SOBIGDATA.it
ITALIAN RESEARCH INFRASTRUCTURE

SoBigData.it @ UnivAQ

Infrastructure Services and Research for Open Data Science

Speakers: Valerio Grossi (CNR-ISTI), Michele Tucci, Giordano d'Aloisio, and Daniele Di Pompeo (UnivAQ)

Coordinator: Roberto Trasarti
L'Aquila node coordinator: Antinisca Di Marco (UnivAQ)
Management team: Valerio Grossi and Michela Natilli
Communication Manager: Daniele Fadda
Institute of Information Science and Technologies (ISTI),
National Research Council (CNR), Pisa, Italy

http://www.sobigdata.eu/
https://sobigdata-univaq.github.io/

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
SOBIGDATA.it ITALIAN RESEARCH INFRASTRUCTURE
Consiglio Nazionale delle Ricerche

SOBIGDATA.it
ITALIAN RESEARCH INFRASTRUCTURE

SoBigData.it @ UnivAQ

Morning sessions

11:00 - Welcome Coffee Break

11:30 - Opening Remarks

11:45 - Presentation of the SoBigData Research Infrastructure

13:00 - Lunch Break

Afternoon sessions

14:00 - Contributions in VL-XAI to Address and Democratize Software Fairness

14:30 - Urban Digital Twin for Territorial Management - VL Disaster

15:00 - Entity Extraction in Clinical Summary of Mammary Malignancy Data using Transformer-Based Models - VL Health

15:30 - How to build a research infrastructure. My experience in the legal unit of SoBigData and my research on data sharing

16:00 - Coffee Break

SOBIGDATA
RESEARCH INFRASTRUCTURE

SoBigData is a distributed **digital** Research Infrastructure with the mission of use social mining and big data to **understand** the **complexity** of our contemporary, globally interconnected **society** and offer **services** to researchers, industry, public bodies, and citizens through the creation of a **multidisciplinary** scientific community according to the **European vision** of on **ethics**, **legality**, and **open science**.

In the ESFRI ROADMAP 2021, currently it relies on competitive projects since 2015

SoBigData is **coordinated by CNR-ISTI** and distributed over 14 Countries

Italy

Sweden

Poland

Netherlands

UK

Greece

France

Finland

Spain

Estonia

Germany

Bulgaria

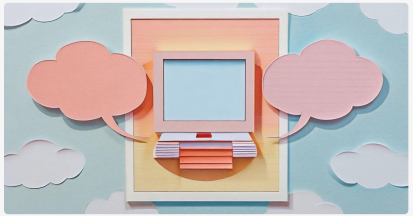
Belgium

Austria

SoBigData is now at the end of **preparation phase** of ESFRI and are going to become an ERIC

Research **Spaces**

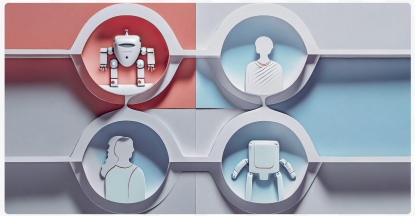
VERTICAL CONTEXTS FOSTERING TANGIBLE PROGRESS TOWARD GRAND SOCIETAL CHALLENGES



Societal debate
and misinformation

Demography, Economy
and Finance 2.0

Sustainable Cities for Citizens



Social Impact of AI and
explainable machine learning

Health Studies

Disaster response
and recovery

Next-Generation Internet
& beyond 5G Networks

Pervasive Intelligence

slices RI + SOBIGDATA

SoBigData RI

SoBigData **Network**

SoBigData has a large network of collaborations:

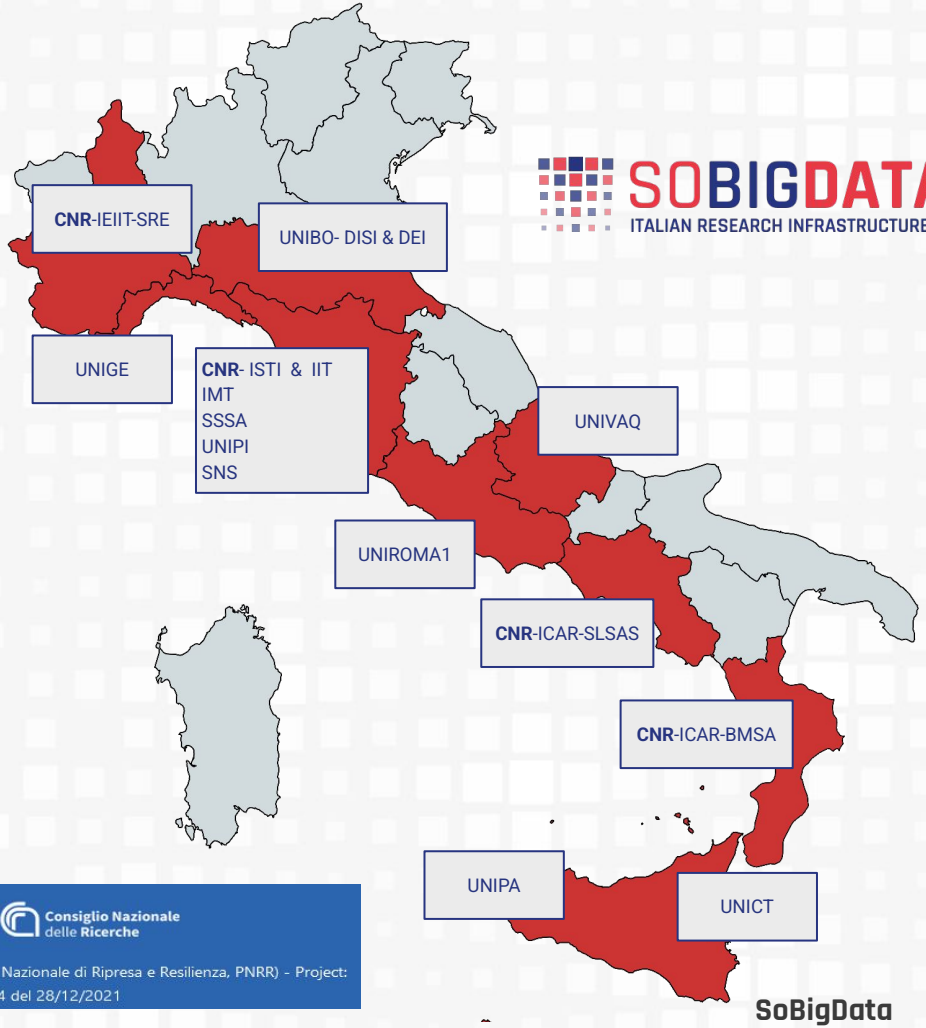
- **Research centers**
- **Computational centers**
- **Experts in social science areas**
- **Industrial partners**

This create a **multidisciplinary** community which is able to generate values and create opportunities for **real data access** and **innovative projects**.

Central Hub
Italian Node

Was also supported by a PNRR project "SoBigData.it" with the objective of increasing the capabilities and the **scientific** and **technological** services **empowering** the **Italian** node.

Connected to CINI and GARR

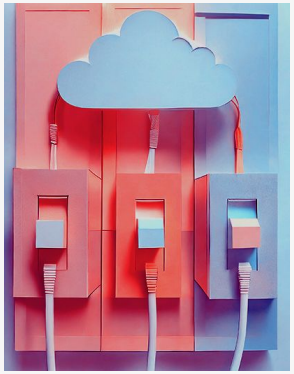


Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani | SOBIGDATA | Consiglio Nazionale delle Ricerche

SoBigData.it receives funding from European Union - NextGenerationEU - National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) - Project: "SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics" - Prot. IR0000013 - Avviso n. 3264 del 28/12/2021

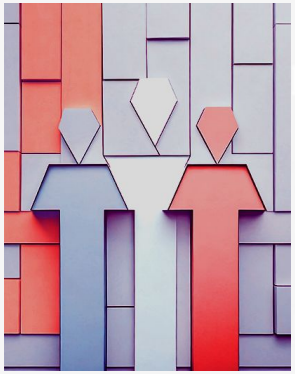
SoBigData RI **Services**

TECHNOLOGY



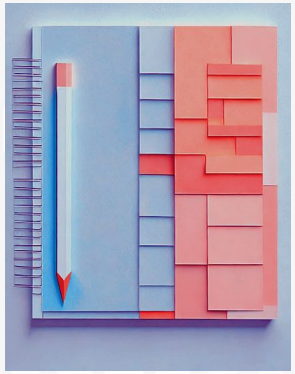
- Cloud Computing Infrastructures
- Access and Storage Management
- Thematic Virtual Research Environments
- SoBigData Lab: Execution Environments

RESOURCES



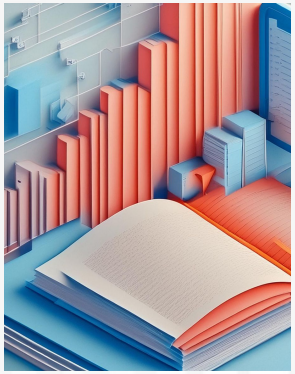
- Catalogue and Resources Management
- Integrated advanced analytics Tools
- Vertical Applications for Specific Domains

ETHICAL AND LEGAL



- Policies for FAIR Data Management
- Ethical and Legal Support for Projects
- Connections with EU authorities

TRAINING & MOBILITY



- Specialized Courses and Workshops
- Summer School
- Master in Big Data
- SoBigData Academy
- Transnational Access program (TNA)

Specialised Services for Policymakers and Businesses

For years, the RI has been collaborating with regional and national entities to process complex data (social, economic, environmental, health), providing indicators, metrics, and insights useful for defining effective, targeted policies.

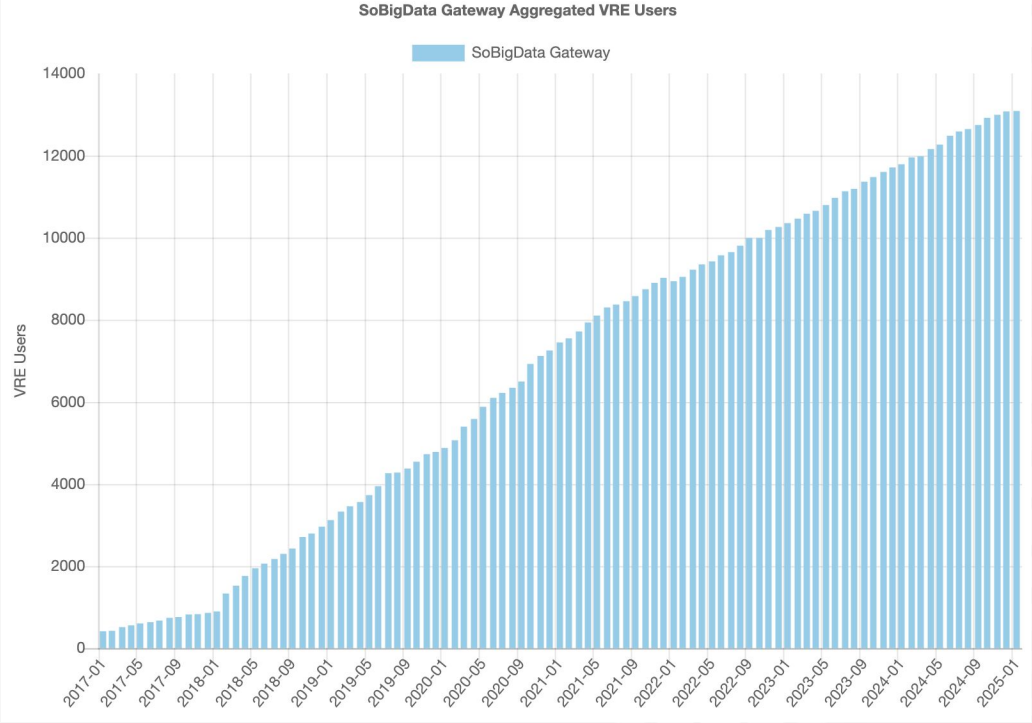
Specific services were implemented for Policymakers and Business:

- **Ex-ante and Ex-post Impact Assessment:** Modeling and simulation of political scenarios, evaluation of expected results and effectiveness of interventions, identification of critical issues or unforeseen effects.
- **Proof-of-Concepts for SMEs (Challenge US):** innovation support for the creation of proof-of-concepts, designed to create collaborations between researchers and industry to stimulate the creation of innovative data-driven solutions.
- **Customized Training Courses for Businesses:** ad-hoc courses for companies on specific topics.

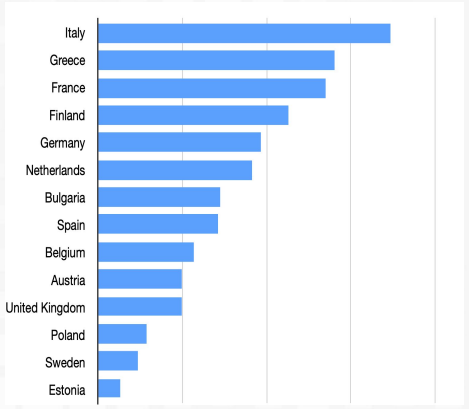
Registered Users

At the beginning of 2020, the users were 4880. In December 2024, the RI has 13076 with an increment of 168% with a constant growth of users (around 100 per month).

Italy is central and it is the principal service provider and consumer.



Services usage by country



SOBIGDATA
RESEARCH INFRASTRUCTURE

SoBigData Pillars

Ethics: Responsible Data Scientist

Commitment to Ethical Practices

- Data Privacy & Security: Rigorous adherence to GDPR and other regulatory frameworks.
- Transparency & Accountability: Clear guidelines on data use, provenance, and processing.

Ethical Framework Integration

- Ethics by Design: Embedding ethical considerations from project inception to implementation.
- Risk Mitigation: Regular ethical reviews and impact assessments to safeguard societal interests.

Building Trust

- Stakeholder Engagement: Continuous dialogue with citizens, industry, and policymakers.
- Responsible Innovation: Ensuring that technological progress aligns with societal values.

Open Science: Democratizing Knowledge

Commitment to Openness & Transparency

- Open Data Initiatives: Promoting free and accessible datasets for research and innovation.
- Open Source Software: Development and sharing of tools that drive collaborative improvements.

Collaboration & Reproducibility

- Interdisciplinary Partnerships: Facilitating cross-border and cross-sector collaborations.
- Reproducible Research: Encouraging methods and results that can be independently verified.

Enhancing Impact

- Community-Driven Projects: Leveraging crowd-sourced ideas and community feedback.
- Broad Dissemination: Ensuring findings reach academic, industrial, and public audiences.

Multidisciplinary and Innovation

Integration Across Disciplines

- Bringing Together Diverse Fields: Combining computer science, social sciences, law, ethics, and more.
- Holistic Problem-Solving: Approaching complex challenges with a blend of technical, social, and ethical perspectives.

Fostering Innovation

- Cross-Pollination of Ideas: Creating an environment where different disciplines inspire new approaches.
- Collaborative Research Models: Encouraging joint projects that leverage strengths from multiple fields.
- Addressing Complex Societal Challenges: From urban planning to public health, using multidisciplinary insights.
- Continuous Innovation: Drive research into next-generation AI techniques tailored to address evolving societal challenges.

Community: A Thriving Ecosystem

Inclusive Engagement

- Diverse Stakeholders: Involving academia, industry, government, and citizens in the research process.
- Collaborative Networks: Building a robust network of experts, practitioners, and innovators.

Knowledge Exchange & Capacity Building

- Workshops & Seminars: Regular events to share best practices, insights, and emerging trends.
- Mentorship & Training: Programs to develop skills in data science, ethics, and technological innovation.

Impactful Collaborations

- Joint Research Initiatives: Facilitating projects that address real-world challenges.
- Policy Engagement: Informing evidence-based policymaking through collaborative research outputs.

SOBIGDATA
RESEARCH INFRASTRUCTURE

SoBigData History

The Idea…

We, as researchers, develop methods and apply them to several context but it is difficult to share them with the community and difficult to use them…

Mobility data mining,

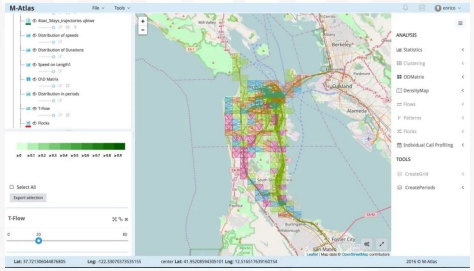
Social network analysis

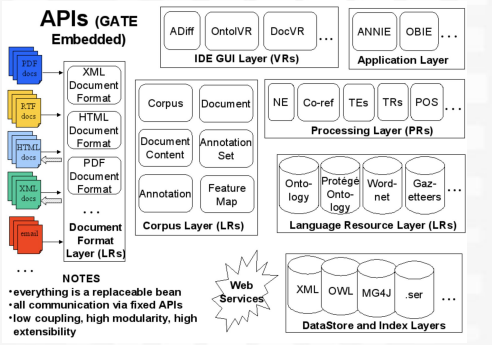
Analytic platform (and their Data & Knowledge infrastructures)

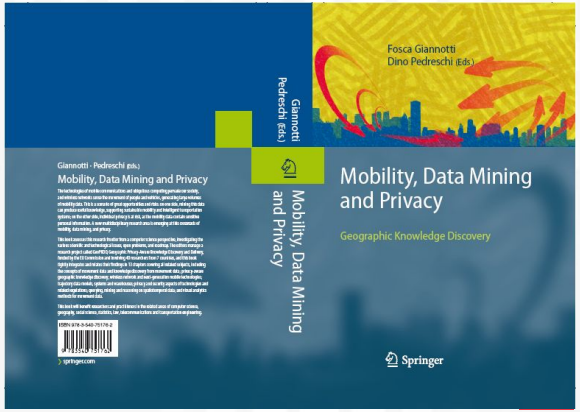
- M-Atlas
- Gate
- Common-GIS
- etc…

Ethics in big data analytics

- Privacy
- Intellectual property







Year 2015-2016

SoBigData (Age 0-1)

INFRAIA-1-2014-2015 - Integrating and opening existing national and regional research infrastructures of European interest

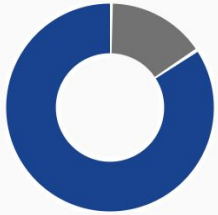
Start date
1 September 2015

End date
31 December 2019

Funded under
EXCELLENT SCIENCE - Research Infrastructures

Total cost
€ 5 917 500,00

EU contribution
€ 5 000 000,00



Coordinated by
CONSIGLIO NAZIONALE DELLE RICERCHE
Italy

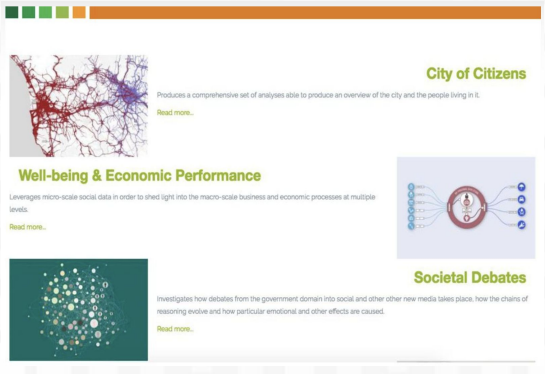
THE UNIVERSITY OF SHEFFIELD United Kingdom	Net EU contribution € 986 125,00	
UNIVERSITA DI PISA Italy	Net EU contribution € 360 000,00	
FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV Germany	Net EU contribution € 616 875,00	
TARTU ULIKOOL Estonia	Net EU contribution € 227 500,00	
SCUOLA IMT (ISTITUZIONI, MERCATI, TECNOLOGIE) ALTI STUDI DI LUCCA Italy	Net EU contribution € 175 000,00	
GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER Germany	Net EU contribution € 407 000,00	
KING'S COLLEGE LONDON United Kingdom	Net EU contribution € 377 500,00	
SCUOLA NORMALE SUPERIORE Italy	Net EU contribution € 175 000,00	
AALTO KORKEAKOULUSAATIO SR Finland	Net EU contribution € 425 000,00	
EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH Switzerland	Net EU contribution € 0,00	
TECHNISCHE UNIVERSITEIT DELFT Netherlands	Net EU contribution € 165 000,00	

Year 2017-2019

SoBigData (Age 2-4)

After 2 years the national initiatives merged into a single platform starting from D4Science digital infrastructure.

- Virtual Access on a unified platform
- First attempt of creating a catalogue of methods and data
- Creation of the Board of Ethics and Legality (BOEL)
- Transnational Access
- Creating “Exploratories” as containers for different topics
- Community grows…



Datasets | Methods & Tools | Workflows | Thematic Clusters | People | Contact

Method	Partner	SoBigData RI - Integration
Urban Profiles	AALTO	Service hosted, Download
Urban Mobility Atlas	CNR	Web Service
Trajectory Builder	CNR	Service hosted, Download
Borders	CNR	
Sociometer	CNR	Download
Trip Builder	CNR	Web Page
Car Pooling	CNR	Service hosted, Download
MyWay	CNR	Service hosted, Download
Privacy Risk	CNR	Service hosted, Download
O/D Matrix	CNR	Web Service
Mobility Profiles	CNR	Service hosted, Download
Exploration of Time	FRH	Download
Statistical Validation	SNS	

Year 2020-2021

SoBigData++ (Age 5-6)

Social Mining & Big Data Analytics
SoBigData
RESEARCH INFRASTRUCTURE ++

INFRAIA-01-2018-2019 - Integrating Activities for Advanced Communities

- Partners from 11 to 31: not only computer scientists but more "specialised" partners: Economics, Social Scientists, Political analysists...
- Better definition of the objectives as RI

Start date
1 January 2020

End date
31 December 2024

Funded under
EXCELLENT SCIENCE - Research Infrastructures

Total cost
€ 9 997 172,50

EU contribution
€ 9 997 172,50

Coordinated by
CONSIGLIO NAZIONALE DELLE RICERCHE

Italy

Year 2022-2024

SoBigData RI (Age 7-9)

ESFRI Roadmap 2021: Becoming one of the first research infrastructure in the DIGITAL RI area.
Objective becoming an legal entity as ERIC in 2027

- Long Term sustainability plan
- Political and Financial support from Countries
- From Exploratories to “Research Spaces”
- Service definitions
- Not only a project

Data di avvio
1 Ottobre 2022

Data di completamento
30 Settembre 2025

Finanziato da
Research infrastructures

Costo totale
€ 3 222 266,25

Contributo UE
€ 3 222 266,00

Coordinato da
CONSIGLIO NAZIONALE DELLE RICERCHE
Italy

SOBIGDATA
RESEARCH INFRASTRUCTURE

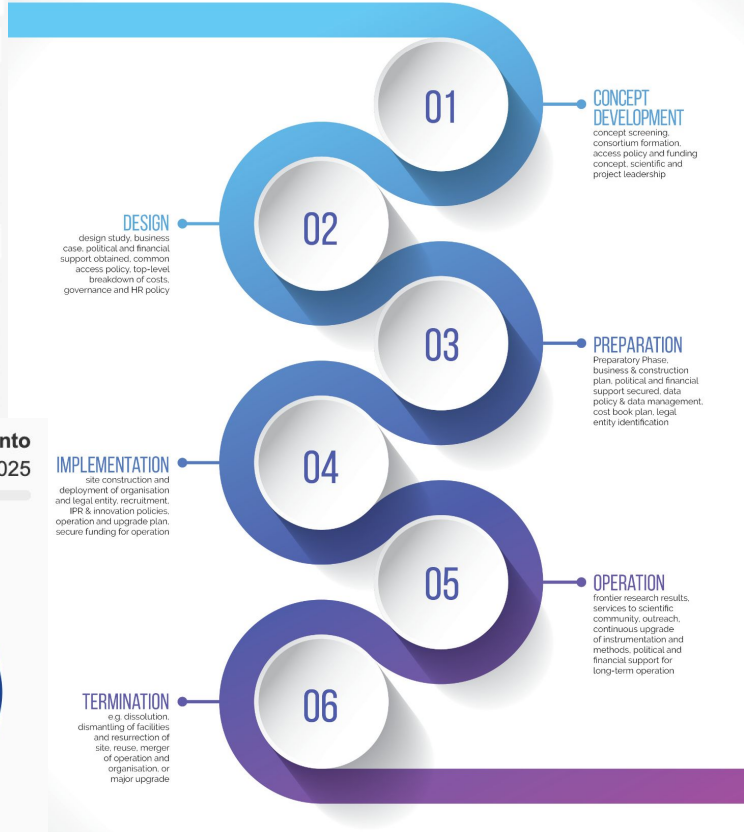


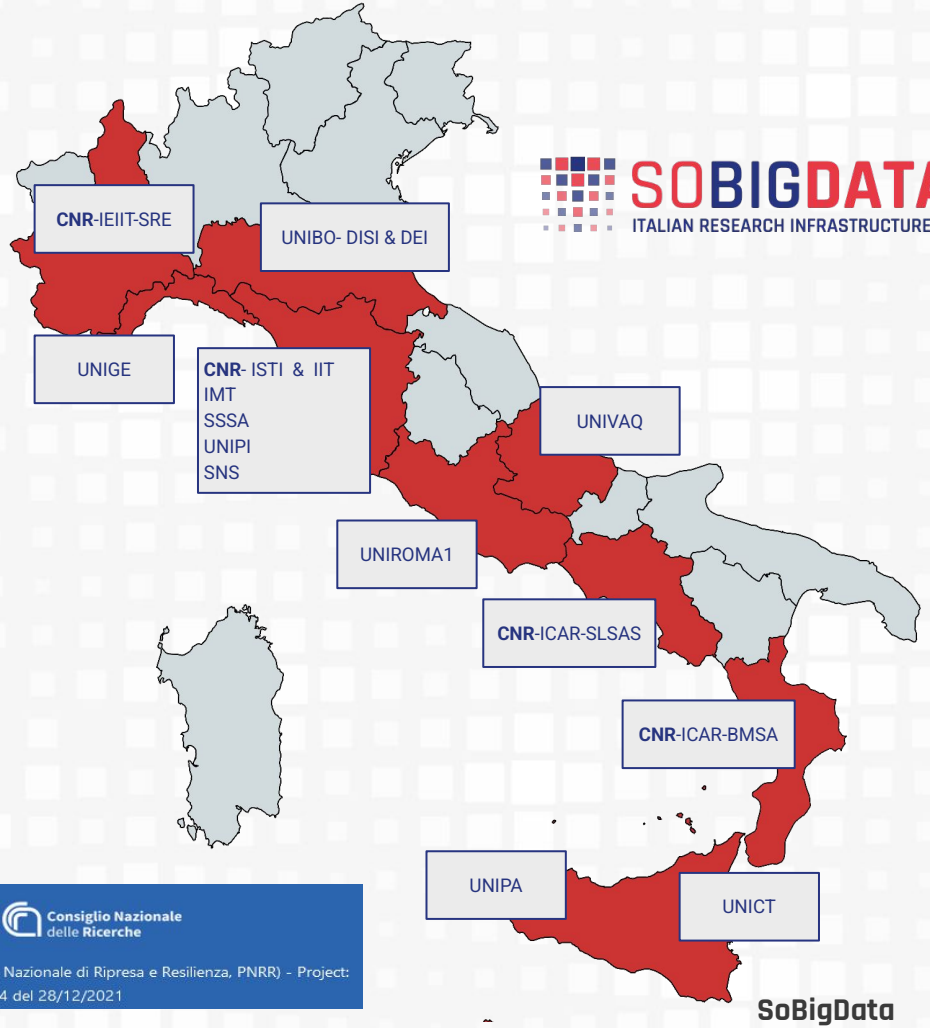
FIGURE 1.
Lifecycle approach

Year 2022-2024

SoBigData RI (Age 7-9)

SoBigData.it PNRR: Strengthening the italian node

- Collaboration with Slices RI Italian node
- Creating new Computational centers
- Enlarge the Research Spaces



SOBIGDATA.it
ITALIAN RESEARCH INFRASTRUCTURE

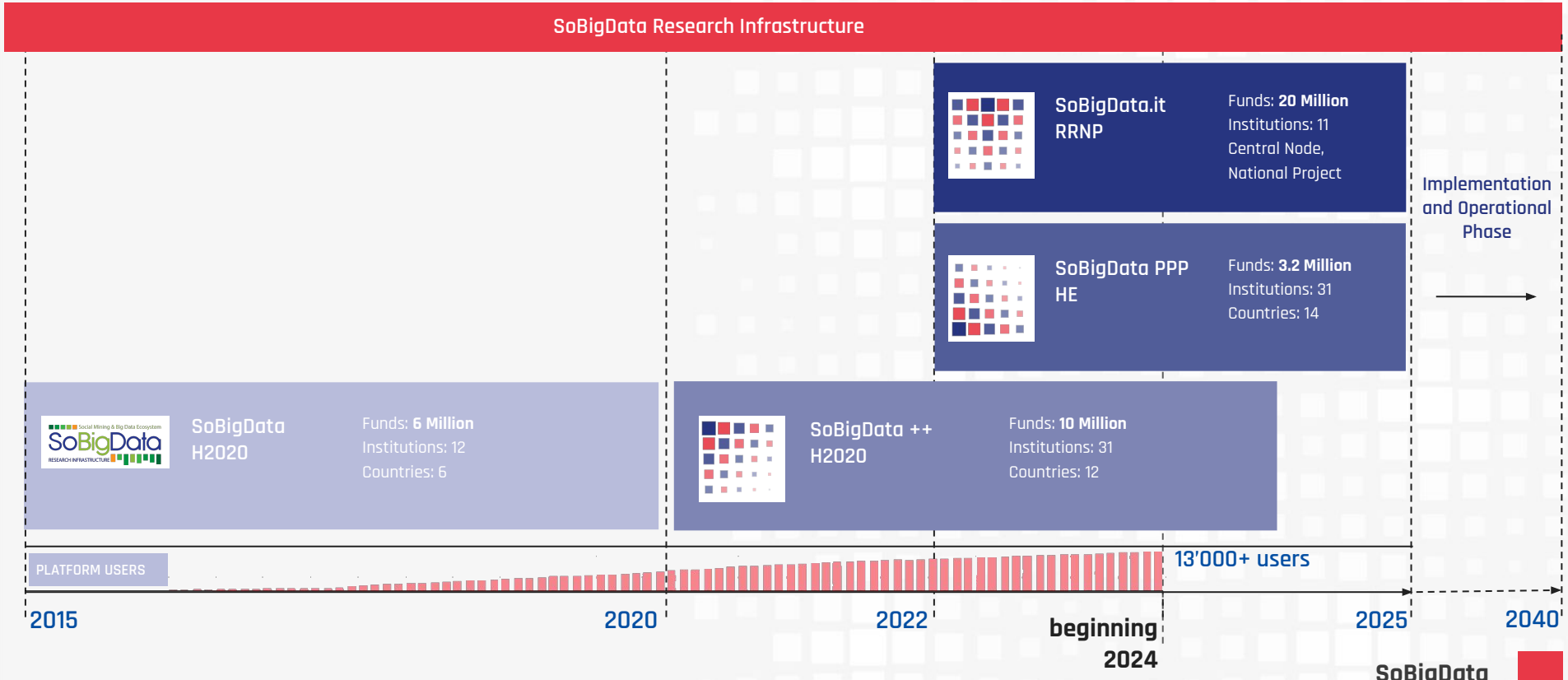
Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani | SOBIGDATA | Consiglio Nazionale delle Ricerche

SoBigData.it receives funding from European Union - NextGenerationEU - National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) - Project: "SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics" - Prot. IR0000013 - Avviso n. 3264 del 28/12/2021

Year 2022-2024

SoBigData RI (Age 7-9)

SOBIGDATA
RESEARCH INFRASTRUCTURE



Year 2025 (Today)

SoBigData RI (Age 10)

SOBIGDATA
RESEARCH INFRASTRUCTURE

- SoBigData++ ended
- Large community around the RI (13.000+ users)
- Submitting at New Calls on Infrastructures:
  - SoBigData **Implementation Phase** ESFRI (SoBigData IP, preparing to become ERIC in 2026-2027)
  - Adapting to new EU scenarios and priorities: **Data Altruism** and **Green Computing** (SoBigData Forward)
  - Exploring **pervasive AI** in collaboration with SLICES RI
  - Expanding tools for **Democracy**, **Discrimination** and **Misinformation**
  - Expanding **physical infrastructure**
- Governmental participation: political and financial support
- Drafting Statute
- Drafting Business plan

SOBIGDATA
RESEARCH INFRASTRUCTURE

The role of Research Infrastructures

The Role of research infrastructures in EU

Facilities that provide **resources** and **services** for the research communities to conduct research and foster innovation in their fields.

These include:

- major equipment or sets of instruments
- knowledge-related facilities such as collections,
- archives of scientific data infrastructures
- computing systems
- communication networks

The Role of research infrastructure in EU

Key objectives:

- reduce fragmentation of the research and innovation ecosystem, avoiding duplication of effort
- establish strategies for pan-European, well-established intergovernmental or national Research Infrastructures
- join forces internationally, foster the innovation and use Research Infrastructures for science diplomacy and build partnerships internationally
- *share EU principles for an open science and an ethical usage of AI*

Implementing strategies in SoBigData RI

Reducing Fragmentation & Avoiding Duplication: SoBigData RI brings together diverse research groups and data resources under one umbrella, ensuring seamless collaboration and reducing redundant efforts across Europe's research landscape.

Open Science tools: offering open science tools, it promoting efficiency, effective and replicable experiments. It also ensure a resource management system to projects and communities

Fostering Innovation & Science Diplomacy: By creating a collaborative ecosystem that spans disciplines and borders, SoBigData RI not only accelerates scientific innovation but also uses research infrastructures as a platform for science diplomacy, enhancing global dialogue and partnership.

National plan: it supports strategies for strengthening national facilities, ensuring a cohesive approach that leverages both innovative and well-established resources.

Ethical principles: Offer a way to train the new generation of Responsible Data Scientist and offer help evaluating experiment and project according to the European principles on AI.

Challenges

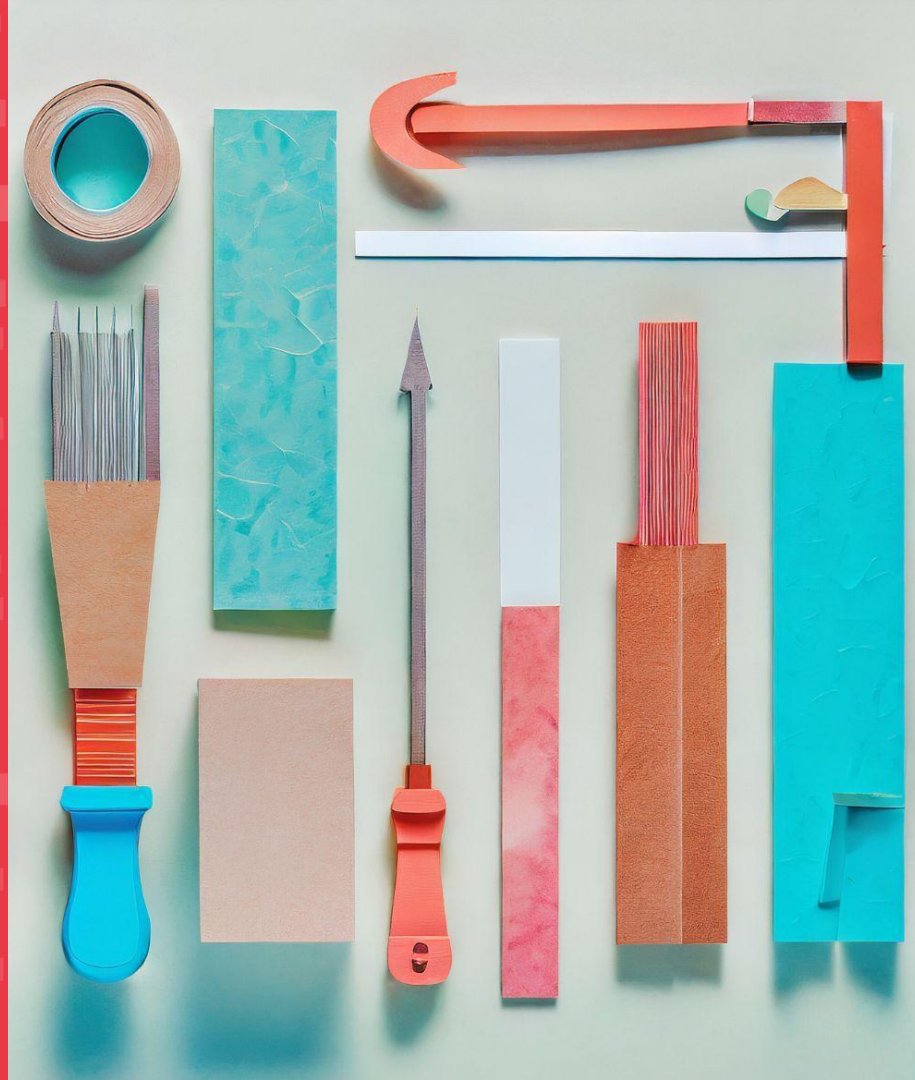
- **Establishing common policies** for data collection, storage, and sharing requires significant time and resources.
- Balancing differing **national priorities, funding cycles, and aligning goals** and processes among a wide range of stakeholders can be complex.
- **Balancing short-term research needs with long-term strategic goals** may be challenging, particularly when innovation cycles differ among institutions.
- Established institutions may **resist changes** due to comfort with existing technologies and systems or don't want to put extra effort in sharing resources
- **Administrative structures** may slow down the decision-making process
- **Communication and participation** of partners

Challenges: SoBigData Approach

- **Diverse Funding Streams**: Combine EU funding, national investments, and industry partnerships to ensure long-term financial sustainability.
- **Organize interdisciplinary workshops and events** to align research agendas among stakeholders.
- **Modular, Scalable Architecture:** Invest in adaptable and modular technical infrastructure that can evolve with emerging technologies and changing research needs.
- **Transparent Decision Processes:** Implement clear protocols and communication channels for decision-making and feedback among all partners.
- **Engage Partners** in projects to stimulate participation

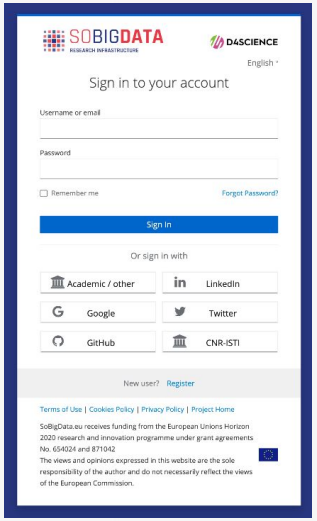
SOBIGDATA
RESEARCH INFRASTRUCTURE

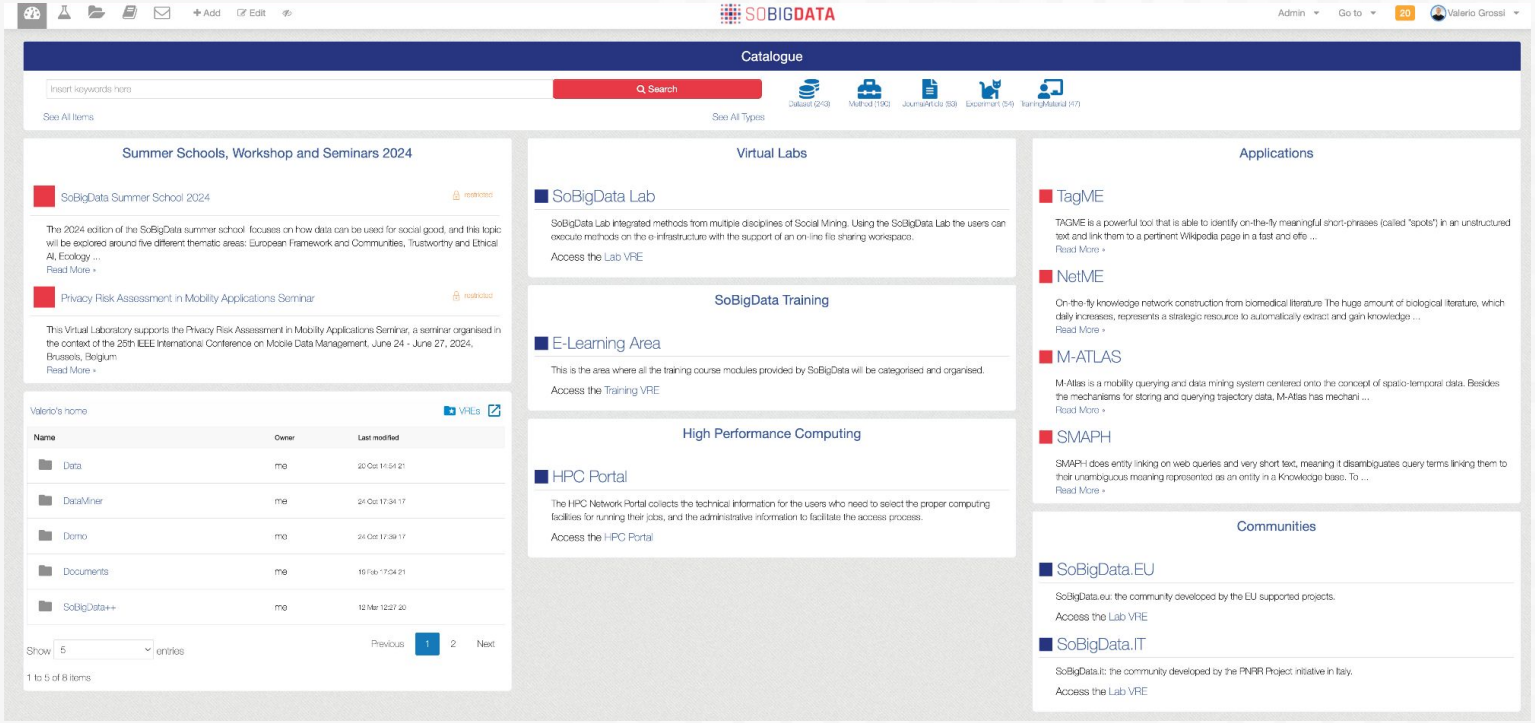
Services Focus



Accessing the e-Infrastructure - **www.sobigdata.eu**

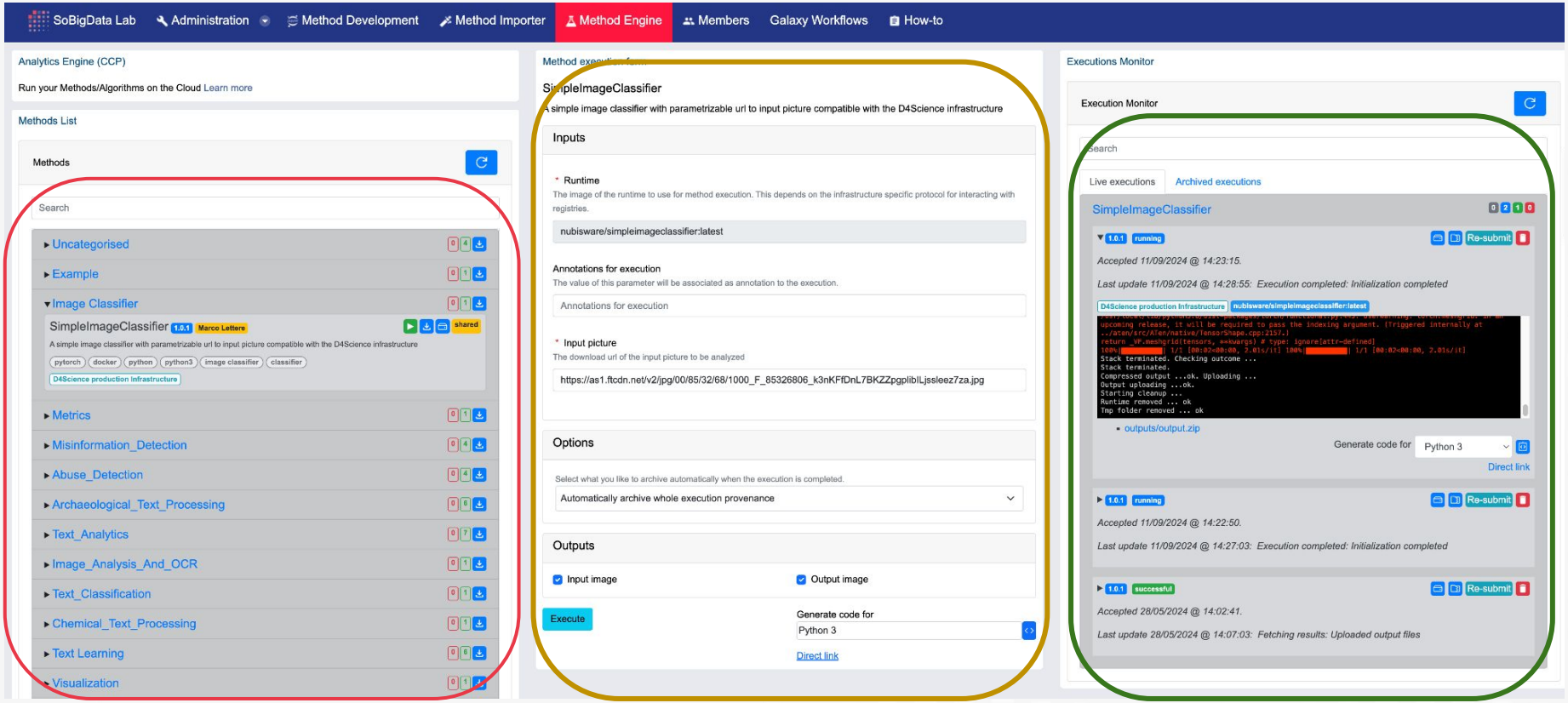
REGISTRATION REQUIRED





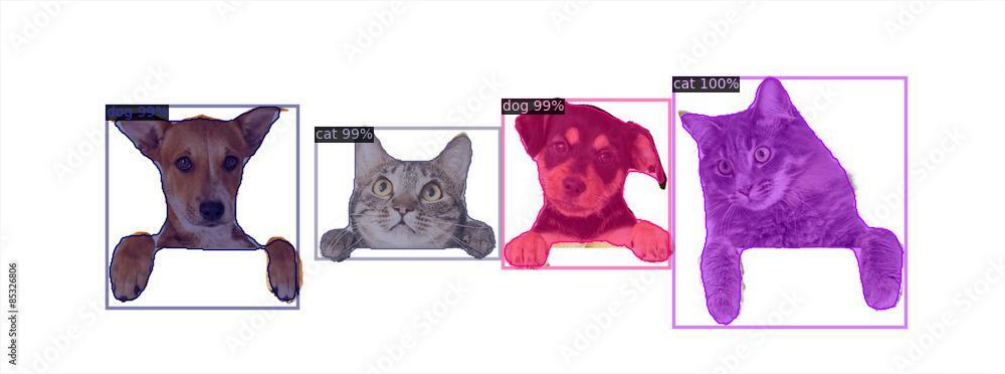
Service Focus 1: SoBigData LAB

CLOUD COMPUTING PLATFORM (CCP)



SoBigData LAB

CLOUD COMPUTING PLATFORM (CCP) - DOWNLOADABLE

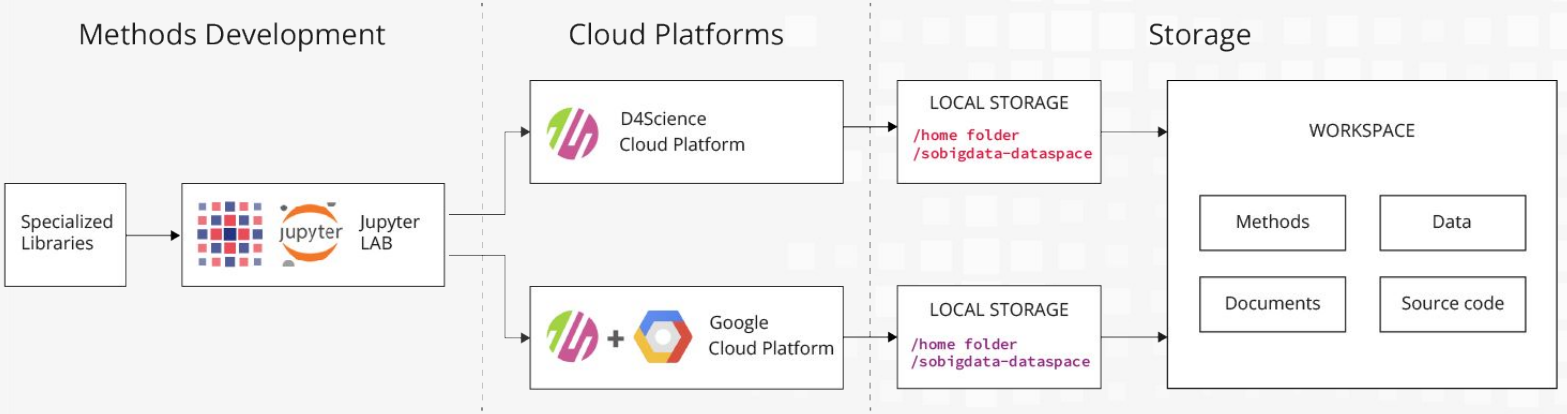



```
model_final_f10217.pkl: 0.00B [00:00, ?B/s]
model_final_f10217.pkl:   0%|          | 106k/178M [00:00<02:50, 1.04MB/s]
model_final_f10217.pkl:   2%|          | 3.79M/178M [00:00<00:08, 20.9MB/s]
model_final_f10217.pkl:   6%|▏         | 10.4M/178M [00:00<00:04, 40.9MB/s]
model_final_f10217.pkl:  10%|▏         | 18.0M/178M [00:00<00:02, 54.6MB/s]
model_final_f10217.pkl:  16%|▏         | 28.5M/178M [00:00<00:02, 72.3MB/s]
model_final_f10217.pkl:  22%|▏         | 39.0M/178M [00:00<00:01, 83.4MB/s]
model_final_f10217.pkl:  29%|▎         | 51.4M/178M [00:00<00:01, 96.4MB/s]
model_final_f10217.pkl:  35%|▎         | 63.0M/178M [00:00<00:01, 103MB/s]
model_final_f10217.pkl:  42%|▍         | 74.7M/178M [00:00<00:00, 107MB/s]
model_final_f10217.pkl:  49%|▍         | 86.3M/178M [00:01<00:00, 110MB/s]
model_final_f10217.pkl:  55%|▌         | 98.1M/178M [00:01<00:00, 112MB/s]
model_final_f10217.pkl:  62%|▌         | 110M/178M [00:01<00:00, 114MB/s]
model_final_f10217.pkl:  68%|▋         | 121M/178M [00:01<00:00, 115MB/s]
model_final_f10217.pkl:  75%|▋         | 133M/178M [00:01<00:00, 115MB/s]
model_final_f10217.pkl:  81%|▊         | 145M/178M [00:01<00:00, 116MB/s]
model_final_f10217.pkl:  88%|▊         | 156M/178M [00:01<00:00, 116MB/s]
model_final_f10217.pkl:  95%|▉         | 168M/178M [00:01<00:00, 116MB/s]
model_final_f10217.pkl: 178MB [00:01, 99.1MB/s]

  0%|          | 0/1 [00:00<?, ?it/s]/usr/local/lib/python3.8/dist-packages/detectron2/structures/image_list.py:88: UserWarning:
__floordiv__ is deprecated, and its behavior will change in a future version of pytorch. It currently rounds toward 0 (like the
'trunc' function NOT 'floor'). This results in incorrect rounding for negative values. To keep the current behavior, use
torch.div(a, b, rounding_mode='trunc'), or for actual floor division, use torch.div(a, b, rounding_mode='floor').
  max_size = (max_size + (stride - 1)) // stride * stride
/usr/local/lib/python3.8/dist-packages/torch/functional.py:445: UserWarning: torch.meshgrid: in an upcoming release, it will be
required to pass the indexing argument. (Triggered internally at  ../aten/src/ATen/native/TensorShape.cpp:2157.)
  return _VF.meshgrid(tensors, **kwargs)  # type: ignore[attr-defined]

100%|██████████| 1/1 [00:02<00:00,  2.00s/it]
100%|██████████| 1/1 [00:02<00:00,  2.00s/it]

  0%|          | 0/1 [00:00<?, ?it/s]/usr/local/lib/python3.8/dist-packages/detectron2/structures/image_list.py:88: UserWarning:
__floordiv__ is deprecated, and its behavior will change in a future version of pytorch. It currently rounds toward 0 (like the
'trunc' function NOT 'floor'). This results in incorrect rounding for negative values. To keep the current behavior, use
torch.div(a, b, rounding_mode='trunc'), or for actual floor division, use torch.div(a, b, rounding_mode='floor').
  max_size = (max_size + (stride - 1)) // stride * stride
/usr/local/lib/python3.8/dist-packages/torch/functional.py:445: UserWarning: torch.meshgrid: in an upcoming release, it will be
required to pass the indexing argument. (Triggered internally at  ../aten/src/ATen/native/TensorShape.cpp:2157.)
  return _VF.meshgrid(tensors, **kwargs)  # type: ignore[attr-defined]

100%|██████████| 1/1 [00:02<00:00,  2.01s/it]
100%|██████████| 1/1 [00:02<00:00,  2.01s/it]
```

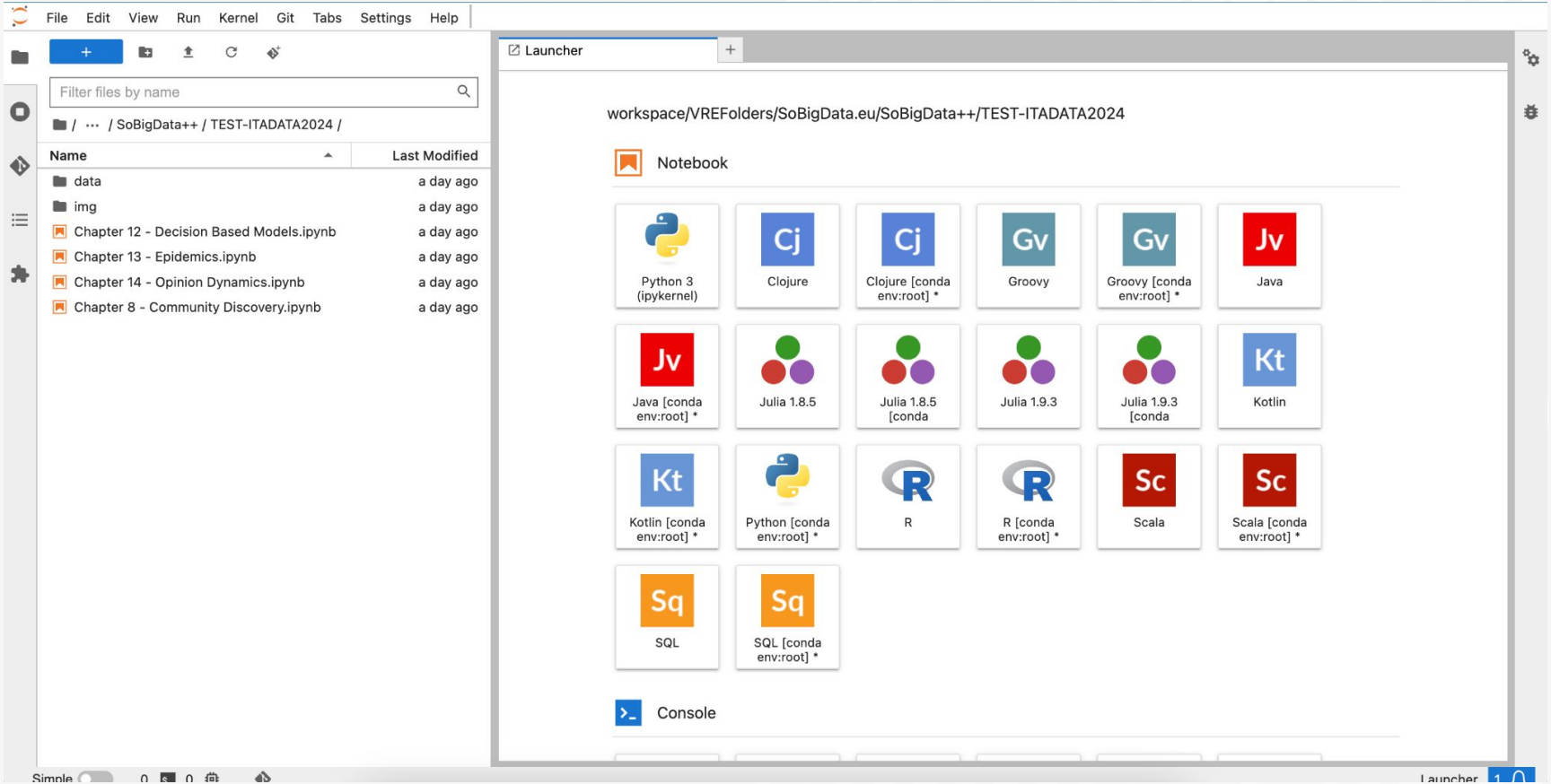
SoBigData LAB: Jupyter environment

This service provides a ready-to-use environment, eliminating the **need to install** and **maintain** software yourself.

It comes **pre-loaded** with popular **data analysis libraries** and **packages**, so you can start working on your projects immediately. Plus, you can easily share data and resources with others through the integrated Workspace.



SoBigData LAB: Jupyter environment



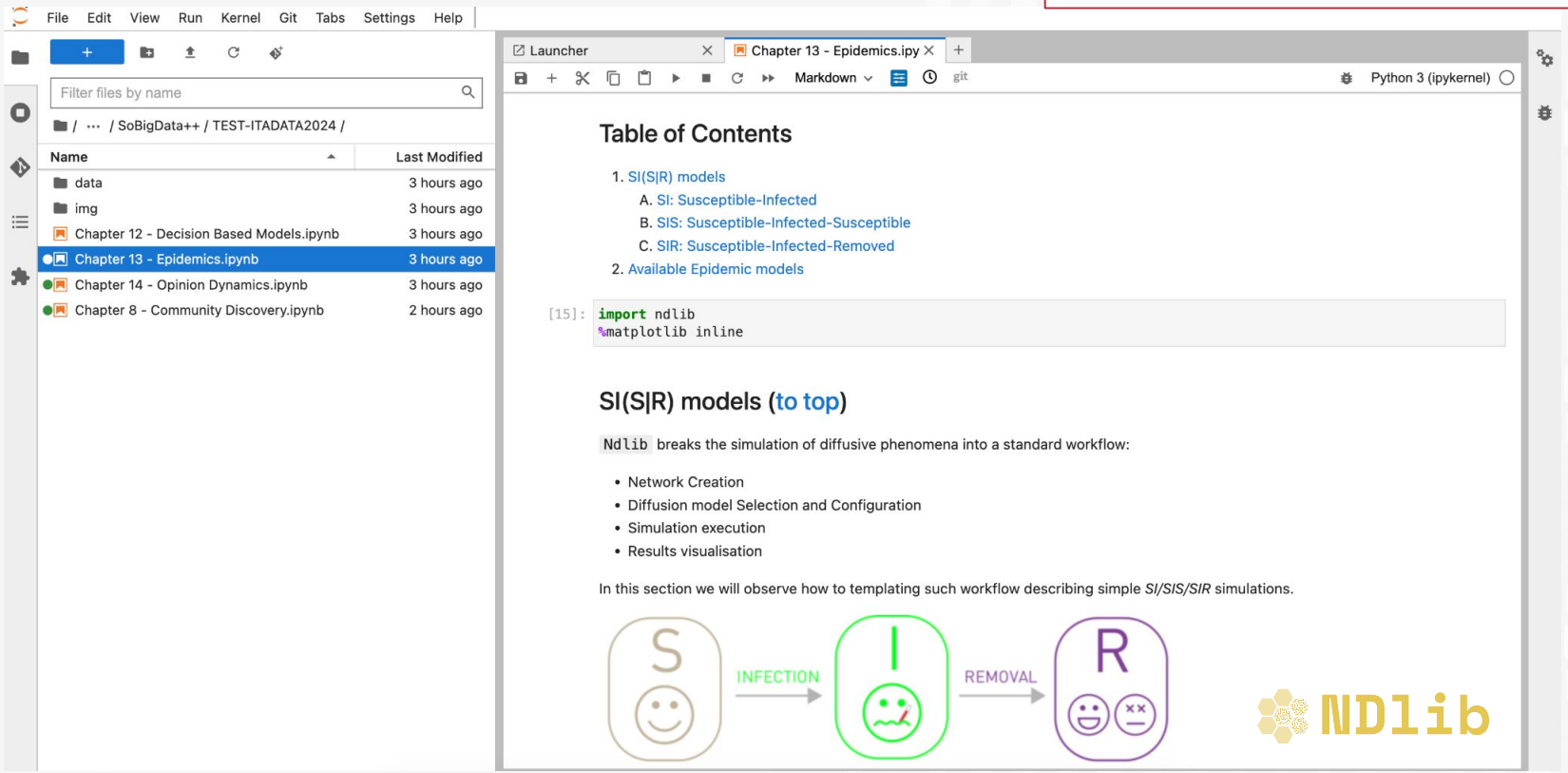
SoBigData Lab: JupyterLAB environment

EXAMPLE OF USE - NDlib

NDlib - Network Diffusion Library

`NDlib` is a Python software package that allows to describe, simulate, and study diffusion processes on complex networks.

Date	Python Versions	Main Author	GitHub	pypl
2024-06-26	>=3.6	Giulio Rossetti	Source	Distribution



SoBigData Lab: Jupyter environment

EXAMPLE OF USE - NDlib

SI: Susceptible-Infected (to top)

Each individual has β contacts with randomly chosen others individuals per unit time.

If there are I infected individual and S susceptible individuals, the average rate of new infection is βsi/N

[17]:
```
model = ep.SIModel(g)
```

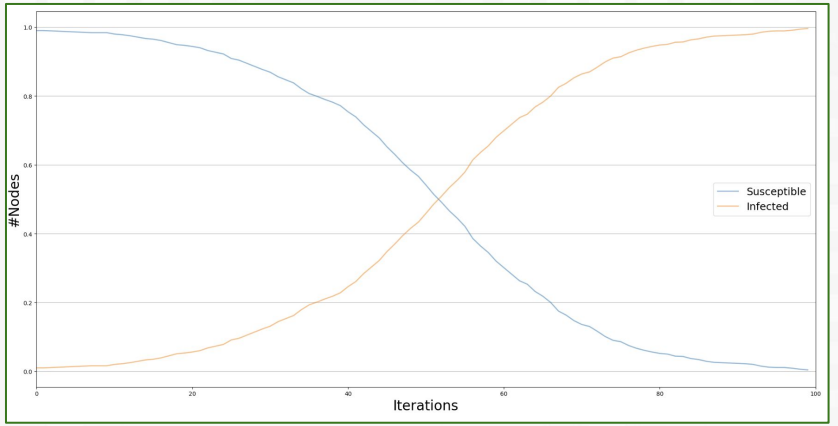
[18]:
```
model.available_statuses
```

[18]:
```
{'Susceptible': 0, 'Infected': 1}
```

[19]:
```
cfg = mc.Configuration()
cfg.add_model_parameter('beta', 0.001) # infection rate
cfg.add_model_parameter("percentage_infected", 0.01)
model.set_initial_status(cfg)
```

[20]:
```
iterations = model.iteration_bunch(100, node_status=True)
trends = model.build_trends(iterations)
```

[21]:
```
%matplotlib inline
from ndlib.viz.mpl.DiffusionTrend import DiffusionTrend
viz = DiffusionTrend(model, trends)
viz.plot()
```



SoBigData LAB: Jupyter environment

EXAMPLE OF USE - NDlib

SIR: Susceptible-Infected-Recovered (to top)

Each individual has β contacts with randomly chosen others individuals per unit time.

Each infected individual has μ probability of becoming immune after being infected.

[29]:
```
model = ep.SIRModel(g)
```

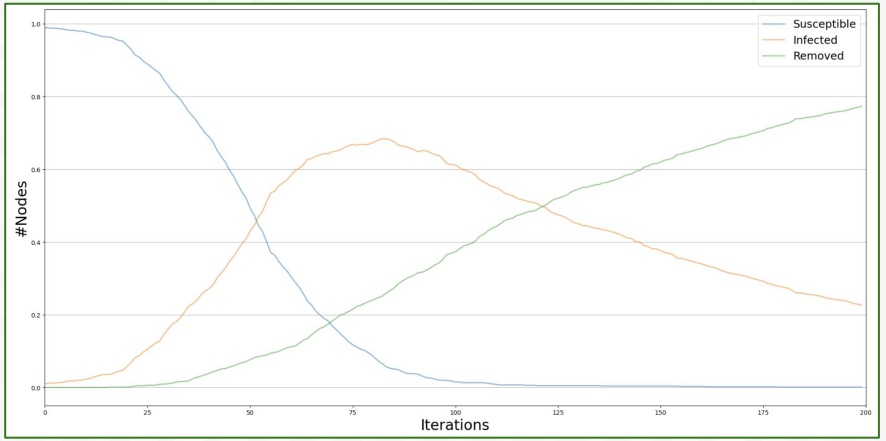
[30]:
```
model.available_statuses
```

[30]:
```
{'Susceptible': 0, 'Infected': 1, 'Removed': 2}
```

[31]:
```
cfg = mc.Configuration()
cfg.add_model_parameter('beta', 0.001) # infection rate
cfg.add_model_parameter('gamma', 0.01) # recovery rate
cfg.add_model_parameter("percentage_infected", 0.01)
model.set_initial_status(cfg)
```

[32]:
```
iterations = model.iteration_bunch(200, node_status=True)
trends = model.build_trends(iterations)
```

[33]:
```
%matplotlib inline
from ndlib.viz.mpl.DiffusionTrend import DiffusionTrend
viz = DiffusionTrend(model, trends)
viz.plot()
```



Available models (to top)

When we talk about epidemics, we think about contagious diseases caused by biological pathogens, like influenza, measles, chickenpox and sexually transmitted viruses that spread from person to person.

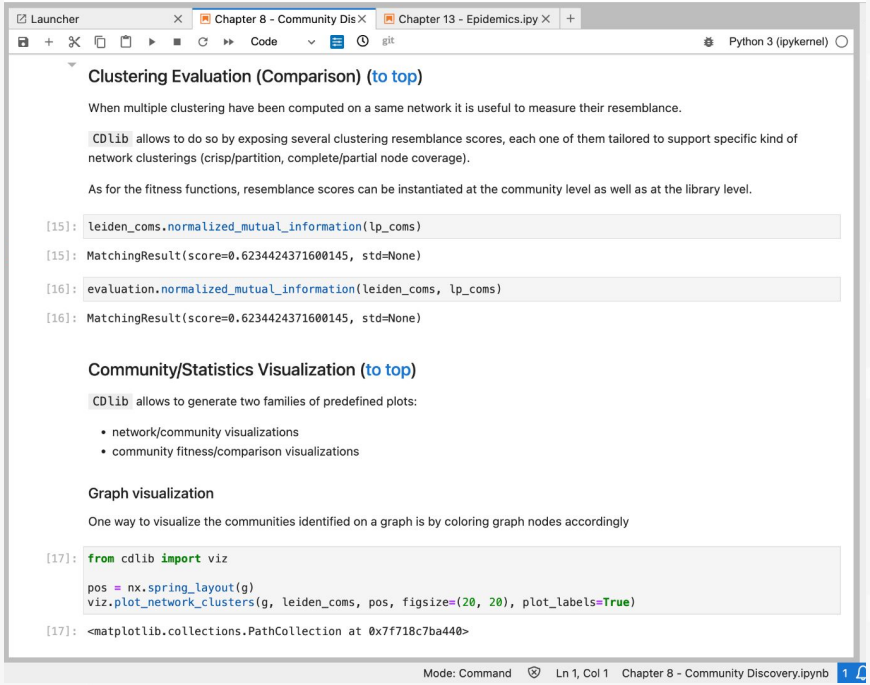
Several elements determine the patterns by which epidemics spread through groups of people: the properties carried by the pathogen (its contagiousness, the length of its infectious period and its severity), the structure of the network as well as the mobility patterns of the people involved.

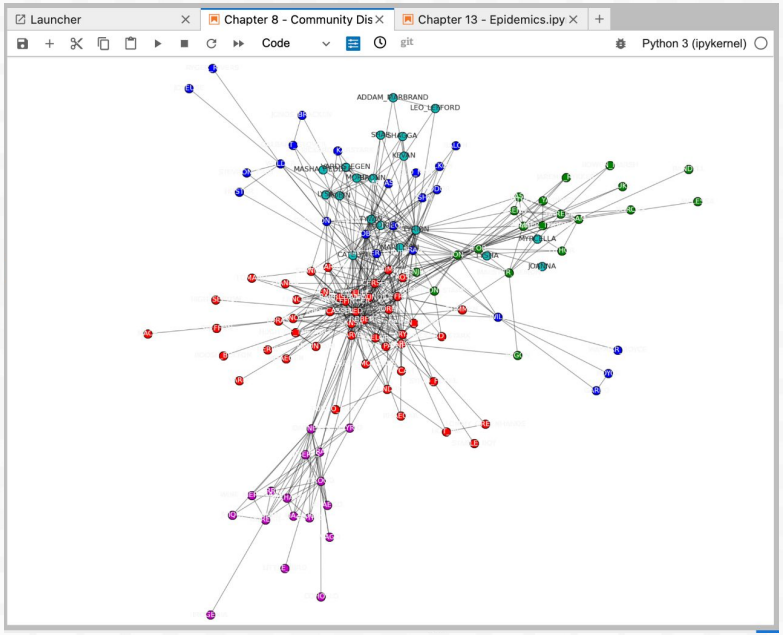
In `NDlib` are implemented the following 12 Epidemic models:

SI	SIS	SIR
SEIR	SEIS	SWIR
Threshold	Generalised Threshold	Kertesz Threshold
Profile	Profile-Threshold	Independent Cascades

SoBigData LAB: Jupyter environment

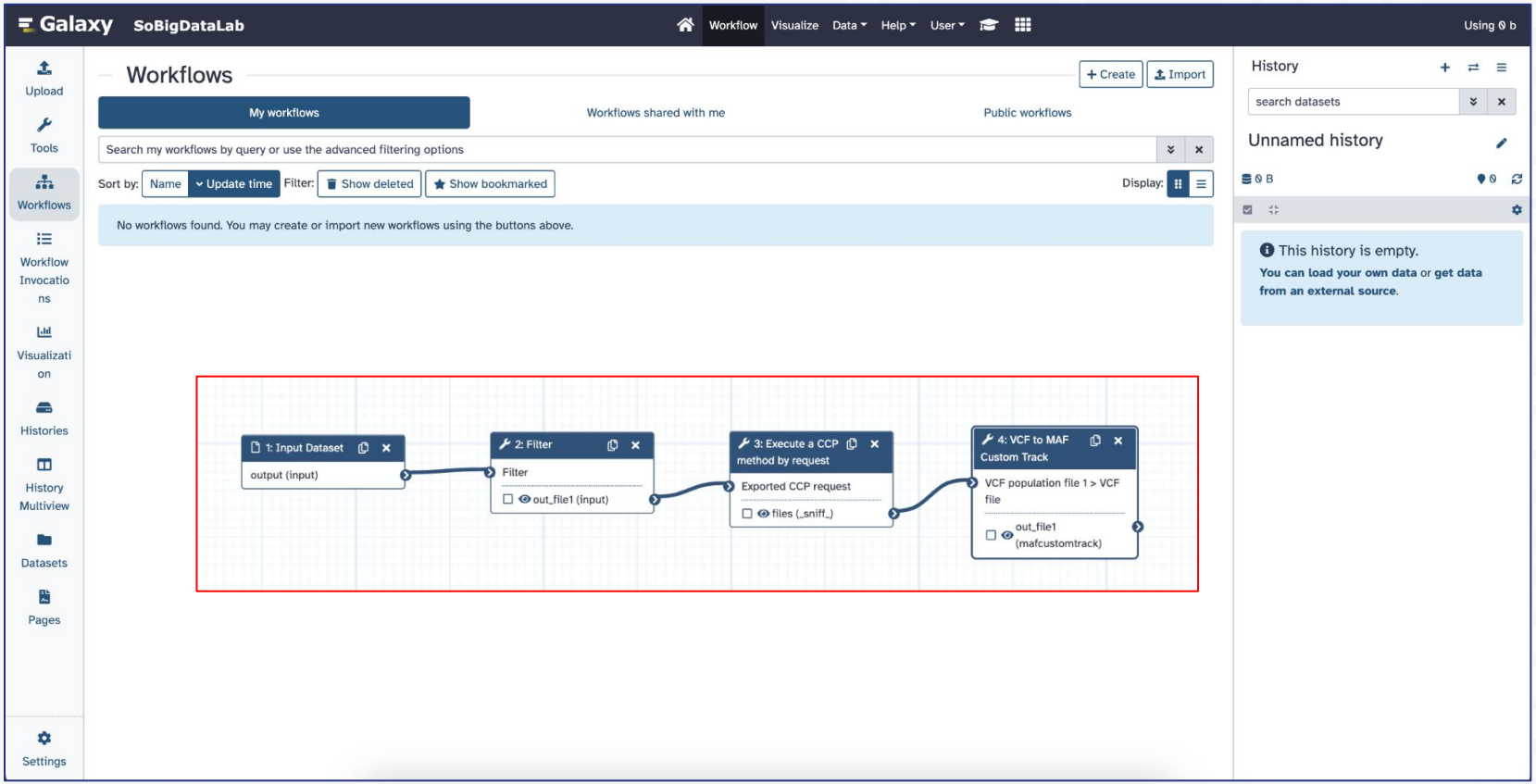
EXAMPLE OF USE - CDlib - Community Detection Library





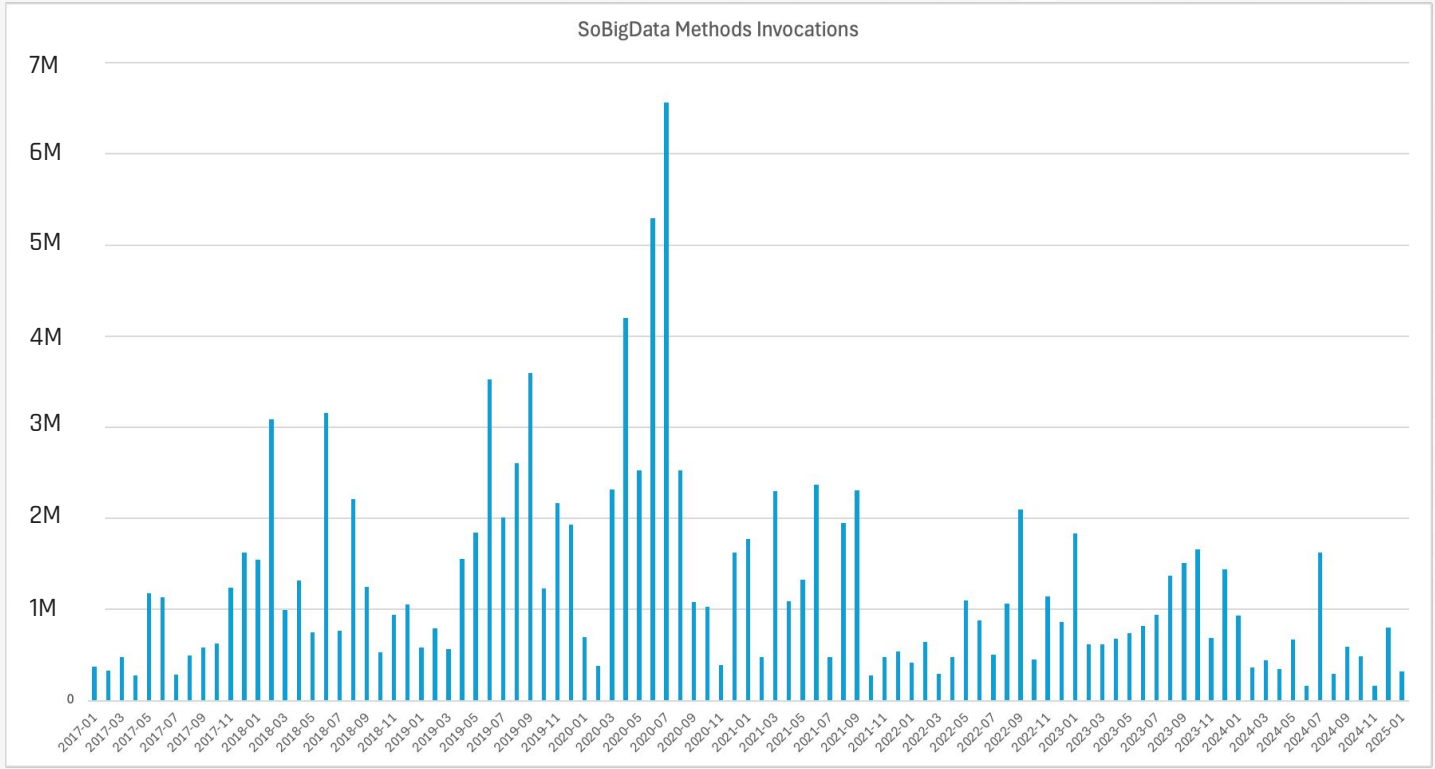
CDlib

SoBigData LAB: Galaxy

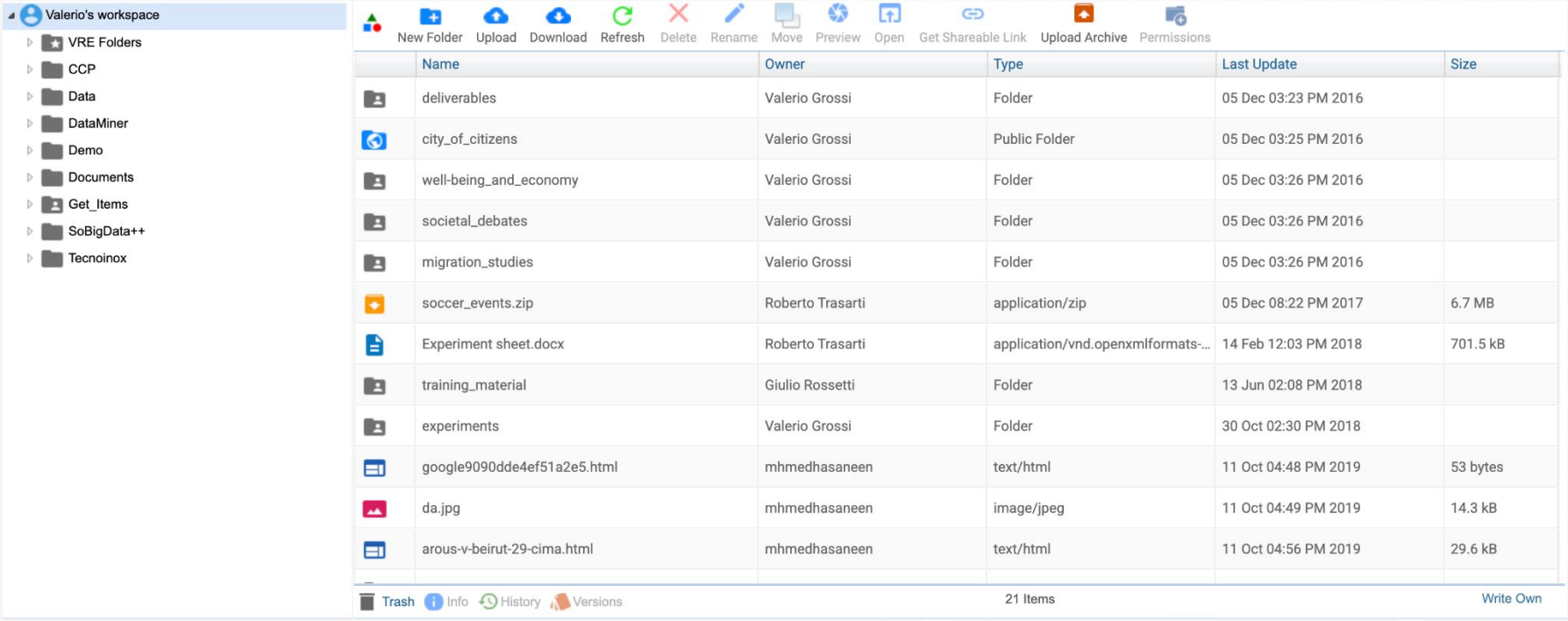


SoBigData Methods invocations

The average number of calls to the RI for algorithms executions are on average **1,265,200 per month**, with an high variability.



Service Focus 2: SoBigData Workspace



Workspace Features

SEARCH

The Workspace search allows to look for any item (file or folder) stored on your workspace (be it your or shared by others). To do this, click the lens icon below then type the name of the file or folder or just part of it.

Searching for “data” will return all the files and folders whose name includes the string 'data'

SHARE

Workspace Share Folders and Files

The quickest way to share something is using the Share Folder. Locate the folder with the files you want to share and then click 'Share'

SHAREABLE LINKS

Link To File as Private

(Only the) Members are enacted to access the file and the shared folder content. Login required

as Public

Anyone with link can download it. No Login required

UPLOAD

Workspace Upload Files and Archives

User can upload files in the Workspace in several ways:

1 - Drop your files from Desktop;

2 - Click 'Upload' and Browse Files;

3 - Upload a zip file to unzip directly its content in the Workspace.i

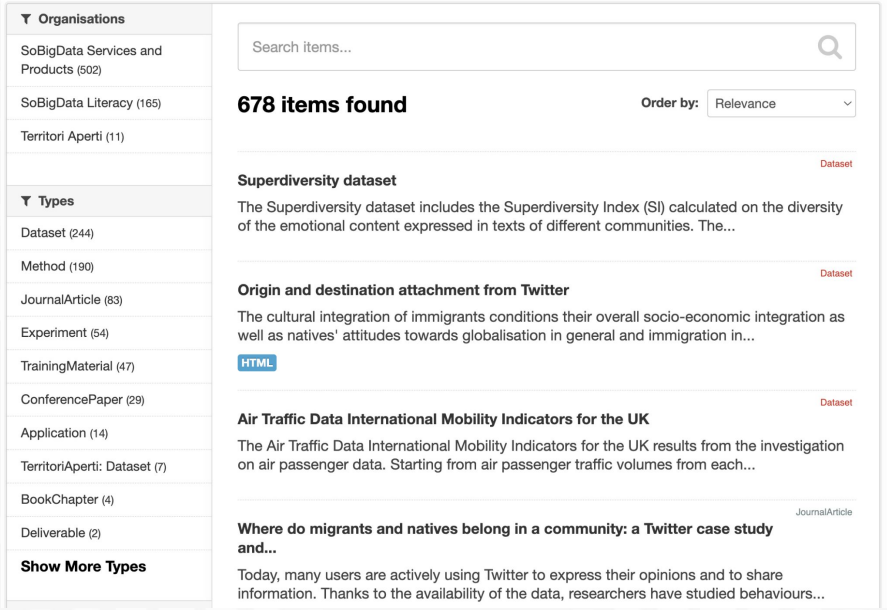
VERSION CONTROL

Workspace version control: Workspace keeps track of any file version, transparently.

To see the other file versions, select a file, then right click on it and click "Versions".

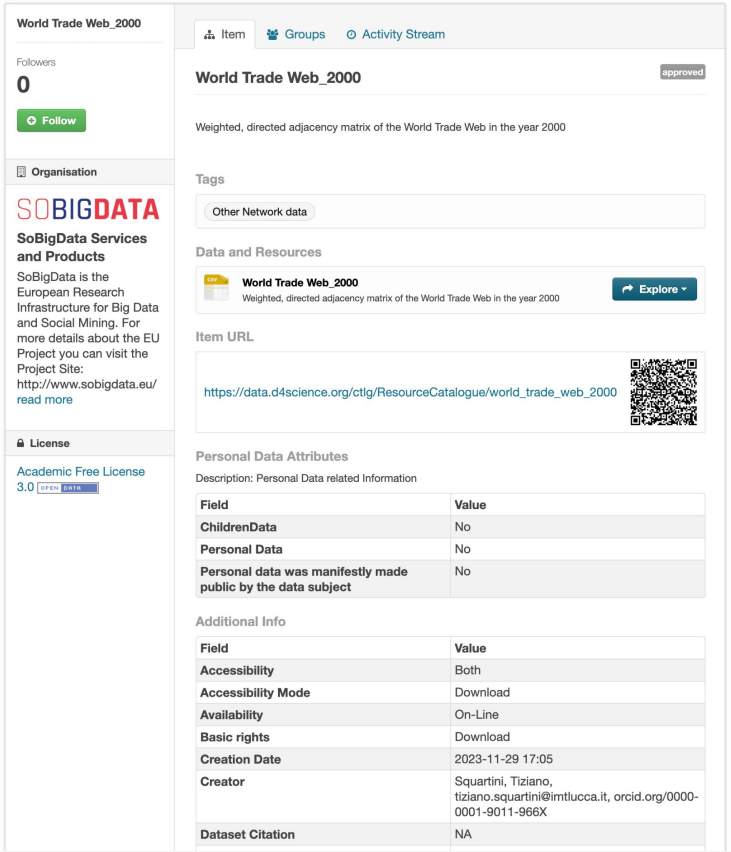
Service Focus 3: Catalogue

REGISTRATION REQUIRED FOR ACCESSING THE RESOURCES



Catalogue

REGISTRATION REQUIRED FOR ACCESSING THE RESOURCES



World Trade Web_2000

Followers
0
Follow

Organisation

SOBIGDATA

SoBigData Services and Products

SoBigData is the European Research Infrastructure for Big Data and Social Mining. For more details about the EU Project you can visit the Project Site: http://www.sobigdata.eu/ read more

License

Academic Free License 3.0

Item | Groups | Activity Stream

World Trade Web_2000 approved

Weighted, directed adjacency matrix of the World Trade Web in the year 2000

Tags

Other Network data

Data and Resources

World Trade Web_2000
Weighted, directed adjacency matrix of the World Trade Web in the year 2000
Explore

Item URL

https://data.d4science.org/ctlg/ResourceCatalogue/world_trade_web_2000

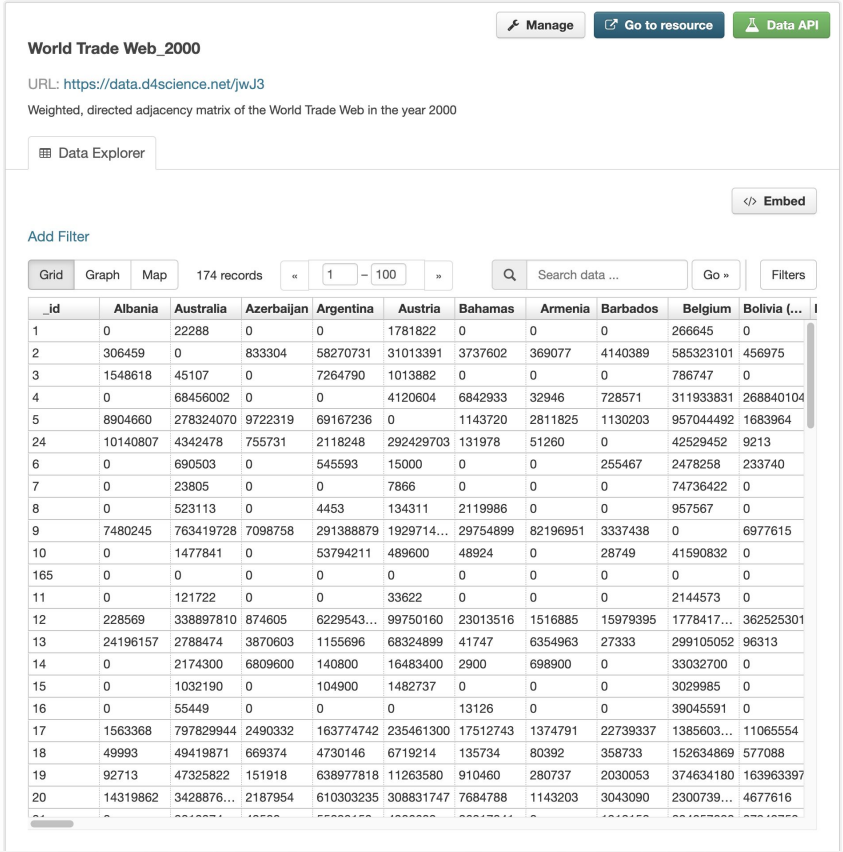
Personal Data Attributes

Description: Personal Data related Information

Field	Value
ChildrenData	No
Personal Data	No
Personal data was manifestly made public by the data subject	No

Additional Info

Field	Value
Accessibility	Both
Accessibility Mode	Download
Availability	On-Line
Basic rights	Download
Creation Date	2023-11-29 17:05
Creator	Squartini, Tiziano, tiziano.squartini@imtlucca.it, orcid.org/0000-0001-9011-966X
Dataset Citation	NA



Manage | Go to resource | Data API

World Trade Web_2000

URL: https://data.d4science.net/jwJ3

Weighted, directed adjacency matrix of the World Trade Web in the year 2000

Data Explorer

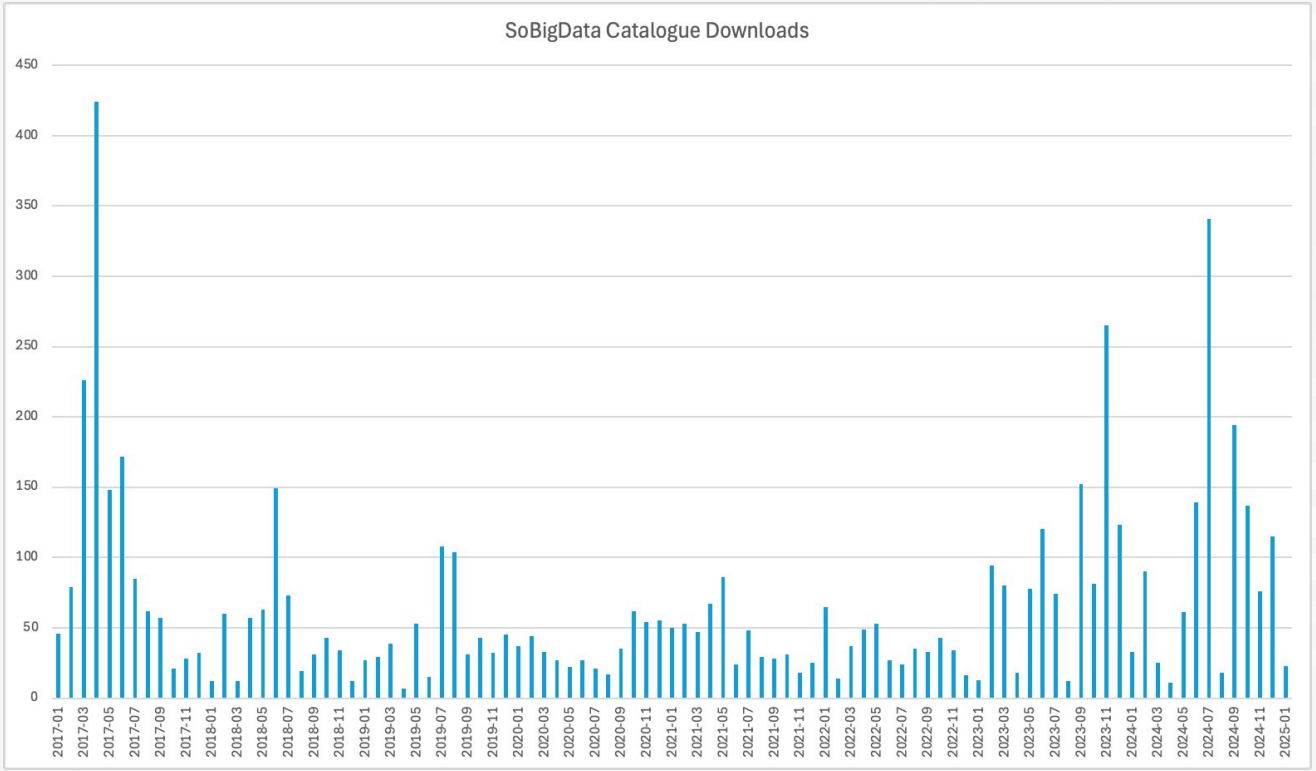
Embed

Add Filter

Grid | Graph | Map | 174 records | « | 1 – 100 | » | Search data ... | Go » | Filters

_id	Albania	Australia	Azerbaijan	Argentina	Austria	Bahamas	Armenia	Barbados	Belgium	Bolivia (...
1	0	22288	0	0	1781822	0	0	0	266645	0
2	306459	0	833304	58270731	31013391	3737602	369077	4140389	585323101	456975
3	1548618	45107	0	7264790	1013882	0	0	0	786747	0
4	0	68456002	0	0	4120604	6842933	32946	728571	311933831	268840104
5	8904660	278324070	9722319	69167236	0	1143720	2811825	1130203	957044492	1683964
24	10140807	4342478	755731	2118248	292429703	131978	51260	0	42529452	9213
6	0	690503	0	545593	15000	0	0	255467	2478258	233740
7	0	23805	0	0	7866	0	0	0	74736422	0
8	0	523113	0	4453	134311	2119986	0	0	957567	0
9	7480245	763419728	7098758	291388879	1929714...	29754899	82196951	3337438	0	6977615
10	0	1477841	0	53794211	489600	48924	0	28749	41590832	0
165	0	0	0	0	0	0	0	0	0	0
11	0	121722	0	0	33622	0	0	0	2144573	0
12	228569	338897810	874605	6229543...	99750160	23013516	1516885	15979395	1778417...	362525301
13	24196157	2788474	3870603	1155696	68324899	41747	6354963	27333	299105052	96313
14	0	2174300	6809600	140800	16483400	2900	698900	0	33032700	0
15	0	1032190	0	104900	1482737	0	0	0	3029985	0
16	0	55449	0	0	0	13126	0	0	39045591	0
17	1563368	797829944	2490332	163774742	235461300	17512743	1374791	22739337	1385603...	11065554
18	49993	49419871	669374	4730146	6719214	135734	80392	358733	152634869	577088
19	92713	47325822	151918	638977818	11263580	910460	280737	2030053	374634180	163963397
20	14319862	3428876...	2187954	610303235	308831747	7684788	1143203	3043090	2300739...	4677616

Catalogue downloads



The average number of downloads from catalogue per month are 64. Peaks usually correspond to big events and/or the release of a new important dataset.

Catalogue: Data Management for projects

SoBigData manage product of research for communities and projects which want to share them in an Open Science environment.

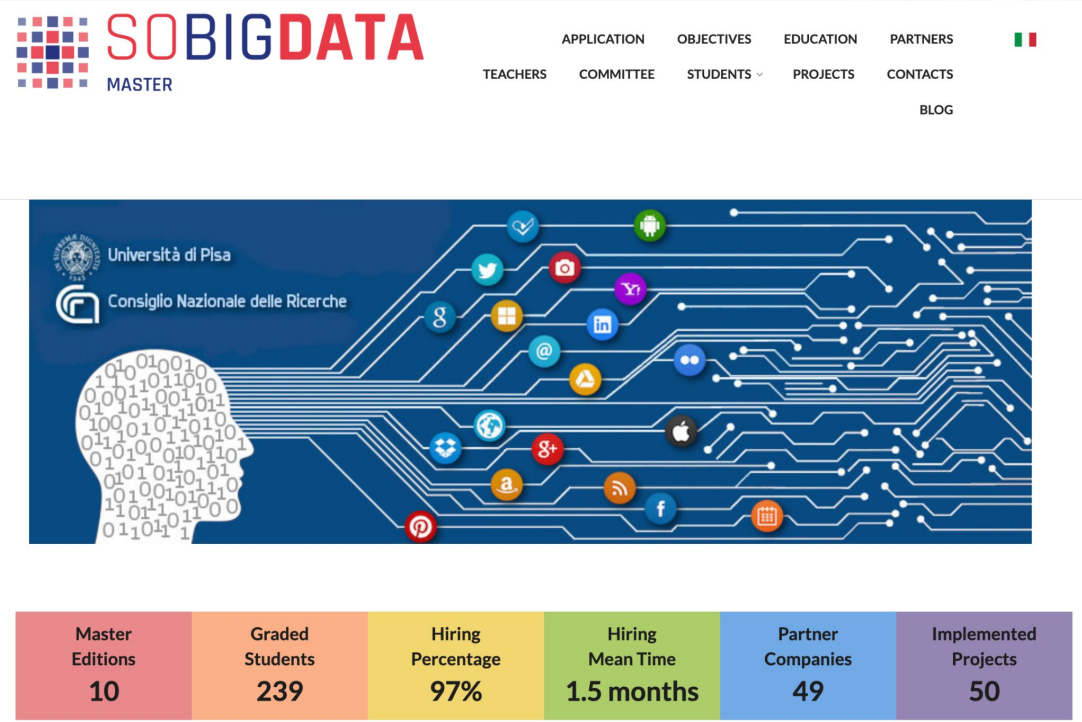
- Public Data
- Private Data (IP, personal data, etc.), sharing only metadata
- Algorithms and tools
- Experiments
- Publications

The catalogue is accessible both from the website and through API.

The resources has a special field called “associated project” which make visible the source and make data management easier for projects (e.g. data management plan information retrieving)

Service Focus 4: Master in Big Data Analytics & AI for Society

Training new generation of responsible data scientists - 11th year



Master Editions	Graded Students	Hiring Percentage	Hiring Mean Time	Partner Companies	Implemented Projects
10	239	97%	1.5 months	49	50

Designed for graduates of all disciplines, it provides the tools to work in the area of data analysis applied to all fields, from business to research.

Multidisciplinary and Industrial link is the key

TECHNOLOGIAL & SCIENTIFIC AREAS

Big Data Technology
Data Management for Business Intelligence, High Performance and Scalable Analytics, NoSQL Big Data Platforms

Big Data Sensing & Procurement
Analytical Web Crawling, Web Search and Information Retrieval, Text Annotation, Big Data Sources and Crowdsensing

Big Data Mining
Data Mining, Machine Learning, Social Network Analysis, Web Mining, Nowcasting, Sentiment Analysis

Big Data Story Telling
Visualization, Visual Analytics and Data Jurnalism

Big Data Ethics
Privacy-by-Design, EU Data Protection Regulation, Data Scientist's Responsibility

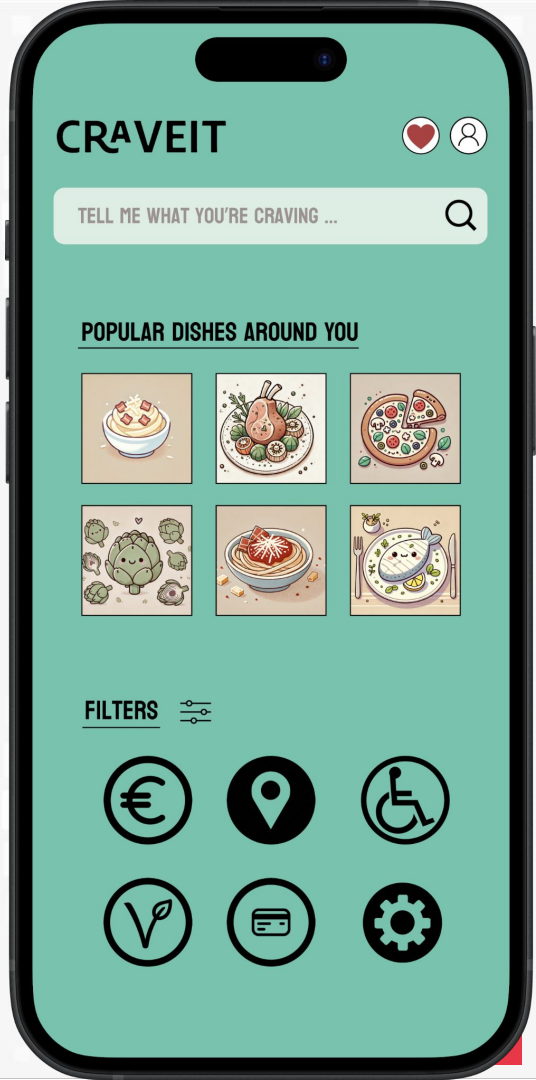
Big Data for Social Good
Mobility Analysis using Mobile Phones Records, GPS Tracks, Smart-City Sensors, etc. Diffusion of Opinions, Reputation, Sentiment and Engagement in Social Media. Big Data and Official Statistics

Big Data for Business
Big Data in Finance and Economics, Recommendation Systems, Novel CRM Applications, Data Journalism and the use of Big Data in Electronic Publishing

Example of Final Project: CraveIt

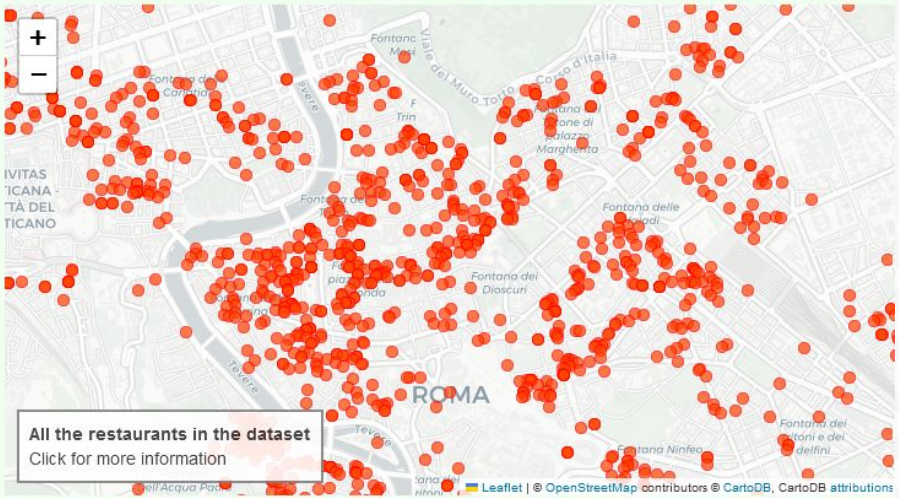
What's the best carbonara in Rome?

- CraveIT is a dish-specific guide to eating
- Unlike traditional apps, instead of suggesting good restaurants, CraveIT focuses on what you actually want to eat.
- Using a custom AI-powered algorithm CraveIT suggests the perfect place.



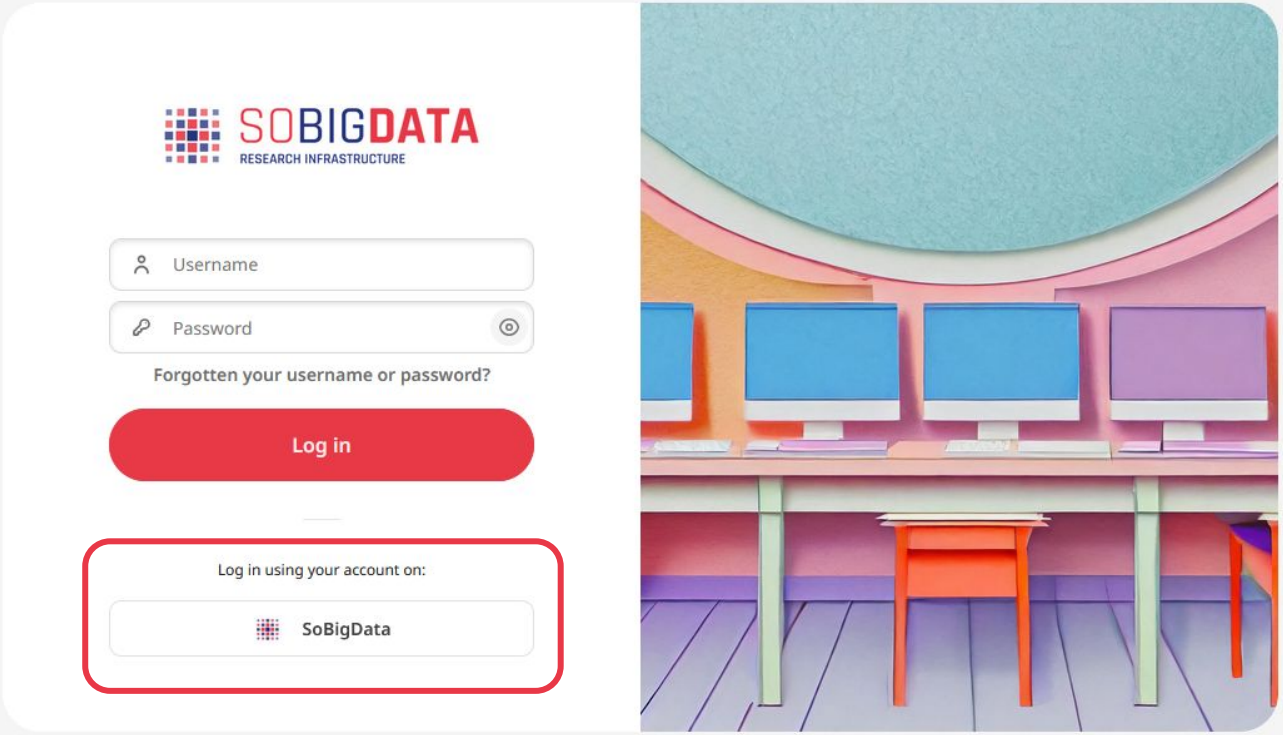
Build a dish-specific recommender

- Focus on Rome
- Data from January 2019 onward
- Over 1.000.000 reviews from the most influential information hubs of the food industry
- 15 target dishes
- Sentiment analysis to pinpoin the sentiment associated with the target dish
- Validation (comparison with articles like on The best carbonara in Rome)



Service Focus 5: SoBigData Academy

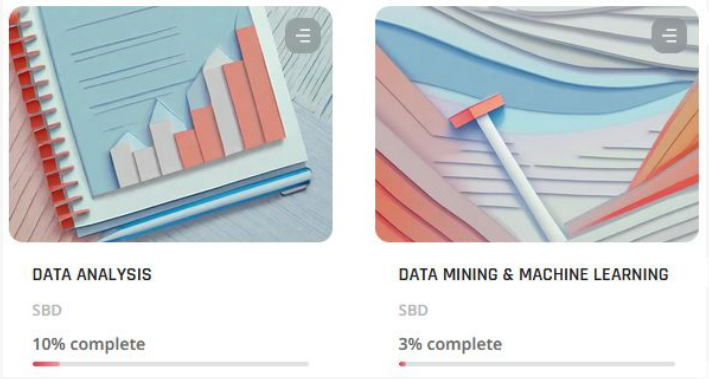
Training new generation of responsible data scientists - Launched in December 2024



The Academy's core includes a series of **MOOCs** (Massive Open Online Courses), with easy enrollment and adaptive learning speed and complexity.

http://www.sobigdata.eu/academy

SoBigData Academy



- **Free and Accessible**: Learn from experts without geographic or time constraints.
- **Network with Leading Institutions**: Connect with European universities and industry leaders.
- **Diverse Curriculum**: Explore a wide range of data science courses.
- **Interactive Learning**: Engage with interactive lessons, videos, and quizzes.
- **Hands-On Practice**: Utilize Jupyter Notebook for practical coding exercises.
- **Personalized Learning**: Choose between "Expert" and "Beginner" paths
- **Certification**: Earn a recognized certificate upon course completion.

SoBigData Academy

LIST OF COURSES

- **Basic Python**
- **Data Analysis**
- **Databases**
- **Data Theory and Society**
- **Legal and Ethical aspects of Data Science**
- **Data Mining & Machine Learning**
- **Information Retrieval**
- **Complex Network Analysis**
- Text Analytics
- Data Visualization and Storytelling
- Neural Network & Deep Learning
- Reinforcement Learning theory & practice

Those courses can be used by project and other training initiatives as **pre-requisite** (e.g. Masters) or as **training materials for teachers** (e.g. Generali internal training)

SOBIGDATA
RESEARCH INFRASTRUCTURE

Engaging stakeholders

The Role of Stakeholder Engagement in Research Infrastructures

Who are the stakeholders?
Stakeholders include academics, industry partners, policymakers, end users, and civil society.

As **European research infrastructure**, the success depends on collaborative efforts of our partners and users

- Diverse perspectives lead to creative problem solving and novel research ideas.
- Aligns research with societal needs, ensuring that outcomes have practical applications.
- Open dialogue builds credibility and fosters long-term support from all parties.

Strategies for Effective Stakeholder Engagement

MAP: Identify key stakeholders across sectors (academia, industry, government, NGOs) and their expertise.

ATTRACT: Organize *workshops*, roundtable discussions, participate to conferences and establish regular updates through *magazine* and webinars. Provide *useful tools* for society.

INVOLVE: Leverages stakeholder insights to propose policies and solutions. *Create joint projects*

SOBIGDATA.it
ITALIAN RESEARCH INFRASTRUCTURE

Summer School 2025:
From Data to Social Innovation

22-28 June 2025 – Baratti (Piombino) – TUSCANY (Italy)

https://summerschool2025.sobigdata.eu/

Coordinator: Roberto Trasarti
L'Aquila node coordinator: Antinisca Di Marco (UnivAQ)
Management team: Valerio Grossi and Michela Natilli
Communication Manager: Daniele Fadda
Institute of Information Science and Technologies (ISTI),
National Research Council (CNR), Pisa, Italy

SoBigData.eu receives funding from the European Union's grant agreements No. 654024, 871042 and 101079043.
Is also funded by the National Recovery and Resilient Programme for the Central Hub: "SoBigData.it"

Topics

The Summer School provides a unique **interdisciplinary mixture** where participants can explore state-of-the-art tools in **data analysis, machine learning, and artificial intelligence**. The school will have lessons from experts during the morning and will leave the afternoon to **group work guided by dedicated tutors** helping them in defining and developing their project which will be evaluated at the end of the school by a panel of experts.

1. European research framework

Helping the new generation of researchers in understanding the new directive and policies about data access and open science principles.

2. Information Dynamics

Methodologies to understand how (dis)information is generated, transferred, transformed, and utilized within a complex system over time.

3. Politic Dynamics

A focus on the political discussion and the interactions between individuals, institutions, ideologies, and external factors that drive political opinion.

4. Social Dynamics

Understanding behavioral patterns of society over time, how individuals and institutions influence each other, how social structures evolve, and how collective behaviors emerge.

Venue

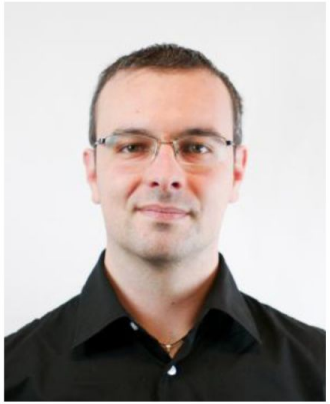
Nested in the gulf of Baratti on the “Etruscan Coast” in Tuscany (Italy), **Poggio all’Agnello** is a fully equipped resort that will host the 2025 edition of the SoBigData Summer school.

The Gulf of Baratti hosts the **Archaeological Park** of Baratti and Populonia , with a necropolis and acropolis.

Moreover, the venue is located on the **Road of Wine** and Oil, an itinerary that comprises some of Tuscany’s most refined wineries and oil productors.

Keynote Speakers

Keynote Speakers



Giulio Rossetti

Institute of Information Science and Technologies (ISTI) of the National Research Council of Italy (CNR)

János Kertész

Eötvös University Budapest

Daniele Quercia

Politecnico di Torino - Nokia Bell Labs Cambridge

Jussara Marques de Almeida

Universidade Federal de Minas Gerais

Full program available on the website: **https://summerschool2025.sobigdata.eu/schedule.html**

Registration

Registration for the SoBigData Summer School 2025

The cost of the summer school is **850€**

The fee is valid until the **30 April 2025 (early registration).**

Registration costs **€250** more between **1 May 2025 and 31 May 2025 (late registration).**

The registration fee comprises:

- **Attendance** to all summer school talks and sessions
- **Meals** (breakfast, lunch, dinner), and coffee breaks for all the duration of the school
- **Lodging (3-4 persons per apartment)**
- Participation to **three social events**
- **Attendance certificate**

Once you have filled and successfully sent this form, you will be contacted by the Summer School's organization in order to finalize the fee payment. You will be officially enrolled once your payment has been validated by the School's administration.

https://summerschool2025.sobigdata.eu/

Example 1: Diversity & Inclusion

seminar, events and tools

Activities carried out and tools developed to study and promote diversity and inclusion in Data Science.

Colorful Seminars series: Enhancing Diversity and Inclusion

- UNIPI initiative to boost **diversity representation**, bridging equity gaps for marginalized groups
- Invite diverse international experts to conduct seminars on SoBigData.it topics, fostering inclusive participation in computer and data science events.
- The seminars are hosted by the Department of Computer Science (also with streaming) or exclusively online

PinKamP 2024

Completely free UNIVAQ initiative for girls passionate about digital tech, exploring computer science, information engineering, and math.

Target Audience:

- Creative and motivated girls
- Interested in digital technologies
- Eager to explore computer science, information engineering, and mathematics

Project Purpose:

- Introduce girls to the disciplines of the digital society
- Overcome gender stereotypes
- Remove barriers and prejudices
- Showcase women's role in shaping future tech via creativity and problem-solving.

Women career in the Italian university

Analysis of the gender equality among researchers in the University panorama in Italy

- Time series on the research staff of the Universities over time disaggregated by positions.
- Percentage distribution of the research staff according to their role and gender among years

  ⇒ **Leaking Pipeline**

CINECA

Using the data from the CINECA database of research personnel we constructed a simple dashboard for helping in visualizing the Italian situation.

Cerca Università

Dashboard

--> LINK TO THE DASHBOARD

Designed by Eleonora Cappuccio, Daniele Fadda & Michela Natilli

Seleziona un ateneo

PISA

Seleziona una Macrosettore

01 Scienze matematiche e infor…

Seleziona un anno

2021

Uomini	Donne
106	29

SOBIGDATA.it
ITALIAN RESEARCH INFRASTRUCTURE

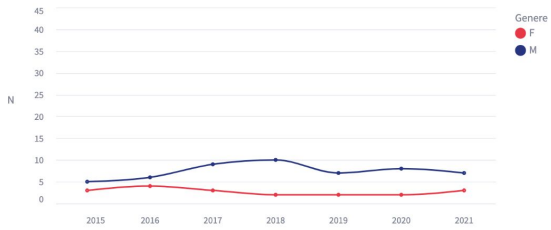
Finanziato dalla UE - NextGenerationEU - Prot. IR0000013 - Avviso n. 3264 del 28/12/2021 - Progetto PNRR: SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics

Gender gap nell'università italiana

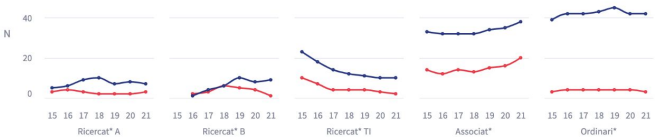
Seleziona una fascia

- Ricercat* A
- Ricercat* B
- Ricercat* TI
- Associat*
- Ordinari*

Numero di **Ricercat* A** in **Scienze matematiche e informatiche** presso **PISA**

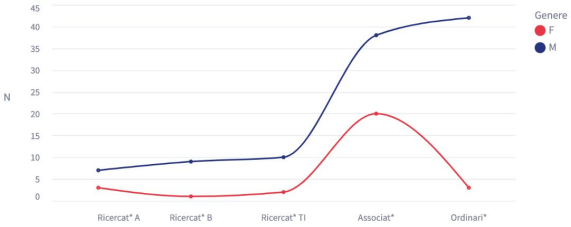


Fascia



Leaking Pipeline

Scienze matematiche e informatiche presso **PISA** nel **2021**



Leaking pipeline, letteralmente *tubo che perde*, è la tendenza generale a perdere consistenti presenze femminili lungo il percorso della carriera scientifica e tecnologica.

Data source

Example 2: Magazine

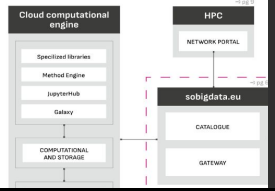
A way to inform and update the community periodically.

http://sobigdata.eu/magazine

01

What is SoBigData?

A distributed, Pan-European, multi-d
research infrastructure aimed at



EVENTS

CONFERENCES AND SCHOOLS 03

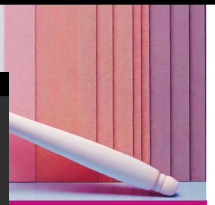
Future events

We have outlined a series of events that delve into "the future of Big Data and Data Science", including various summer schools exploring the role of data science and the societal impact of AI.

JUNE 16–22
Empowering Data for Social Good
Summer School to be held in Baratti, Tuscany

JUNE 23–29
AI & Society 2024 summer school
PhD in AI Summer School to be held in Capo Vaticano (Vibo Valentia), Calabria

JULY
Lipari School on Computational



06 Research Spaces
PAGE 28
An overview of the eight research areas in which SoBigData's work is focused on

07 Research Highlights
PAGE 32
Six of the most recent scientific outputs connected to the RI

Editorial

Welcome to the 10th edition of the SoBigData Magazine, a milestone that marks our journey in disseminating knowledge within the research infrastructure.

This edition is a turning point for the SoBigData Magazine. This is the first issue open to a wider audience beyond the SoBigData consortium, shifting from its traditional role of internal communication instrument to a communication tool open to a wider audience with the aim of fostering connections between academia, public sector, and industry for a mutually beneficial exchange of knowledge and innovation.

As a distributed infrastructure, SoBigData is composed of several nodes around Europe. We would like to use this and the next editions to present each node and illustrate the unique contributions it brings to the research infrastructure. In this edition, we take a closer look at the Italian node, the central hub of the infrastructure, giving you a peek into the ecosystem that drives our research initiatives.

Moreover, in each edition, we will focus on a specific theme: for this edition, our focus is the European legislative landscape regulating the digital world and how various

by Katia Genovali

KNOWLEDGE TRANSFER

A data-driven future for European business

The case of OCTO-SoBigData collaboration on smart mobility

05

Since its launch in 2015, our research infrastructure dedicated to data analysis and social mining has sparked numerous collaborations with industry partners in a dynamic exchange of expertise between public research and business sectors.

In this renewed edition of the SoBigData Magazine, we will explore in depth one of our the most established collaborations, featuring an exclusive interview with Tina Martino, the Head of Marketing at OCTO Telematics, and Mirco Nanni, Head of KDD Lab at CNR-ISTI.

We present how the partnership between OCTO Telematics, a company operating in the smart mobility industry, and SoBigData is shaping the future of data-driven innovation.

The two worlds of OCTO Telematics and the "Sustainable Cities for Citizens" SoBigData Research Space found a way to match their respective goals and objectives combining innovative 'data driven' business models enabled by the huge amount of OCTO mobility data and many competencies to analyse and extract knowledge and value from them.

"Our collaboration with OCTO Telematics started around 15 years ago", Nanni says. "Our research group was leading a pioneering project on mobility data analysis (GeoPKDD), and OCTO showed immediate interest in novel ways to enhance their data use beyond the Insurance Market, which was their core business."

By designing and developing several R&D projects over the years, OCTO and KDD Lab worked on projects through multiple engagement models to develop

Tina Martino, Head of Strategic Marketing and Intelligence at Octo Telematics

Mirco Nan
CNR-ISTI in

novel data-driven services in mobility and sectors such as Insurance, Telco operators (Italian and foreign), energy providers, nationwide retail sellers, car makers and urban mobility policy managers. Thus, this collaboration perfectly showcases how applied research impacts society by sharing and transferring competencies and technologies.

Results and impacts of a successful partnership

"Big Data is a game changer for the mobility transformation. Patterns and trends can be identified analysing large datasets, and novel insights create a valuable layer of shared knowledge, enabling more efficient resource allocation and easing informed decision-making across various sectors and stakeholders, i.e., car makers, policymakers, fleet managers from which citizens benefit first and foremost", Martino says. One of the most noticeable results obtained by the partnership has been the publication in the prestigious scientific journal Nature Sustainability, offering valuable insights into non-trivial behaviours. The study, carried out in 2022 by Mirco Nanni and colleagues based on OCTO data, highlighted that the primary pollution from city traffic results not as much from the large number of cars in circulation but from a few highly polluting vehicles: "Having a few extreme cases in any population can be considered normal, yet it is not trivial to understand when their impact is significant. Access to real data makes it possible to quantify that impact, and thus provide better insights to the domain experts and decision-makers."

The application of these

SECTION 09

Experience TNA

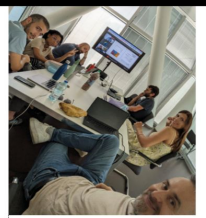
A Memorable Journey into Mobility Research: Leo Ferres Experience at Pisa CNR

Leo Ferres, a Computer Science Professor at IDS UDD/Telefónica/ISI Foundation, embarked on a 7-week visit to CNR in Pisa, exploring mobility research's potential. Engaging in discussions, new projects, datasets, and Pisa's culture defined Ferres' enriching experience. Welcomed warmly by Dr. Luca Pappalardo and his team, the visit promised valuable contributions to the Mobility field.

This is an extract of his experience; the complete article can be read on the SoBigData website.

Exploring mobility: meetings that sparked innovation

Throughout my visit, I had numerous meetings with experts from various disciplines, all centered around the theme of mobility. The discussions were inspiring, and witnessing how each researcher brought their unique perspective to the table was fascinating. These meetings acted as catalysts, giving rise to two novel projects addressing real-world mobility challenges.



Traffic and urban mobility through data-driven solutions

Two projects aimed at enhancing urban mobility were undertaken. The first utilized public transportation, ride-sharing GPS data, and anonymized mobile phone data to analyze traffic patterns and optimize transit routes. Its potential to reduce congestion and enhance efficiency was significant. The second project delved into the correlation between traffic and human mobility, using mobile phone datasets to understand how traffic influences daily routines. Insights gleaned could aid smarter traffic management, improving urban life quality.

Discovering Pisa's delights

In Pisa, my time spanned beyond research. Captivated by the Leaning Tower, I explored historic streets and relished local cuisine, cherishing the flavors of Italy. Discovering Gusto Giusto's authentic scene, I joined musicians, fostering unity. Spontaneous park jam sessions united us, transcending research boundaries

JOIN OUR PROGRAM

TransNational Access Grant

Transnational access (TNA) is an opportunity for researchers and professionals to carry forward their projects as visitors of the SoBigData Research Infrastructure.

WHY YOU SHOULD CONSIDER APPLYING

Through TNA, researchers and professionals gain access to extensive computing platforms, social data resources, and cutting-edge computational methods within selected Exploratories.

This opportunity facilitates multi-disciplinary social mining experiments using SoBigData's assets: vast datasets, analytical tools, services, and expertise.

TNA participants have a range of opportunities:

1. Interacting with local experts.
2. Engaging in discussions on research queries
3. Conducting experiments using non-public big social datasets and algorithms.
4. Presenting their findings at workshops or seminars.

WHO CAN APPLY

Applications are open to individuals with scientific interests, professionals, startups, and innovators seeking benefits from data science and social media analytics training.

WHAT'S PROVIDED

Participants can access funding of up to €5,000 covering expenses for daily subsistence, accommodation, and economy travel, enabling them to fully immerse themselves in this collaborative research environment.

More info on sobigdata.eu/calls/

Example 3: Challenge Us

An opportunity offered by SoBigData to companies who want to exploit the potential of data in their business.

Developing proof-of-concept to evaluate and understand the potentialities hidden in their (big) data.

Challenge 1:
Fashion design & luxury

- The company supports Retail and rands to operate in the B2B2C fashion design and luxury sector.
- The company has designed and developed a **data customer platform**
- The company would like to accompany the client with advice on omnichannel strategy and digital transformation

Business Questions

- The in cloud data customer platform supports customers profiling, manage relationship and all physical and digital sales processes.
- With this platform they collect thousand information about customer purchasing and their habits.
- They would like to:
  1. **Compare** physical store purchases with digital sales
  2. **Identify** sales trends across time, categories, and customer preferences.
  3. **Explore** customer demographics and geographic distributions
  4. **Analyze** profits and spending patterns

From **business** questions to **research** questions

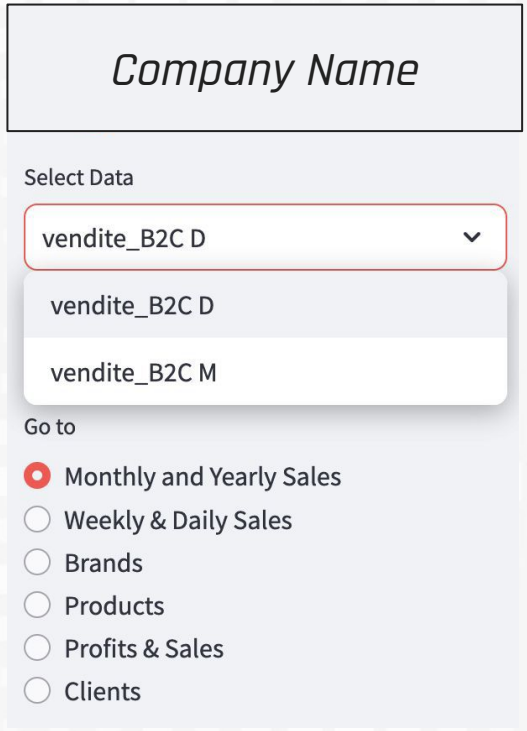
Task: **Data mining approach** to create an **interactive dashboard** for exploring and comparing selling trends and customer behavior across online and physical stores

Main Steps:

- Preliminary data analysis, cleaning, and standardization;
- Mapping market questions (e.g., sales by time/day, top items/brands) to data mining goals;
- Extracting key variables of interest;
- Dashboard creation and visualization assessment.

Interactive dashboard: Characteristics

- **Modular** dashboard structure for easy update;
- Dedicated interactive visualizations for each market question;
- Dynamic data selection (e.g., physical vs. online sales) with automatic linkage to related customer data;
- Six thematic pages:
  - Monthly and Yearly Sales,
  - Weekly & Daily Sales,
  - Brands,
  - Products,
  - Profits & Sales,
  - Clients.



Conclusion

- **Sales Insights**: Interactive dashboard analyzes trends, top products, and customer behaviors.
- **Flexibility**: Modular design for easy future updates.
- **AI-Ready Insights**: Helps identify key data inputs to inform Company Management on future strategy.