BDVA labelled i-Spaces

General Information

i-Space Name	TeraLab
Label Category	Silver

Teralab is a Big Data platform developed within the framework of the PIA. It has the role of hosting efforts to enhance the value of industrial data in partnership with laboratories or collaborative projects. It has been operational for more than 3 years and offers state-of-the-art infrastructure and tools. TeraLab's infrastructure is secure, sovereign and neutral. It provides the necessary security guarantees for industrial partners so that they can make available, within a defined framework, their high-value data for research or innovation projects. Beyond the unique technical characteristics of the infrastructure TeraLab facilitates the link between data providers and data scientists thanks to the network formed over the three years of existence. The support of the platform does not stop with the provision of workspaces but according to the requests a help to the formalization of use cases, the choice of tools and design of the architecture, skills On regulations such as the GDPR or the rules specific to health data and more generally on the governance of hosted data. Currently 50 research or innovation projects benefit from the platform. TeraLab is recognized by its technical excellence and its impact in the eco system, both at national level and at European level in the PPP Big Data, in particular by its silver i-Space labelling in December 2016 by the BDVA. It is the only French platform among the 4 European laureates who have achieved this highest level of labelling. In addition, an H2020 project on the PPP call Factory of the Future positions TeraLab as the French competence center for a Digital Innovation Hub (DIH) on sovereignty and cybersecurity. The cost of using the platform depends on the size of the workspace provided and aims to balance operating costs. The platform is non-profit; Break even on operating cost is reached in 2017.

Address	Institut Mines telecom 46 rue Barrault 75013 Paris France
Contact Mail	anne-sophie.talandier@mines-telecom.fr
Partner organizations	Institut Mines Telecom is coodinator with support from INSEE an GENES
Web Site	https://www.teralab-datascience.fr/fr/

BDVA labelled i-Spaces

Platform and Services Information

Platform(s)

Resource	Value
Max number of cores usable in parallel for one a single project	520
GPU Accelerators	M40 from Nvidia
online storage	400TB
internal network (Gbit/sec)	> 1 Gyte/sec
external network (Gbit/sec)	> 1 Gyte/sec

Provided Services

TeraLab is a founding and contributing member of this community and has recently been awarded the highest label recognizing excellence of TeraLab assets and impact
The i-Spaces community is growing into a pan European loosely coupled federation with ability to support data innovation across sectors and boundaries.

- ability to perform experiments on more data-sets from different sectors across EU;
- access for industry from any member state to SotA experiment platforms & tools;
- market-realistic use-conditions to test & validate new tool concepts from Academia;
- enabling cross-regional access to SotA Academic knowhow;
- sharing best governance and incubation support methods between existing i-Spaces;
- wider access to industry data and challenges for education and training of students.

BDVA labelled i-Spaces

Selected Projects and/or Success Stories

DATAIKU and a Large Health Insurance Company

Early stage Data experiment prototype scenario: A Large French Mutual Health Insurance (Name is confidential) company is considering an important strategic move toward novel Big Data techniques to improve knowledge of their subscriber behaviour. The business lines had identified several use cases, involving heavy machine learning algorithmic. They requested support to IT division, which evaluated the necessary investment. At this stage the Business Lines were unable to provide ROI evaluation without concrete experimentation to allow authorization of such investment. TeraLab proposed to run a "Data Experiment", lowering entry barrier for the proposed experiment. After discussion the Legal department authorized transfer of a data corpus to the TeraLab Trusted, Secure Data Incubator, and data access agreements "for research" were signed. The insurance company contributed to the funding of the experiment, complementing funding from national innovation initiatives. A call for Data Innovator was performed and one of the leading startups was chosen to work on the data experiment: Dataiku. The experiment was successful and the partners were able to assess a ROI model. The Insurance company prepared a call for tender for the industrialization of the solution. Major Commercial solutions made a proposal. The Startup which had been performing the Experiment was chosen to run the first phase of Industrialization, main reason of this choice, beyond cost, was that the proposed solution, because open and interactive, allowed to accelerate take up.

CAP ITEA3 European project TeraLab a exploratory sandbox for Thales

Thales TCS: Social media is a widespread and now common open ecosystem that allows citizens to share opinions and information very quickly. As such, from a crisis management perspective, they offer new opportunities for two-way communication between PPDR (public protection and disaster relief) officials and citizens. These features, among other things, social media are a potential tool for improving disaster response and crisis response efforts. However, given the very large volumes of data that should not be analysed in near real-time, automated tools capable of triggering instant alerts based on the discovery of hidden and rare, but relevant information is required to ensure capabilities Quasi-real time processing while integrating advanced and complex algorithms such as those included in the GeoIntelligence application. We decided to implement the CAP architecture architecture "Lambda" which allows to combine the two algorithms Batch processing and quasi-real-time data flow. Because social media activity varies considerably depending on the evolution of the crisis and the magnitude of its consequences, it is important to offer an elastic deployment of such a Big Data application. That is, a deployment that can adjust the processing capacity as a function of the load. This is particularly important in restricted environments where hardware resources may be limited, such as those encountered by TCS clients. To realize the elasticity mechanism, we exploited the mechanisms of dynamic allocation of resources offered by the RAR (the Hadoop resource manager). We have integrated a control loop into our GeoIntelligence application, monitoring the activity of social networks, and adding or releasing processing nodes according to the load. In addition, we have docked the SpeedLayer in order to have finer granularity and more flexibility in the process of resource allocation, a key element in constrained environments. To demonstrate this process of elasticity, we have developed a tweet simulator which allows to inject into the application of the tweets already collected while choosing the speed of injection. Thanks to this simulator, we are able to reproduce and accelerate the different stages of a crisis. This allowed us to create artificial peaks of activity to test the elasticity of the platform GeoIntelligence to effectively monitor such an ecosystem.

BDVA labelled i-Spaces

CAP ITEA3 European project TeraLab LAMANE spinoff on anonymization

• Business development / financing: creation of a spin-off (LAMANE society), which is incubated in a school of the IMT (IMT Atlantic)
• Ecosystem building: contact with experts research laboratories of IMTon the topic of data anonymization
• Technology services: safeguarding of the data of a corpus of data of the confidential 15To (flashing of mail in France for a year from the sender to the recipient)

Analysis of the use of the Gallica digital library (French Component of Europeana)

Since 2013 Collaboration BnF + Telecom ParisTech Department of Economics and Social Sciences:

- 2014 observation of Gallica users (V. Beaudoin, J. Denis);
- 2015: Study of the use of funds digitized by amateurs of the Great War. (V. Beaudoin, Z. Pehlivan).
- 2017: Log-mining study on the behavior of users of the BnF: 40 m visitors / days, 20 M lines of logs / day O "Regex" to extract for each line of logs: O IP anonymity date request (HTTP): HTML / design / ARK (unique identifier for document).

Contribution of Teralab: Implementation of an ElasticSearch database; Easy recovery for treatments; Benefits for the BnF: Data security, Data stored in France.

Benefits for Telecom ParisTech: Flexibility, adaptability (changing needs) Analyzes produced; Simple statistics (popular documents, average times, entry point ...); Analysis and optimization of the impact of mediation (blog, facebook) on the Gallica audience; Characterization of uses: how Gallica users navigate the site.

MIDIH, project H2020 on the industry of the future, the Commission with the EIT Digital KIC , funded starts end 2017

• Business development / financing: Acceleration of the start-ups in connection with EIT Digital, FIWARE and Industrial Data Space initiatives.
• Ecosystem building: networking by Teralab of data industry providers with experts big data, IoT and cyber security in Europe.
• Technology Services: distinctive expertise on securing data and cybersecurity for industrial tools.