Social Mining & Big Data Analytics

SoBigData

RESEARCH INFRASTRUCTURE++

SoBigData SoBigData

www.sobigdata.eu

Pan-European, Multidisciplinary Research Infrastructure

SoBigData++ will deliver a distributed, Pan-European, multi-disciplinary research infrastructure for big social data analytics, coupled with the consolidation of a cross-disciplinary European research community, aimed at using social mining and big data to understand the complexity of our contemporary, globally-interconnected society. SoBigData++ is set to advance on such ambitious tasks thanks to SoBigData, the predecessor project that started this construction in 2015. It resulted in creation of a distributed platform of interoperable social data mining tools, methodologies and services for obtaining, analyzing, and visualizing massive datasets, together with associated data scientists' skills for the ethically safe deployment of big data analytics.

Vision

The main objective of SoBigData++ is to mature the infrastructure of SoBigData to become a research infrastructure recognized by ESFRI RoadMap 2021 and sustained by a SoBigData Association which will design an appropriate form of legal entity. In the process, SoBigData will become an easy-to-access platform for execution of complex social mining processes, compliant with the European Open Science Cloud (EOSC). It will also become a wide and diverse community engaged in challenging research questions, advocating a critical data literacy aimed at facilitating data citizenship and data democracy, and capable of empowering the next generation of responsible data scientists. It will develop concrete tools to operationalize ethics with value-sensitive design, incorporating values and norms for privacy protection, fairness, transparency and pluralism. It will facilitate collaborations with industries to develop pilot projects and proofs-of-concept and in this way accelerate data-driven innovation.

Academic impact

2 book chapters
1 monograph

60 journal papers
90 conference papers

100 researchers (including PhD) in the project

91 unique big datasets
98 social mining methods
6 applications

Papers in high impact journals

PNAS,
Phys. Rev. X,
ICML, ICLR,
WWW, AAAI

EXCELLENCE

Events

REAL TIME EPIDEMIC ATATHON
Epidemics + Hackathon + Data as open-source real time forecasting challenge
(6/3/20 – 31/5/20)

FORC
Symposium on Foundations of Responsible Computing FORC
(1/5/20 – 3/5/20)

Data Science Colloquium
Winter 2020
(20/5/20 – 8/6/20)

Data Science in Techno-Socio-Economic Systems Online Workshop 2020
(10/6/20 – 11/6/20)

Leading scientists

Five SoBigData scientists are PIs of ERC grants in the field of artificial intelligence.

erc

Fosca Giannotti *CNR*
Science and technology for the explanation of AI decision making.

Aristides Gionis *KTH*
Designing algorithms to reduce filter bubbles in social media.

Marlon Dumas *UT*
The Process Improvement Explorer: Automated Discovery and Assessment of Business Process Improvement Opportunities.

Dirk Helbing *ETHZ*
Urban Innovations and Smart Cities.

Stefano Leonardi *UNIROMA*
Algorithmic and Mechanism Design Research in Online Markets.

SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics - GA. 871042

FUNDING

Initial design — 2015

2015 – SoBigData, the predecessor project, received the first grant n. 654024 of 6M Euro under the H2020 programme INFRAIA 2014-2015.

Current design — 2020

2020 – SoBigData++ grant n. 871042 of 10M Euro under the H2020-EU.1.4.1.2 programme.

Activities

- Joint Research Activity 41.4%;
- Network Activity 27%;
- Transnational Access 14.17%;
- Virtual Access 4.96%;
- Management 4.87%;
- Partners Travels and Consumables 7.58%

of the total budget.

ESFRI RoadMap 2021 — 2020

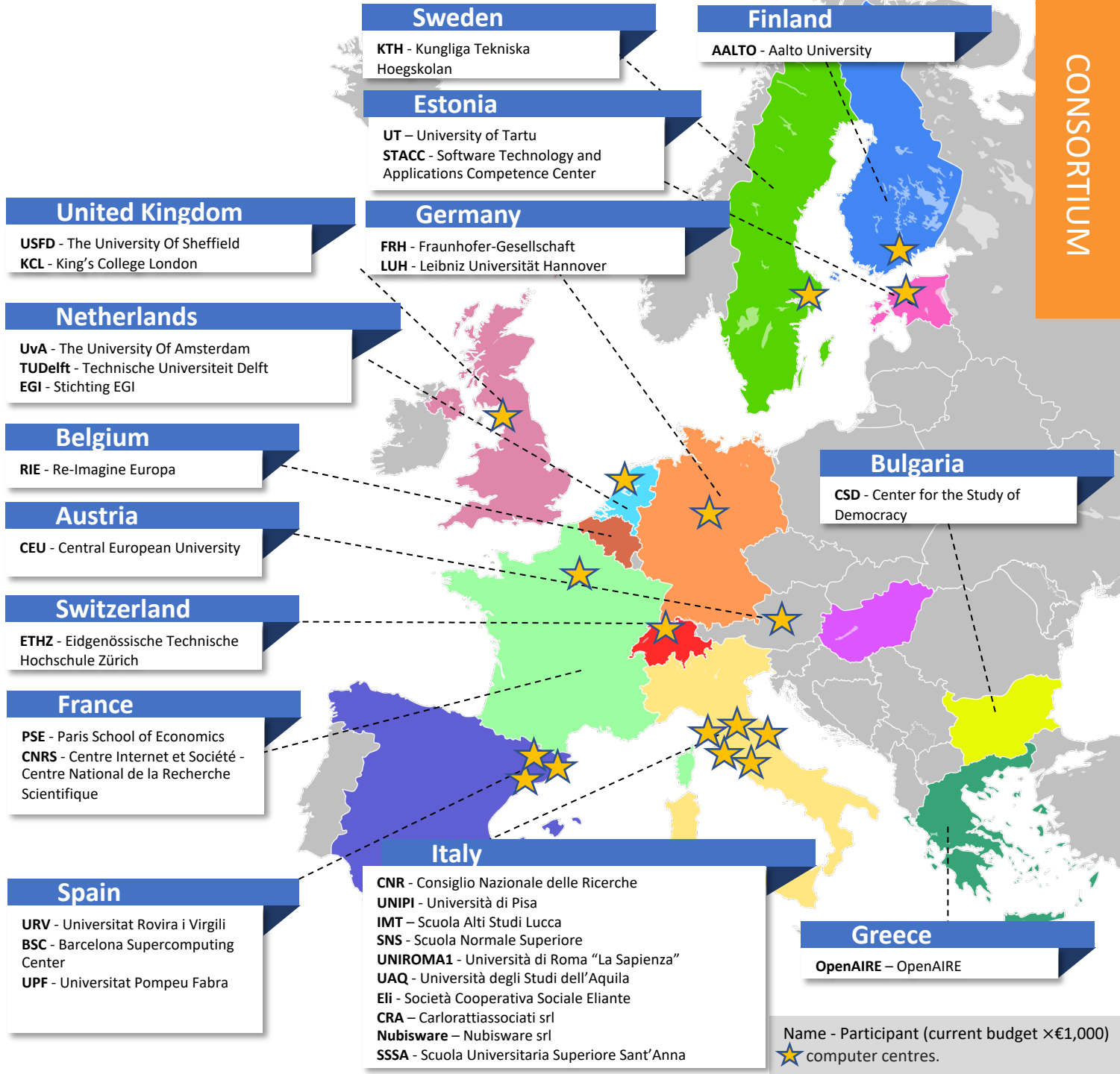
- All institutions of the consortium are supportive of the initiative.
- Italy, Germany, Spain, Finland, Estonia, Netherlands are supporting the initiative.

Preparation 2020-2025

Implementation 2025-2029

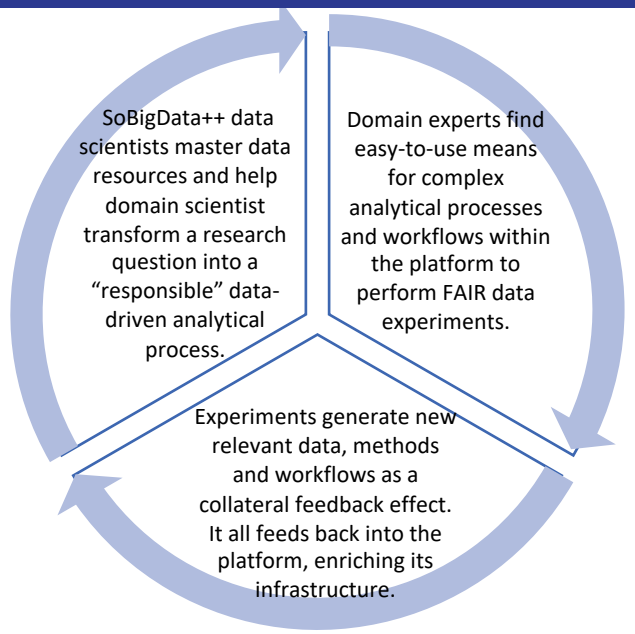
Operation 2030-2050 — 2050

CONSORTIUM



INFRASTRUCTURE

A circular structure of human and digital resources



SoBigData ECOSYSTEM ++

Three lines of activity

1. Networking

Community building activities to promote cooperation of developers and platform users.

- **Dissemination** – active presence on social media and key science events, a newsletter, production of communication material, production of high-level scientific publications.
- **Training**
  - Summer schools.
  - Datathons, including critical data literacy datathon.
  - E-learning area, composed of open-source modules developed during training events: e.g. Machine Learning for Social Science.
  - SoBigData Master in Big Data Analytics and Social Mining course.

2. Access

Virtual Access (VA)
Online front-end comprising the Resource Catalogue, Virtual Research Environments and the E-learning area. VA currently has **6000+ users**.

Transnational Access (TA)
On-site accesses for targeting research questions bound to multidisciplinary themes specified in the call, or targeting researchers wishing to explore the infrastructure for their own research projects. TA currently has **50+ users**.

3. Joint Research

Joint Research happens within the social mining platform, which has **98 social mining methods**, clustered in **6 thematic clusters**: *Text and Social Media Mining (TSMM); Social Network Analysis (SNA); Human Mobility Analytics (HMA); Web Analytics (WA); Visual Analytics (VA); Social Data (SD)*. The platform also has **91 big datasets** and **6 applications**. They are used in **6 exploratories** which aim is to emphasize and promote research of current socio-economic challenges.

Exploratories

Societal Debates and Online Misinformation. Investigation such as on the impact of Brexit on the scientific relationships between EU and UK.

Sustainable Cities for Citizens. Studies of smart cities. E.g. text/opinion mining of online news to define reliable indicators of happiness and peace. Initial results indicate that the level of crime of a territory can be approximated by analyzing the news related to that territory.

Demography, Economics & Finance 2.0, for example studies of models will forecast the effects of policy choices on poverty and well-being and will reveal insights about their composite indicators.

Migration Studies, including studies of macroscopic human flows, social behavior analysis.

Sport Data Science: studies of indicators of training and performance predictors.

Social Impact of Artificial Intelligence and Explainable Machine Learning: will include research on the impact of AI on future societies, e.g. withing labor workforce.

Ethical framework

- SoBigData++ adheres to the EU vision on Responsible Research and Innovation, and operationalises values that are driving the ongoing reform of the EU Data Protection and Fundamental Rights legislation. It will maintain ethical and legal boards which will accommodate within-infrastructure ethical issues as well as will aim answering new questions on the scope, interpretation, and application of the GDPR along with the expanding role of AI, machine learning and data mining.

Findability
Accessibility
Interoperability
Reproducibility

- SoBigData++ infrastructure is compliant with **FAIR** data principles.

- SoBigData++ aims to become a **peace room** (Helbing, D., Seele, P. *Nature* **549,** 458 (2017)) environment that intends to contribute to the agenda of advancing pro-social uses such as advancing peace, sustainability and resilience using big data, artificial and human intelligence. Peace room is characterized by features such as:

- higher degree of transparency,
- democratic framework of operation,
- interdisciplinary teams,
- supervision by ethical experts.