



Deliverable D3.7

## Sustainability Plan 3



## DOCUMENT INFORMATION

PROJECT	
PROJECT ACRONYM	SoBigData-PlusPlus
PROJECT TITLE	SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics
STARTING DATE	01/01/2020 (60 months)
ENDING DATE	31/12/2024
PROJECT WEBSITE	<a href="http://www.sobigdata.eu">http://www.sobigdata.eu</a>
TOPIC	INFRAIA-01-2018-2019 Integrating Activities for Advanced Communities
GRANT AGREEMENT N.	871042
DELIVERABLE INFORMATION	
WORK PACKAGE	WP3 - NA2 - Dissemination, Impact, and Sustainability
WORK PACKAGE LEADER	USFD
WORK PACKAGE PARTICIPANTS	All
DELIVERABLE NUMBER and TITLE	D3.7 Sustainability Plan 3
AUTHOR(S)	Valerio Grossi (CNR), Roberto Trasarti (CNR), Michela Natilli (CNR), Ilaria Barsanti (CNR), Marco Braghieri (KCL)
CONTRIBUTOR(S)	
EDITOR(S)	Valerio Grossi (CNR), Michela Natilli (CNR)
REVIEWER(S)	Kalina Bontcheva (USFD), Joanna Wright (USFD)
CONTRACTUAL DELIVERY DATE	31/12/2024
ACTUAL DELIVERY DATE	03/01/2025
VERSION	1.1
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	23
KEYWORDS	Long-term sustainability, ESFRI, legal framework, financial aspects, business model, cost book, ERIC

## EXECUTIVE SUMMARY

This deliverable completes the sustainability plan and strategy for the future development of the project results. The document updates the previous deliverables D3.5 - Sustainability Plan 1 and D3.6 - Sustainability Plan 2. It describes how the original task T3.5 “Sustainability Planning” changed through time with the entrance of the research infrastructure in the ESFRI RoadMap 2021 and the consequent more ambitious objective of creating an ERIC. The document reports how the SoBigData++ project supported this process.

## DISCLAIMER

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871042.

SoBigData++ strives to deliver a distributed, Pan-European, multi-disciplinary research infrastructure for big social data analytics, coupled with the consolidation of a cross-disciplinary European research community, aimed at using social mining and big data to understand the complexity of our contemporary, globally interconnected society. SoBigData++ is set to advance on such ambitious tasks thanks to SoBigData, the predecessor project that started this construction in 2015. Becoming an advanced community, SoBigData++ will strengthen its tools and services to empower researchers and innovators through a platform for the design and execution of large-scale social mining experiments.

This document contains information on SoBigData++ core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as SoBigData++ Board members. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The content of this publication is the sole responsibility of the SoBigData++ Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

Copyright © The SoBigData++ Consortium 2020. See <http://www.sobigdata.eu/> for details on the copyright holders.

For more information on the project, its partners and contributors please see <http://project.sobigdata.eu/>. You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: "Copyright © The SoBigData++ Consortium 2020."

The information contained in this document represents the views of the SoBigData++ Consortium as of the date they are published. The SoBigData++ Consortium does not guarantee that any information contained herein is error-free, or up to date. THE SoBigData++ CONSORTIUM MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

## GLOSSARY

EC	European Commission
ESFRI	European Strategy Forum on Research Infrastructures
ERIC	European Research Infrastructure Consortium
EU	European Union
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
MOOC	Massive Open Online Course
MS	Member State
PPP	Preparatory Phase Project

# TABLE OF CONTENTS

1	Relevance to SoBigData++ .....	7
1.1	Relevance to project objectives .....	7
1.2	Relation to other work packages .....	8
1.3	Structure of the document.....	8
2	SoBigData RI Objectives.....	9
2.1	SoBigData Services .....	10
2.2	Past and (next) Future Financing of the RI.....	13
3	Sustainability plan & SoBigData ERIC .....	15
3.1	Strategic vision of the RI.....	15
3.2	The establishment of the ERIC .....	18
3.2.1	Governing Board - Tentative structure .....	18
3.2.2	The relationship between the central hub and the national nodes .....	21
3.2.3	Access policy for users .....	22
4	Conclusions .....	23

## 1 Relevance to SoBigData++

This document updates and completes the description of the main actions performed in task T3.5 in the period from July 2023 to December 2024. During this period, many activities on the definition of the future governance of the RI were performed since SoBigData RI became part of the ESFRI RoadMap 2021. During the mentioned period all the partners were informed about the steps required to become an ERIC and the possible future scenarios. As already reported in deliverable D3.6 - Sustainability Plan 2<sup>1</sup>, the objective of creating the SoBigData Association has changed into the creation of the ERIC within the synergic work with the SoBigData RI PPP<sup>2</sup> and SoBigData.it<sup>3</sup>.

### 1.1 Relevance to project objectives

One of the key objectives of the SoBigData++ project was to reinforce and expand the SoBigData RI. This infrastructure is enabling innovative research in social mining, data science and AI. Achieving this objective requires not only technical and scientific excellence but also a strong focus on ensuring the sustainability and long-term sustainability. Careful and strategic planning has been essential to provide a clear and actionable vision for the future of the SoBigData RI.

From July 2023 to December 2024, SoBigData++ dedicated efforts to a series of activities aimed at securing the long-term success of the RI. All the partners of the consortiums refined the definition of the project's key exploitable results (KERs) and the services that each institution can provide to the infrastructure. This step was critical to ensure that the most valuable outputs of the project are effectively leveraged and can deliver lasting impact to the research community and society at large and was fundamental for the definition of the business aspect, the cost book and for the design overall RI. In this context, the SoBigData++ project consortium contributed on designing a comprehensive and robust financial plan for establishing the proposed legal entity under the ERIC framework. This financial plan was developed to ensure not only the immediate operational stability of the RI but also its long-term sustainability. This planning process included foresight activities to anticipate future funding needs and opportunities, ensuring that the RI remains adaptable in a rapidly evolving landscape.

Finally, significant attention was given to defining appropriate legal and governance solutions for the RI. The defined board were carefully crafted to uphold the original objectives and vision of the infrastructure while ensuring compliance with relevant legal and regulatory requirements. The governance model was designed to foster inclusivity, transparency, and effective decision-making, facilitating collaboration among stakeholders and ensuring alignment with the overarching goals of the SoBigData++ project.

---

<sup>1</sup> <https://data.d4science.net/EQtK>

<sup>2</sup> Project supported by the European Union under the scheme HORIZON-INFRA-2021-DEV-02-01 – Preparatory phase of new ESFRI research infrastructure projects, Grant Agreement n.101079043, “SoBigData RI PPP: SoBigData RI Preparatory Phase Project” – <https://ppp.sobigdata.eu/>

<sup>3</sup> 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 - <https://pnrr.sobigdata.it/>

Through these concerted efforts, SoBigData++ were fundamental for the creation for a sustainable and impactful RI that will continue to serve as a hub for innovation, collaboration, and excellence in social data science for years to come.

## 1.2 Relation to other work packages

Since the sustainability plan have the highest importance for the RI future development, there is a clear interaction between all work packages. In particular:

- WP3 T3.4 Outreach to Policy Makers and the Public at large: reaching relevant stakeholders, including policymakers, MEPs, civil society organizations, NGOs, government officials, and representatives of other organizations, is considered an act of the highest importance for the future of the RI.
- WP5 - Accelerating Innovation, where, in T5.1, partnerships with industry and contributions to policymaking are the core actions.
- WP9 E-Infrastructure and Supercomputing Network, where the goal is enhancing quantitatively and qualitatively the services offered by the SoBigData RI.

## 1.3 Structure of the document

This report contains three main sections:

- Section 2 outlines the SoBigData RI objectives and how the RI will interact with the projects funding the RI.
- Section 3 updates the long-term sustainability and outlines the sustainability plan.

## 2 SoBigData RI Objectives

Before introducing several technical aspects related to the sustainability of the RI, it is important to highlight that the sustainability plan is related to well-defined and established objectives that the RI will implement in the next year.

SoBigData++ has identified a set of objectives to strengthen the position RI as a leading research infrastructure in Europe and considering the European Research Area. These objectives are already driving SoBigData RI's development and operations of and can be summarized as follows:

- **Creating a Multi-disciplinary Scientific Community:** SoBigData RI brings together researchers from diverse fields, facilitating cross-disciplinary collaboration and the exchange of ideas. This community-driven approach enhances the scope and impact of research initiatives and increases the opportunity of creating high-impact and competitive project proposals for National and International calls.
- **Becoming the Reference Infrastructure in Europe for Big Data Analysis and Social Mining:** SoBigData RI aims to be the go-to infrastructure for big data analysis and social mining in Europe. By providing state-of-the-art tools and platforms, SoBigData RI supports data scientists and AI researchers in designing and developing innovative big data experiments in social sciences.
- **Supporting Enterprises in Data Exploitation:** in line with the EU Strategy for Data objectives, SoBigData RI helps enterprises to leverage data to create new products and services, explore new markets, reduce costs, and enhance profitability. By offering consulting, data analysis services, and tailored training courses, SoBigData RI empowers businesses to harness the power of data for strategic advantage and operational efficiency.
- **Being a Reference Point for Public Bodies and Policymakers:** SoBigData RI aspires to be a central resource for public bodies and policymakers on data management, AI, and their ethical use. By promoting ethical and legal standards in data practices and providing expert consulting services, SoBigData RI directs public entities toward concrete solutions that enhance governance, policy-making, and public service delivery.

To fulfil these objectives, SoBigData RI provides a wide range of services tailored to meet the needs of various stakeholders, including researchers, companies, and public institutions. SoBigData++ played a key role on the definition of these services, that now are designed to enhance research capabilities, promote technological transfer, and facilitate the creation of new business opportunities, all while adhering to ethical and legal standards.

Another important aspect, that guided the design of the RI sustainability plan (and that will play a key role in the next future) is the growing importance of the data access. Data access is essential for scientific progress and societal advancement. However, alongside data accessibility's benefits comes a critical responsibility to safeguard individuals' privacy and maintain the integrity of sensitive information.

SoBigData++ committed to embedding these principles in the development of innovative technologies. SoBigData RI incorporates ethical guidelines and legal compliance measures in all its data-related activities. This includes obtaining informed consent from data subjects, ensuring data anonymization, and implementing secure data storage and access protocols. By adhering to the General Data Protection Regulation (GDPR) and other relevant legal standards, SoBigData RI ensures that data handling practices meet the highest ethical and legal benchmarks.

SoBigData is implementing this vision by facilitating responsible data access through several key initiatives and practices:

- **Awareness and Education:** SoBigData actively promotes awareness about the ethical and legal implications of data usage. Through workshops, webinars, and guidelines, researchers are equipped with the knowledge to handle data responsibly, ensuring compliance with regulations like GDPR and maintaining ethical standards.
- **Data Infrastructure:** by providing a robust data infrastructure, SoBigData simplifies access (online and on-site) to vast and complex datasets.
- **Security Measures:** SoBigData includes encryption protocols, access controls, and anonymization techniques to protect against unauthorised access and breaches.
- **Privacy Preservation:** SoBigData employs cutting-edge techniques to ensure that personal information remains protected while still allowing meaningful analysis.
- **Collaboration and Governance:** SoBigData fosters a collaborative environment where researchers, institutions, and data providers work together transparently.

Synergies with EOSC and European Data Spaces: SoBigData++ maintains a constant connection with EOSC workgroups to develop commons and define data and services sharing policies. Moreover, the RI community participates in various data space initiatives and projects.

## 2.1 SoBigData Services

To achieve its objectives, SoBigData RI, in collaboration with the European network of institutions that are part of the research infrastructure, offers a wide range of services tailored to the needs of different stakeholders—researchers, businesses, and public institutions. These services are designed to enhance research capabilities, promote technology transfer, and facilitate the creation of new commercial opportunities, all while complying with ethical and legal standards.

The services provided have been carefully selected and designed to strengthen research capabilities by offering access to cutting-edge technologies, state-of-the-art facilities, and expert support. These aspects enable researchers to conduct high-quality studies and achieve breakthrough discoveries. Additionally, another pivotal role is in technology transfer by facilitating the translation of scientific findings into practical applications, thereby bridging the gap between research and industry. Moreover, the RI will support the creation of new business opportunities by promoting entrepreneurship, nurturing startups, and encouraging collaboration between academia and industry.

### *Technological Services*

The technological services of the RI node aim to provide a robust, scalable, and secure infrastructure for accessing, processing, and storing large-scale data. They include:

- **Cloud Computing Infrastructures:** access to high-performance computing resources for processing large volumes of data and running complex algorithms. Specifically, SoBigData offers three types of access: through an online Jupiter Notebook, via the Galaxy framework, and using a Cloud Computing Platform (CCP).

- **Containers and Virtualization:** it is possible to request the creation of virtual environments defined through containers, ensuring flexibility, resource isolation, and adaptability to the diverse needs of individual projects.
- **Storage and Data Management Services:** solutions for data archiving, geographic replication, versioning, and secure data management, with particular attention to regulatory compliance and protection of sensitive data

### *Resource-Sharing Services for Social Mining*

Application services focus on offering ready-to-use data and tools, integrated environments, and vertical solutions supporting specific research and innovation activities:

- **Access to Datasets and Thematic Catalogues:** provision of certified, high-quality data collections from multiple domains (e.g., research spaces), complete with metadata and documentation provided by publishers.
- **Advanced Analytics Tools (AI, Machine Learning, Data Mining):** ready-to-use libraries, frameworks, and applications for predictive modelling, knowledge extraction, interactive visualization, and natural language processing.
- **Vertical Applications for Specific Domains:** preconfigured platforms and tools for analysing social, economic, health, environmental, or cultural data, integrating models, indicators, and dashboards adapted to the needs of various sectors.

### *Support Services for Research Communities and Projects*

The research support services are designed to facilitate scientific activities and the work of experimenters, providing expertise, methodologies, and tools:

- **Policies for Data Management According to FAIR Principles:** support in defining data analysis strategies, selecting statistical and machine learning methods, and adopting best practices for dataset collection and documentation.
- **Data Management Platform:** by allowing data publication in its resource catalogue, SoBigData enables European and national projects to easily manage and publish their data, as well as obtain information via APIs for their periodic data management plans.
- **Ethical-Legal Support for Projects:** the RI offers a dedicated consulting and support service to help researchers and scientific communities address ethical and legal aspects related to accessing and using Big Data and Artificial Intelligence in research and innovation projects. Thanks to an expert committee on data protection, privacy, licensing, intellectual property, and research ethics, the RI can provide opinions and evaluations.
- **Creation of Virtual Research Environments (VREs):** establishment of virtual research environments for communities and events. These VREs provide tools for sharing data sources and results, in addition to all the technological services listed above. VREs can be public or restricted to a list of authorized users.

### *Services for Policy Makers*

The RI offers ad-hoc services to support policy makers, public administrations, and organizations in designing, evaluating, and monitoring data-driven policies. Through multidisciplinary expertise, cutting-edge technologies, and an ethical approach, the infrastructure allows policy makers to access reliable information, forecasting models, and interactive decision-support tools.

- **Evidence-Based Analyses:** the RI has long collaborated with regional and national institutions in processing complex data (social, economic, environmental, health) to provide indicators, metrics, and insights useful for defining effective and targeted policies.
- **Ex-ante and Ex-post Impact Assessment:** modelling and simulation of policy scenarios, assessment of expected outcomes and intervention effectiveness, and identification of potential critical issues or unforeseen effects. Examples of such collaborations include mobility modelling for metropolitan cities or optimization of health resources at the regional level.

### *Services for Enterprises*

The RI offers businesses a set of initiatives designed to facilitate access to advanced technologies, high-quality data, interdisciplinary expertise, and strategic consulting in the field of big data analytics and artificial intelligence. These services aim to support innovation processes and create added value for companies, regardless of their sector.

- **Master in Big Data:** a second level master's program for capacity building targeted at enterprises. The central hub of the RI manages and implements this Master in Big Data, designed for professionals and recent graduates. The program offers comprehensive training on the most advanced technologies and methods for managing, analysing, and interpreting large volumes of data required by Industry 4.0.
- **Proof-of-Concepts for SMEs:** support for innovation through the creation of proof-of-concepts, intended to foster collaborations between researchers and industry, stimulate the development of innovative data-driven solutions, promote projects, and strengthen the visibility of participating companies.
- **Ad-Hoc Courses for Enterprises:** tailored courses for companies interested in exploring specific topics. Thanks to its network across Italy, the node can involve experts in the desired field to meet the enterprise's needs (e.g., social network data analysis, the use of technologies for distributed computing, or the use of large language models).

### *Training Services*

The RI is particularly committed to training new skills and continuously updating researchers, analysts, and industry professionals:

- **Specialized Courses and Workshops:** training events on data science, artificial intelligence, data security, privacy-by-design, data lifecycle management, ethics, and accountability.
- **Summer School:** intensive programs with international experts, dedicated to exploring emerging topics and learning how to use the tools provided by the infrastructure.

- **SoBigData Academy:** the RI has developed MOOCs within the SoBigData Academy to train the new generation of responsible data scientists. These courses follow a self-learning educational methodology.

Additionally, the RI has implemented a mobility program based on “transnational access” (TNA), allowing international or national institutions and researchers to access various nodes of the research infrastructure through cross-border cooperation agreements. This type of access encourages participation in national and European projects, promoting the exchange of expertise, data, and advanced methodologies. Transnational access provides visitors with the opportunity to discover other research environments and, above all, to gain access to resources that may be subject to certain sharing and/or data transfer restrictions.

## 2.2 Past and (next) Future Financing of the RI

Figure 2.2.1 details time schedule of funding, including both historical and projections of RI funding for the coming years. This timeline provides a comprehensive overview of the financial trajectory of the RI, highlighting the secured funding from previous years and the anticipated resources required to sustain and expand operations in the future.

As part of this forward-looking planning, particular consideration has been given to the potential availability of Fondo Ordinario per gli Enti e le Istituzioni di Ricerca (FOE)<sup>4</sup> funds for the year 2026:

1. The FOE inclusion will be active by the year 2026 during the last step of SoBigData ERIC establishment. This inclusion reflects a strategic approach to ensuring financial stability and leveraging potential funding opportunities to support the long-term goals of the RI.
2. FOE and the ESFRI implementation project will support the RI during the next phase (expected 2026 - 2028). This phase includes the establishment of the ERIC (planned for mid-2027).
3. The operational phase of the ERIC is expected from 2029 (onward when the implementation phase is completed), and will be covered by several aspects, included the fees of the member states, the involvement of the RI in National and European projects, by the collaboration with companies and public bodies and by the funds guaranteed by the services.

Moreover, the RI is planning to participate to the future competitive calls in the Horizon Europe Work Programme 2025 to gather additional fundings and plan new services and collaborations.

---

<sup>4</sup> Since the RI is coordinated by CNR, and it is the only one RI in the area digit of ESFRI coordinated by Italy, it can access to special fund called “Fondo Ordinario per gli Enti e le Istituzioni di Ricerca”.

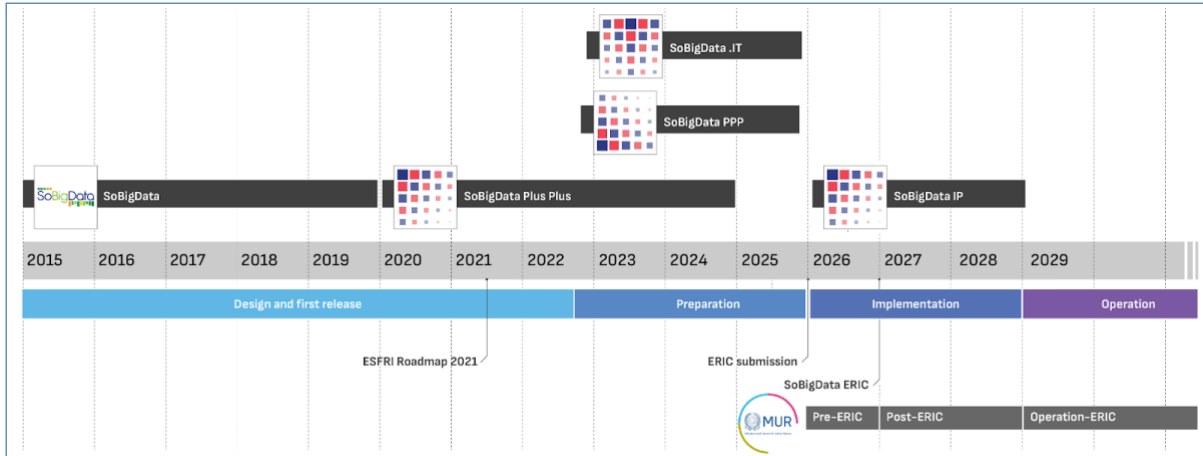


Figure 2.2.1 Time plan of funding

### 3 Sustainability plan & SoBigData ERIC

The time schedule shown in Section 2 represents a critical tool for aligning financial planning with the operational and developmental needs of the project, ensuring that adequate resources are allocated to achieve its objectives while maintaining adaptability to future funding landscapes. In brief we can state that, the SoBigData IP (Implementation project) will have the following aims:

- **ERIC Submission:** the submission process for establishing the ERIC is a critical milestone in formalizing the governance and operational framework of the SoBigData RI. This process will start in 2026, and it involves the preparation and submission of comprehensive documentation to the European Commission, detailing the objectives, structure, financial plan, and governance model of the RI. Achieving ERIC status will provide a robust legal framework for ensure long-term sustainability by enabling access to additional funding opportunities and formal recognition within the European research landscape (see Section 3.2 for details).
- **Services Implementation** is a cornerstone of the operational phase, aimed at delivering a comprehensive suite of tools, platforms, and support systems to meet the needs of the research and education community. This aspect includes deploying advanced social mining, AI and big data analytics services, ensuring seamless integration with existing infrastructure, and providing user-friendly interfaces for diverse stakeholders. A key focus will be on ensuring the scalability, reliability, and interoperability of these services, as well as maintaining robust user support mechanisms to foster widespread adoption and impactful research outcomes (more details in Section 3.1).
- **Start-Up of the Governing Bodies:** the establishment of the governing bodies marks the transition to a fully operational governance model, ensuring effective oversight and strategic direction for the SoBigData RI. This includes constituting key entities such as the General Assembly, the Scientific Advisory Board, and the Executive Board. These bodies will be responsible for decision-making, policy-setting, and operational management, ensuring alignment with the RI's mission and long-term objectives. The start-up phase involves defining roles, responsibilities, and workflows, as well as fostering collaborative dynamics among stakeholders to create a transparent, efficient, and inclusive governance structure (more details in Section 3.1.2).

#### 3.1 Strategic vision of the RI

During the transition towards the establishment of the ERIC, it is imperative to delineate the strategic vision and sustainable plan considering the various costs and revenues. This section presents the status of the work in defining the sustainability plan and proposes a projected budget balancing and listing the costs and services that will be enabled in the next phase the RI<sup>5</sup>.

The RI commitment to sustainability is multifaceted, focusing on **financial stability**, **resource optimization**, and **service scalability**. Projections for the budget reveal a well-balanced financial plan, ensuring that SoBigData ERIC remains robust and adaptable to the evolving demands of the research landscape. This financial strategy considers a combination of public funding, competitive grants, and possible other

---

<sup>5</sup> This work has been done in synergy with the projects SoBigData RI PPP and SoBigData.it. The costs reported here represent a preliminary study that will be refined during the 2025 and can be rediscussed and modified during the negotiations with the member state representatives.

partnerships, ensuring a diversified and resilient revenues stream. The future of SoBigData ERIC is characterized by an expansion of our service portfolio, designed to meet the needs of our user community and granting a boost to research development to the state members.

Achieving ERIC status will significantly enhance SoBigData RI's capacity to support high-impact research, foster international collaboration, and contribute to the European Research Area's strategic objectives. It will also strengthen SoBigData RI's role as a central hub for Social Mining and AI research, driving innovation and societal benefits across Europe.

Currently, there are 14 countries involved in SoBigData RI consortiums which are **potential nodes** of the future SoBigData ERIC, namely: Italy, UK, Sweden, Poland, Netherlands, Greece, France, Finland, Spain, Estonia, Germany, Bulgaria, Belgium and Austria. As stated in the previous section, according to the relationships of each node with their governmental bodies and their national roadmaps, from this set of potential nodes we expect to have two possible scenarios for initial set of members:

- Scenario 5 Countries - Italy, United Kingdom, Netherlands, Germany and Spain
- Scenario 7 Countries - Italy, United Kingdom, Netherlands, Germany, Spain, Greece and France

The other countries will probably join after the ERIC will be established progressively as observers or members. Based on those two scenarios, our projections for sustainability take into consideration and based on the following aspects:

- **Member Fee:** the minimum fee paid by a member is computed by a fixed amount (15.000 Euros) plus a variable part (0.010% of the 2021 GDP of the country). This amount is capped at 200.000 Euros.
- **Node In-Kind contribution:** each node (intended as the institutions in the country) will give an In-Kind contribution. The minimum is proportional to the Member State Fee paid<sup>6</sup>.
- **Services Cost:** each service to be provided by SoBigData ERIC was defined in terms of costs (Implementation, maintenance and provision, and scalability). From this list, we selected the most important and sustainable ones to be enabled and provided to the users. In the projections, we will show the overall costs for the ERIC to sustain them. This is one of the principal ways in which the institutions from member states will get funded.
- **Base Research Funding:** to fund the base research, we guarantee a payback of 25% of the member state fee to the institutions.
- **Governance costs:** according to the draft statute of the ERIC (see summary in annex 3), we identified a governance cost as the personnel required for the ERIC management. This includes 5 roles in the Board of Directors as well as a part-time administrative and a part-time researcher (0.5 FTE) for each member.
- **Operation and Maintenance of Technological Infrastructure:** the overall cost of maintaining the technological infrastructure and data centers.
- **Travels:** the cost for travel to be covered for coordination and management meetings.
- **Annual Events:** the cost of organizing annual events for dissemination and a plenary meeting.
- **Staff Training:** periodical training of the SoBigData ERIC personnel for innovations in the services and technological infrastructure.
- **Unexpected costs:** other possible costs not planned in the budget.

---

<sup>6</sup> It is important to notice that the Member Fees and In-Kind contributions are only the **minimum** amounts required. It doesn't exclude a bigger contribution from a member, which leads to a bigger involvement and political "power" in the General Assembly of the RI.

- **Project Participation:** the revenues coming from the participation in national and international calls of the SoBigData ERIC intended as a portion of funding dedicated to the maintenance of the infrastructure (not the research done by the single nodes, which will be funded for their research). In this, we include funds coming from the future ESFRI call for the implementation phase.
- **Service Market:** we envisage the possibility of asking our users for a fee for specific services (e.g., ad hoc training courses or specific consultancy services).

Tables 3.1.1 and 3.1.2 report our simulations consider the two potential scenarios. The proposed projections are meant to be indicative and represent the budget (one year) of the SoBigData ERIC planned to be established in the next phase of the RI.

Description	Costs	Description	Revenues
Governance Personnel			
- <i>General Director</i>	€99,000	Fees	€1,102,059
- <i>Infrastructure Manager</i>	€99,000	In-Kind Administrative	€330,000
- <i>Chief Financial Officer</i>	€82,500	In-Kind Researchers	€1,089,000
- <i>Chief Technology Officer</i>	€82,500	Project Participations	€200,000
- <i>Chief Legal Officer</i>	€82,500	Service Market	€10,000
- <i>Other Personnel</i>	€198,000	In-Kind Infrastructure maintenance	€550,000
- <i>Administrative from nodes</i>	€132,000		
- <i>Researchers PI from nodes</i>	€132,000		
Operation and Update of Technological Infrastructure	€140,000		
Maintenance of Infrastructures	€700,000		
Travels	€18,000		
Annual Events	€30,000		
Node Basic Research fundings (25% fee)	€275,515		
Internal Training (HR training)	€20,000		
Services	€1,154,944		
Unexpected costs	€32,000		
<b>Total costs</b>	<b>€3,277,959</b>	<b>Total revenues</b>	<b>€3,281,059</b>
Balance	<b>€3,100</b>		

Table 3.1.1 - Simulation 5 countries - Italy, United Kingdom, Netherlands, Germany and Spain

Description	Costs	Description	Revenues
Governance Personnel			
- <i>General Director</i>	€99,000	Fees	€1,522,438
- <i>Infrastructure Manager</i>	€99,000	In-Kind Administrative	€415,800
- <i>Chief Financial Officer</i>	€82,500	In-Kind Researchers	€1,339,800
- <i>Chief Technology Officer</i>	€82,500	Project Participations	€260,000
- <i>Chief Legal Officer</i>	€82,500	Service Market	€17,500
- <i>Other Personnel</i>	€297,000	In-Kind Infrastructure maintenance	€550,000
- <i>Administrative from nodes</i>	€231,000		
- <i>Researchers PI from nodes</i>	€231,000		
Operation and Update of Technological Infrastructure	€170,000		
Maintenance of Infrastructures	€700,000		
Travels	€24,000		
Annual Events	€30,000		
Node Basic Research fundings (25% fee)	€380,609		
Internal Training (HR training)	€35,000		
Services	€1,383,304		
Unexpected costs	€56,000		
<b>Total costs</b>	<b>€3,983,413</b>	<b>Total revenues</b>	<b>€4,105,538</b>
Balance	<b>€122,124</b>		

*Table 3.1.2 - Simulation 7 countries - Italy, United Kingdom, Netherlands, Germany, Spain, Greece and France*

### 3.2 The establishment of the ERIC

Since 2021, SoBigData RI has been following a strategic plan outlined in the European Strategy Forum on Research Infrastructures (ESFRI) roadmap to become an ERIC by 2026. ESFRI is a key framework guiding the development and integration of research infrastructures across Europe, ensuring they meet high standards of excellence, sustainability, and innovation.

The ERIC is a specific legal form that facilitates the establishment and operation of Research Infrastructures with European interest. This form has been selected by the SoBigData consortium as legal for the sustainable RI related to the entrance into ESFRI RoadMap 2021. Regulation 723/2009 defines that the activities, the geographical scope, the rules for membership and their implications (including contributions and benefits), the governance, and the seat are part of the statute of ERIC.

In the following we describe the preliminary work done on the definition of the governing bodies and policies done for the future sustainability.

#### 3.2.1 Governing Board - Tentative structure

The governance of the future SoBigData ERIC is currently being defined in line with the regulation track for developing ERICs. To become a node, each country (Member State of the Union; associated country; third country other than associated country or intergovernmental organization) must be a member. Each member will appoint one Representing Entity to be part of the General Assembly, which will be the supreme governing

body of the SoBigData ERIC. This means that being part of SoBigData ERIC will lead to becoming an active participating entity, influencing both the scientific direction and the innovation investments at a European and node/national level.

#### 3.2.1.1 General Assembly

The General Assembly (GA) will be the highest decision-making body and includes one or two representatives for Member countries. The GA is the body that is responsible for approving strategic priorities, taking decisions on the budget on policies, and making statute changes. It convenes at least annually. The weight of the votes will also be dependent on the contribution of the node. The GA in its operational rules will also decide on the quorum for appointing the Board of Directors. The final statutes will envisage the general rule for a quorum for decisions and eventual special higher or lower quorum. Different quorums can consider both membership and contribution to the ERIC.

#### 3.2.1.2 Board of Directors

The Board of Directors will be led by a General Director who will manage the daily operations, finances, and administration of SoBigData ERIC. The General Director will be supported by a professional team overseeing various functions such as finance, administration, and communication.

This General Director is the legal representative of the ERIC. The Board of Directors votes based on the rules established by the General Assembly. The ERIC will reserve the right to create new advisory bodies, Virtual Competence Centres, and Working Groups if deemed necessary and pursuant to the Operational Rules.

#### 3.2.1.3 Board Roles

A tentatively and a non-exhaustive list of the roles could be:

- *General Director*: this role is responsible for the overall leadership and strategic direction of the ERIC. They manage relationships with stakeholders, represent the ERIC externally, and ensure the achievement of the consortium's objectives. They manage the scientific aspects of the ERIC, including VCC, and collaborate with the scientific community to design specific objectives. The General Director will also liaise with the National Nodes Council (NNC).
- *Infrastructure Manager* assists the General Director in discharging the duties and may act as a second-in-command. They could assume responsibilities for specific topics as delegated by the General Director.
- *Chief Financial Officer* oversees administrative, financial, and operational functions of the ERIC, including human resources and general administration. They will manage day-to-day operations to ensure efficiency.
- *Chief Technology Officer* manages the technological aspects of the ERIC, including research infrastructure IT systems, and data management. The Chief Technology Officer will collaborate with the scientific community to align technological capabilities with research needs.
- *Chief Legal Officer* provides legal advice and support to the ERIC, ensuring compliance with relevant regulations and statutes. The Chief Legal Officer manages legal affairs, contracts, and agreements and will also act as Secretary General of the Board.
- *Chief Communications Officer* manages the internal and external communications for the ERIC. The Chief Communications Officer will ensure effective communication with member countries, stakeholders, and the public. They will lead the Communications Group.

- *Chief Acquisitions Officer* is responsible for the engagement of new institutions and countries as well as the procurement of new datasets and methods to be integrated into the RI for new services. The Chief Acquisitions Officer will also be responsible for creating consortiums to participate in European and national calls among the members (priority) and associated and cooperative partners and lead the Acquisitions Group.
- *Training Officer* is responsible for the production and quality of the training material and training events and courses provided by the RI. It is led by a training chief. This role will be nominated by the Board of Directors.

#### Other bodies

- *Advisory Board* is nominated by the General Assembly and it is composed of external experts with the aim of advising the ERIC. It should meet once per year and produce a report.
- *National Nodes Council (NNC)* is composed of the scientific representatives of the state members and associates. Its role will be to advise and prepare the topics of discussion for the General Assembly. It will also propose the composition of the Board of Directors to the General Assembly. It will be elected by the General Assembly.
- *Scientific Advisory Board*: SoBigData RI will establish a Scientific Advisory Board composed of leading experts in digital services research. The Scientific Advisory Board will advise on scientific priorities, research quality, and emerging trends, contributing to the strategic direction of the ERIC. It will be formed by experts selected from the SoBigData members. Efforts will be made to ensure diversity in the expertise represented on the Scientific Board, as well as gender equality and cultural representation. This may include experts from various scientific disciplines, different geographical regions, and different sectors. The Scientific Advisory Board will be proposed by the Board of Directors to the General Assembly for approval and led by the General Director. The Scientific Advisory Board (in cooperation with the Technical Advisory Committee) shall evaluate proposals for new partner facilities, and the operation of existing ones, advising the General Assembly on acceptance and continuation.
- *Administrative and Financial Group*: this group oversees the consortium's financial aspects, ensuring compliance with budgetary decisions, financial sustainability, and transparency. The Chief Financial Officer will lead it.
- *Technical Advisory Committee* oversees the consortium's technical aspects, ensuring compliance with budgetary decisions and RI requirements and standards to provide services. The Chief Technology Officer will lead it. The Technical Advisory Committee (in cooperation with the Scientific Advisory Board) shall evaluate proposals for new partner facilities, and the operation of existing ones, advising the General Assembly on acceptance and continuation.
- *Communication Group* manages internal and external communications for the ERIC. They will ensure effective communication with member countries, stakeholders, and the public. The Chief Communications Officer will lead it.
- *Virtual Competency Centre (VCC)* is a virtual team, made up of people from partner institutions and eventual cooperating partners, focused on a specific research or technological topic, and will represent a cross-node body. The VCC must report its activities to the Scientific Board and Technical Committee periodically.
- *Working Groups*: In addition to VCC, where needed, the Chair of the General Assembly or the General Director may establish working groups to analyse or follow specific issues for which a special effort is needed, which cannot be performed by existing bodies alone.

These bodies will be established by the Board of Directors and confirmed by the General Assembly. The Board of Directors may leverage on external experts for consultancy on specific aspects such as an ethical, legal, or business evaluation.

### 3.2.2 The relationship between the central hub and the national nodes

A **Node** of the infrastructure is a composition of local sites (institutions) in the same country with political and financial support from the local government. Each node is a **service provider** which maintains the responsibility for the provision, quality, and intellectual property of its services. The central hub is the node that also has the responsibility of providing a unified access point to those services.

In this context, a service can be defined as a specialized offering that supports and enhances the capabilities of the RI enabling its users in conducting digital research. Such a service aims to provide the necessary tools, resources, and support to facilitate efficient and effective digital research activities (e.g. infrastructure facilities, training courses, AI algorithms, etc.) A service is defined by its characteristics: target users, modality (online or on-site) if requires expert support or not, and its costs.

Each node will provide its set of unique services and is committed in adopting the required RI standards and policies suggested by the central hub to the Board of Directors and approved by the General Assembly. The central hub monitors and evaluates the services usage suggesting a strategical roadmap to the administrative and financial group and Board of Directors in terms of service provision. Some services can be developed and maintained across different nodes, in this case, an agreement between them defining responsibilities is required (e.g., intellectual property, costs, etc.). Under these premises, the nodes are heterogeneous in terms of both social mining, AI, and computational services. To guarantee a minimum level of quality and long-term sustainability, the following requirements are defined:

- *Administrative requirements:* the requirement from the administrative point of view is to ensure SoBigData RI (at least) one Full-Time Equivalent (FTE) administrative personnel per year for managing the financial and organizational aspects of the node and linking with the central offices to develop strategies and implement policies at the national level.
- *Research requirements:* the requirement from the research point of view is to ensure that SoBigData RI (at least) has one FTE research personnel per year to monitor the activities of the node and develop common strategies on the RI research lines. This research unit will also be responsible for participation in European, national, and internal RI projects, which are critical for obtaining additional funding.
- *Technological requirements:* the nodes have the possibility of adding their data centres, they must respect technological standard for interoperability required by the RI.

Countries, and their institutions, that join the SoBigData RI network gain access to a wealth of opportunities and resources that enhance their research and innovation capabilities. Being part of this prestigious network offers several key advantages:

- *Access to an Advanced Infrastructure:* nodes will benefit from state-of-the-art data analysis tools and platforms, enabling cutting-edge research and technological innovation.
- *Collaboration Opportunities:* nodes will collaborate with leading researchers and industry experts, fostering interdisciplinary partnerships and joint projects.

- *Funding and Grants*: being part of the SoBigData network increases eligibility for European funding opportunities and research grants, providing financial support for high-impact projects. It enables the participation of dedicated calls for RIs from the European Commission.
- *Influence in European Research Agendas*: nodes can actively shape European research agendas and policymaking, ensuring that the institution's research priorities align with broader European goals.
- *Capacity Building*: continuous training and development opportunities for researchers and staff, enhancing their skills and expertise in data science and AI.

Investing in SoBigData RI not only strengthens the capabilities of individual institutions and countries but also enhances their influence in European research initiatives and policymaking. Through active participation in European Calls, such as the European Research Area, SoBigData RI institutions can secure funding, shape research agendas, and drive strategic initiatives that align with European priorities. This investment return extends beyond financial gains, fostering deeper stakeholder engagement and strengthening the role of European institutions in global research and innovation networks.

SoBigData ERIC will sustain the scientific initiatives, services offered, and its governance mainly thanks to the contributions coming from Members. In addition, it will drive funding bids from European/National projects and collect the fees generated by some ad hoc services, which will be charged to end users.

### 3.2.3 Access policy for users

The tools and services offered by the RI shall in principle be freely available for use by the scientific and education community. This means to provide access, within the applicable legal and ethical framework, to resources and/or associated data under a common set of standards and conditions. Generally, data, tools, and services offered by SoBigData ERIC shall be open to all employees and students in research institutions such as universities, research centres, museums, and research libraries, according to the license terms of the content/service providers and through an agreed upon authentication and authorisation system.

Access to data produced by SoBigData ERIC shall, wherever possible consider third-party licences and any pre-existing arrangements, be free and open to all members of scientific institutions and other stakeholders. SoBigData RI data access policy already follows the best international and European practices with respect to public data, such as those established by the European Union and shall recognize the rights and obligations of the data holders, and the RI will continue to encourage open access following the FAIR principles to research data, data products, services, and software, as well as research facilities. However, limitations can be envisaged to guarantee the protection of fundamental rights and liberties, individual rights, technological sovereignty of the EU and cyber-physical security.

## 4 Conclusions

This deliverable outlined the necessary next steps to ensure the sustainability of the SoBigData RI. The SoBigData++ consortium has served as a framework for generating new research initiatives, particularly in the areas of service definition and integration, knowledge transfer, and the facilitation of collaborative efforts across both public and private entities, emphasizing the importance of collaborative, interdisciplinary, and transnational perspectives in advancing research and fostering the effective archiving and utilization of data. Through such efforts, SoBigData++ aims to position the RI as a vital resource for tackling pressing societal challenges and promoting innovative, data-driven solutions.

In conclusion, we can state that the SoBigData RI is a critical asset in contemporary society, as it provides tools and platforms necessary to harness data's full potential. SoBigData RI drives innovation, economic growth, and societal advancement by fostering synergy between research and companies. Its role in enhancing Europe's leadership in Social Mining and AI underscores the strategic importance of investing in data infrastructure. By supporting SoBigData RI, each country and Europe will remain at the cutting edge of research and innovation, delivering tangible benefits to society and driving progress across multiple domains. Furthermore, the integration of ethical and legal standards in data analysis practices strengthens responsible and sustainable development of innovative technologies, safeguarding public trust and promoting long-term societal benefits.

Joining the SoBigData network as a node offers institutions opportunities to expand their research capacities, secure funding, and influence in the European research agendas. This affiliation not only enhances the institution's research but also elevates its position within the European research community, ensuring a lasting impact on the advancement of data science and AI.