business process model) and European (for example, European Statistical System (ESS) Enterprise Architecture Reference Framework - EARF) .
* Innovative outputs: new navigation techniques, information discovery and visualization, integration of different data sources, open data, linked open data and administrative archives.
* Use of new ICT technologies: to improve performance, security, architectures and new ICT services.

## Project lifecycle

The innovation program is implemented adopting a phased approach that starts from the birth of the idea to its realization.

The realization of the laboratory took place following the steps below:

1. **Project Call**

All interested colleagues can present a project during the call of the innovation laboratory, of the three-month period, by filling out a pre-compiled form and following a computerized submission procedure present on the intranet of our NSI

1. **Project Selection**

The selection of ideas follows a well-defined process:

* first phase: the first evaluation sees a joint work - coordinated by the Innovation Laboratory staff - of methodologists, IT experts and thematic experts in order to evaluate feasibility, sustainability and technological and IT needs.
* Second phase: the ideas that pass the first evaluation phase are submitted to the final evaluation of the Research Committee to verify consistency and the compliance of the idea respect to the innovation program and takes into consideration a series of aspects aimed at evaluating the feasibility, complexity, risks and safety associated with its implementation.

1. **definition of needs and skills**
2. In this phase are defined the IT needs, the skills necessary to the development of the projects and the resources project teams . The team can include resources from the organizational and / or methodological and IT sectors to ensure and improve the integration and exchange of resources and skills in order to activate a virtuous circle of knowledge in the Institute.**Project Implementation**

In this phase the team's resources work in the innovation lab for a maximum period of six months with the aim of developing the proposed idea.

1. **Dissemination Result**

The results obtained by the teams will then be disseminated on special days dedicated to training and communication to share results and to facilitate the production of outputs.

1. **Future developments**

The first year the laboratory will be accessible on an experimental basis only to institute employees, subsequently they will be able to propose their innovative ideas also students, researchers and subjects coming from external structures such as universities, other research institutions and institutions, private organizations with which our institute will activated collaborations and/or agreements. In the long term, training and learning paths, specific internships for ex. for new graduates (lab training). The external collaborations will promote and facilitate the profitable interaction process already started through the establishment of a network of knowledge, research and innovation that strengthens the virtuous circle in which the Institute is already inserted thanks to partnership agreements, agreements with other bodies and organizations of national and international research.

# IT Architecture

The general computer architecture of the laboratory was designed in collaboration with IT management experts, starting from the assumption of creating a platform with high computing power, flexible and adaptable to the different specificities of each individual project. In the planning, we also took into account the technological need driven by the innovation program, that is to have a platform suitable to operate in Big Data and Machine Learning projects. For these reasons an IT platform has been set up based on Hyper-Convergent technology from the manufacturer NUTANIX, allows you to isolate the environment to make the most of the HW / SW characteristics by multiplying the values ​​of IOPS. Furthermore, the use of solid state storage technologies (high quality disks / SSD-NVME) and integration with dedicated GPUs (nVidia TESLA P40) guarantee high performance with deep learning and parallel computing workloads.



Figure 1: Schematic representation of the IT platform of the Innovation laboratory

# Results

Thanks to the work done in collaboration with various departments of our institute, the Innovation Lab was inaugurated in March 2018 and activities have been started on the projects selected in the first call of the innovation laboratory. The first call, completed in September 2017, was a success with 27 projects submitted, proposed by 33 researchers.

Of these, 6 ideas has been selected were more focused on the production of an experimental statistical output in accordance with the areas of the innovation program such as Big-Data, machine learning and Data Integration.

The projects are as follows:

* Use of Deep Learning algorithms on spatial images for the recognition and automatic classification of objects on maps for statistical production such as parks and green areas.
* Web scraping on active companies for Big-Data production to be integrated with the Statistical Register of companies through record linkage technologies.
* Integration between cartographic data and the administrative data of the cadastre for the recognition of the areas of agricultural holdings for the production of useful maps at the end of the census of the agriculture
* Estimation methods for the production of statistical indicators by integrating data from sample surveys, administrative data and census data

* Web Scraping on the main job search engines for the collection of information related to job vacancies for the production and distribution of time series of data on the labour market
* Use of Metropolis-Hastings parallelization algorithms on GPU machines for estimation of confidence intervals on census data of the population size

Some of these projects are over and others will end by the end of 2018.

The finished ones produced excellent results, presented at the national statistics conference.

The second call was concluded in September 2018 with the submission of over 30 proposals and the request for collaboration with universities and start-ups.

A specific institutional research protocol has been defined to activate the requests for collaboration that have arrived.

These results are confirming the role defined for the innovation laboratory, promoting it in the institute as a tool for research and innovation, as a channel to promote collaborations and partnerships and as infrastructure to encourage colleagues and researchers to develop innovation.

# References

[1] <https://www.researchgate.net/profile/Owen_Abbott/publication/304252081_ONS_Innovation_Laboratories/links/576ab09708aef2a864d1f72a/ONS-Innovation-Laboratories.pdf?origin=publication_detail>

[2] <https://www.oecd.org/naec/NAEC_Innovation_Lab_blueprint.pdf>