



Deliverable D3.5

Sustainability Plan 1



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	http://www.sobigdata.eu
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	D3.5
DELIVERABLE TITLE	Sustainability Plan 1
AUTHOR(S)	Valerio Grossi (CNR), Roberto Trasarti (CNR), Beatrice Rapisarda (CNR)
CONTRIBUTOR(S)	Michela Natilli (CNR)
EDITOR(S)	Beatrice Rapisarda (CNR), Valerio Grossi (CNR)
REVIEWER(S)	Kalina Bontcheva (USFD), Alice Tucker (USFD)
CONTRACTUAL DELIVERY DATE	31/12/2021
ACTUAL DELIVERY DATE	26/04/2023
VERSION	V1.2
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	19
KEYWORDS	Sustainability, ESFRI, Legal Framework, Financial aspects

EXECUTIVE SUMMARY

This deliverable reports the sustainability plan and strategy for the future development of the project results. The former plan was to create a not-for-profit (NPO) entity at the end of the first year, which would have been named SoBigData Association, with the task, at the end of the third year, to investigate the possibility of transferring and integrating this legal body into the newly established SoBigData Foundation which should gather partners from Research, Industries, Public Sector, NGOs, Venture Capital with core project partners acting as founders. The entrance to ESFRI RoadMap2021 facilitates this aim and enables SoBigData RI to reach legal status. This document updates the original plan and reports the next steps that are required from ESFRI.

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
EU	European Union
ERDF	European Regional Development Fund
ESFRI	European Strategy Forum on Research Infrastructures
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
MS	Member State
PPP	Preparatory Phase Project

TABLE OF CONTENTS

- 1 Relevance to SoBigData++ 7
 - 1.1 Purpose of this document 7
 - 1.2 Relevance to project objectives 7
 - 1.3 Relation to other work packages..... 7
 - 1.4 Structure of the document..... 7
- 2 European Strategy Forum on Research Infrastructures (ESFRI) 8
 - 2.1 Lifecycle approach of RIs 8
- 3 SoBigData RI Long-term Sustainability..... 11
 - 3.1 SoBigData RI Preparation Phase Project (PPP) 13
 - 3.2 SoBigData RI PPP Impacts..... 15
- 4 Conclusions 18
- References..... 19

1 Relevance to SoBigData++

1.1 Purpose of this document

The effective investment in and use of RIs is one of the priorities for realizing the European Research Area (ERA). This deliverable outlines the main SoBigData++ consortium actions to guarantee accurate planning of the RI. The aspects considered include: *i)* **the Legal and Financial Stability**. This aspect covers all the activities for the management, the creation of legal and financial stability of the RI (including the preparation of the ERIC and the Business Plan). *ii)* **Sustainable Operation**, i.e., the definition of strategies for improving the impact of the RI in the European scenario and the creation of a competitive service catalog. *iii)* **the Technical Maturity** with the definition of technical standards and governance rules for the organization of the RI at the national and international levels.

1.2 Relevance to project objectives

One of the main SoBigData++ objectives is to consolidate the SoBigData RI, a platform for designing and executing large-scale social mining experiments. Careful planning of the sustainability actions is of the highest importance to providing a clear vision of the future of the SoBigData RI and the future development of the project results.

1.3 Relation to other work packages

Being the sustainability plan of highest importance for the project's 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 action 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.

1.4 Structure of the document

This report contains two main sections:

- Section 2 provides an overview of the lifecycle approach of RIs.
- Section 3 outlines the plan for SoBigData RI long term sustainability.

2 European Strategy Forum on Research Infrastructures (ESFRI)

2.1 Lifecycle approach of RIs

Research Infrastructures (RIs), from large facilities such as CERN to distributed data networks, play a vital role in Research and Innovation. Recent years have seen strong growth in the number of RIs that are operational or planned across Europe. While these RIs offer great opportunities, they also present the challenge of ensuring that they can be operated sustainably at a high level. As reported in [1], “history shows that a robust long-term vision is the most important prerequisite to successfully and sustainably build and operate a RI.” RIs are typically operational for several decades, requiring continuous and stable support. Sufficient time and support must be given to the RI to unfold and develop its full potential fully. This support cannot be reduced to financial considerations alone, though very important, but rather be founded on a broader consensus (nationally and EU-wide) as it is typically well beyond any electoral or standard budgetary planning period. The effective investment in and use of RIs is one of the priorities for realizing the European Research Area (ERA).



Figure 2.1.1 Lifecycle approach of RIs

To summarize, we can state that RI sustainability requires **stability**, **continuity**, and **predictability**, including robust and flexible governance. As in the case of SoBigData, a new RI typically emerges from the scientific communities clustering around well-identified scientific needs and goals. Such a concept can originate from completely novel approaches to answer scientific questions or respond to the need for enhanced capacity at pan- European. Figure 2.1.1 reports the phases recognized by ESFRI for defining and guaranteeing long-term sustainability (LTS) to a RI. As reported in [2], the defined phases include:

- *DESIGN*: it covers the proof of the scientific concept and technical feasibility of the RI, the analysis of the potential user community – both science and innovation-oriented; the outline of a business case, and the rationale for the international consortium. The feasibility study can be conducted with institutional, national, or international support. The design also includes an initial analysis of its position in the RI landscape, e-Infrastructure requirements, and (open)-data management and policy. Importantly, the RI also foresees government and funding agencies' financial and political support necessary for the Preparatory Phase.
- *PREPARATION* – carried out at institutional, national, European, or international level – is directed towards developing the RI fully-edged organization. Completion of preparation for the RIs in the Roadmap is often carried out through a **Preparatory Phase contract** under FP, resulting in a business plan, a legal entity, an agreed role for the RI also in the context of the landscape of existing RIs at European and global level, and secured funding safeguarding the financial sustainability for the Implementation Phase and also extending for the Operation Phase. Some projects face a gap of funding between the end of their Preparatory Phase contract and the final decisions for implementation – legal, funding, and construction – which can lead to the establishment of ad hoc interim legal entities and governance to ensure appropriate funding to complete the preparation and start construction.
- *IMPLEMENTATION* is different for single-sited and distributed RIs. In the first case, it corresponds to an intense investment period of several years for construction engaging human and financial resources with a big impact on the market – suppliers of goods and technologies. Longer-term benefits are generated to the hosting territory: employment, upgrade of services, internationalization, and up-skilling of the population, increased demand on high-level services – schools, communication, financial services for international employees – and joint development of novel technologies that remain as a competitiveness legacy to the procuring firms. In case of distributed RIs, the implementation implies intense negotiations as both the Central Hub and the national nodes require specific commitments. Developing a successful governance and management structure may be of higher complexity than for single-sited RIs. Nevertheless, distributed RIs have been quite efficient at establishing their legal entities and launching services to the user community in several cases.
- During their *OPERATION*, RIs produce frontier research and deliver advanced services for excellent science satisfying the users' demand, boosting brain circulation of early-career scientists and trainees, therefore improving the ranking of their academic and research institutions. RIs can create spin-offs and start-ups and attract corporate partners generating a high potential for innovation. The operational costs of RIs range from 8 to 12% of the initial capital investment per year. A twenty-year operation cycle may develop before major upgrades, requiring new substantial capital investment, are needed. The upgrade cycles in the case of e-Infrastructures are typically much shorter.
- The *TERMINATION* may encompass dissolution of the organization, the dismantling of facilities, related safety aspects, and resurrection of the original site, but it does not apply in these identical terms in all research domains. The Termination Phase could also result in new infrastructure development as part of the field's evolution. Re-orientation of RI sites has already occurred, e.g., in nuclear research or high-energy physics, where outdated RI have been transformed into analytical facilities with new science missions built upon technological infrastructure, logistics, and human resources organization.

- 1** **Establish and maintain excellence** through the entire lifecycle of RIs by all appropriate means, by securing adequate framework conditions, and by opening the RIs up to the world.
- 2** **Ensure that RIs have the right people in the right place at the right time** by strengthening and harmonising national research and educational systems to make sure that all essential skills are available.
- 3** Harmonise and **integrate a vision for convergent operation of RIs and e-Infrastructures** in Europe to ensure cost-effective service provision to the user communities.
- 4** **Fully exploit the potential of RIs as innovation hubs** by incorporating strategies for their development into national and European innovation policies.
- 5** Set up effective means of **determining the economic and wider social value of RIs**, and incorporate these benefits into science-policy-society dialogues.
- 6** Establish adequate framework conditions **for effective governance and sustainable long-term funding for RIs at every stage in their lifecycle**, together with effective management.
- 7** Foster broader **coordination at National and European levels** when designing processes for planning and supporting national and pan European RIs and so enhance their strategic value.

Figure 2.1.2 ESFRI Main Recommendations for LTS of an RI

The recommendations for LTS reported in Figure 2.1.2 are also relevant for international organizations or significant national RIs with meta-regional outreach. Similarly, some of the recommendations go a little beyond traditional research agendas, as strong overlaps exist with several other policy domains – for example, education, cohesion, competitiveness, and social policy. Finally, Figure 2.1.3 reports the list of the new RI project selected in 2021.

List of new pan-European RI Projects

			Building Europe's Research Infrastructure ROADMAP 2021
	ACRONYM	TITLE	POLITICAL SUPPORT
DIGIT	EBRAINS	European Brain ReseArch INfrastructureS	FR, BG, EL, IT, NL, NO, ES, SE, CH, DK
	SLICES	Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies	FR, CY, FI, EL, IT, LU, NL, ES, CH, PL
	SoBigData++ RI	European Integrated Infrastructure for Social Mining and Big Data Analytics	IT, EE, CH
ENE	MARINERG-I	Offshore Renewable Energy Research Infrastructure	IE, ES, PT, BE, UK
H F	EIRENE RI	Research Infrastructure for EnvIRonmental Exposure assessment in Europe	CZ, AT, NL, SK, EL, IT, IS, DE
PSE	ET	Einstein Telescope	IT, ES, NL, BE, PL
	EuPRAXIA	European Plasma Research Accelerator with Excellence in Applications	IT, PT, HU, CZ, UK
SCI	GGP	The Generations and Gender Programme	NL, HR, EE, DE, HU, ES, SE, PL, AT
	GUIDE	Growing Up in Digital Europe-EuroCohort	IE, HR, EE, LV, PT, ES, HU, UK
	OPERAS	Open Access in the European Research Area through Scholarly Communication	FR, HR, IT, PT, DE, EL, NL, PL, UK
	RESILIENCE	Religious Studies Infrastructure: Tools, Innovation, Experts, Connections and Centers	IT, IL, BA, BG, NL, AL, EL

Figure 2.1.3 Pan-European RI Project entered in ESFRI RoadMap 2021

3 SoBigData RI Long-term Sustainability

The European Strategy Forum on Research Infrastructures has selected SoBigData RI for the ESFRI Roadmap 2021. This ambitious recognition given to the most important AI and big data research infrastructure in Europe lays the groundwork to receive the largest funding for a research consortium dedicated to improving artificial intelligence for social good. ESFRI ensures long-term sustainability to SoBigData RI and opens new opportunities to grow and establish itself more and more as a reference point for research in AI and big data, not only in Europe but worldwide. Indeed, SoBigData aims to become a world-leading institution in its field. As reported in Figure 2.1.3 SoBigData RI is included in the DIGIT area with SLICES and EBRAINS.

SoBigData++ sent the ESFRI application in September 2020. To be eligible for ESFRI RoadMap 2021, SoBigData has gained political support from Italy, Estonia, Switzerland, and Bulgaria. The Italian Ministry of Research, the Italian National Research Council (CNR), and ten other European entities have been granted financial support, while the consortium comprises 27 partners. The proposal focuses on creating a Central Hub in Italy with ten nodes in the following countries: Netherlands, Estonia, Switzerland, Finland, Sweden, Austria, Germany, France, Spain, and the United Kingdom. At the moment of submission, the overall cost of the ESFRI SoBigData RI was estimated at more than 150 million €, which includes both the build-up and operational phase. The preparation phase started in 2020, and the RI will be operative until 2050. Figure 3.1 reports the costs and timeline expected to provide an LTS to SoBigData RI.

PHASE PERIOD	EURO (M€)	SUMMARY OF ACTIONS PERFORMED
DESIGN 2015 – 2020	7	This cost was partially covered by H2020 Projects SoBigData (GA. 654024) and SoBigData++ (GA. 871042), and national funds.
PREPARATION 2020 – 2024	10	This phase includes the cost of becoming a legal ERIC entity with a registered office in Italy. The consortium will decide both the final form of this legal entity and national funding from the partners during this phase. Expected €4M from the Preparatory Phase Project supported by ESFRI.
IMPLEMENTATION 2025 – 2029	12	SoBigData RI will be a distributed research infrastructure with different nodes in Europe. This cost covers the ERIC establishment, software development, physical spaces, competencies for implementing each national node, and the central hub.
AVERAGE ANNUAL OPERATION 2030 – 2050	5 (per year)	Participants cover this cost at various levels and support the national nodes of the RI (minimum expected: 50k per year, i.e., the equivalent of a one year of postdoc contract or the same in-kind contribution or in hardware and service provision). This cost is also covered by national funds and by the financial support of the countries that participate in the ERIC.
TERMINATION 2050	0,5	The average cost for dismantling each of 11 nodes can be around 40 – 50k per node. Since we do not have any physical dismantling, this cost includes the closure of legal entities plus the reconfiguration of hardware and software, data management, and relative competencies.
TOTAL INVESTMENT	129.5	

Figure 3.1 SoBigData ESFRI Phases and Costs

SoBigData++ Project fully supported the design phase of the RI requested. In early 2020, SoBigData++ Project delivered new tools, techniques, and methods, allowing researchers to execute analytical workflows, and to reproduce experiments, all within a robust legal and ethical framework. At the same time, the SoBigData++ has created a strong network of institutions and individuals from Europe and beyond that are committed to the SoBigData RI. These aspects enable us to provide a complete design of the future sustainable Ri, considering the design phase completed by the end of 2020. Figure 3.2 reports the initial planning of the long-term operation of the RI with the main ESFRI phases and the main outputs expected for each phase. As we can see from the figure the preparation phase started in 2021 with the acceptance of the ESFRI proposal that was entirely supported by SoBigData++. Due to the entrance into the RoadMap 2021, the EU launched a specific call to support the preparatory phase of new ESFRI research infrastructure projects in 2022. The complete design of the RI and the definition of the SoBigData RI PPP project has been done by SoBigData++ with the tasks related to RI sustainability and innovation. The ESFRI projects have been selected for their scientific case's excellence and strategic importance for the European Research Area and the structuring of the European research infrastructure ecosystem. SoBigData++ will apply to the call "Developing and consolidating the European research infrastructures landscape, maintaining global leadership (2021) (HORIZON-INFRA-2021-DEV-02)", with the type of action HORIZON-CSA HORIZON Coordination and Support Actions. The deadline is 20 January 2022. Project results are expected to contribute to several of the following expected outcomes: structuring effect on ERA; the scientific excellence of the European landscape of sustainable RI enhances problem-solving capacities to address challenges in science, industry, and society; solid ground for the decision-making on new research infrastructures, is available to MS/ACs, their funding bodies and other relevant stakeholders (e.g., international organizations, third countries; foundations; etc.); long-term perspective for RI investments; consistent and well-functioning European research infrastructures ecosystem through the development of synergies and complementarities between new and existing research infrastructures, including technology infrastructures and infrastructures financed by ERDF.

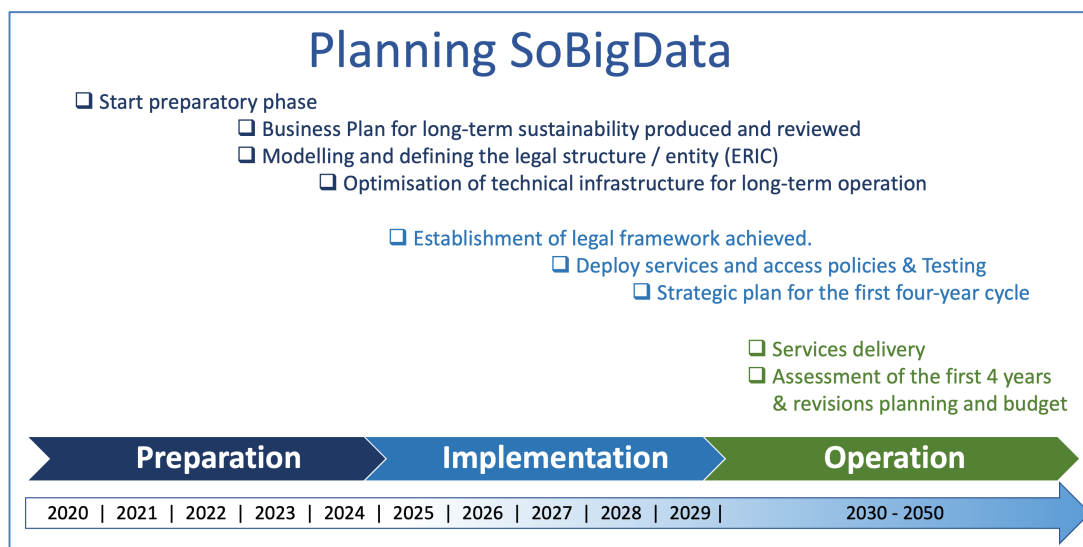


Figure 3.2 SoBigData RI planning

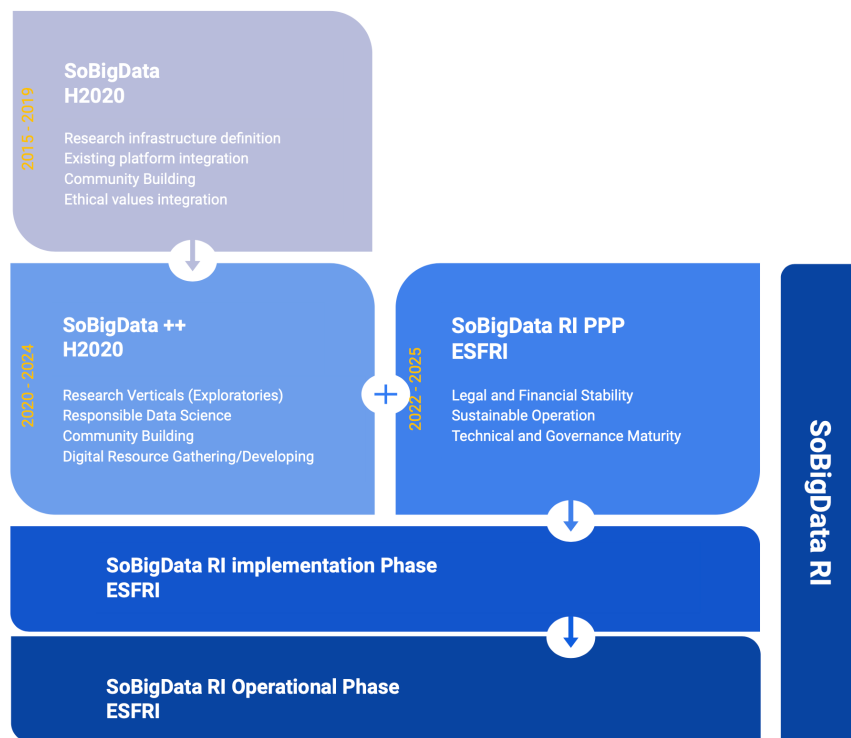


Figure 3.3 Overview of SoBigData RI construction and related projects

3.1 SoBigData RI Preparation Phase Project (PPP)

Figure 3.3 reports the projects that support the SoBigData RI, that with its tools and services, empowers researchers and innovators through a platform for the design and execution of large-scale data science and social mining experiments, open to users with diverse backgrounds, accessible on the cloud (aligned with EOSC guidelines), and exploiting supercomputing facilities. SoBigData RI will render social mining experiments more efficiently designed, adjusted, and repeatable by not data scientists' domain experts pushing the FAIR (Findable, Accessible, Interoperable) and FACT (Fair, Accountable, Confidential, and Transparent) principles. SoBigData RI will orient resources from multiple perspectives:

- E-infrastructures and online services developers.
- Big data analytics and AI.
- Complex systems focussed on modelling social phenomena.
- Ethics and legal aspects of data protection.
- Privacy-preserving techniques.

SoBigData RI PPP will move our RI forward from the simple awareness of ethical and legal challenges in social mining to the development of concrete tools that operationalize ethics with value-sensitive design, incorporating values and norms for privacy protection, fairness, transparency, and pluralism. The main objectives of SoBigData RI PPP are to define operative strategies for:

1. Modeling and definition of the ERIC legal entity, then **acquiring legal status**. The preparation phase of SoBigData RI will define and design all the aspects related to the definition of a European Research Infrastructure Consortium (ERIC) with a registered office in Italy. The first aim is to start all the negotiations required to become members and extend the countries supporting the RI. At the end of SoBigData RI PPP (September 2025), we expect to have defined the statutes, internal rules for implementation, and rules governing relations with users and external stakeholders, ensuring their alignment with HR policies and the selected governance structure. The legal status will be acquired at the beginning of the implementation phase, expected in 2026.
2. Preparing the **financial, legal aspects** (for both central hub and national nodes). With this objective, we plan to develop strategies for establishing partner agreements; to develop an effective and durable governance structure for both central and national hubs; to define the involvement of the Member States and Associated Countries in the management structure (including the Observer states). One of the actual aims in this context is to produce key legal documents (statutes, internal rules for implementation, and rules governing relations with users and external stakeholders) and a ready-to-sign agreement for stakeholder involvement to be used in the implementation phase (2026).
3. Producing and reviewing a **Business Plan** for long-term sustainability. Designing and engineering a formal business plan that describes the nature of our core business (related to RI services), background information on the organization, the RI financial projections, and all the strategies we intend to implement to achieve the stated targets. In the context, beyond the definition of the operational budget, we will produce a detailed cost book modeling the accesses and the services related to the SoBigData++ and the respective costs. Finally, we establish robust risk and performance management systems for the RI. The business plan definition will start with the beginning of SoBigData RI PPP and begin by analyzing and sharing the experience matured by SoBigData++. The business plan will be active starting from the implementation phase (2026).
4. Engineering, planning, and optimization of **technical infrastructure**. The organizational/technical challenge is to deliver state-of-the-art dynamic digital assets to remote sites without the need for expensive on-site expertise. From the technical point of view, the SoBigData RI will be based on the technologies developed during the SoBigData and SoBigData++ projects.
5. Defining strategies for **service design**, community involvement, and partnerships with third parties. In this context, the objective is to develop a sustainable data and method integration strategy that enables the discovery and use of heterogeneous services. For this reason, we will identify analysis and plans to identify and involve all stakeholders in the technical integration work of the RI. We will develop specific communication strategies to involve new stakeholders and disseminate and advise our service beyond our reference community and stakeholders. SoBigData++, particularly WP3, WP5, and WP7, will have strict collaboration in service design, stakeholder involvement, result dissemination, and boosting. The objective is to boost and transfer the research done in SoBigData++ to create long-lasting results and services, also creating durable and sustainable connections with EOSC, Service providers, Research Infrastructure, European projects, and training initiatives. SoBigData++ and SoBigData RI PPP will carry out these activities during 2023 and the first part of 2024, with the involvement of SoBigData++ in result boosting and innovation aspects.

These objectives pose several key challenges that must be addressed during the SoBigData RI PPP. Due to the multidisciplinary nature of our RI, we need to identify and develop sustainable services, responding to the demand for cross-disciplinary data-driven research and innovation. Furthermore, we need an open ecosystem with an adequate means for accessing big social data and algorithms for extracting knowledge from them without forgetting the aim of democratizing the benefits of data science within an ethical framework that harmonizes individual rights and collective interests. Another challenge is related to RI governance, which imposes the development of a practical and durable governance structure, paying particular attention to the relationship between the central hub and the national nodes, the involvement of the Member States and Associated Countries in this structure, (and the taking in of new partners). It also includes establishing robust systems for risk and performance management of the RI. Moreover, in this context, we also consider strategies for the mobility of knowledge and/or researchers within the ERA and the definition of channels of communication for the dissemination of the services and optimization of the results. The definition of legal framework implies challenges in creating agreements on the hosting of the central hub and its relationship to national nodes (and national sites); in preparing the ground to become ERIC, and finding financial support from countries involved in the consortium using specific agreements.

The key challenges SoBigData RI addresses are of such magnitude that they can only be approached from a long-term perspective. A sustained, well-coordinated, and pan-European effort is required to address this need, and a long-term RI with a pan-European outlook is best positioned to undertake it. At the moment, to the best of our knowledge, no research infrastructure covers the aims of SoBigData RI. The uniqueness of SoBigData RI is represented by its capability to make heterogeneous scientific communities on data science and AI a unique synergistic work. The above aspects make SoBigData RI a competitive and valid candidate to become a fundamental resource for the research on AI in Europe in line with the current requirements of the HE work program in AI, and its design has been conducted in close alignment with the European Research Area. SoBigData RI PPP consortium is the instrumental setting for new impulses regarding service integration, knowledge and expertise transfer, and collaborative efforts across disciplinary and geographic boundaries. Approaching these challenges is vital for using social mining and big data to understand the complexity of our contemporary, globally interconnected society research and archiving from collaborative, interdisciplinary, and transnational perspectives. The formalization of the existing links between the two past H2020 SoBigData project consortia into a long-term RI consortium offers a unique opportunity to deepen and widen the current collaboration making our objectives realistically achievable.

3.2 SoBigData RI PPP Impacts

SoBigData RI PPP will consider all the research results from the SoBigData++ Project inside the exploratories. It will select and assess the importance of some research lines (for each exploratory) to improve the quality of the results, the visibility, the usability in terms of resources available in the RI, and a strategic plan for the research activities in the next phases of the ESFRI lifecycle approach.

SoBigData RI has the ambition to support the rising demand for cross-disciplinary research and innovation on the multiple aspects of social complexity from combined data-driven and model-driven perspectives and the increasing importance of ethics and data scientists' responsibility as a pillar of trustworthy use of Big Data and analytical technology. The necessary starting point to tackle the challenges is to observe how our society

works. The big data originating from the digital breadcrumbs of human activities offer a considerable opportunity to scrutinize the ground truth of individual and collective behavior at an unprecedented detail and on a global scale. This increasing wealth of data is a chance to understand social complexity, provided we can rely on social mining, i.e., adequate means for accessing big social data together with models for extracting knowledge from them. SoBigData RI, with its tools and services, empowers researchers and innovators through a platform for the design and execution of large-scale social mining experiments, open to users with diverse backgrounds, accessible on the cloud (aligned with EOSC), and exploiting supercomputing facilities. Pushing the FAIR (Findable, Accessible, Interoperable) and FACT (Fair, Accountable, Confidential and Transparent) principles further, SoBigData RI will render social mining experiments more easily designed, adjusted, and repeatable by domain experts that are not data scientists.

SoBigData RI will impact the ERA at: 1) **training** the next generation of responsible social data scientists engaged in the challenging research questions of the exploratories and ambassadors of critical data literacy to facilitate data citizenship and data democracy. 2) providing an **accelerator** of data-driven innovation that streamlines the collaboration with industries and start-ups to develop pilot projects and proofs-of-concept. 3) to **democratize** the benefits of AI, data science, and Big Data through a network of excellence nodes within an ethical framework that harmonizes individual rights and collective interest.

SoBigData RI PPP will implement and make operative an advanced social mining platform characterized by: 1) a continuously growing, distributed data ecosystem for procurement, access, and curation of big social data within an ethic-sensitive context. This ecosystem is based on innovative strategies to acquire big social data for research purposes, using opportunistic means provided by social sensing technologies and participatory means based on user involvement as prosumers of social data and knowledge. 2) A continuously growing, distributed platform of interoperable social data mining tools, methodologies, and services for distributed platform of interoperable social data mining tools, methodologies, and services for mining, analyzing, and visualizing massive datasets, together with associated data scientists' skills for the ethically safe deployment of big data analytics. All services will be made available through cloud-based services and/or high-performance computing software packages. The platform will be aligned with the EOSC. Another action is the opalization of ethical and legal principles. SoBigData RI will pursue the EU views on Responsible Research and Innovation and will especially uphold the values and norms of the EU Data Protection law to strengthen personal data protection as a fundamental right, combined with boosting the free flow of personal data as a common good.

Figure 3.2.1 summarizes the general architecture of the future RI. The nodes include Italy (as the central hub), Switzerland, the UK, Germany, and Spain; the second group comprises Bulgaria, Estonia, France, and the Netherlands; and the third group comprises Finland, Greece, Austria, Netherlands, and Poland. One or more sites can serve each node.

Becoming a National Node:

- Finding Political and Financial support from your country
- Becoming part of the ERIC and becoming a decisional part in the construction of the RI with the other members
- Create a sustainable future for your research “excellence”
- A way to offer your research results as service for the EC

What is NOT a node:

- A replica of a “standard node”
- Only a computational resource allocation
- A way to fund **only** basic research



Figure 3.2.1 Distributed SoBigData RI structure

With the entrance of SoBigData RI PPP at the end of 2022, a complete redefinition of task T3.5 is needed. RP2 will report details about these changes, and more details about strict deadlines and results obtained by the collaboration of the two projects will be reported in deliverable D3.6 - Sustainability Plan 2 expected for June 2023. Furthermore, SoBigData RI PPP will capitalize on the position of SoBigData RI within the broader network of research projects to foster engagement and exploit synergies. The work done in the SoBigData++ already serves and collaborates with many H2020 projects in humanities, social science, and AI (the list includes HUMAN AI Net, HUMMINGbird, AI4EU, Tailor, XAI).

4 Conclusions

This deliverable updated the original sustainability plan and reported the next steps that are required from ESFRI to guarantee the sustainability of SoBigData RI.

Section 2 has described the ESFRI Roadmap 2021, outlining the three main phases of design, preparation, and implementation of the RI. Section 3 has described the Preparation Phase Project. It has delineated the path taken to join ESFRI and ensure long-term sustainability for SoBigData RI. Specifically, the preparation phase started in 2020, and the PP (Preparatory Phase) Project and the plan for long-term sustainability have been detailed. The main objectives of SoBigData RI PPP will be, in fact, modelling and definition of the ERIC legal entity and the acquisition of legal status (for both the central hub and national nodes), defining a financial plan and a business plan for long-term sustainability and a strategy for services design, community involvement, and partnerships; furthermore, there will be the engineering, planning, and optimization of technical infrastructure. The Preparatory Phase includes the cost of becoming a legal entity (ERIC - European Research Infrastructure Consortium) with a registered office in Italy. This phase will end at the end of 2024.

The next step, after the Preparatory Phase, will be the implementation phase's whose main aims are to make operative all the management and legal structures expected, implement the cost book, and define and model the accesses and the services related to the SoBigData RI. This phase will start in 2025 and will end in 2029. The implementation phase covers the ERIC establishment, software development, physical spaces, and competencies for implementing each national node and the central hub. The operation phase will start in 2030 and will end in 2050. The SoBigData RI mission is expected to remain relevant for at least two decades.

The ESFRI SoBigData project is a boost for creating new working positions, recruiting people, and creating high-level staffing. During preparation and implementation, we will gradually build up the staffing at the central hub. Most key staff at the national nodes are already in place at the RI consortium institutions. Furthermore, the central organization requires a professional office for administrative, legal, financial, communication, and project management support. Scientific leadership is also required for the services that cover SoBigData RI's core activities.

In conclusion, we can state that SoBigData RI will be a strategic resource in Europe for sharing datasets, methods, experiments, research skills, and computational resources for supporting the comprehension of the current societal transformation also including studies and practices on ethical and privacy issues that current changes imply.

References

[1] European Strategy Forum on Research Infrastructures Long-Term Sustainability Working Group. ESFRI Scripta Volume II, Long-Term Sustainability of Research Infrastructures.

https://www.esfri.eu/sites/default/files/u4/ESFRI_SCRIPTA_SINGLE_PAGE_19102017.pdf, 2017.

[2] European Strategy Forum on Research Infrastructures. ESFRI RoadMap 2021 - Public Guide.

https://www.esfri.eu/sites/default/files/ESFRI_Roadmap2021_Public_Guide_Public.pdf, 2019.