UNDERSTANDING THE COMPLEXITY OF OUR CONTEMPORARY GLOBALLY-INTERCONNECTED SOCIETY

Number 10
February 2024

SOBIGDATA

MAGAZINE

Content

DIRECTOR Roberto Trasarti ■ SCIENTIFIC MANAGER Michela Natilli ■ CHIEF EDITOR Daniele Fadda ■ EDITORIAL STAFF Marco Braghieri, Katia Genovali, Valerio Grossi, Michela Natilli, Beatrice Rapisarda ■ CONTRIBUTORS Riccardo Guidotti, Roberta Savella ■ DESIGN AND GRAPHIC LAYOUT Eleonora Cappuccio, Daniele Fadda ■ IMAGE PRODUCTION AND GRAPHIC APPARATUS Daniele Fadda, Laura Pollacci

All the illustrations are designed by Daniele Fadda with the support of Adobe Firefly.

SOBIGDATA Magazine *is not for sale but is distributed for purposes of study and research and published online at www.sobigdata.eu/newsletter*

Printed on 2 february 2024 11:05 a.m.

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 actors interact with data

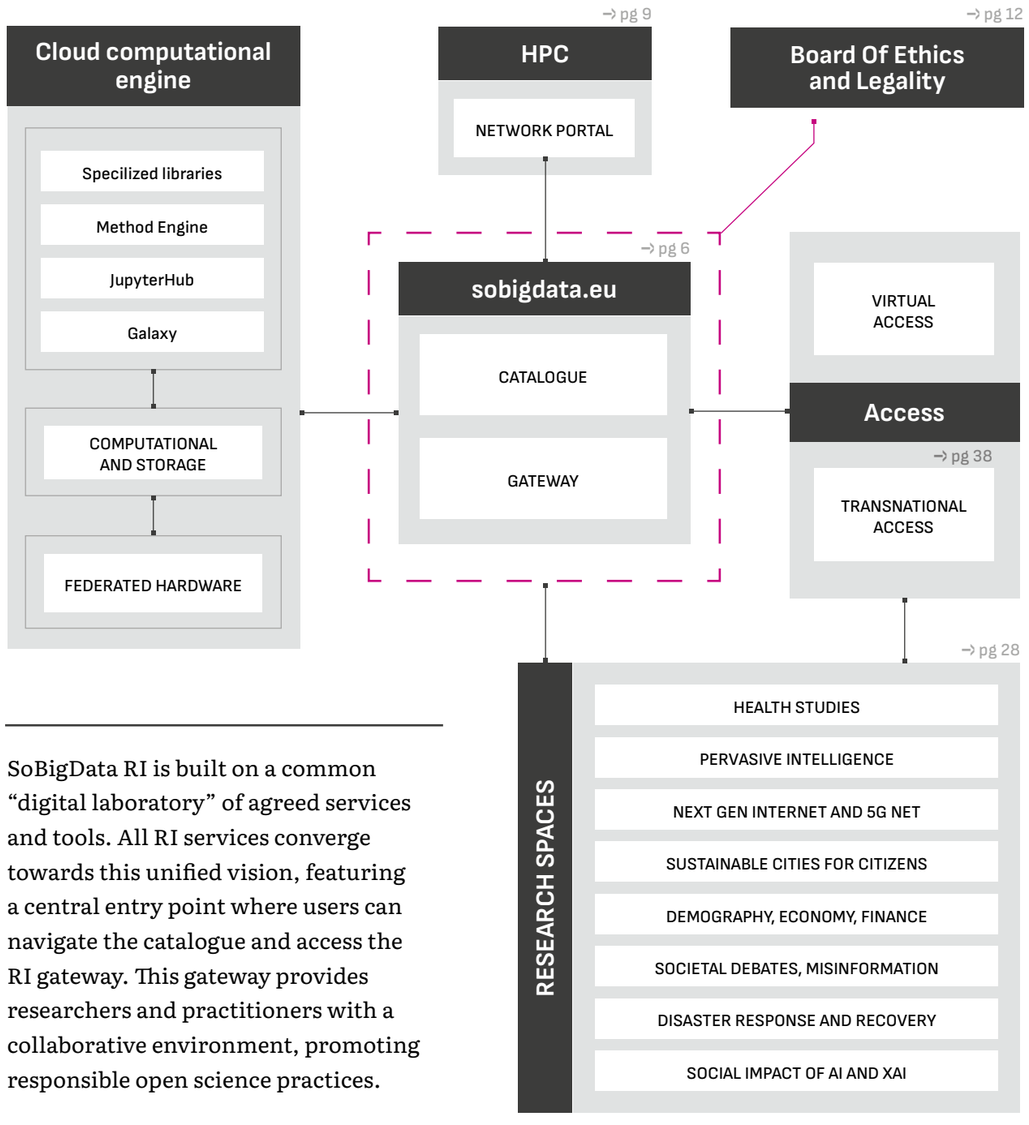
We are publishing this edition concurrently with the organization of an event called "Empowering Enterprises through BigData and AI – The European Data Strategy 2024", because we believe that establishing relations with the private sector, small and medium enterprises and large industries is a key factor in developing an impactful research infrastructure.

We're excited to grow with our readers and build a community that shares our enthusiasm for data-driven innovation.

Roberto Trasarti

SOBIGDATA RESEARCH INFRASTRUCTURE

01 What is SoBigData?



SoBigData RI is built on a common "digital laboratory" of agreed services and tools. All RI services converge towards this unified vision, featuring a central entry point where users can navigate the catalogue and access the RI gateway. This gateway provides researchers and practitioners with a collaborative environment, promoting responsible open science practices.

A distributed, Pan-European, multi-disciplinary research infrastructure aimed at using social mining and big data to understand the complexity of our contemporary, globally interconnected society

In today's complex society, data has become a strategic asset capable of influencing multiple aspects, including research, innovation, business strategy, operational efficiency, and creating services for citizens. The ability to collect, manage, and extract value from data has become a key skill for successful organisations. Leveraging data, organisations can make more informed, reality-based decisions and gain an in-depth understanding of society and its dynamics. This understanding is crucial for developing effective strategies.

Data also plays a crucial role in enabling organisations to take reality-based decisions. It provides information needed to gain insights into society and its complexities, using concepts like digital twins. Using digital twins, organisations can create virtual representations of real-world systems and use data to analyse and simulate various scenarios. This allows a deeper understanding of society's dynamics and enables organisations to make more informed decisions.

Furthermore, the synergy between research and industry is vital for fostering innovation, economic growth, and societal advancement. By bringing together researchers from various disciplines and industry experts, new ideas can be generated, leading to innovative solutions and products. This collaboration between research and industry drives economic growth by creating new opportunities and markets.

However, the need for transparency in private companies can pose several challenges and concerns. Stakeholders, including investors, employees, customers, and the general public, require transparency to ensure accountability and trust. Balancing the benefits of transparency with potential drawbacks is essential. The EU recognises the importance of transparency and is working on specific regulations to address this need. These regulations aim to strike a balance between transparency and protecting sensitive information, ensuring that private companies operate ethically and responsibly.

The SoBigData research infrastructure aims to leverage big data to understand the complexity of society. By providing no-profit services to researchers, industry, public bodies, and citizens, SoBigData aims to create a multidisciplinary scientific community that adheres to the EU's vision of ethical, legal, and open science. Through its goals and collaborations, SoBigData strives to become the reference infrastructure in Europe for big data analysis and social mining, supporting data scientists, AI researchers, enterprises, and policymakers in their data-driven endeavours.

What is Social Mining?

1. Social mining is a **toolbox** aimed at understanding and modelling **complex social phenomena** and their evolution.

2. It combines data-driven and model-driven analytic approaches (network science, data mining, and AI): *what is* and *what if*.

SoBigData's MISSION

To use big data and AI to understand the complexity of society and offer no-profit services to researchers, industry, public bodies, and citizens through the creation of a multidisciplinary scientific community according to the EU's vision of an ethical, legal, and open science

GOALS

By pursuing these goals, the SoBigData research infrastructure contributes to advancements in big data analysis, social mining, and the responsible use of data in various sectors.

1

SoBigData aims to become the reference infrastructure in Europe for big data analysis and social mining. It supports data scientists and AI researchers in designing and developing big data experiments in social science and innovative, cutting-edge research.

2

SoBigData strives to support enterprises in data exploitation for new products and services, new markets, cost reduction, and profitability. These efforts align with the objectives of the EU Strategy for Data.

3

SoBigData aims to be a reference point in the EU for public bodies and policymakers on data management, AI, and their ethical use. It directs its focus toward concrete solutions, facilitating responsible and ethical practices.

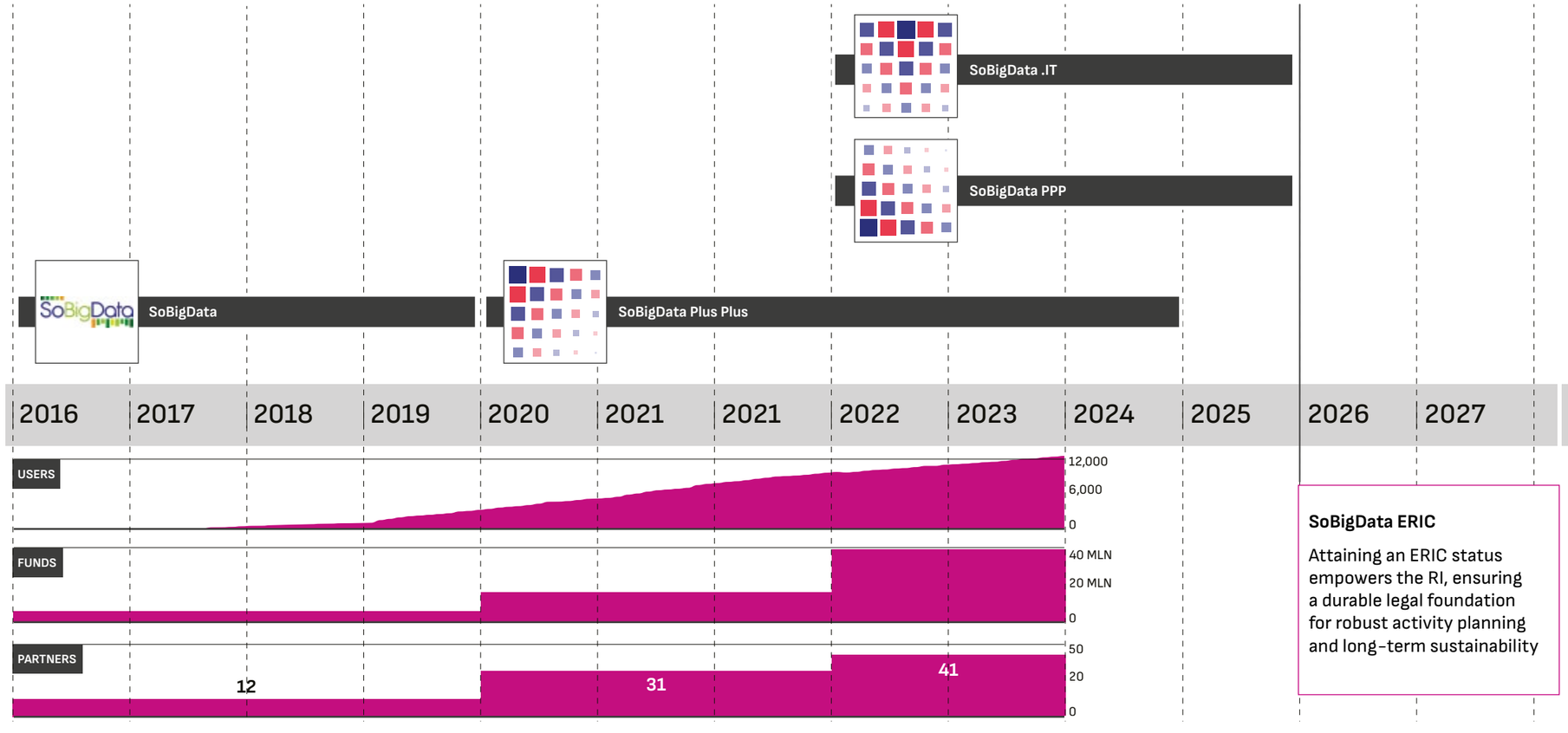
SoBigData History

The Research Infrastructure was born in 2015 with the EC H2020 SoBigData project (g.a. 654024) to integrate existing EU platforms on various social mining topics. Initially, the consortium was formed by academics and researchers from twelve institutions in seven countries. The first research infrastructure release came out in 2016, including some of the functionalities inherited by the initial platforms.

In 2020 the new project EC H2020 SoBigData++ started (g.a. 871042), enlarging the consortium to the experts in different fields collaborating during the first project. **Thirty one partners from 14 countries form the new consortium.**

2021 the SoBigData Research Infrastructure was selected to be part of the ESFRI roadmap as relevant digital infrastructure in the European Research Area (ERA).

In 2022 SoBigData was included in the top relevant infrastructure in Italy and won a call part of the National Recovery and Resilience Plan call (NRRP) to strengthen its capabilities.



SoBigData ERIC

Attaining an ERIC status empowers the RI, ensuring a durable legal foundation for robust activity planning and long-term sustainability

The SoBigData Research Infrastructure IN NUMBERS

Impact on scientific communication

900+	Scientific publications
5200+	Young researchers trained
120+	Researchers hosted
175+	Pilot projects

User community

12'000+	Research platform users
2mln	Peek access App
50+	Number of events
8.000+	Events Audience

Available Resources

220+	Datasets
200+	Methods and Apps

Updated on 8 January 2024

CONSORTIUM

SoBigData RI consists of 14 national nodes, involving 41 partners. The Infrastructure collaborates with over 250 researchers. For details on the institutions, please refer to the back cover.

14 National nodes

AT BE BG BE DE IT EE ES FR GR NL PL FI SE

41 Partners

250+ Researchers

HPC

High Performance Computing facilities

The HPC facility registry managed by SoBigData complements the SoBigData infrastructure for the management and execution of data-intensive and compute-intensive applications. While the SoBigData computing facilities are general-purpose systems that use virtualization and distributed computing to provide on-demand access to scalable and elastic resources over the internet, the HPC computing facilities exploit clusters of powerful processors, working in parallel, to process massive multi-dimensional datasets and solve complex problems at extremely high speeds.

This HPC facility registry is realized through an online portal where users can find detailed procedures for requesting the use of specific resources. The HPC Network Portal compiles technical information, allowing users to choose the appropriate computing facility for their tasks. It also provides administrative details, including contacts for access conditions and assistance in the access process. Selecting an institution to host the user's jobs is designed to be a straightforward self-service process. The access procedure is then facilitated by an institution mediator, guiding the user through the administrative process.

The HPC portal offers computational and storage facilities from eight institutions across Europe associated with the Research Infrastructure. For instance, users can request access to MareNostrum4, provided by Barcelona Super Computing, with its 165,888 CPU cores, 3,456 nodes, 390 terabytes of main memory. Alternatively, they may opt for a HADOOP cluster from Gottfried Wilhelm Leibniz Universitaet Hannover, featuring 40 nodes, 900 CPU cores, 2.4PB HDFS storage, and 6TB main memory. Overall, the portal facilitates access to a diverse range of computational nodes with varying features, accommodating researchers from different backgrounds and research communities, thus supporting a wide range of applications and research studies.

An overview of the European way to tackle systemic risks and the role of SoBigData

COVER STORY

Another FAIR digital world is possible

02

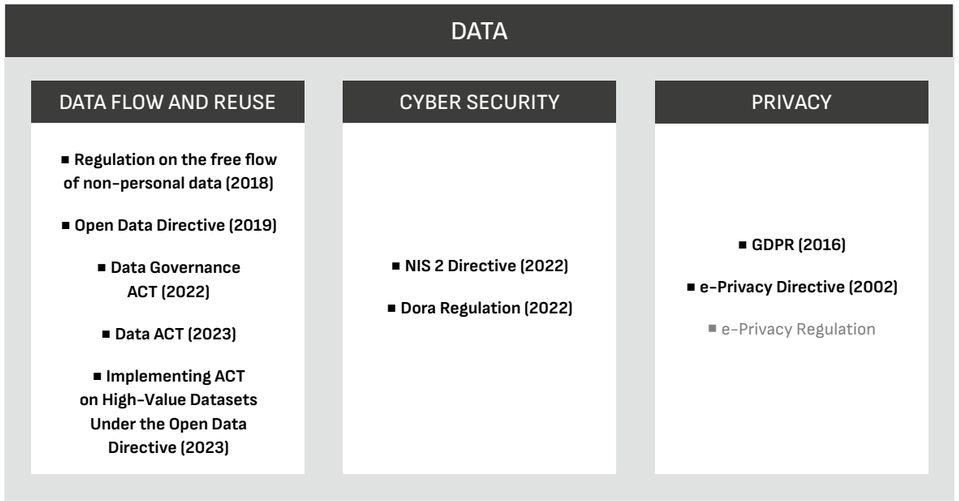
by Roberta Savella

Every researcher knows it: it's all about the data. They describe our world and ultimately shape it through new digital technologies, while in Western Countries people's existence grows *onlife*, as philosopher Luciano Floridi defines the experience of the inextricable blending between online and offline spheres in human life.

We understand the influence of online platforms and services on society, as we have learned with scandals such as *Cambridge Analytica*, when Big Data analysis was used to swing elections by targeting people with personalized political advertisements.

Therefore, we need data: information about trends and fake news, but also about online security, discrimination, and gender-based violence, just to name a few. Data gives us knowledge of the risks which develop in the digital world, and hence the power to reveal them and to suggest solutions when revealing is not enough to make them disappear.

European digital legislative framework | FOCUS



In this focus, a classification of the EU's digital legislative landscape is presented. Each act is colored differently depending on whether it is:

- **Approved**
- Under discussion
- Proposed

Ensuring Ethical and Legal Compliance: SoBigData's Board for Operational Ethics and Legality (BOEL)

SoBigData ensures an ethical and legal framework, facilitated by the Board for Operational Ethics and Legality (BOEL). BOEL offers an internal service that is able to address ethics and legal problems in sharing data and executing social experiments on them, guaranteeing the compliance with the applicable laws and regulations. In the meanwhile, it collects experiences in addressing problems generalizing the methodologies and making them effective in and for society.
Its primary functions include acting as a gateway for dataset publication, providing advice to researchers, and conducting ethical and legal evaluations for external projects upon request.

We need the online sphere to become a transparent environment, which allows researchers to understand its intricate mechanisms and risks

Accessing data to fight systemic risks: new instruments for researchers

The European Union has developed one of the most revolutionary regulations of the European Digital Strategy providing mechanisms which allow researchers access to the data of the *Very Large Online Platforms* (*VLOPs*) and *Very Large Online Search Engines* (*VLOSEs*). The Digital Services Act was in fact designed to tackle the systemic risks of the online world and foster a human-centered digital environment, shaped on the basis of the European values. We are talking about fighting the dissemination of illegal content, but also protecting fundamental rights such as human dignity, freedom of expression, privacy, non-discrimination, and ensuring public security and health, fair electoral processes and civil discourse, the protection of minors in the digital sphere. Thanks to the Digital Services Act, *VLOPs* and *VLOSEs* are now compelled to grant access to their data to **vetted researchers** who work on projects aimed at contributing to the detection, identification and understanding of systemic risks in the EU and at assessing the adequacy, efficiency and impacts of the risk mitigation measures adopted by intermediary services. The procedure is complex and involves the Digital Services Coordinator (a new authority responsible for all matters relating to supervision and enforcement of the Digital Services Act) of the Member State where the *VLOP* or *VLOSE* is established. This body is the one giving the status of "vetted researcher" when specific conditions are met, and for a determined research project aimed at tackling the systemic risks addressed by the Digital Services Act.

Moreover, for the same purposes, this Regulation compels providers to grant some non-vetted researchers (who meet some of the conditions established for the vetted ones, but not all of them) access to their data when it is publicly accessible in their online interface. This last provision allows for a more limited instrument which can contribute in the fight against the digital systemic risks without the need for researchers to go through the vetting process.

Building the path towards a common European Data Space

Aside from the novel access possibilities granted by the Digital Services Act regarding *VLOPs* and *VLOSEs*, if you work in the digital field, it is highly probable that you have recently come across the buzzword **data altruism**. And, regarding data altruism, the EU has implemented the new Data Governance Act. This is one of the pillars of the European Data Strategy and focuses on fostering data flow, enabling the reuse of certain public-sector data and the work of data intermediaries and registered data altruism organizations. Other aspects regarding the circulation and exploitation of data are disciplined in the **Open Data Directive** (for public sector data), as well as in the recently adopted **Data Act**, which regulates the access and reuse of data generated using digital products and services. This approach is also consistent with the pre-existing framework that comprises the General Data Protection Regulation (GDPR). EU institutions reaffirm the importance of data flow, while builiding a regulatory framework to ensure safeguards to individuals' rights and freedoms are put in place.

The ultimate goal is to build a European common Data Space, much like the single market. In this perspective, it becomes paramount to make the most of new data-based economies so that the EU can be a hub for innovation and growth. And thus, we must go back to the Digital Services Package and to the role of researchers: for the EU's vision to materialize into a futuristic utopia, we need **safeguards against systemic risks** and excessive powers of the Big Data companies. To create such safeguards, we must detect the threats which lie in the digital world and design possible solutions. We need the online sphere to become a transparent environment, which allows researchers to understand its intricate mechanisms and their risks. The other side of the coin is Surveillance Capitalism, maybe something even worse. Therefore, the complex (and growing) European regulatory framework for the digital sphere can be seen as a composite group of checks and balances designed to reach a fragile **equilibrium between data use and fundamental rights.**

We have written a *White Paper* outlining our efforts as a community of researchers in focusing our interest on the interaction between the digital sphere and the real world, and how the first can shape the second, for better or for worse

How SoBigData fits in this scenario

SoBigData operates in research spaces which regard topics aimed at exploiting the **benefits of new technologies** to build a future society which is **human-centered** and based on **fundamental values** such as those recognized by the European Union. Many of our projects are focused on understanding and finding solutions to risks such as the systemic ones identified in the Digital Services Act.

To present some of these works we have written a *White Paper*, where we outline our activities in fields such as misinformation, smart cities and mobility, finance 2.0, and online societal debates. This is part of a broader initiative: the SoBigData Social Observatory, which is aimed at collecting, updating, and creating indicators in order to publish periodic reports to monitor the evolution of systemic risks.

This White Paper demonstrates how SoBigData can contribute to civic goals with currently available data and gives a perspective of how our Research Infrastructure can help in fighting systemic risks as an **actor involved in data access mechanisms** designed by the Digital Services Act. To provide an overview of SoBigData's capability of tackling the most pressing issues in the digital world, we presented our resources and research spaces, which provide a common ground for researchers to share ideas, data, methodologies, and results equipped with open science tools and a mobility program to allow them to travel around Europe to collaborate. In the *White Paper* we also highlight that we strive to be always in line with the **Open Science** values, so our solutions are *"as open as possible, as closed as necessary"*, also consistent with the **Findability, Accessibility, Interoperability, and Reusability (FAIR) and Fair, Accurate, Confidential and Transparent (FACT) principles**.

The white paper showcases SoBigData's experience regarding very actual issues in the digital world, such as the detection and characterization of polluted online debates, opinion dynamics and polarization in the digital sphere, coordinated behaviors of social networks users, the relation between social media information and financial markets and how the latter can be also influenced by publicly available online news, the impact of navigation services' recommendation on urban environments, and bot detection on social media platforms. These examples give an overview of our efforts as a community to focus our interest on the **interaction between the digital sphere and the real world**, and how the first can shape the second, for better or for worse. As we have seen above, this kind of work can be essential to create a future society where technology benefits all.

SoBigData embraces *fair* and *fact* principles

FOCUS

FAIR

To maximize benefits of data science for society, it is important for it to be Open. Therefore, the data and metadata used should respect the FAIR guiding principles which foster the reusability of information both by humans and machines. This means that data should be Findable, Accessible, Interoperable, and Reusable.

F **Findable**

Data and metadata must be easy to find by both humans and machines, also in an automated way using machine-readable metadata.

A **Accessible**

Once found, it must be possible to access to the data/metadata in an easy way, with clearly defined conditions.

I **Interoperable**

The data/metadata must be in such a format that it is possible to combine it with other data, also using automated means.

R **Reusable**

It must be possible to reuse the data/metadata for future research and to further process it also using automated means.

FACT

Data science activities should be carried out in an ethical and secure manner so they can benefit society without violating fundamental values and individual' rights. To reach this goal, we should foster FACT data science, which is Fair, Accurate, Confidential, and Transparent.

F **Fair**

The results must be impartial and non-discriminatory, and efforts must be made to avoid unfair results even if they are computationally correct.

A **Accurate**

The results must be correct and up to date, so misleading conclusions are avoided.

C **Confidential**

Privacy, trade secrets and confidentiality are respected throughout all the process of data analysis.

T **Transparent**

The results can be explained and understood in a transparent manner, so they can be trustworthy.

Data access in practice: misinformation, coordinated online behaviors and a code of conduct which should not be underestimated

The study of misinformation and social debates is one of the focuses of SoBigData: as the Internet and social networks have redefined lifestyles worldwide, it is fundamental to understand their influence on society also in the political and democratic discourse and news circulation.

We have talked about this with Serena Tardelli, researcher at the Institute of Informatics and Telematics (IIT) of the CNR in Pisa, who has researched coordinated online behaviors on the formerly called Twitter (now "X"). She and her colleagues chose to study this phenomenon as it usually reflects social dynamics which previously existed in the real world and amplifies them in the online sphere, also influencing how people perceive certain events or even some political issues. Therefore, their aim was to find coordinated online behaviors which could influence public debates, also understanding when some groups tried to promote certain narratives, maybe with specific goals (which could be economically or politically driven), and whether these behaviors materialized in some real-world consequences, like protests that were initiated and coordinated online. This also helps to understand when coordinated online behaviors carry risks for society, as it is clear now that this phenomenon is not intrinsically negative but can have very serious repercussions.

During our talk with Serena Tardelli, it was immediately apparent that having access to Twitter's data via its API was fundamental for her research. She and her colleagues choose this social network because, for its user base and dynamics, it is a particularly interesting platform to analyze coordinated online behaviors: everybody can talk about specific subjects on the same feed, use hashtags to gather all conversations about a certain topic, communicate very quickly (also because of the characters limitation for each tweet). Discussion on Twitter is public, open and often updated instantaneously when a certain event happens; for example, for these reasons this platform is used a lot during election times to exchange real time updates. However, the platform's own features also favor misinformation, so X has become one of the most studied platforms to research this issue. When Serena and her colleagues started their project, Twitter's data was quite easy to access for researchers, and this enabled their work to reach the form it has now. However, with the new management of the platform, not only its iconic name was changed, but also its policy about API's accessibility, so now this kind of research would be significantly more complicated to carry out. This could have very negative repercussions on misinformation and disinformation: dynamics such as coordinated online behaviors could go on unobserved on this platform, polluting democratic discourse and thus generating social effects which could pose threats to society.

> Without data access, dynamics such as coordinated online behaviors could go on unobserved, polluting democratic discourse and thus generating social effects which could pose threats to society

It is notable that in June 2023 X left the Strengthened Code of Practice on Disinformation, a voluntary code of conduct which was delivered in 2022 by major players of the online sector (including Twitter) following recommendations by the European Commission. This document is designed to become one of the codes of conduct envisaged by the Digital Services Act to mitigate systemic risks in the digital sphere and, like the Regulation, it contains specific measures to empower researchers allowing them to access platforms' data. Of course, in this case, the focus is disinformation: to foster research on this topic, the Signatories of the Code commit to provide access, wherever safe and practicable, to non-personal data and anonymized, aggregated, or manifestly made public data through automated means such as APIs or other open and accessible technical solutions, allowing the analysis of said data. Like in the Digital Services Act, "vetted researchers" are provided with a broader form of access, and the Signatories commit to developing, funding, and cooperating with an independent, third-party body that can vet researchers and their proposals. Even though it is a voluntary instrument, which binds only its Signatories (among which we can find most of the VLOPs and VLOSEs), this Code provides for additional safeguards against the systemic risks linked to online disinformation, which should not be underestimated, given the Code's connection to the Digital Services Act. This has become crystal clear after X's withdrawal from the Code, when senior figures among the EU institutions heavily criticized this action as a sign of possible non-compliance to the European Regulation, and in the past few weeks, because now the platform faces infringement proceedings (officially opened in December 2023) over potential breaches of the Digital Services Act, including the provisions regarding data access to researchers.

We end our discussion with Serena Tardelli on a hopeful note: institutions and private players are designing instruments to empower researchers who work to fight disinformation and misinformation, like the Digital Services Act and the Strengthened Code of Practice on Disinformation, so that it will be possible to carry out studies, like the one about coordinated online behaviors, which pave the road to a better digital world.

Network cordinated behavior ■ Each node in the network is a user. The network's edges represent how coordinated these users are. Connections form various communities, each colored differently

Future events

Schools and conferences promoted by SoBigData RI

Registrations will open soon on *www.itadata.it*

SoBigData is dedicated to fostering an open innovation culture in data science. Events, such as conferences, summer schools, datathons, and tailored workshops, are central to the success of Research Infrastructure (RI) activities.

These actions also enable the RI to bridge the gap between research and application, striving not only to advance the field of data science but also to ensure that the knowledge generated is accessible and beneficial across various disciplines and industries.

The convergence of experts from diverse fields at conferences and other gatherings allows for the exchange of ideas, methodologies, and knowledge. Moreover, it serves as a training ground for the next generation of data scientists, promoting a new sense of responsibility in handling and interpreting social data.

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 14–20

Lipari School on Computational Complex and Social Systems

The Summer School on Computational Complex and Social Systems to be held in Lipari, Italy

JULY 1–12

Digital Methods Summer School

The Summer School about Visual methodologies and analytics. Media Studies, University of Amsterdam

SEPTEMBER 17–19

ITADATA conference

The third Italian Conference on Big Data and Data Science will take place in Pisa, Italy

OCTOBER 14–16

Discovery Science conference

27th International Conference on Discovery Science 2024 | Pisa, Italy

AUTUMN TBA

ML of Dynamic Processes and Time Series Analysis

The third edition of the Machine Learning of Dynamic Processes and Time Series Analysis school to be held at Scuola Normale in Pisa, Italy

TRAINING

04

Master in Big Data Analytics & AI for Society: a bridge between Academia and Industry

10 Years of SoBigData Master

by Riccardo Guidotti

The Master in Big Data Analytics & AI for Society at the University of Pisa (SoBigData Master), powered by the SoBigData RI, is an annual postgraduate course for aspiring data scientists and AI experts. Tailored for graduates across diverse disciplines, this program serves as a transformative pathway, providing essential tools to navigate the expansive field of data analysis applied across sectors, from business to cutting-edge research. The SoBigData Master is inclusive by design and extends its benefits to professionals seeking to enhance their skill set and elevate their professional standing. Unlike programs narrowly focused on specific applications of data science and AI, this Master presents a panoramic view of the field. It offers a spectrum of courses, spanning across business and research, empowering students to enhance their expertise in the sector aligned with their academic background and aspirations. Commencing with foundational courses in databases and programming, students gradually immerse themselves in advanced modules covering a diverse range of topics encompassing data mining, deep learning, information retrieval, data crawling, decision support systems, and more, providing a comprehensive understanding of the multifaceted landscape of data science and AI. The SoBigData Master ensures that graduates not only gain technical skills but also the flexibility to specialize according to their individual preferences and career objectives.

Beyond traditional classrooms, students in the SoBigData Master are immersed in hands-on, practical laboratories that translate theoretical concepts into tangible skills. Embracing the "learning-by-doing" model, the program goes beyond knowledge acquisition, fostering the development of competencies—teaching students not just what to know but how to apply their knowledge effectively. With an enriching learning experience, the Master incorporates seminars led by industry experts, offering insights into the utilization of Big Data technologies and AI tools within industrial and innovation processes. This exposure to real-world applications enhances students' understanding and prepares them for the challenges of the professional landscape. The course's structure is thoughtfully organized, balancing six months of intensive lectures with an equally crucial six-month segment dedicated to internships and bootcamps. These hands-on experiences take place in collaboration with esteemed partner companies and institutions, guided by both corporate and academic mentors. Remarkably, the SoBigData Master transcends the conventional scope of academic programs by instilling a profound sense of responsibility in its graduates. Students are encouraged to explore the altruistic potential of Big Data and AI, contemplating ways these technologies can be leveraged for the betterment of society. This socially conscious approach sets the program

Embracing the "learning-by-doing" model, the program goes beyond knowledge acquisition, fostering the development of competencies

The overarching goal of the SoBigData Master is to cultivate professional data scientists and AI experts

apart, instilling in students a sense of responsibility and an understanding of the ethical implications of their work.

The overarching goal of the SoBigData Master is to cultivate professionals named "data scientists and AI experts" — individuals possessing a rich mosaic of multidisciplinary skills. This unique skill set allows them not only to acquire and process extensive datasets but also to extract meaningful insights that underpin AI-driven decision-making and the conception of innovative services. Crucially, these professionals are well-versed in navigating the ethical and legal considerations entwined with the utilization of such services. The complexity of training data scientists lies precisely at the convergence of technological, analytical, narrative, and ethical competencies. As such, the program strategically integrates knowledge from diverse disciplines, encompassing machine learning, data analysis and visualization, complex systems science and networks, computational sociology, as well as ethics, data journalism, and story-telling. This holistic approach ensures that graduates are not only proficient in the technical aspects but also possess a nuanced understanding of the ethical implications inherent in their work.

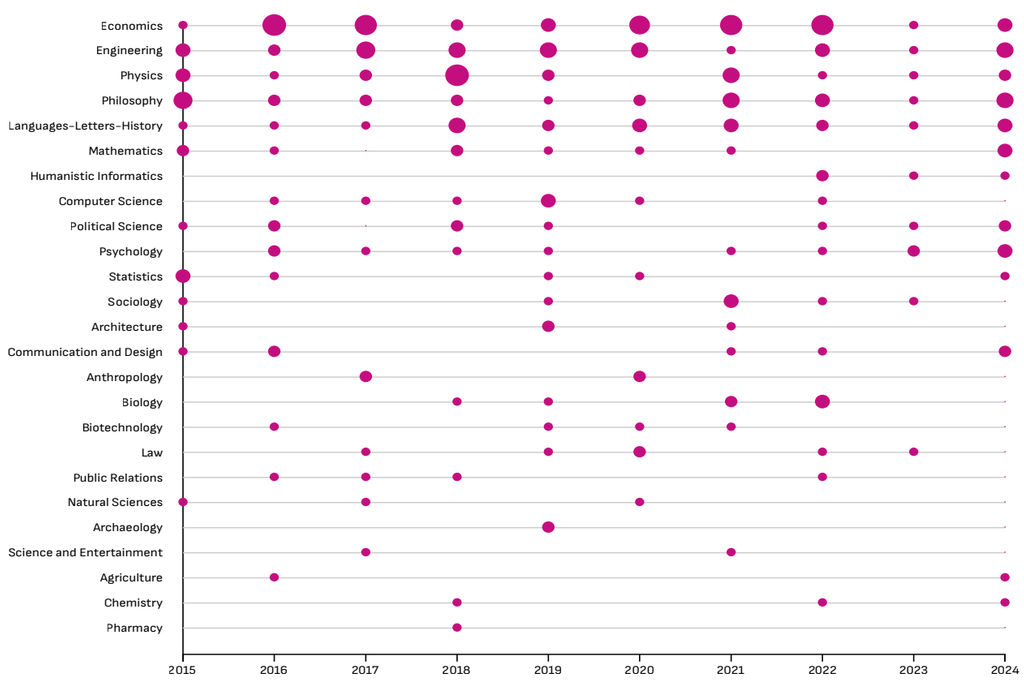
Dino Pedreschi, the inaugural Director of the Master, and Fosca Giannotti, both visionaries, embarked on a groundbreaking journey a decade ago. Fueled by the support of the SoBigData EU project, they conceived the SoBigData Master with a steadfast goal that endures today — shaping adept "data scientists." A decade ago, The Economist coined this profession as "the sexiest job of the 21st century," a name that remains pertinent. Contemporary reports from prominent job platforms like Linkedin underline the continued and heightened demand for Data Scientists and AI experts, forecasting their pivotal roles in the professional landscape. In anticipating the evolving employment needs, the SoBigData Master uniquely positioned itself as an avant-garde training ground, attuned to the burgeoning importance of AI and the growing emphasis on ethical considerations within the EU community. The foresight demonstrated a decade ago has proven invaluable, aligning the program with current industry demands. A key facet of the program's success lies in the decision to assemble a teaching team comprising experts from diverse institutions. While spearheaded by the University of Pisa, the Master collaborates seamlessly with ISTI-CNR, IIT-CNR, Scuola Normale Superiore, Scuola Superiore Sant'Anna, and IMT Lucca. This strategic alliance brings together the collective excellence of these institutions, creating a learning environment that draws from the best in the region's research area.

An integral part of the success of the SoBigData Master is its continuous collaboration with companies and institutions, a facet meticulously fostered since 2014. This strategic dimension owes much to the adept guidance of Anna Monreale, the second Director of the Master, who not only initiated but also expanded partnerships from the initial few to a remarkable 40 and beyond. The essence of this collaboration transforms the Master into a vital bridge, seamlessly connecting the realms of academia and enterprise. This symbiotic relationship ensures that theoretical and practical insights imparted by the Academia find real-world applications within industrial and research processes during the internships. Notably, the bridge extends beyond the learning phase, as students often find employment with the companies hosting their internships. A noteworthy achievement is the profound impact on the industry landscape, as most partner companies have established Data Science departments through the contributions of Master's students. These alumni, now leading these departments, continue to reciprocate by collaborating with and contributing to the ongoing success of the Master. In essence, the close cooperation with companies and institutions not only enriches the learning experience but also fosters a sustainable and mutually beneficial ecosystem that continues to thrive and evolve.

Thus, the SoBigData Master stands as a bridge, connecting the theoretical rigor of academia and the practical demands of industries and companies. Graduates emerge not only as skilled data scientists and AI experts but as professionals who understand the broader implications of their work on society. This unique blend makes the Master a valuable contributor to the ever-evolving landscape of technology and its transformative impact on the world.

GRADED STUDENTS	PARTNER COMPANIES	PROJECTS
265	47	49

FEMALE PRESENCE	HIRING MEAN TIME	HIRING RATE
92	1.5 MONTHS	97%



On the right: distribution of students' background by year

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.

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

Mirco Nanni, permanent researcher at CNR-ISTI in Pisa and Head of the KDD Lab

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 **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 seen 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 results can find applications in other sectors, such as Fleet companies, to improve their ecological footprint in the transition to green mobility supporting mobility managers in understanding how to mitigate their daily routines through less polluting solutions. But the scientific results provide further insight, as incremental analysis done by OCTO shows how more contextual information as individual predictive risk and eco-scoring lay the foundation for the driver to measure his driving style and then learn how to improve towards a safer and greener behaviour.

"From the analysis of OCTO data acquired by connected vehicles, the collaboration with CNR-ISTI has created many valuable synergies with immediate implications" according to Martino. "Just to mention some of them, knowing the mobility of people based on real mobility data

How SoBigData supports Industry

SoBigData offers a range of services to support businesses in harnessing the power of AI, big data, and cloud computing. Our consulting services provide expertise in AI and big data analysis methods, while our storage and computational services ensure secure and efficient data management.

We also offer evaluation services to address ethical and legal risks, as well as partnership opportunities for project development. Additionally, our staff training and education programs and digital integration services help companies stay at the forefront of technology.

1. Consulting service on AI and Big Data methods and algorithms
2. Storage and computational services, cloud computing, secure storage, and access control
3. Evaluation of ethical and legal risk consulting service for process design
4. Partnership in project development, support for company project design and co-participation in EU calls
5. Business-oriented training and educational initiatives along cutting-edge research
6. Tailored training courses designed to the company's needs
7. Digital integration among public platforms and services

allows to design infrastructure roads, crossings, traffic lights, and public transportation based on actual and measurable needs of transfer; define emergency planning and way-out for crowded spaces; reduce the impact of traffic after roadwork sites; support the EV chargers planning based on the energy demand of the drivers, on trip mileage, class of vehicle, time of stops".

Different goals, shared benefits

The possibility of working within OCTO's massive database of mobility was an excellent opportunity for the research team as well, Nanni underlines: "OCTO's needs exposed us to novel technical issues and challenging objectives that stimulated our research. Also, working on OCTO's great data assets allowed us to develop considerable expertise in big mobility data".
"On the company side, we were able to put the research results in place at the industrial level, raising the company's competitiveness with advanced technology, guaranteeing the quick applicability of the results on the market," adds Martino. Each company could find a well-consolidated and real useful technical partner in the public research team to carry out complex, cutting-edge technology projects.
Moreover, a public research mission often lets researchers explore problems and applications beyond what exists in the market, suggesting novel ideas that may have potential for the company: "Some results allow us today to provide solutions to our business customers to face the challenges due to their daily activities: manage the Total Cost of Ownership for Fleet Managers, reduce the risk and the frauds in the Insurance Market, provide more customized and valuable services based on the single user experience" Martino underlines"Last but not least, the results of the cooperation have provided a new lens to understand the society that will lead us to have a greater capacity for risk prevention for the benefit of human safety and health".

The lesson from the OCTO-SoBigData partnership

These impacts exemplify how data can reveal surprising insights into social phenomena in different data contexts and for other industry sectors to foster innovation. According to Martino:"Economic growth is moving with a fast pace of technological innovation is a big challenge for each company and requires considering different operating models to be competitive and sustainable in the broader sense: economic sustainability, regulations' compliance, profitability for all the stakeholders, and grafted into ethical principles."
Moreover, this collaboration highlights the relevance of the public research sector to the private one. For Martino: "Against the monopoly of a few big technology companies, cooperation with the Public research on Big Data and AI is necessary for a company looking to stay at the forefront of technological advancements and foster innovation. This collaboration is a gain for the effectiveness of the investment, builds constructive partnerships with open sharing ideas, is helpful for the company growth and the sane competition, contributes to creating a bridge for the skills upgrades of their employees, and facilitates the students' entry into the job market."

Sharing a vision

Data can drive company innovation, and the public research sector can support the industry in society's transformation in a customized way, embracing the ethical and legal principles of Big Data and AI use. Moreover, a public research institution such as SoBigData aims to improve social lifestyles by providing policymakers and company leaders the tools to make data-driven solutions with a tangible impact on societal benefit. At this point, OCTO's and SoBigData's visions meet together, aligned with the overarching goals of global sustainability, as Martino explains:
"The role of both public research and private companies is crucial in facilitating rapid adoption, necessitating widespread competence and skill upgrades, extensive education, and a cultural shift.
Additionally, responsible dissemination and support for policymakers are essential for creating regulations and policies that encourage innovation while addressing potential risks and challenges associated with Big Data and AI. Being part of a broader ecosystem aids in faster learning and mitigates risks through access to the right information and skills. Shaping the future together, make it happen: we aim to put in action our vision of zero (Zero pollution, Zero traffic and Zero accidents) encompassing all activities, including investing in research and development" well aligned with OCTO mission "Smart Analytics for a better connected world".

Sustainable Cities for Citizens: Stories about cities and their flows' sustainability

Unlocking insights from mobile phone traces, vehicular GPS, and social media data provides a wealth of Big Data, serving as invaluable proxies for understanding human behaviour. In the mobility sector, analysing these diverse data sources holds the key to making informed decisions and steering data-driven innovation.

Delving into the heart of this data-driven revolution is the SoBigData Research Space "Sustainable Cities for Citizens," spearheaded by the research team from IMT in Lucca, Italy, and KDD Lab at CNR-ISTI Pisa, Italy. Focused on smart mobility, this dynamic team has collaborated over the years in partnership with industry leaders to pioneer innovative solutions in the mobility and smart cities industry.

OCTO-Telematics: Zero Crashes, Zero Congestion, Zero Pollution

With a 5.7 million connected users network and one of the most significant global telematics database, OCTO Telematics stands at the forefront of the connected mobility sector. Recognized as a world leader in insurance telematics, OCTO extends its influence into diverse markets such as Fleet Management, shared mobility, vehicle diagnostics, real-time traffic monitoring, public transport, and mobility services.
Thanks to scalable and modular Internet of Things (IoT) platforms and the leverage of Machine Learning and Artificial Intelligence technologies, OCTO provides innovative solutions in Insurtech and mobility markets to manage and expand the businesses of partner companies.
OCTO has implemented a roadmap of concrete actions to disseminate best practices towards OCTO Vision Zero (Zero pollution, Zero traffic, and Zero accidents), aligned with the European program dedicated to reducing CO2 emissions and road safety.

SECTION

06

SOBIGDATA RESEARCH TOPICS

Research spaces

The scientific collaboration in SoBigData is organised in specific research topics aimed at applying data science to reality

1 Demography, Economy and Finance 2.0

Exploring changes and correlations in people's and companies' behavior

This research space focuses on examining both established intricate socio-economic financial systems and emerging ones, notably blockchain and cryptocurrency markets. Research delves into their diverse applications, including smart property, the Internet of Things (IoT), energy trading, and smart contracts. Through comprehensive research, the interplay between conventional systems and these domains is explored, seeking to understand their implications and potential impacts on various sectors.

2 Disaster Response and Recovery

Environmental health, reconstruction

This research space focuses on methods and tools to analyze, monitor, and improve post-disaster reconstruction processes in socio-economic areas, spatial planning, environmental health in cooperation with national and international institutions. ICT-enhanced solutions to the response and the recovery phases of the disaster management cycle can improve the efficacy of "search & rescue" activities, emergency relief, reconstruction, and rehabilitation. This includes the study of methodologies to create evacuation plans from an enriched city graph, methodologies based on predictive approaches to guide the decision-makers in the definition of (urban and socio-economic) reconstruction policies, and the definition of collaborative virtual spaces for the monitoring of the territory.

3 Health Studies

Network Medicine, sports data science, food analytics

This research space focuses on medical, nutrition, and environmental research to address health issues. It employs machine learning and AI, including Network Medicine, combining systems biology and network science for disease understanding and personalized treatments. The field builds BioMedical knowledge graphs using specialized tools for research papers. Studying personal activities' impact on health and performance involves sensing technologies for reliable data. Integrating big data with AI unveils complexities and achieves challenging goals.

4 Pervasive Intelligence in Cyber-Physical Systems for Future Society

Cyber-physical systems, human-centric AI, IoT and edge devices

This research space focuses on leveraging extensive traffic data to develop "Zero Touch" networking for the growing number of devices in society. It aims to study AI models for dynamic resource allocation, emphasizing responsiveness to changes in services, application requirements, and user-generated traffic. The research space also investigates Cyber-Physical systems emulating biological brain mechanisms for self-control in scenarios like industrial settings. The focus is on decentralized computation paradigms using networking technologies (e.g., IoT and edge devices) to make AI pervasive, moving intelligence from centralized data centers to users' devices. The goal is to foster advanced services based on human-centric AI.

Research Spaces

The scientific SoBigData collaboration operates in specific research topics aimed to apply data science to reality world

5 Social Impact of AI and explainable ML

Tracking changes from human coding to algorithms automatically learning to solve tasks

This research space focuses on the rise of the 'black box' society in the last decade. By observing many examples of the expected input/output behaviour, often algorithms' internal reasoning is obscure even for their own developers. The research space assesses black box AI systems for automated decision making, which are often based on machine learning trained with big data and which map out user features into a class predicting his or her behavioural traits.

6 Societal and Industrial Impact of Next-Generation Internet & beyond 5G Networks

5/6G networks, next-generation Internet, beyond 5G Networks

This research space focuses on 5/6G networks and next-generation Internet to enable new services impacting society and manufacturing. This research space investigates specifically the adoption of smart radio environments, open-RAN paradigms and novel network virtualisation technologies. Thanks to the integration of social data and industrial activities, it focuses on the impact of communication services on specific applications, such as mobility, vehicular traffic management, and manufacturing processes. In addition, it investigates the effect of Next Generation Internet paradigms by developing user- and device-centric services for groups of mobile users based on direct communication between far-edge devices, validating them in the same application areas.

7 Societal Debates and Misinformation

How do people discuss on online social networks? Which users take part in the political debate and how?

This research space focuses on public debates analyzing both discussions on social media and journalistic production to understand which are the most debated topics. These analyses can identify themes, assessing how they are discussed and tracking them in time and space. This research space analyses individuals' discussion using data from online social networks in order to investigate topics that are relevant for society.

8 Sustainable Cities for Citizens

Stories about cities and their flows' sustainability

This research space focuses on mobile phone traces, vehicular GPS and social data, all of which are Big Data sources and proxy of human behavior. If analysed, results can be useful both for administrators and citizens. Local administrators can better understand urban mobility using a tool that quantifies traffic and the city's usage, thus taking better decisions in mobility management. Citizens can access real-time traffic information and choose the best and fastest route.

RESEARCH

07

Highlights

Our mission, as SoBigData researchers, is to explore the complexity of society through the lens of massive datasets, uncovering hidden patterns, trends, and insights that shape our interconnected world.

In this section, we showcase some research highlights as examples of how SoBigData is able to provide valuable perspectives on human behaviour, societal dynamics, and the impact of digital technologies using Big Data and AI.

Big Data and AI are not solitary entities; they are symbiotic, each enhancing the capabilities of the other. Big Data, characterized by the exponential growth of datasets, provides the raw material, while AI algorithms offer the means to extract meaningful insights from this large set of information. The convergence of these two forces requires a collaborative multi-disciplinary research approach involving the public and private sectors.

In the following highlights, we will navigate the intersection of computer science, social science, and data analytics with the research and industrial worlds, pushing the boundaries of understanding and contributing to the advancement of science for social good. These highlights offer a glimpse into the transformative potential of social big data, showcasing the SoBigData RI commitment to shaping a data-driven future.

1

How routing strategies impact urban emissions

Giuliano Cornacchia, Matteo Böhm, Giovanni Mauro, Mirco Nanni, Dino Pedreschi, Luca Pappalardo

SUSTAINABLE CITIES FOR CITIZENS

The widespread use of GPS navigation services like TomTom, Google Maps, and Waze has raised concerns about their unintended consequences on urban well-being. While these services aim to optimize travel time, they may compromise safety and increase pollution. SoBigData researchers developed TraffiCO2, an open-source simulation framework, to rigorously assess the impact of GPS navigation apps on urban well-being regarding carbon dioxide (CO2) emissions. TraffiCO2 utilizes GPS data, TomTom and OpenStreetMap APIs for routing suggestions, and SUMO for comprehensive urban mobility simulation. The study, conducted in Milan, revealed that scenarios where all or no vehicles follow navigation suggestions result in the highest CO2 emissions. Conversely, when around 50% of vehicles follow navigation suggestions, emissions decrease, and distribution becomes more even across the road network. The study also highlighted the spatial impact, increasing pollution on Milan's external ring road while reducing emissions on internal roads. TraffiCO2 provides a valuable tool for assessing and comparing routing strategies, aiding drivers, institutions, and policymakers in understanding and mitigating the impact of navigation apps on the urban environment.

2

Rebalancing social feed to minimize polarization and disagreement

Federico Cinus, Aristides Gionis, Francesco Bonchi

SOCIETAL DEBATES AND MISINFORMATION

The research paper addresses the escalating concerns of polarization and echo chambers within social networks. Acknowledging the role of recommender systems in shaping user exposure and potentially exacerbating polarization, the paper introduces the LcGD algorithm. This innovative approach optimizes the frequency of users encountering opinions to strike a balance between reducing polarized content and ensuring alignment with users' interests.
The study emphasizes the distinction between expressed and innate opinions, aiming to gradually expose users to a broader spectrum of perspectives without overwhelming them.
The research underscores the significance of addressing polarization's implications and proposes strategic interventions to mitigate its effects. The LcGD algorithm proves efficient in large social networks, outperforming existing recommendation methods.
The reduction of polarization within social networks is crucial for promoting understanding, empathy, and constructive dialogue. The research provides insights and tools to mitigate polarization's negative consequences, enabling social networks to play a positive role in shaping public discourse.

3

Ask “Who”, Not “What”: Bitcoin Volatility Forecasting with Twitter Data

M. Eren Akbiyik, Mert Erkul, Killian Kämpf, Vaiva Vasiliauskaite, Nino Antulov-Fantulin

DEMOGRAPHY, ECONOMY AND FINANCE 2.0

Understanding the variations in trading price (volatility), and its response to exogenous information, is a well-researched topic in finance. In this study, the authors focus on finding stable and accurate volatility predictors for a relatively new asset class of cryptocurrencies, in particular Bitcoin, using deep learning representations of public social media data obtained from Twitter. Authors extracted semantic information and user statistics from over 30 million Bitcoin-related tweets, together with 15-minute frequency price data over a horizon of 144 days. Several deep learning architectures were built. For each model, ablation studies to assess the influence of different components and feature sets over the prediction accuracy were conducted. The authors found evidences for the hypotheses that: (i) temporal convolutional networks perform significantly better than both classical autoregressive models and other deep learning-based architectures in the literature, and (ii) tweet author meta-information, even detached from the tweet itself, is a better predictor of volatility than the semantic content and tweet volume statistics. They demonstrate how different information sets gathered from social media can be utilized in different architectures and how they affect the prediction results.

4

XGDAG: explainable gene–disease associations via graph neural networks

Andrea Mastropietro, Gianluca De Carlo, Aris Anagnostopoulos

HEALTH STUDIES

Disease gene prioritization consists of identifying genes that are likely to be involved in the mechanisms of a given disease, and providing a ranking of such genes. Recently, the research community has used computational methods to uncover unknown gene—disease associations; these methods range from combinatorial to machine learning-based approaches. In particular, approaches based on deep learning have provided superior results compared to more traditional ones. Yet, the problem with these is their inherent black-box structure, which prevents interpretability. The authors propose a new methodology for disease gene discovery, which leverages graph-structured data using graph neural networks (GNNs) along with an explainability phase for determining the ranking of candidate genes and understanding the model’s output. The approach is based on a positive—unlabeled learning strategy, which outperforms existing gene discovery methods by exploiting GNNs in a non-black-box fashion. The proposed methodology is effective even in scenarios where a large number of associated genes need to be retrieved, in which gene prioritization methods often tend to lose their reliability.

5

Industry 4.0 base technologies and business models: a bibliometric analysis

Giulio Ferrigno, Nicola Del Sarto, Andrea Piccaluga, Alessandro Baroncelli

INDUSTRY

The fourth industrial revolution, driven by Industry 4.0 technologies such as IoT, cloud services, big data, and analytics, has profoundly transformed business practices. Despite extensive literature on the subject over the past 12 years, a comprehensive overview of how these technologies impact firms’ business models is lacking. This study addresses this gap through a bibliometric analysis of nearly 500 papers. The key research themes include smart products, business model innovation, technological platforms, value creation and appropriation, and digital business models. The study identifies future research trajectories such as digital transformation, servitization, opportunities for SMEs, circular economy systems, and disruptive supply chain logistics. The authors stress that Industry 4.0 extends beyond the studied technologies, suggesting the influence of others like 3D printing and cybersecurity. The research underscores the importance of adaptable digital business models for firms to thrive in the evolving technological landscape, emphasizing the critical role of investing in both technological innovation and new business models.

6

The Propagation of Misinformation in Social Media. A Cross-platform Analysis

Richard Rogers

SOCIETAL DEBATES AND MISINFORMATION

There is growing awareness about how social media circulate extreme viewpoints and turn up the temperature of public debate. Posts that exhibit agitation garner disproportionate engagement. Within this clamour, fringe sources and viewpoints are mainstreaming, and mainstream media are marginalized. This book takes up the mainstreaming of the fringe and the marginalization of the mainstream. In a cross-platform analysis of Google Web Search, Facebook, YouTube, Reddit, Twitter, Instagram, 4chan and TikTok, we found that hyperpartisan web operators, alternative influencers and ambivalent commentators are in ascendency. The book can be read as a form of platform criticism. It puts on display the current state of information online, noting how social media platforms have taken on the mantle of accidental authorities, privileging their own on-platform performers and at the same time adjudicating between claims of what is considered acceptable discourse.

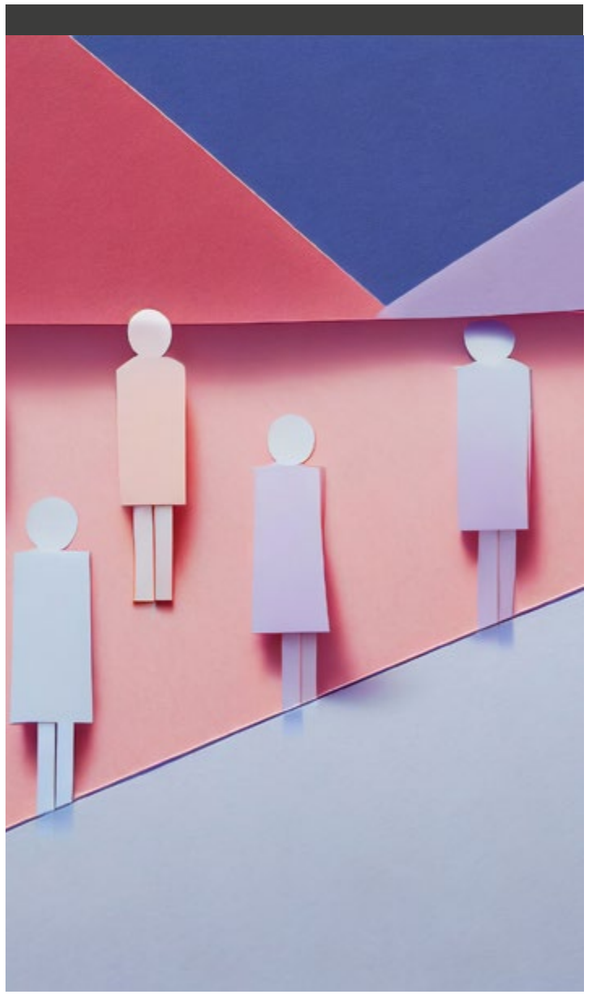
by Michela Natilli

RESEARCH

08

Diversity & inclusion

SoBigData actively promotes diversity and inclusion in data science



Raise awareness about data science job opportunities, with a focus on women in STEM fields.

SoBigData is committed to disseminating information regarding the numerous career opportunities within data science, with a specific emphasis on encouraging women in STEM fields. The objective is to showcase the compelling aspects of data science careers, intending to inspire increased participation of women in these opportunities and foster their contributions to the field.

Plan dedicated initiatives to identify under-represented categories.

In order to prioritize diversity and inclusion, SoBigData is dedicated to strategically planning and executing initiatives tailored to recognize and redress underrepresented categories. By proactively identifying individuals traditionally marginalized within the field of data science, SoBigData aims to contribute to the establishment of a more inclusive and diverse community.

Promote positive actions at various levels of education (high school, undergraduate, graduate, and PhD students).

At SoBigData, the belief is held that the cultivation of diversity and inclusion must commence at all levels of the educational trajectory. The commitment extends to the implementation of affirmative measures designed to foster diversity and inclusion, encompassing not solely higher education but also secondary schools. Through the provision of support, resources, and opportunities to students at each stage, the goal is to cultivate a more diverse and inclusive community within the domain of data science.

1

SoBigData Award for Diversity and Inclusion

It aims to promote a more diverse participation to computer and data science events by financially aiding individuals that identify with a minority group to attend selected conferences

2

PinKamP 2023

UNIVAQ's initiative to empower girls in digital fields, break stereotypes, eliminate barriers, and highlight women's impact on future tech.

3

Colorful Seminars series: enhancing diversity and inclusion

To enhance diversity and address equity gaps, UNIPI implements a series of seminars presented by global experts.

AI-GAP

4

AI - GAP Algorithmic Biases in Artificial Intelligence from Interdisciplinary Perspectives

This workshop addresses the social impact of Artificial Intelligence (AI) technologies that reproduce human biases and inequities.

SECTION 09

Experience TNA

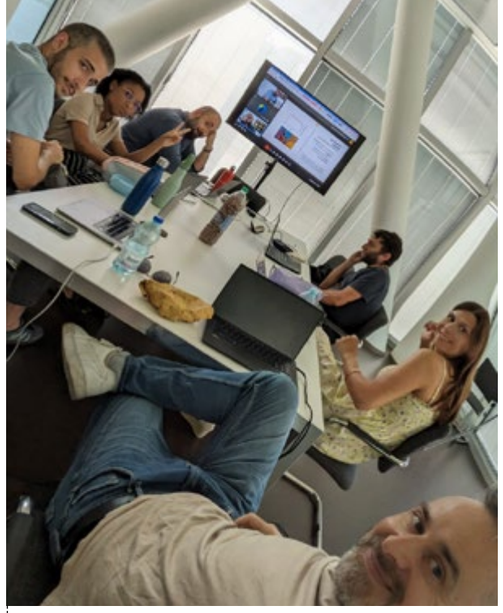
A Memorable Journey into Mobility Research: Leo Ferres Experience at Pisa CNR

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 music scene, I joined musicians, fostering unity. Spontaneous park jam sessions united us, transcending research boundaries through music's universal language.

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/transnational-access-2024*

OUR CONSORTIUM

1. Aalto University **2.** University of Bologna **3.** Athena **4.** Barcelona Supercomputing Center **5.** Carlo Ratti Associati **6.** National Research Council (Italy) **7.** Center for the study of democracy **8.** National Center for Scientific Research (France) **9.** Cental European University **10.** EGI - Advanced Computing Services for Research **11.** Eliante **12.** Erasmus University Rotterdam **13.** ETH Zurich **14.** L3S Research Center **15.** Fraunhofer Institute **16.** ICHB Pan **17.** IMT Lucca **18.** King's college London **19.** KTH Vetenskap Och Konst **20.** Nubisware **21.** Open Aire **22.**Paris School of Economics **23.** Reimagine Uerope **24.** Rephrain **25.** Sapienza University of Rome **26.** Scuola Normale Superiore **27.** Sant'Anna School of Advanced Studies **28.** SmartLex **29.** Stacc **30.** University of Tartu **31.** University of Sheffield **32.** Tu Delft **33.** University of Palermo **34.** University of L'Aquila **35.** University of Catania **36.** University of Pisa **37.** University of Genoa **38.** University of Amsterdam **39.** Universitat Pompeu Fabra Barcelona **40.** Universitat Rovira i Virgili **41.** Utrecht University

Where to find us

To keep in touch with the SoBigData Research Infrastructure please use one of the channels below.
Are you are a PhD student or member of academia and want to explore our researcher mobility program? Or perhaps work in a start-up or small or medium enterprise and would like to explore a possible collaboration with the reseach infrastructure? Or are generally curious about SoBigData? Please get in touch, we'd love to hear from you, either via email or by following us on our social media channels.

SOBIGDATA RESEARCH INFRASTRUCTURE

Let's keep in touch

www.sobigdata.eu

twitter.com/sobigdata

linkedin.com/company/sobigdata

info@sobigdata.eu

SoBigData RI receives funding from the European Union's: •Project SoBigData++ No: 871042 | Program: H2020-INFRAIA 2018/2019•Project SoBigData PPP No:101079043 | Program: HORIZON-INFRA-2021-DEV-02 •Project: "SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics"NextGenerationEU - National Recovery and Resilience Plan (Piano Nazionale di Ripresa e Resilienza, PNRR) - Prot. IR0000013 - Avviso n. 3264 del 28/12/2021