

# SoBigData RI Data & AI Services for Digital Sciences

---

GIULIO ROSSETTI - CNR

[giulio.rossetti@isti.cnr.it](mailto:giulio.rossetti@isti.cnr.it)

SoBigData RI [INFO@SOBIGDATA.EU](mailto:INFO@SOBIGDATA.EU)



Funded by  
the European Union

DIGITAfrica project has received funding from the EU Horizon Europe research and innovation Programme and Switzerland under Grant Agreement No. 101187966. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

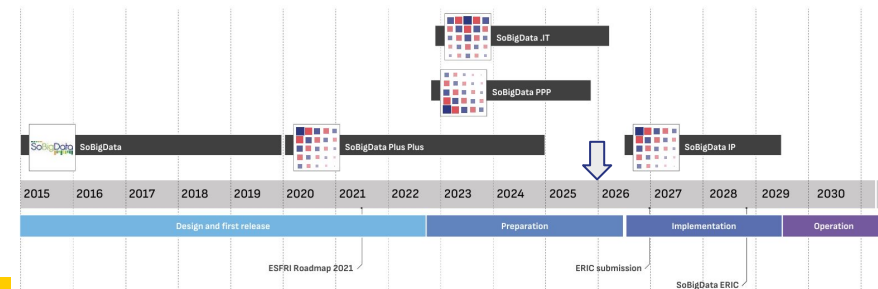
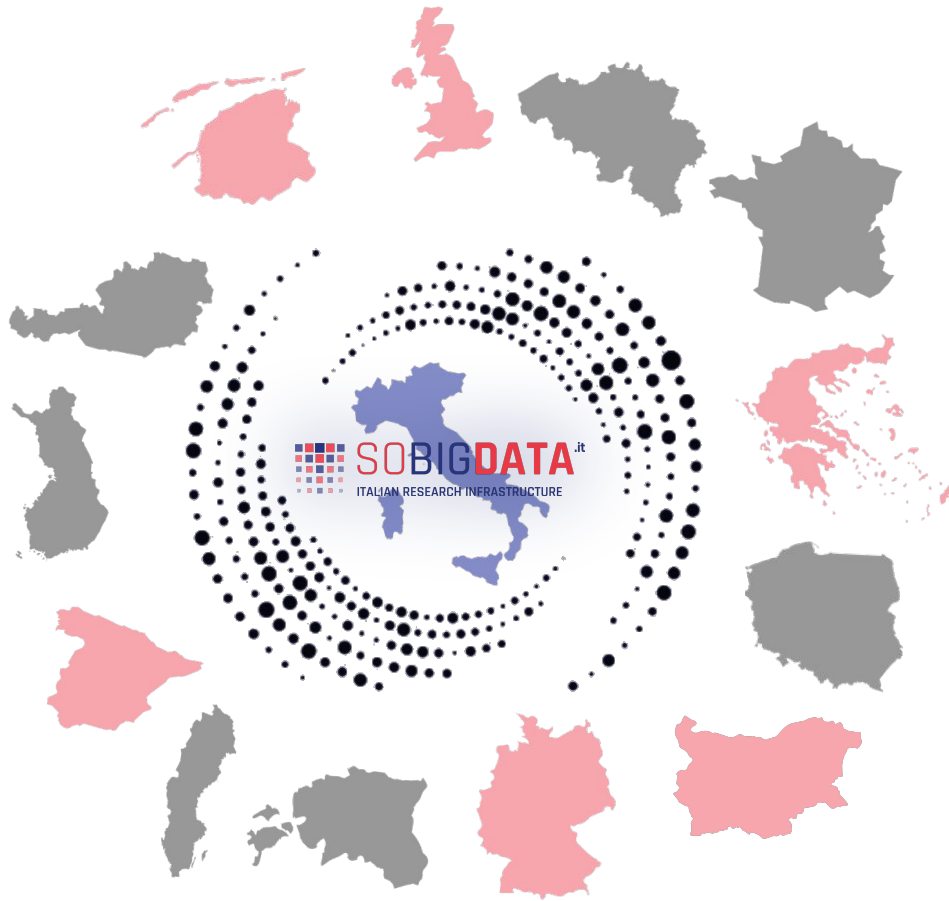




# SoBigData RI



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 on ethics, legality, and open science.



# SoBigData RI Overview

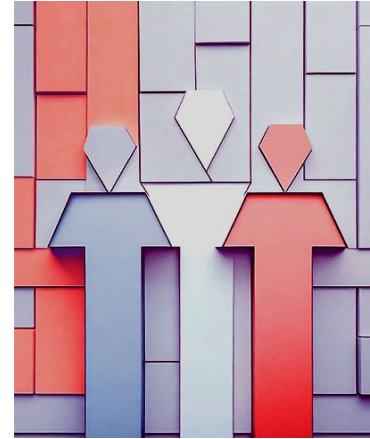


## INFRASTRUCTURE



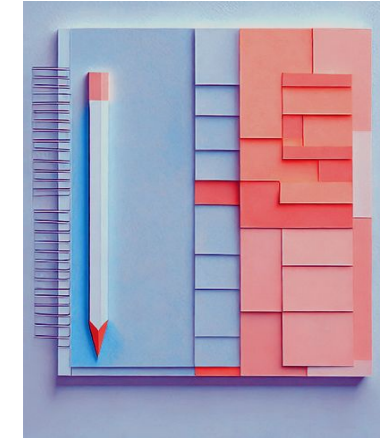
- **Computational Power**
- Open Science tools
- Resources
- Datasets
- Methods
- Technologies
- Training

## COMMUNITY BUILDING



- **Researcher communities**
- Industries and PA involvement
- Connecting side projects
- Connecting to other realities

## GUIDELINES



- **Ethical aspects**
- **Legal Aspects**
- Open Science principles

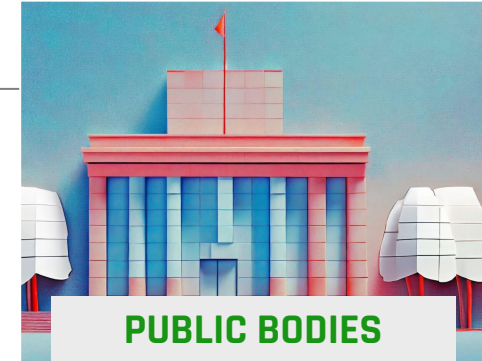
# SoBigData RI Activities



- Creating a multi-disciplinary scientific **community**
- **Boosting** research quality among the nodes
- Increasing the opportunities in **EU partnerships** and calls
- Embracing the European ethical and legal principles



- Promoting the **technological transfer**
- Offering **consulting** and data analysis **services**
- Tailored **training** courses
- Facilitating the creation of new **business opportunities**



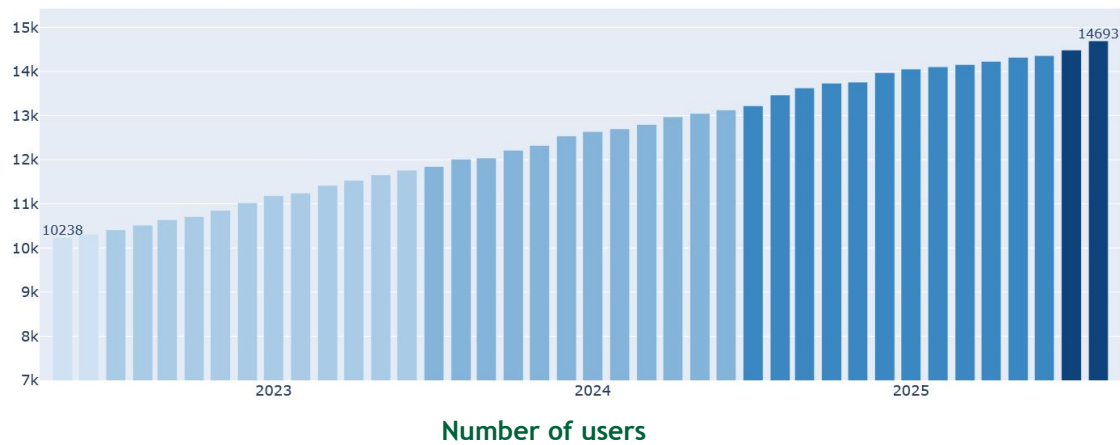
- Giving **value to public data** to understand the territory
- Creating **indicators** for policy makers
- Designing new **services for citizens**

# 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.



# Research Spaces

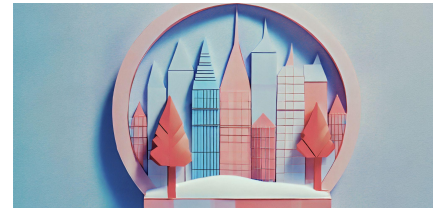
Vertical contexts fostering tangible progress toward societal challenges.



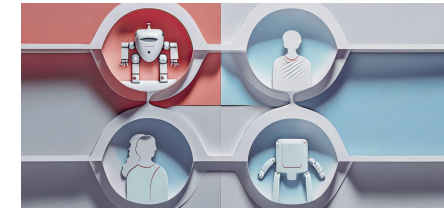
Societal Debates  
and Misinformation  
Analysis



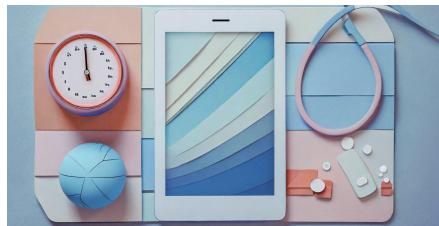
Demography, Economy  
and Finance 2.0



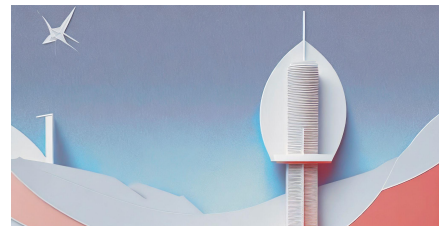
Sustainable  
Cities for Citizens



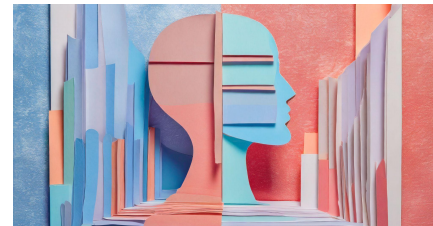
Social Impact of AI  
and explainable machine  
learning



Health Studies



Societal and Industrial  
Impact of Next-Gen.  
Internet & beyond 5G  
Networks



Pervasive Intelligence in  
Cyber-Physical Systems  
for Future Society

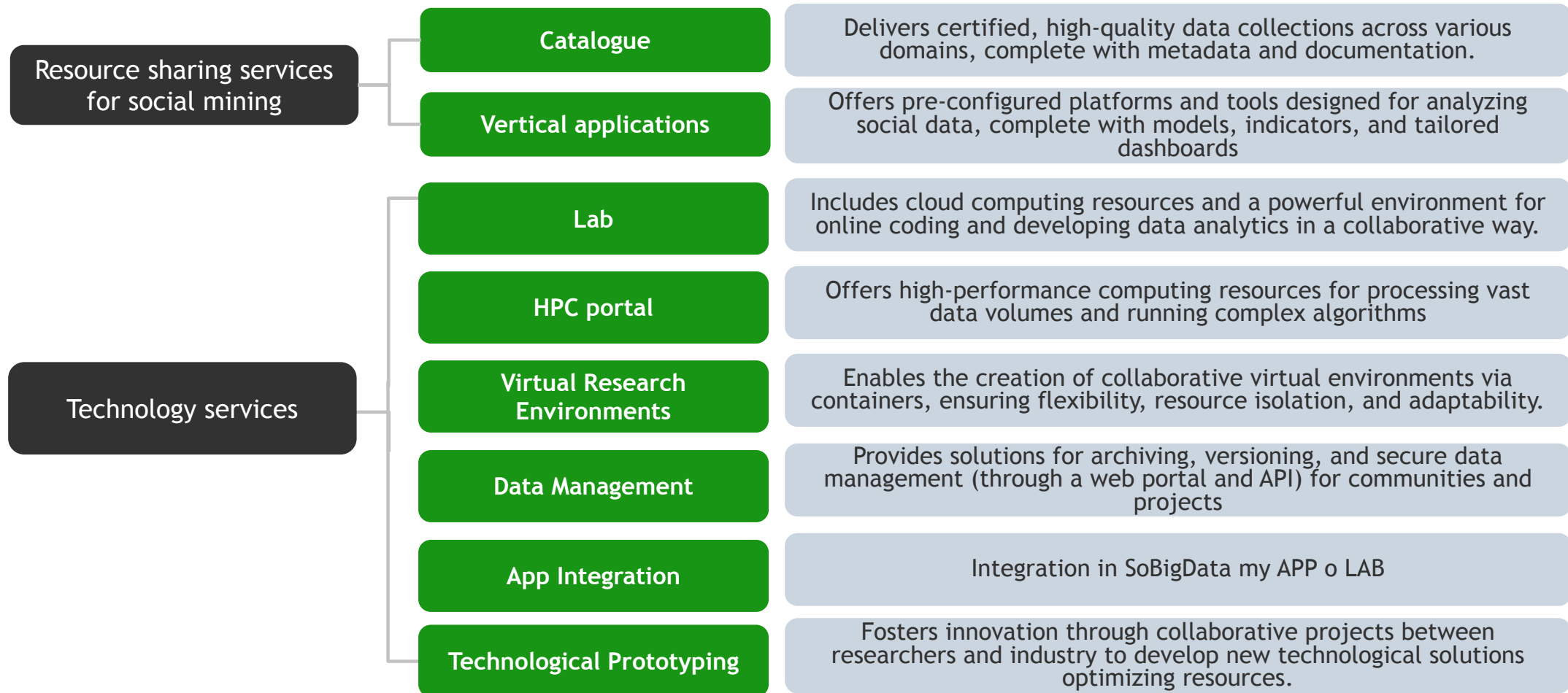


Disaster response  
and recovery

More info at: [https://www.sobigdata.eu/research\\_spaces](https://www.sobigdata.eu/research_spaces)



# SoBigData services portfolio





# SoBigData services portfolio





# SoBigData Catalogue

Resource sharing services  
for social mining

The Catalogue delivers a complete list of the resources offered by the RI:

- Data Collections
- Methods and Applications
- Experiments
- Other Resources

They can be browsed by various tags and filters for an easy discovery. All the resources are described by metadata.

They follow the **FAIR principle** and an open science approach. Anyway some resources are available on line some others are accessible upon request to the data owner for IP reasons.

The screenshot displays the SoBigData Catalogue interface. At the top, there is a search bar and navigation tabs for 'Dataset', 'Method', 'JournalArticle', and 'Experiment'. Below this, a section titled 'Browse the SoBigData now!' features a grid of resource cards, including 'SoBigData Services and Products' and 'SoBigData statistics'. A world map filter is visible, along with a list of 'Organisations' and 'Types'. The main content area shows a detailed view for the resource 'y/Politics 1k', which is a social simulation dataset. It includes a description, author information, a QR code, and a table of 'Personal Data Attributes'. The table lists attributes such as 'ChildrenData', 'Personal Data', and 'Personal data was manifestly made public by the data subject'. A 'Followers' section shows 0 followers and 0 items. The 'Additional Info' section provides details on accessibility, availability, and creation date.

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

Attribute	Value
Accessibility	Both
Accessability Mode	Download
Availability	On-line
Distribution	Distribution
Creation Date	2024-09-13 10:50
Creator	Rossini, Giulio, gullorossini@lists.cnr.it
Dataset Citation	Giulio Rossini, y/Politics 1k dataset, 2024
Dataset Use Safeguards	Name
Fields/Scope of Use	Any use
Group	Societal Debates and Misinformation

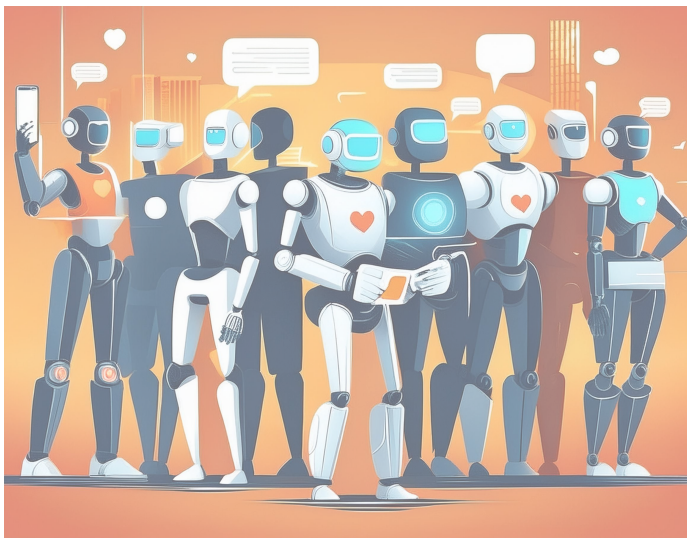


# Example of Vertical Application: Y Social

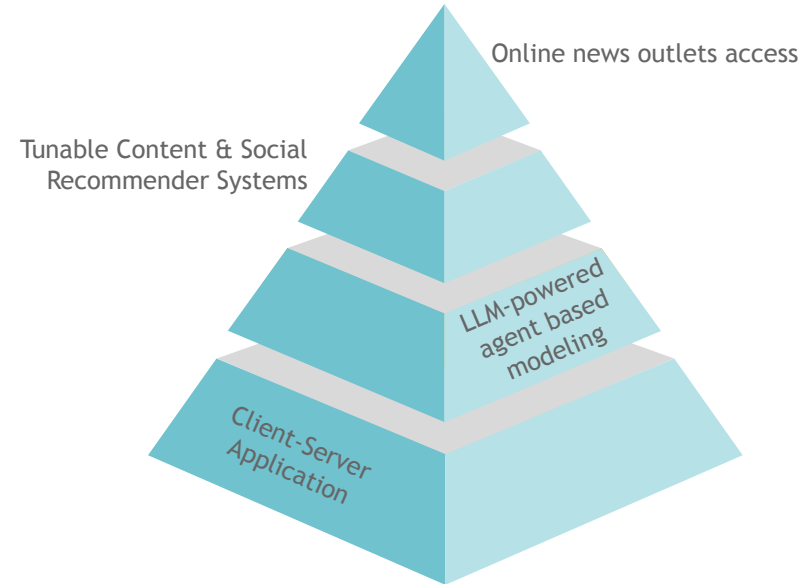
Vertical Application

## Digital Twin?

“[...] a set of **adaptive models** that **emulate** the behaviour of a real system in a virtual one getting **real time data** to update itself along its life cycle”



An **online social platform replica** that allows to design **realistic social simulations** in a **controlled environment**

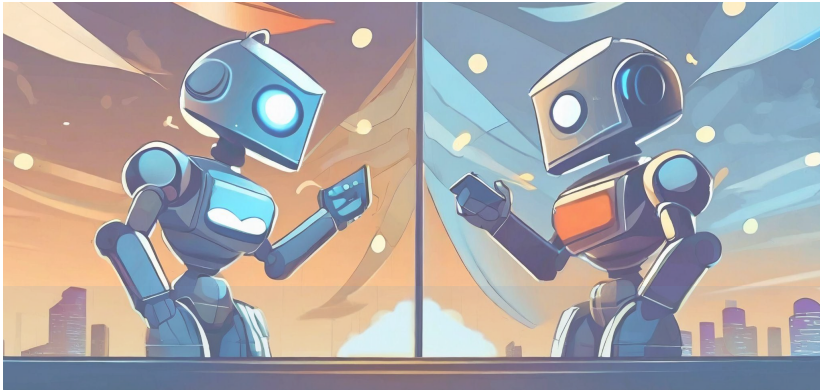


Digital Twin based  
**Social Simulations on Demand!**



# y/Politics - Simulating political debates

Vertical Application



## Political Debates

Topics: Reddit r/politics extracted topics  
Political leanings: 9 political alignments



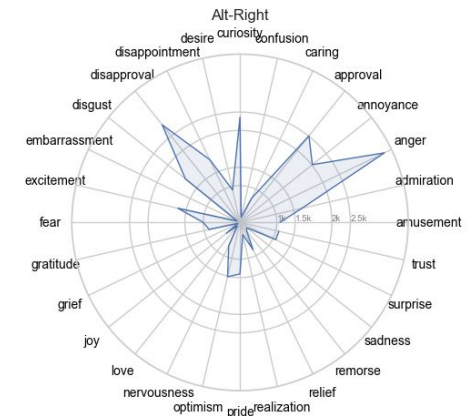
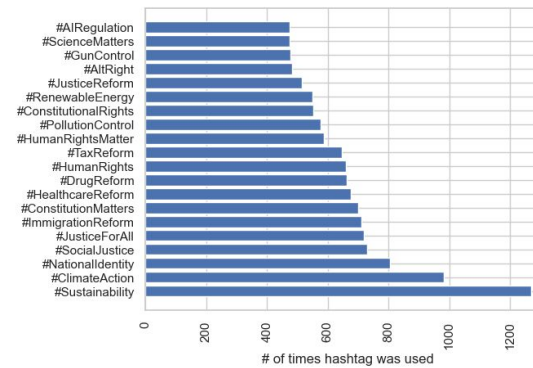
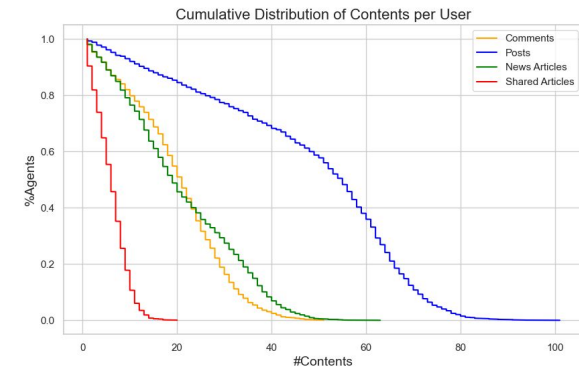
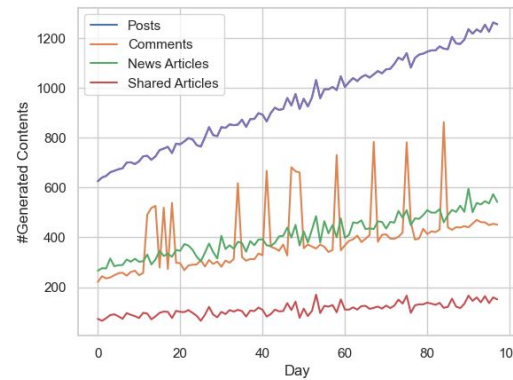
## RecSys

Contents. Popularity & Follower  
Social. Preferential Attachment



## Simulation parameters

1k starting agents  
Activity rates fitted on BlueSky  
100 days of simulation





# SoBigData Lab

Technology services

Deliver a set of methodologies and algorithms integrated in the cloud environment for developing new data analytics processes.

The SoBigData Lab supports experimentation, prototyping, and validation of AI big data applications offering a list of tools libraries for social mining.

The screenshot shows the SoBigData Lab interface. On the left, a 'Methods List' is visible with a red box around the 'SimpleImageClassifier' method, labeled 'Select'. The main area shows the configuration for 'SimpleImageClassifier', with a blue box around the 'Inputs' and 'Options' sections, labeled 'Configure'. On the right, the 'Execution Monitor' shows a list of executions, with a blue box around the execution details, labeled 'Execute'.

Method invocations in SoBigData



The screenshot shows a file explorer window with a list of files and folders. A blue box labeled 'Storage' is overlaid on the table.

Name	Owner	Type	Last Update	Size
deliverables	Valerio Grossi	Folder	05 Dec 03:23 PM 2016	
city_of_citizens	Valerio Grossi	Public Folder	05 Dec 03:25 PM 2016	
well-being_and_economy	Valerio Grossi	Folder	05 Dec 03:26 PM 2016	
societal_debates	Valerio Grossi	Folder	05 Dec 03:26 PM 2016	
migration_studies	Valerio Grossi	Folder	05 Dec 03:26 PM 2016	
soccer_events	Valerio Grossi	File	05 Dec 08:22 PM 2017	6.7 MB
Experiment sheet	Valerio Grossi	Application/vnd.openxmlformats-offic	14 Feb 12:03 PM 2018	701.5 kB
training_material	Valerio Grossi	Folder	13 Jun 02:08 PM 2018	
experiments	Valerio Grossi	Folder	30 Oct 02:30 PM 2018	
google9f90d4e4f51a	mhmedhasaneen	text/html	11 Oct 04:48 PM 2019	53 bytes
da.jpg	mhmedhasaneen	image/png	11 Oct 04:49 PM 2019	14.3 kB
arous-v-beirut-29-cma.html	mhmedhasaneen	text/html	11 Oct 04:56 PM 2019	29.6 kB



# HPC portal

Technology services

SoBigData provides a way of submitting **proposal** for accessing SoBigData federated HPC resources and run **Big Data experiments** requiring specialised systems (Storage, CPUs, GPUs, Networks)

The proposal are evaluated by the **Board of Ethics and Law** and the **Scientific committee** then allocated by priority and availability.





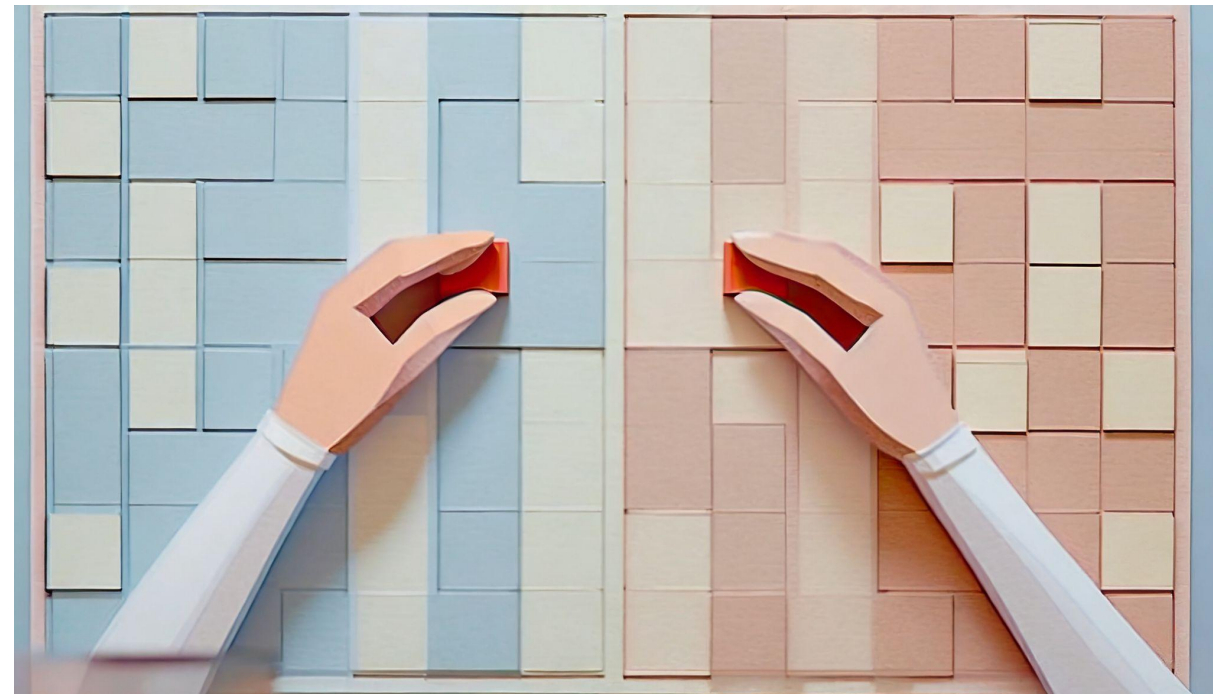
# Virtual Research Environment

---

They implement a **virtual shared space** where users (everyone, a community, a company or a research project) share resources and tools.

They can be “temporary” or “permanent”. They have a shared workspace and all the **services offered by the RI can be specialised** and attached to it.

This virtual research environments are build with the community and are specialised for their needs.



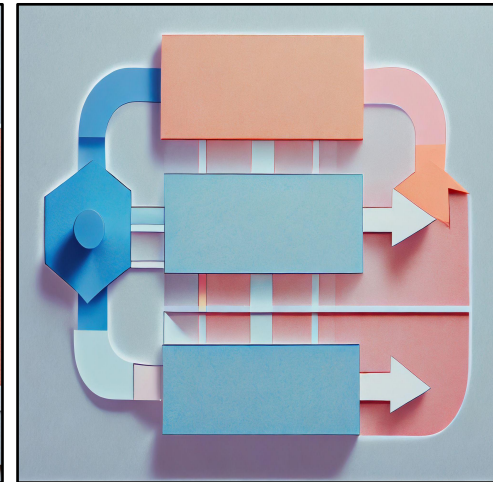
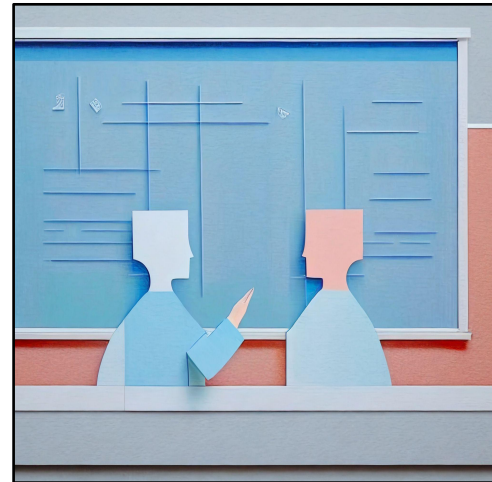
Technology services



Technology services

# Technological Prototyping

**Co-creation with companies and policy makers to develop proof-of-concepts and prototypes giving values to their data.**



**What i can do with my data?**

Can I create a new service with my data?

**How can i use my data to optimize my activity?**

I have an Idea, is it ethical/legal to use my data in this way?

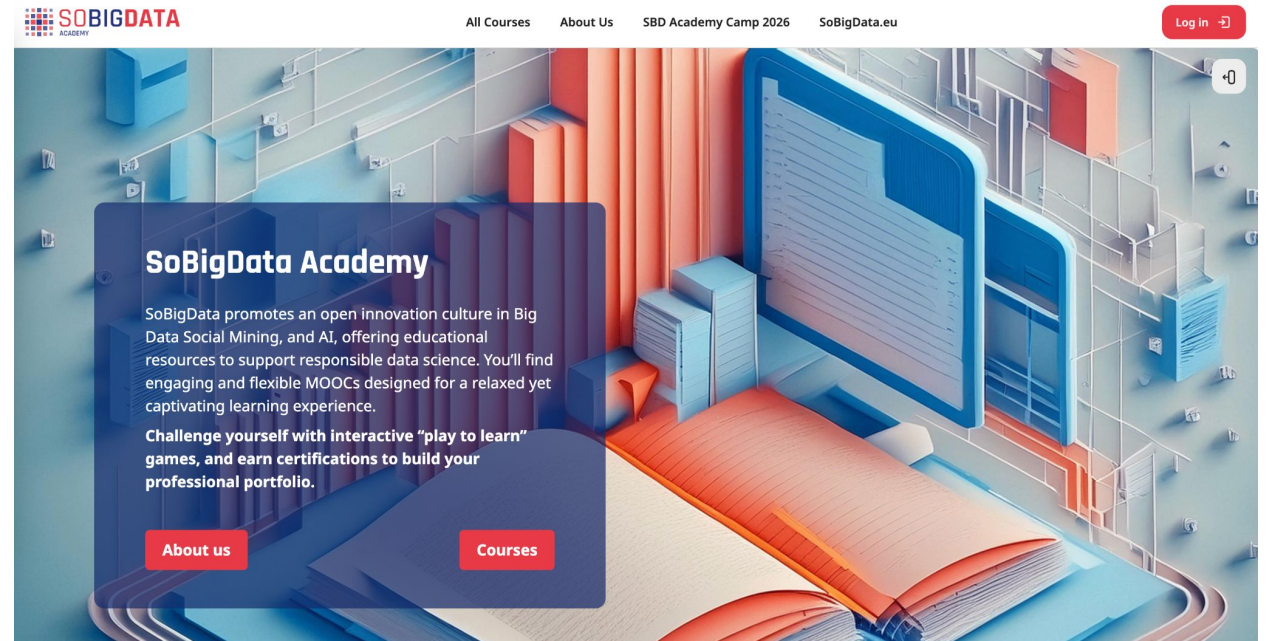


# SoBigData Academy

Training

The training hub bringing together researchers, professors, and professionals in co-creating open online courses in Big Data, Social Mining and Artificial Intelligence.

**Not a simple repository of slides, but interactive courses with practical exercises, forums and badges.**



Synergy with:



MOOC  
Interactive  
Courses

First year  
750+ Students  
(1 year activity)

Students from  
33  
Countries



# SoBigData Academy

Training



**Basic Python**  
Difficulty level: 1  
Estimated time: 3h



**Databasa**  
Difficulty level: 1  
Estimated time: 1.5h



**Data Analysis**  
Difficulty level: 1  
Estimated time: 1.5h



**Legal and ethical aspects of Data Science**  
Difficulty level: 1  
Estimated time: 2.5h



**Fairness in AI Recruitment**  
To access the course, use these credentials:  
user: guest\_user  
pwd: Gons123e



**Big Data Storage**  
Difficulty level: 1



**Container Based Virtualisation: Docker/Swarm**



**Container Based Virtualisation: Kubernetes**



**Information Retrieval**  
Difficulty level: 2  
Estimated time: 1.5h



**Data Mining & Machine Learning**  
Difficulty level: 2  
Estimated time: 7h



**Data Analysis with Spark**  
Difficulty level: 2  
Estimated time: 4h



**Data Theory and Society**  
Difficulty level: 1  
Estimated time: 2.5h



**Data Modelling, Data Formats & Protocols**



**Elasticsearch**



**BitHub**



**JSON**



**Complex Network Analysis**  
Difficulty level: 2  
Estimated time: 2h



**Neural Networks & Deep Learning**  
Difficulty level: 2  
Estimated time: 6h



**Reinforcement Learning**  
Difficulty level: 2  
Estimated time: 2.5h



**Business Models Data-Driven Innovation**  
Difficulty level: 1  
Estimated time: 1.5h



**XML**



**YAML**

Soon available:



**Data Visualization and Visual Analytics**  
Difficulty level: 1  
Estimated time: 1.5h



**Large Language Models**  
Difficulty level: 2  
Estimated time: 6h



**Explainable AI**  
Difficulty level: 2  
Estimated time: 2.5h



**Text Analytics**



**Computer Vision**



**Community detection with spectral techniques**



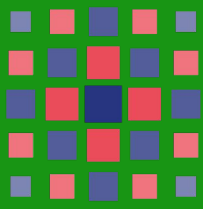
**Open Science Basics**



350+	<b>researchers and PhDs</b>
1 100+	<b>scientific publications</b>
14,500+	<b>RI users</b>
6,200+	<b>young researchers trained</b>

320+	<b>big datasets</b>
240+	<b>methods and apps</b>
2 mln+	<b>peak access to apps</b>
200+	<b>pilot projects</b>

*Updated on 01 April 2026*



SOBIGDATA

RESEARCH INFRASTRUCTURE



# Thank you

---

Giulio Rossetti - CNR - SoBigData RI

[giulio.rossetti@isti.cnr.it](mailto:giulio.rossetti@isti.cnr.it)

[info@sobigdata.eu](mailto:info@sobigdata.eu)