



Deliverable D3.1

Stakeholder analysis



DOCUMENT INFORMATION

PROJECT

PROJECT ACRONYM	SoBigData RI PPP
PROJECT TITLE	SoBigData RI PPP: Preparatory Phase Project
STARTING DATE	01/10/2022 (36 months)
ENDING DATE	30/09/2025
PROJECT WEBSITE	ppp.sobigdata.eu
TOPIC	HORIZON-INFRA-2021-DEV-02
GRANT AGREEMENT N.	101079043

DELIVERABLE INFORMATION

WORK PACKAGE	WP3 - Financial Framework
WORK PACKAGE LEADER	SSSA
WORK PACKAGE PARTICIPANTS	CNR, UNIPI, EGI, SSSA
DELIVERABLE NUMBER and TITLE	D3.1 Stakeholder analysis
AUTHOR(S)	Saverio Barabuffi (SSSA), Giulio Ferrigno (SSSA), Enrico Marcazzan (SSSA), Andrea Piccaluga (SSSA)
CONTRIBUTOR(S)	Valerio Grossi (CNR), Michela Natilli (CNR), Roberto Trasarti (CNR)
EDITOR(S)	Valerio Grossi (CNR)
REVIEWER(S)	Margaret Varilek (SmartLex), Michela Natilli (CNR)
CONTRACTUAL DELIVERY DATE	30/09/2023
ACTUAL DELIVERY DATE	23/11/2023
VERSION	1.0
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	71
KEYWORDS	stakeholder, innovation, collaboration, industry

EXECUTIVE SUMMARY

This deliverable provides an in-depth stakeholder analysis of the SoBigData RI PPP. The primary objective of this document is to identify, categorize, and analyze the stakeholders associated with the SoBigData RI PPP project. Furthermore, we seek to evaluate the impact that these stakeholders have on the daily activities and overarching mission of the Research Infrastructure (RI). This report serves as a resource for stakeholders, project managers, and decision-makers who aspire to navigate the relationships within the SoBigData RI and steer the project towards a successful future. The insights and actionable suggestions presented here are designed to empower stakeholders and project leaders with the knowledge necessary to make informed decisions, foster enhanced collaboration, and bolster the project's continued growth and prosperity.

DISCLAIMER

This project has received funding from the European Union's - Horizon Europe research and innovation programme under grant agreement No 101079043.

SoBigData RI PPP's main objectives are to define operational strategies for modelling and defining the ERIC legal entity and then obtaining legal status; to prepare the financial and legal aspects for both the central hub and the national nodes of RI; to prepare a business plan for long term sustainability; to define strategies for service design, community engagement and third-party partnerships.

This document contains information on SoBigData RI PPP core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as SoBigData RI PPP 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 RI PPP 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 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 RI PPP Consortium 2022. See <http://www.sobigdata.eu/> for details on the copyright holders.

For more information on the project, its partners and contributors please see <http://ppp.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 RI PPP Consortium 2022."

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

GLOSSARY

ESFRI	European Strategy Forum on Research Infrastructures
PPP	PPP Preparatory Phase Project
SMEs	Small and Medium Enterprises

TABLE OF CONTENTS

1	Relevance to SoBigData PPP	8
1.1	Relevance to project objectives	8
1.2	Relation to other work packages	8
1.3	Structure of the document	8
2	Data collection	10
2.1	Phase 1 - The Questionnaire compiled by SoBigData partners	10
2.2	Phase 2 - The interviews with key stakeholders of completed projects.....	11
3	SoBigData RI Stakeholders	12
3.1	The importance of Open Science	12
3.2	Stakeholders' diversity	12
3.3	Stakeholders' distribution	12
3.4	Status of collaborations	13
3.5	Types of collaborations	14
3.6	Understanding collaborations.....	15
3.7	SoBigData's stakeholders' sectors.....	17
3.7.1	SoBigData stakeholders' main sectors.....	18
3.7.2	SoBigData stakeholders' secondary sectors	20
3.8	SoBigData stakeholders' researchers.....	21
3.8.1	Researchers' involvement based on stakeholder types	21
3.8.2	Researchers' involvement based on collaboration types	22
3.9	SoBigData stakeholders' domain	24
3.9.1	SoBigData stakeholders' domain per stakeholder type.....	25
3.9.2	SoBigData stakeholders' domain per collaboration type	26
3.10	SoBigData RI's research spaces	28

3.10.1	SoBigData RI's research spaces per stakeholder types.....	30
3.10.2	SoBigData RI's research spaces per collaboration types	31
4	SoBigData Partners' Stakeholders National nodes	33
4.1	Status of collaboration per each national node.....	33
4.2	Typologies of stakeholders per each national node	34
4.3	Singular typologies of collaborations per each national node	36
4.3.1	Compound typologies of collaborations per each national node	37
4.4	Stakeholders' sectors per each national node.....	38
4.5	Stakeholders' problem-solving domain for each national node.....	41
4.6	SoBigData RI's research spaces for each national node	42
5	SoBigData Partners' Stakeholders National nodes in detail	44
5.1	Number of stakeholders within each national node	44
5.2	Status of collaboration between SoBigData RI's partners and the national nodes	46
5.3	Typologies of stakeholders engaged within each national node	47
5.4	Singular typologies of collaborations within each national node.....	50
5.5	Compound typologies of collaborations within each national node.....	51
6	SoBigData partners' interviews	54
6.1	Preparation	54
6.2	Execution.....	55
6.3	Analysis	55
7	SoBigData RI future developments.....	56
8	Conclusions	58
	References	59
	Appendix 1. Questionnaire "SoBigData PPP Stakeholder Analysis"	60
	Appendix 2. Interviews' questions	64
	Appendix 3. Interviews' transcripts.....	65

1 Relevance to SoBigData PPP

This report aims to provide an initial stakeholder analysis of the SoBigData RI PPP project, a collaborative work led by the WP3 team at the Institute of Management within the Sant'Anna School of Advanced Studies, situated in Pisa, Italy. This analysis is guided by the principles outlined in Freeman's seminal work, titled "Strategic Management: A Stakeholder Approach", which defines stakeholders as any group or individual that has the capacity to influence or be influenced by an organization's objectives (Freeman, 2010). The acknowledgment of the pivotal role that stakeholders play in determining the success or failure of a project underscores the importance of our aims.

1.1 Relevance to project objectives

This report serves as a resource for stakeholders, project managers, and decision-makers who aspire to navigate the intricate web of relationships within the SoBigData RI PPP and steer the project towards a successful future. The insights and actionable suggestions presented here are designed to empower stakeholders and project leaders with the knowledge necessary to make informed decisions, foster collaboration, and bolster the project's continued growth.

The primary objective of this document is to identify, categorize, and analyse the stakeholders associated with the SoBigData RI PPP project. Furthermore, we seek to shed light on the profound impact these stakeholders have on the daily activities and overarching mission of the Research Infrastructure (RI).

1.2 Relation to other work packages

The stakeholder analysis and deliverable can be considered pivotal for several activities in all the WPs defined in the project.

1.3 Structure of the document

Section 2 includes a comprehensive overview of the data collection process, offering transparency into the methods and sources used to identify and gather information about the stakeholders. This lays the foundation for a robust analysis that delves into the complex web of relationships between stakeholders and the RI.

Section 3 outlines a critical aspect of the analysis that involves examining stakeholders in two distinct dimensions. First, we assess their relevance in the context of the services provided by the RI. This entails a detailed examination of how different stakeholders interact with and influence the services offered by the RI, which is invaluable for understanding the dynamics of the project.

Sections 4 and 5 extend the analysis to the national nodes, where we explore the network of stakeholders who are closely connected to these nodes. By understanding their relationships and influence within the national nodes, we gain valuable insights into the broader reach and impact of the SoBigData RI PPP project.

Section 6 (and Appendix 3) include a comprehensive presentation of the findings from interviews that were conducted with specific stakeholders. These interviews provide a qualitative dimension to our analysis, offering insights and perspectives directly from individuals who are engaged with or impacted by the SoBigData RI PPP. This qualitative aspect complements the results of the survey, allowing us to gain a more nuanced understanding of the experiences and viewpoints of stakeholders involved in the project. Our specific focus in these interviews revolves around stakeholders who have been closely associated with completed projects under the SoBigData RI PPP initiative. This approach enables us to not only gauge the satisfaction levels of stakeholders regarding the services provided by the SoBigData RI PPP project but also to gain deeper insights into the challenges and complexities faced during the execution of these completed projects. Consequently, the insights drawn from these interviews will serve two fundamental purposes. Firstly, they will assist us in identifying areas for potential improvement within the SoBigData RI PPP project. This includes pinpointing aspects where enhancements can be made to enhance the overall effectiveness and efficiency of the project. Secondly, the information gleaned from these interviews will play a pivotal role in adjusting both ongoing and future projects that will be launched under the SoBigData infrastructure. By aligning project objectives and features with the preferences and experiences of stakeholders, we can ensure that the SoBigData RI PPP continues to meet the evolving needs of its stakeholders, fostering a more successful and harmonious partnership between the project and its key stakeholders.

Finally, Section 7 concludes this report presenting a comprehensive set of prospective pathways for future development. These recommendations are intricately interwoven with our understanding of stakeholders and the pivotal roles they assume within the SoBigData RI PPP project. Our overarching objective is to provide a holistic perspective on the stakeholder landscape, accentuating the critical significance of their contributions to the project's success and long-term sustainability.

2 Data collection

The data collection process was orchestrated, following a robust approach that unfolded over an extended period. To ensure the utmost precision and rigor in our data collection, we undertook a multifaceted procedure that spanned several months, beginning in May 2023 and concluding in late October 2023. We articulated the data collection phase in two different phases: 1) *Questionnaire phase*; 2) *Interviews phase*. Below we report the details that characterized the aims, the approach, and the procedures we followed.

2.1 Phase 1- The Questionnaire compiled by SoBigData partners

In the initial phase, we devised a comprehensive questionnaire that was specifically tailored to address the SoBigData RI public-private partnership. This questionnaire was a product of thorough deliberation, revisions, and refinements throughout the months of May and June 2023. To further solidify the reliability and validity of our questionnaire, it underwent a rigorous evaluation process. We engaged with the CNR leading team, which is renowned for its expertise in this domain. The questionnaire was launched in July 2023 and remained open until September 8, 2023, allowing us to cast a wide net and gather data from a diverse range of respondents. This timeframe allowed us to capture a more comprehensive snapshot of the landscape within the SoBigData RI PPP, ensuring that our findings would be robust and representative. The questionnaire, entitled “SoBigData PPP Stakeholder Analysis”, was disseminated to all partners within the SoBigData RI PPP consortium. It comprises the following questions:

- 1) Name of the SoBigData partner who is filling the survey;
- 2) The e-mail of the SoBigData partner who is filling the survey;
- 3) Name of the stakeholder involved in a project;
- 4) Whether or not the stakeholder was relevant for SoBigData RI in terms of collaboration and development of the Research Infrastructure;
- 5) Type of collaboration (Training/Development/Consulting);
- 6) Status of the collaboration (Completed/On-going/Planned);
- 7) Type of stakeholder (Researchers, Industry, Policy and law makers, Public administrations, Non-profit organizations, Teaching institutions, Funders, Data analysts, Journalists, Students, Other);
- 8) The sector of the stakeholder (Computer Science, Economy and Finance, Social Science, Humanities, Medicine, Biology, Law and Ethics, Telecommunications, Insurance, Retail, Transport, Sport analytics, Energy, Statistics, Other);
- 9) The related research space (if applicable) (Demography, Economy and Finance 2.0, Disaster Response and Recovery, Health Studies, Pervasive Intelligence in Cyber-Physical Systems for Future Society, Societal and Industrial Impact of Next-Generation Internet & beyond 5G Networks, Societal Debates and Misinformation Analysis, Social Impact of AI and Explainable Machine Learning, Sustainable Cities for Citizens, None);

- 10) The problem-solving domain: Data acquisition, Data analysis, Data storage/curation, Model development, AI Application, Other;
- 11) Number of SoBigData Partner's researchers involved.

For a detailed breakdown of the content and structure of the questionnaire, interested parties can refer to Appendix 1, which provides an overview of its components and questions. At the end of this phase, we collected 86 responses from different stakeholders that engage with SoBigData RI PPP.

2.2 Phase 2- The interviews with key stakeholders of completed projects

With the data collection phase successfully concluded, we turned our attention to identifying and selecting projects that had been completed. For these selected projects, we embarked on a careful and methodical process of reaching out to the pertinent stakeholders who had played pivotal roles in these initiatives. This task involved the acquisition of the contact details of these "crucial" stakeholders. The collection of this information was carried out to guarantee accuracy and completeness. Subsequently, we drafted and disseminated personalized invitations to these stakeholders, extending an offer to participate in brief interviews. These interviews were thoughtfully designed to provide the selected stakeholders with a dedicated platform to share their insights, reflections, and feedback. The overarching aim of these interviews was to gain a deeper understanding of the challenges, opportunities, and perspectives of the stakeholders within the SoBigData RI. Their valuable input was instrumental in shaping the future trajectory of the initiative. The interview phase was in October 2023 and the interviews were conducted ensuring that the insights obtained were robust and reliable.

Following the conclusion of the interview phase, the data analysis process was initiated. This entailed a thorough and comprehensive examination of the collected data, with a focus on identifying patterns, trends, and insights that could inform the research. The culmination of these rigorous and methodical processes was the production of the final report. This report encapsulated the findings, insights, and recommendations that arose from the data collected, the interviews conducted, and the subsequent analysis. For the interviews with crucial stakeholders that completed projects we developed an interview protocol that followed the criteria of qualitative research (Gehman et al., 2018). The interviews' questions are reported in Appendix 2. So far, we have conducted 2 interviews, which lasted 45 minutes each, with two different stakeholders of completed projects. The interviews have been recorded and transcribed. Some key insights that emerged from these interviews will be discussed in more detail in Section 6 of the report.

3 SoBigData RI Stakeholders

In this section, we delve into an analysis of SoBigData stakeholders and their diverse characteristics. SoBigData RI serves as a distributed, Pan-European, multi-disciplinary research infrastructure, engaging with a wide array of stakeholders within the European Union, the United Kingdom, and the United States of America. This infrastructural backbone has successfully cultivated relationships with 86 projects, resulting in an extensive network of stakeholders. SoBigData RI has engaged with a spectrum of stakeholders through its provision of an integrated ecosystem of tools, which empower data scientists and AI researchers to design and execute data science experiments.

3.1 The importance of Open Science

Open science represents a milestone in the transformation of how we conceive and conduct scientific research. This approach, based on the open sharing of data, methodologies, and research findings, holds critical importance for several reasons. Firstly, it promotes transparency and reproducibility in research, reducing the risk of unethical or fraudulent scientific practices. Furthermore, it fosters collaboration and knowledge-sharing among researchers from diverse disciplines and parts of the world, thus accelerating scientific progress, effectively mitigating the risks associated with unethical or fraudulent scientific practices. Open science democratizes access to scientific information, allowing a broader audience to benefit from research findings. This approach is essential for addressing global challenges such as climate change and pandemics, as it enables a faster and coordinated scientific response. Additionally, it encourages the creation of new opportunities for innovation and the development of more effective solutions to current problems. In summary, open science is a fundamental pillar in advancing scientific progress, knowledge, and global well-being (Grossi 2021).

3.2 Stakeholders' diversity

SoBigData RI engages with stakeholders from diverse backgrounds, spanning industries, researchers, non-profit organizations, public administrations, policymakers, international organizations, teaching institutions, journalists, and students. This diversity highlights the societal impact and relevance of the infrastructure's work, from aiding industries in problem-solving to empowering students with cutting-edge research tools. In the next sub-section, we provide more details about the distribution of these stakeholders.

3.3 Stakeholders' distribution

As illustrated in Figure 3.3.1, SoBigData RI's stakeholder distribution showcases a significant concentration among industry and researchers. This underlines the pivotal role of SoBigData RI in fostering 30 partnerships with the industrial sector, leveraging its expertise in data science to address complex problems faced by businesses. Moreover, one of the core missions of SoBigData RI is to promote collaboration among experts,

hence the substantial involvement of 29 researchers as stakeholders. In this context, Figure 3.3.1 underscores SoBigData RI's engagement with 10 public administrations, policymakers, and non-profit organizations, highlighting the practical application of data science and AI research in addressing societal challenges.

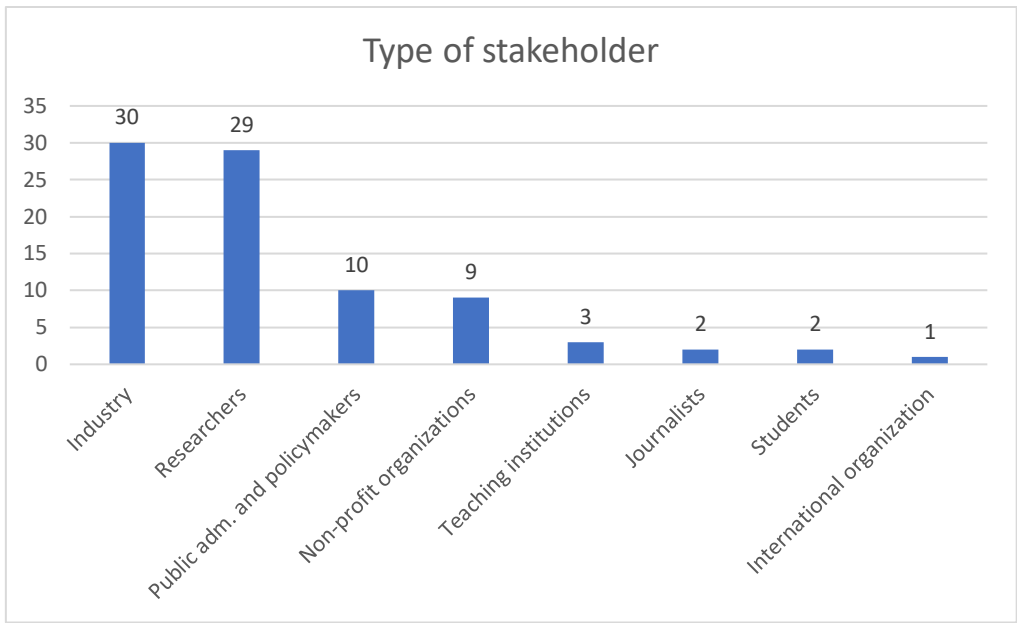


Figure 3.3.1. Distribution of stakeholder's types

3.4 Status of collaborations

Figure 3.4.2 offers insights into the status of collaboration among SoBigData RI's projects. The 77.91% of these projects remain in progress, as evidence of the ongoing commitment to scientific advancement and collaboration. Additionally, 15.12% of projects have been successfully completed, reflecting the tangible outcomes of this collaboration. This indicates that SoBigData RI's partnerships lead to concrete results that advance scientific knowledge and innovation.

Furthermore, the report highlights that a portion of projects is planned for future development, underlining the long-term vision and potential for continued growth and innovation within the SoBigData RI.

3.5 Types of collaborations

SoBigData RI partners engage in various forms of collaboration with their stakeholders, which can be categorized into four distinct types: 1) *Training*; 2) *Development*; 3) *Consulting*; and 4) *Joint Research*. These collaboration categories exemplify the multifaceted nature of SoBigData's engagement with stakeholders and underscore the diverse ways in which they contribute to the research infrastructure's mission. In other terms, this categorization serves as a foundational framework for understanding the varied roles that stakeholders play in shaping the trajectory of SoBigData RI's activities. In this regard, Table 3.5.1 provides a detailed

description of the different types of collaborations in which SoBigData RI's partners are actively involved, further emphasizing the rich web of relationships and activities that drive the success and impact of this research infrastructure.

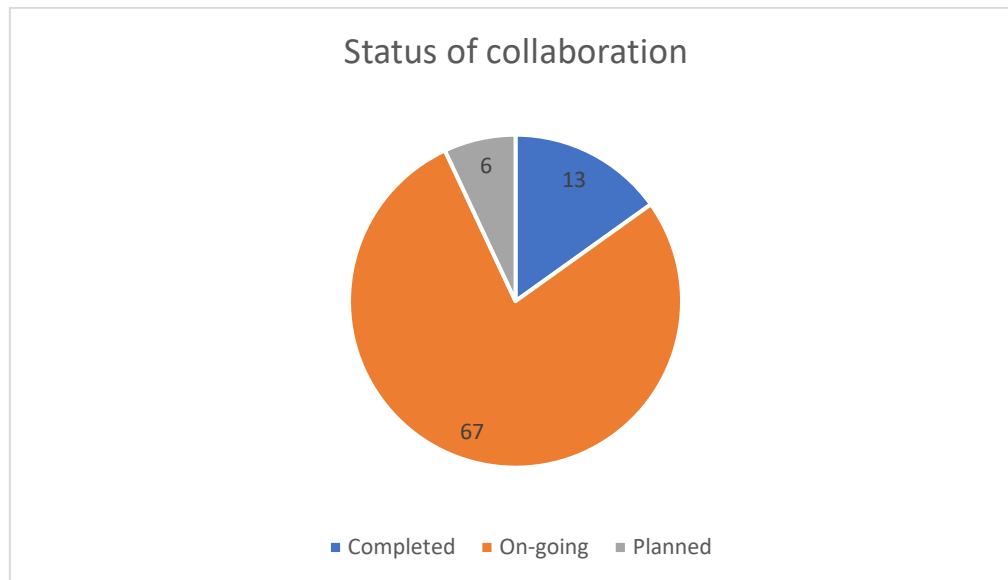


Figure 3.4.2. Distribution of completed, on-going, planned projects

Types of collaborations	Description
Development	SoBigData partner develops a method, an experiment, or deploys a technological infrastructure for the stakeholder
Consulting	SoBigData partner supports the stakeholder with experts in the field to improve/develop an internal project. This may also include the development of a proof of concept
Training	SoBigData partner organizes training materials or courses for the stakeholder
Joint research	SoBigData partner develops a mutual research project with the stakeholder

Table 3.5.1. Types of collaborations

To better assess the complexity of collaborations in which SoBigData RI partners are involved, we decided to subdivide the type of collaborations into singular and compound collaborations. The former represents the 4 categories represented in Table 3.5.1, while the latter represents combinations among the 4 singular categories.

3.6 Understanding collaborations

To gain a more comprehensive understanding of the types of collaborations and their nuances, Figure 3.6.1 offers a detailed breakdown of singular and compound collaboration types. Among the singular types of collaborations, "Development" and "Consulting" emerge as the most requested services. This highlights the significance of these areas in SoBigData RI's ecosystem. While "Training" and "Joint Research" seem to have a more marginal impact when considered individually, their collective contributions should not be underestimated. These types of collaborations continue to play essential roles within the infrastructure's network. When we delve into the compound types of collaborations, it becomes evident that the categories "Training and Consulting" and "Development and Consulting" represent the most sought-after combinations. This signals that "Consulting" holds a vital role not only in isolation but also in conjunction with other collaboration categories. It underscores the multifaceted nature of SoBigData RI's services, and the essential role "Consulting" plays in enhancing the impact and effectiveness of other collaboration types. This insight reveals the complexity and interdependence of the infrastructure's services, resulting in a holistic approach to problem-solving.

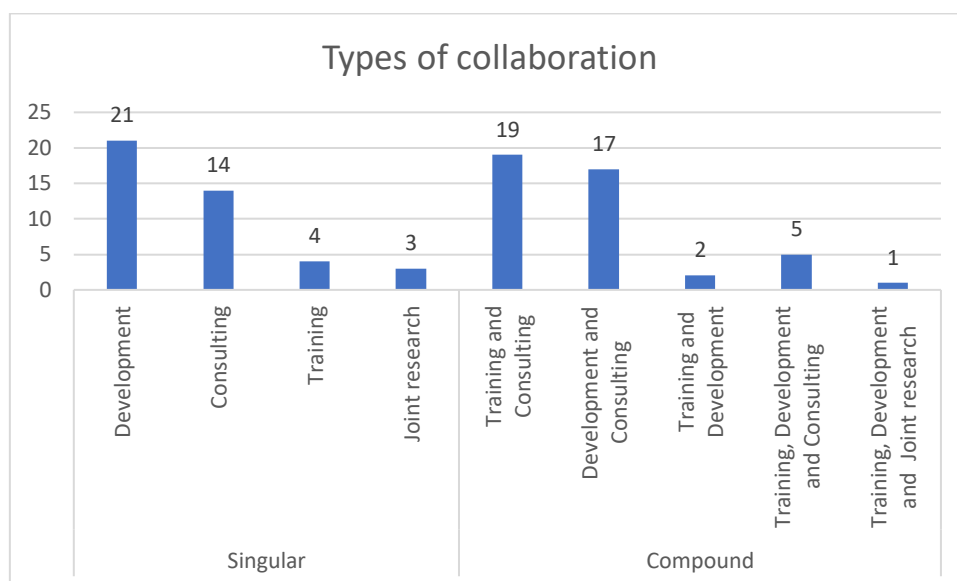


Figure 3.6.1. A comparison between singular and compound collaborations

The landscape of collaborations within SoBigData RI underscores its impact, not only in catering to sector-specific needs but also in nurturing a multidisciplinary approach to research and data science. Nevertheless, while this analysis provides a valuable overview of stakeholders and their respective collaborations, it falls short in delivering an in-depth assessment of the connections between them. Going beyond this categorical summary, Figure 3.6.2 delves into the types of collaborations associated with each specific stakeholder group. In the subsequent discussion, we consider the significance of prominent collaboration types within each stakeholder category:

- Industry stakeholders, they are actively involved in a range of activities, including "Development," "Consulting," "Training and Consulting," "Development and Consulting," and "Training, Development and Consulting." This engagement signifies their focus on improving and developing their projects, showcasing the infrastructure's impact in the business sphere.
- Researchers, representing a large segment of stakeholders, participate in "Development," "Joint Research," "Training and Consulting," and "Development and Consulting." Their wide-ranging engagement underscores the multifaceted nature of their needs, indicating the infrastructure's capability to address various research and data science requirements.
- Public administrations, policymakers, and non-profit organizations primarily necessitate "Development" and "Consulting" services. This implies that these stakeholders rely on the infrastructure for support in project development and advisory services;
- Teaching institutions require "Training" and "Training and Consulting" activities, highlighting their need for educational and capacity-building services. Journalists are mainly involved in "Development" and "Training and Development" activities, revealing the role of SoBigData RI in empowering the media with data-driven insights;
- Students actively participate in "Development" and "Training" activities, reflecting their engagement in hands-on learning and research initiatives. International organizations are mainly involved in "Training activities," showcasing the role of SoBigData RI in providing valuable training resources to these institutions.

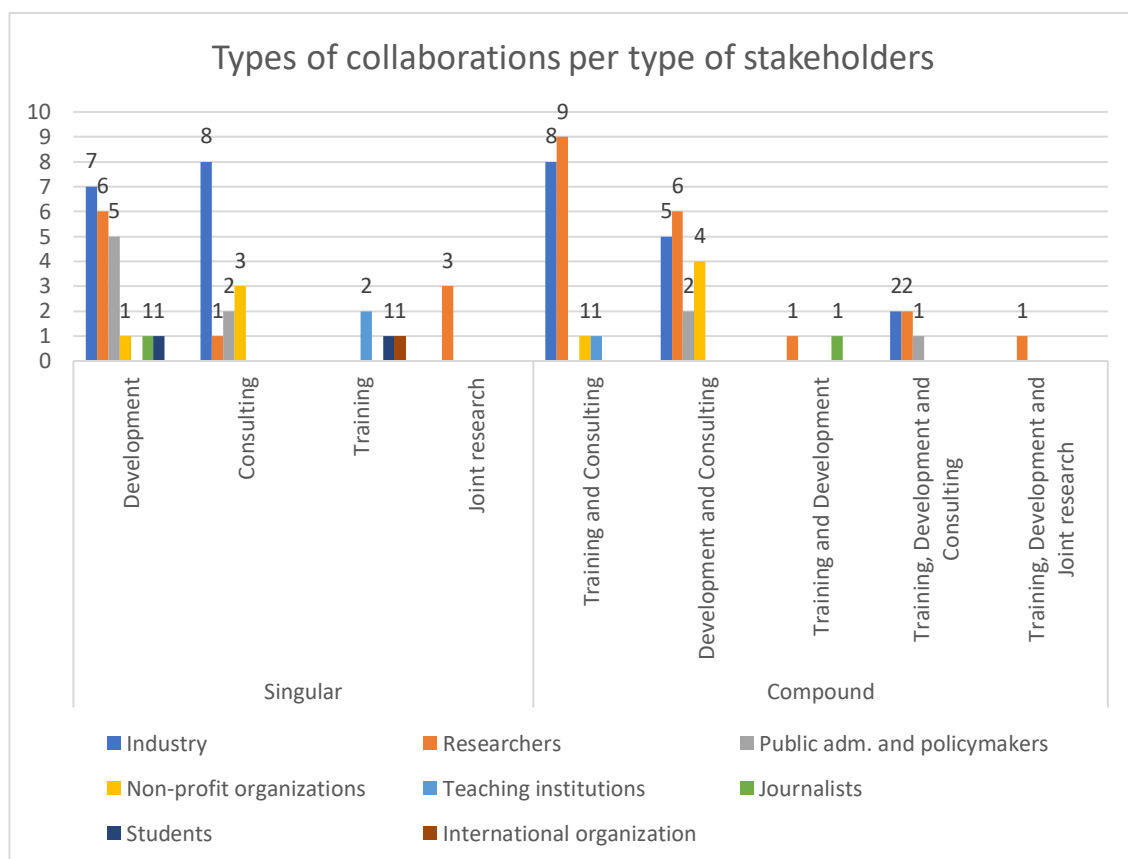


Figure 3.6.2. Types of collaborations per types of stakeholders

3.7 SoBigData's stakeholders' sectors

In Figure 3.7.1, we present an informative overview of the sectors within which SoBigData's stakeholders are actively engaged. This depiction provides insights into the landscape of sectors that are connected to the SoBigData Research Infrastructure.

Several noteworthy insights emerge from this figure. Firstly, the sector most prominently represented among the stakeholders is "Computer Science," showcasing the significant role of technology and data within the SoBigData network. This is followed closely by sectors such as "Economics and Finance" and "Social Sciences," underlining the interdisciplinary nature of the infrastructure's collaborations.

Furthermore, we find an equitable presence of stakeholders from diverse sectors, including "Transport," "Humanities," "Media and Telecommunications," and "Biology and Medicine." These sectors reflect the broad spectrum of expertise and interests that contribute to the richness of SoBigData's stakeholder ecosystem.

Beyond these sectors, we also observe active engagement from stakeholders in "Energy," "Retail," "Insurance," and "Industry 4.0 and e-Infrastructures," signifying the infrastructure's relevance in addressing challenges across various domains.

In the more specialized arenas, we have "Statistics," "Public engagement and government services," "Disinformation research," "Law and Ethics," "Open Science, Science, Technology & Innovation Analysis/Research," and "Sport analytics." While these sectors may be less prominent in terms of stakeholder representation, their active involvement highlights the diverse and comprehensive approach of SoBigData RI.

In conclusion, the SoBigData RI's stakeholder network spans a multitude of sectors, reflecting its inclusive and multifaceted engagement with a wide array of industries and research areas. This diversity not only underlines the significance of data science and AI in contemporary research but also showcases the infrastructure's pivotal role in promoting cross-disciplinary collaboration and knowledge exchange. In the following sub-section, we enrich this analysis by distinguishing between main sectors and secondary sectors.



Figure 3.7.1. Types of stakeholders per sectors

3.7.1 SoBigData stakeholders' main sectors

Figure 3.7.1.1 undertakes a more comprehensive analysis by focusing on the principal sectors. This detailed examination reveals the sector-specific distribution of stakeholders within the SoBigData network, offering a nuanced understanding of their involvement. Below, we discuss each main sector:

- **Computer Science:** Undoubtedly, the most prominent sector across all types of stakeholders, except for Journalists. This underscores the central role of computer science and technology within the SoBigData ecosystem. The diversity of stakeholders engaged in this sector reflects the pervasive relevance of data science and AI across various fields;
- **Economics and Finance:** This sector holds relevance for Industry, Public administrations and policymakers, Researchers, and Students. The intersection of economics and finance with SoBigData signifies its importance in addressing real-world economic challenges and advancing financial research;
- **Social Sciences:** With a broad presence across all types of stakeholders, except Journalists, the social sciences sector highlights the interdisciplinary nature of SoBigData collaborations. It underscores the utility of data-driven insights in addressing societal issues and advancing social research;
- **Transport:** This sector finds its distribution primarily among Industry stakeholders and public administrations and policymakers. It showcases the infrastructure's role in addressing transportation challenges and fostering collaborations with key industry players and policymakers;

- **Humanities:** While being a key sector for International organizations, Humanities also features prominently in Public administrations and policymakers, Researchers, Students, and Teaching institutions. This reflects the infrastructure's role in supporting humanistic research and promoting the integration of data science in the humanities domain;
- **Media and Telecommunications:** This sector prominently engages Journalists, Non-profit organizations, and Public administrations and policymakers. It signifies the importance of data in shaping media and telecommunications practices, aligning with the informational needs of journalists and policymakers;
- **Biology and Medicine:** Actively present in Industry, Researchers, and Students, this sector highlights the significant contributions of data science and AI in advancing biological and medical research. It exemplifies the infrastructure's role in facilitating breakthroughs in healthcare and life sciences;
- **Energy:** Relating to both Industry and Researchers, the energy sector underscores SoBigData contribution to addressing energy challenges and advancing research in this critical domain;
- **Retail:** This sector is exclusively featured among Industry stakeholders. It signifies the infrastructure's role in supporting retail businesses and addressing their specific data-related needs.

In summary, the sectors within the SoBigData stakeholder network attest the infrastructure's inclusive and cross-disciplinary approach. The distribution of stakeholders across these diverse sectors emphasizes the infrastructure's role in promoting knowledge exchange, collaboration, and innovation, encompassing a wide spectrum of industries and research areas.



Figure 3.7.1.1. Types of stakeholders per main sectors

3.7.2 SoBigData stakeholders' secondary sectors

In Figure 3.7.2.1, we repeat the analysis, this time focusing on the secondary sectors that are of relevance among SoBigData stakeholders. This granular exploration provides a comprehensive overview of how these secondary sectors intersect with the broader SoBigData network, shedding light on their specific roles and contributions. Below we provide more details:

- **Insurance:** Notably, this sector is exclusively represented among Industry stakeholders. This singular presence underscores the specialized relationship between the insurance industry and SoBigData, signifying the infrastructure's tailored support for this sector;
- **Industry 4.0 and e-Infrastructures:** This sector finds a significant presence within Industry stakeholders, and it also extends to Non-profit organizations. It highlights the alignment of Industry 4.0 and e-Infrastructures with SoBigData mission and the collaborative efforts undertaken in these areas;
- **Statistics:** Stakeholders engaged in statistics are primarily found among public administrations and policymakers, as well as Researchers. This sector emphasizes the application of statistical methodologies within the context of SoBigData, reflecting the demand for data-driven insights in these domains;
- **Public Engagement and Government Services:** This sector features prominently among Non-profit organizations and public administrations and policymakers. It underscores the relevance of SoBigData in public engagement and government services, aligning with the mission to address societal challenges;
- **Disinformation Research:** Researchers take the lead in this sector, highlighting the critical role of data science in the realm of disinformation research. This involvement signifies the importance of SoBigData in supporting research efforts to combat misinformation;
- **Law and Ethics:** Within the SoBigData stakeholder network, this sector is exclusively featured among public administrations and policymakers. It underscores the specific focus on legal and ethical considerations related to data science and AI;
- **Open Science, Science, Technology and Innovation Analysis/Research:** This sector is solely represented by Non-profit organizations. It underscores the integral role of these organizations in promoting open science and advancing research in the fields of science, technology, and innovation;
- **Sport Analytics:** This sector is primarily concentrated within Sport Analytics. It highlights the specialized nature of sports-related data analysis and its relevance within SoBigData ecosystem in general and to the health research space in particular.

In summary, the secondary sectors within the SoBigData stakeholder network offer a diverse and specialized perspective on the infrastructure's collaborations. The presence of these sectors among stakeholders illustrates the multifaceted engagement of SoBigData, addressing sector-specific needs while promoting cross-disciplinary knowledge exchange. These relationships underscore the infrastructure's capacity to contribute to diverse research domains, industries, and societal challenges.

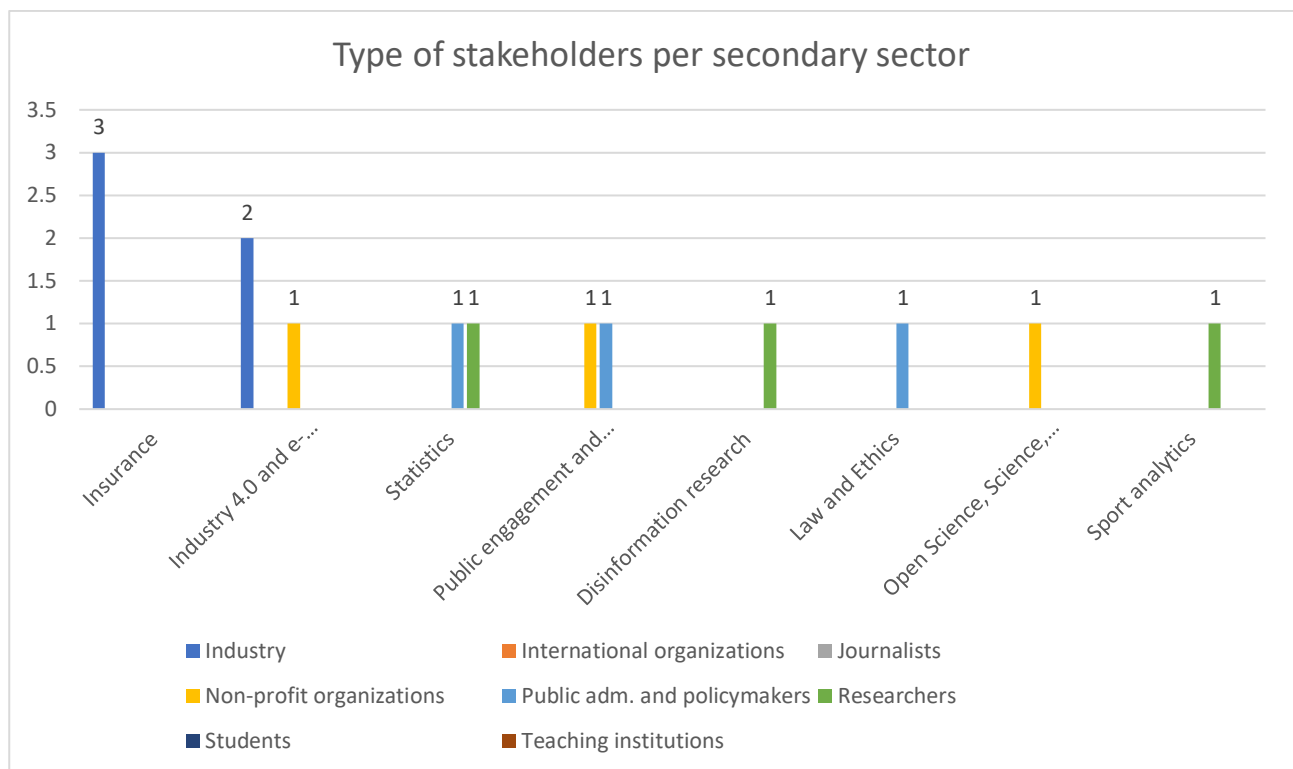


Figure 3.7.2.1. Types of stakeholders per secondary sectors

3.8 SoBigData stakeholders' researchers

This section provides an analysis of the researchers involved in the execution of the projects with the stakeholders of the SoBigData ecosystem. We develop this analysis by following two different directions: 1) *researchers' involvement based on stakeholder types*; 2) *researchers' involvement based on collaboration types*.

3.8.1 Researchers' involvement based on stakeholder types

To gain deeper insights into the SoBigData stakeholder landscape, Figure 3.8.1.1 presents an analysis of the average number of researchers engaged in projects associated with specific types of stakeholders. This analysis sheds light on the extent of researcher involvement across various stakeholder categories. Three key insights emerge:

- **Students, Public Administrations, and Policymakers:** On average, the highest number of researchers are involved in projects that engage with Students, Public Administrations, and Policymakers. This points to the robust collaboration between researchers and these stakeholders, emphasizing the significance of research in academia and the public sector;
- **Industry, Non-profit Organizations, Researchers, Teaching Institutions, and International Organizations:** These stakeholder categories also demand a considerable number of researchers, reflecting the depth and complexity of projects within these domains. The active involvement of

researchers underscores the multifaceted contributions made by these stakeholders and the critical role played by research in addressing their specific needs;

- **Journalists:** In contrast, projects associated with Journalists require the least number of researchers on average. This may indicate a more streamlined collaboration structure, where researchers provide targeted support in areas relevant to journalism and media;

In summary, the average number of researchers involved with different stakeholder types reflects the diversity and depth of engagement within the SoBigData network. It underscores the essential role of researchers in advancing projects across a wide spectrum of stakeholders, from academia to industry, public administrations, and the media. This collaborative framework highlights the adaptability of SoBigData in tailoring its support to meet the distinct requirements of various stakeholder groups.

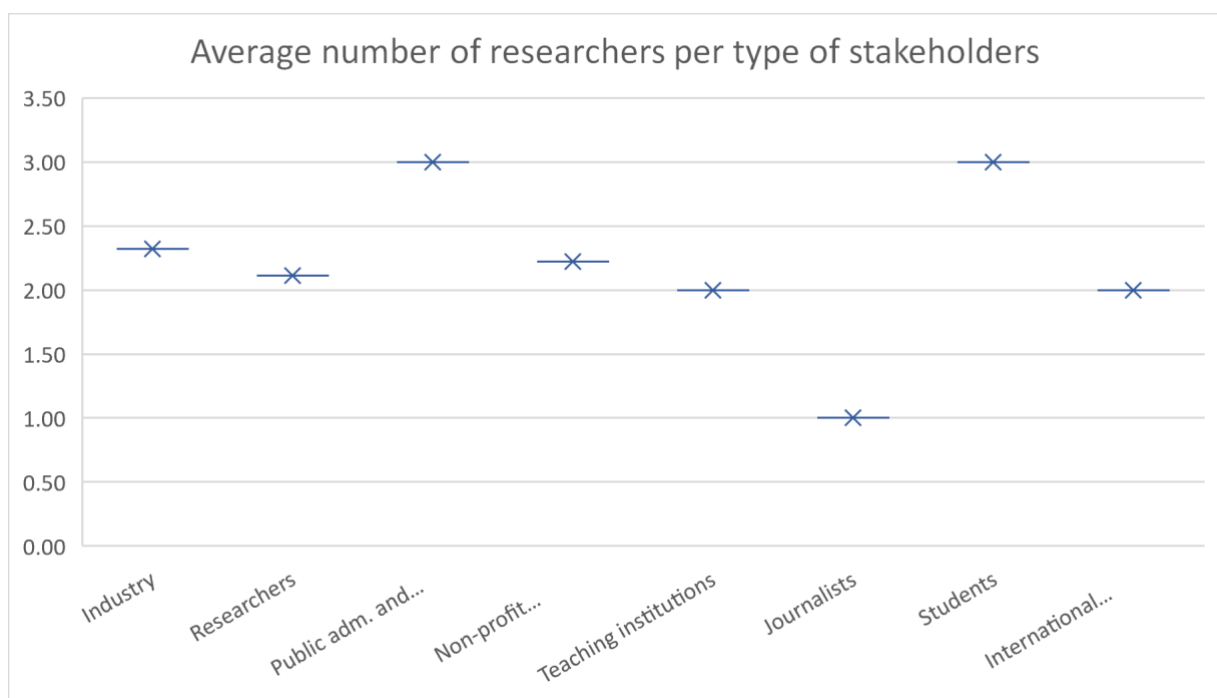


Figure 3.8.1.1 - Average number of researchers per type of stakeholders

3.8.2 Researchers' involvement based on collaboration types

To gain a better understanding of researchers' involvement in SoBigData projects, we explore the average number of researchers required based on the types of collaborations. This analysis provides insights into the depth and complexity of collaboration structures within the SoBigData RI. Two key insights emerge from this analysis:

- **High Researcher Involvement in Specific Collaboration Types:** Notably, collaboration types such as "Development and Consulting," "Training, Development and Joint Research," and "Training, Development and Consulting" require the highest average numbers of researchers. This implies that

as the complexity and diversity of collaboration types increase, so does the demand for a greater number of researchers. ;

- **Similar Involvement in Other Collaboration Types:** In contrast, other collaboration types exhibit similar average numbers of researchers involved. This suggests that they may have a more consistent and standardized framework in terms of researcher participation. These collaborations maintain a stable level of research involvement, which may align with their specific goals and objectives.

In summary, the average number of researchers engaged in SoBigData projects varies depending on the types of collaborations. The higher the number of collaboration types involved, the greater the demand for researchers, reflecting the complexity and comprehensiveness of these collaborations. This analysis underscores the adaptability of SoBigData in tailoring its research resources to meet the diverse needs of collaborative efforts, ensuring that projects benefit from the appropriate level of research expertise.

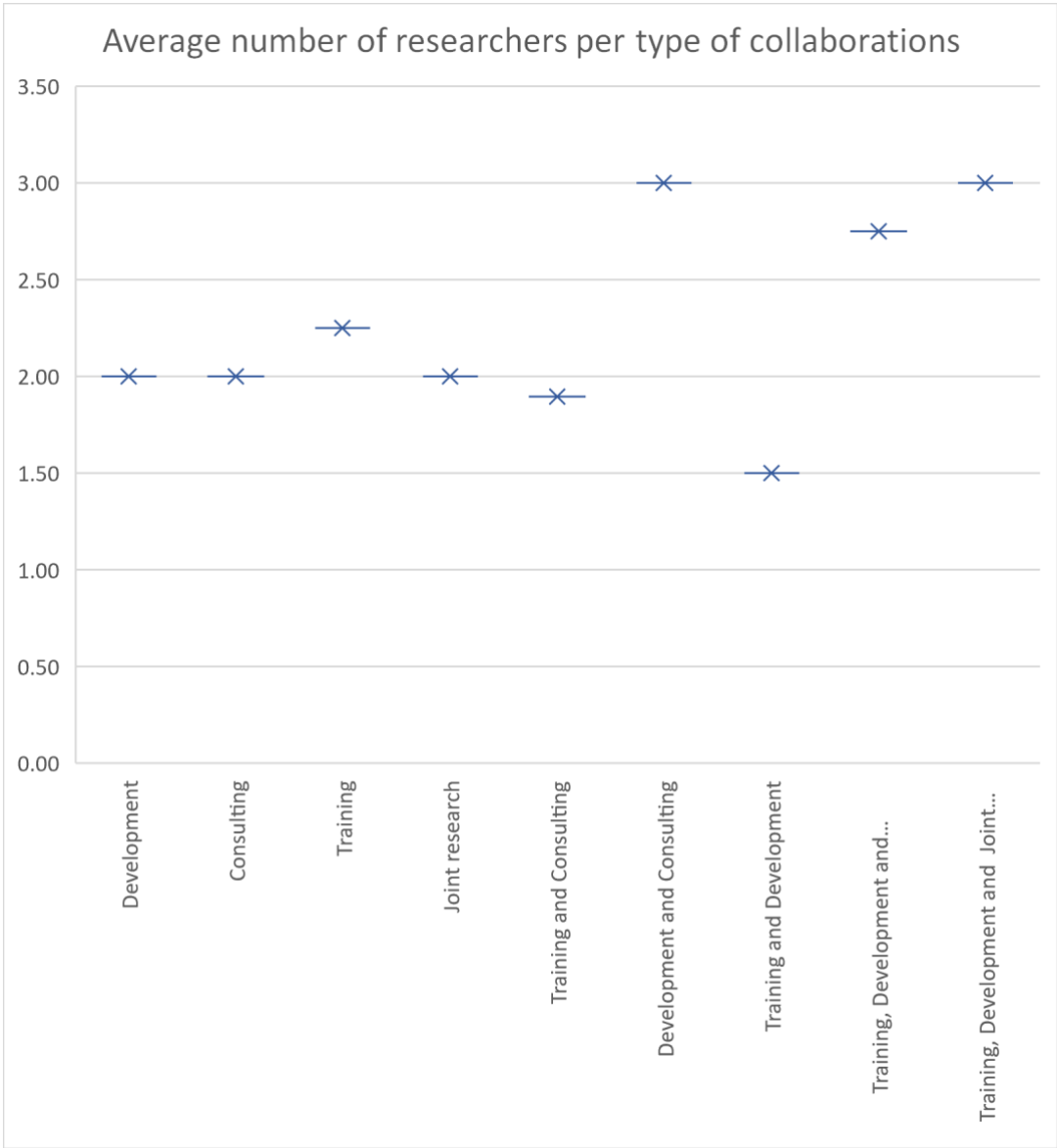


Figure 9 - Average number of researchers involved with specific type of collaborations

3.9 SoBigData stakeholders' domain

Figure 3.9.1 provides a comprehensive overview of the specific problem-solving domains in which SoBigData RI has actively collaborated with its stakeholders. This analysis highlights the most pressing areas of interest and expertise within the SoBigData community. Notably, two prominent domains emerge as the most requested and crucial for stakeholders: "Model Development" and "Data Analysis." This underlines the importance of leveraging advanced models and data-driven insights to address contemporary challenges.

Following closely behind these primary domains, we find "AI Application," indicating a growing interest in harnessing artificial intelligence to drive solutions. "Data Acquisition" and "Data Storage/Curation" also feature prominently, emphasizing the need for robust data infrastructure and management in the pursuit of effective research and innovation. Beyond these vital domains, it's worth highlighting the significant role that stakeholders have played as active partners in the SoBigData Master program. Their involvement has enriched the collaborative landscape, fostering innovation and driving the research agenda forward. Furthermore, SoBigData has taken the initiative to provide "Training Resources" to equip its stakeholders with the necessary skills and knowledge to excel in their roles. This commitment to education and capacity-building underscores the organization's commitment to sustainable progress and success. In addition to these critical aspects, SoBigData has shown a commitment to addressing "Ethical Issues About AI." This proactive stance towards ethical considerations in the development and application of artificial intelligence demonstrates a responsible and forward-thinking approach.

In summary, Figure 3.9.1 provides a comprehensive snapshot of the multifaceted collaboration between SoBigData RI and its stakeholders, highlighting key areas of focus and illustrating the organization's dedication to responsible innovation and effective problem-solving in today's data-driven world.

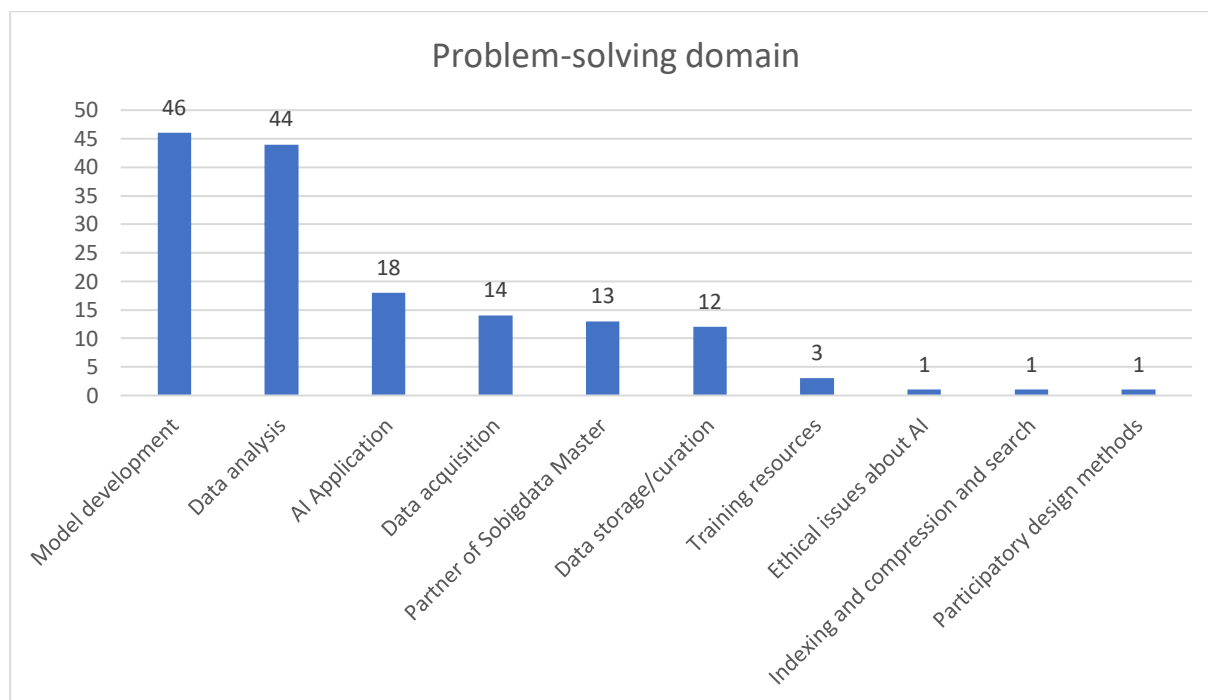


Figure 3.9.1 – Problem-solving domain

3.9.1 SoBigData stakeholders' domain per stakeholder type

In Figure 3.9.1.1, we delve into a detailed breakdown of the problem-solving domains with respect to different types of stakeholders. This segmentation provides useful insights into the specific needs and priorities of various stakeholder groups. Below we present the key findings:

- "Model Development" has emerged as a highly sought-after domain across multiple stakeholder categories, including industry, researchers, public administration, and policymakers. Additionally, non-profit organizations, teaching institutions, journalists, and students have expressed a substantial demand for activities related to "Model Development." This widespread interest underscores the versatile applicability and importance of modelling techniques in addressing diverse challenges;
- Similarly, "Data Analysis" has garnered significant attention from industry, researchers, public administration, and policymakers. This reflects the universal need for data-driven insights and analytical tools to inform decision-making and problem-solving across various sectors;
- The deployment of "AI Application" has been a collaborative effort, involving a broad spectrum of stakeholders. Industry, researchers, public administration, policymakers, non-profit organizations, journalists, and students have all actively engaged in implementing AI applications, underscoring the pervasive influence of artificial intelligence in today's problem-solving landscape;
- "Data Acquisition" activities have primarily involved industry, researchers, public administration, and policymakers, as well as non-profit organizations. This highlights the importance of data sourcing and collection for a wide range of stakeholders in different sectors;
- Conversely, "Data Storage/Curation" has found favour with industry, researchers, non-profit organizations, and students. These stakeholders recognize the significance of robust data management and curation in their respective domains;
- Furthermore, the "Partner of SoBigData Master" program has seen active participation from both industry and researchers, reflecting their integral role in shaping the SoBigData community;
- "Training Resources" have been a valuable asset for teaching institutions and international organizations, serving as a resource for capacity-building and knowledge dissemination;
- Non-profit organizations have shown a particular interest in addressing "Ethical Issues About AI" and "Indexing, Compression, and Search." This focus underscores their commitment to responsible AI practices and data management;
- Researchers, in addition to their extensive involvement in various domains, have also exhibited a strong emphasis on "Participatory Design Methods," highlighting their dedication to involving end-users and stakeholders in the development process.

Finally, Figure 3.9.1.1 provides a view of how distinct stakeholder groups engage with and prioritize various problem-solving domains.

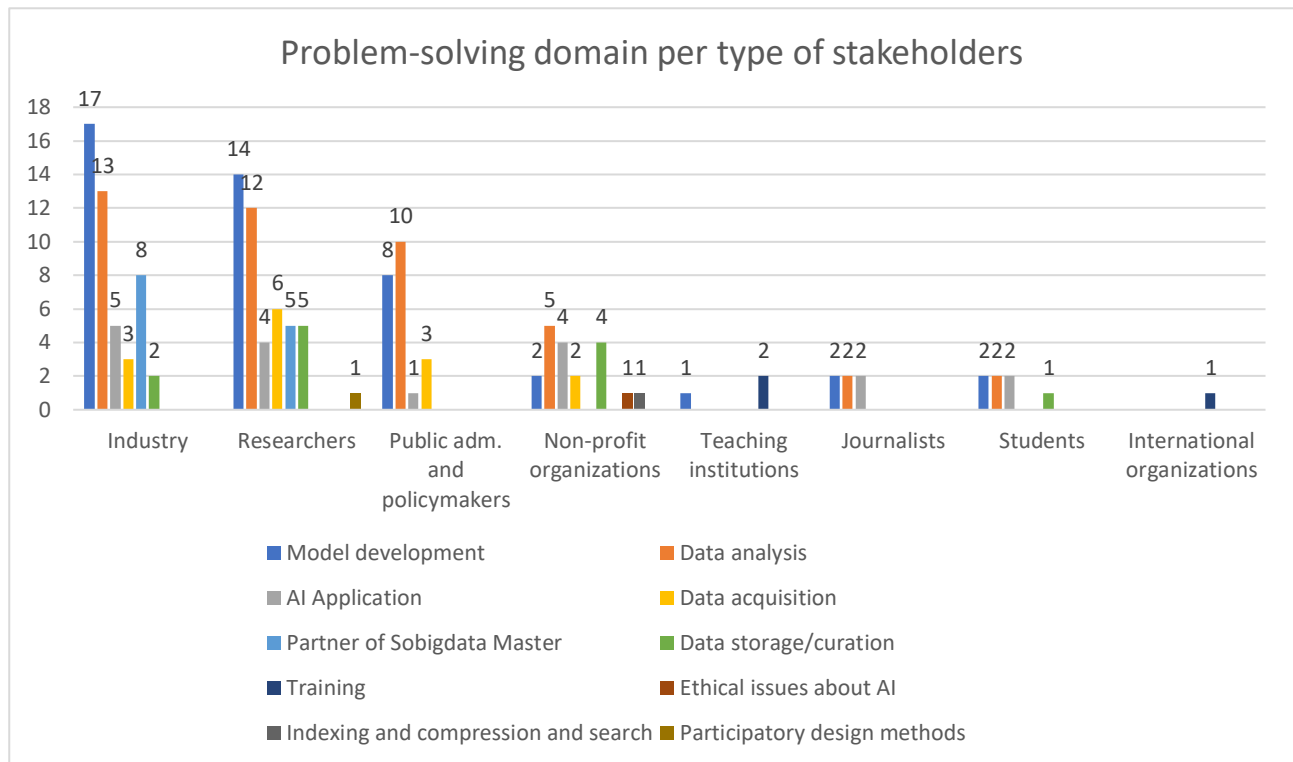


Figure 3.9.1.1 – Problem-solving domain per type of stakeholders

3.9.2 SoBigData stakeholders' domain per collaboration type

After discussing problem-solving domain per stakeholder type, we proceed the analysis with discussing the problem-solving domain per type of collaboration. In this regard, Figure 3.9.2.1 provides a comprehensive view of the problem-solving domains categorized by the type of collaboration. Key findings emerging from this analysis are summarized below:

- In every type of collaboration, "Model Development" and "Data Analysis" emerge as crucial domains. This widespread demand underscores the importance of modelling and data analysis in addressing diverse challenges across collaborative tasks;
- The implementation of "AI Application" has been a recurring theme across various types of collaborations, except for "Joint Research" and "Training, Development, and Joint Research." This reflects the versatility of AI applications in problem-solving efforts across different collaborative contexts;
- "Data Acquisition" activities have been notably absent only in the "Training" and "Training and Development" categories. This highlights the unique nature of these collaborations, where data acquisition may not be as central to the objectives;

- The designation "Partner of SoBigData Master" is exclusively found in the context of "Training and Consulting." This suggests a close association between this role and the specific type of collaborative efforts aimed at training and consulting;
- Conversely, "Training Resources" are made available in both "Training" and "Training and Consulting" collaborations, underscoring their significance in capacity-building and consultation activities;
- "Data Storage/Curation" is a versatile domain, finding relevance in multiple types of collaborations, including "Consulting," "Development," "Training," "Training and Consulting," "Development and Consulting," and "Training, Development, and Consulting." This widespread adoption emphasizes the importance of effective data management and curation in diverse collaborative initiatives.

Figure 3.9.2.1 offers a detailed perspective on the interplay between problem-solving domains and the various types of collaboration within the SoBigData framework. It highlights the adaptability of certain domains across diverse collaborations while underscoring the unique roles and requirements specific to each collaborative context.

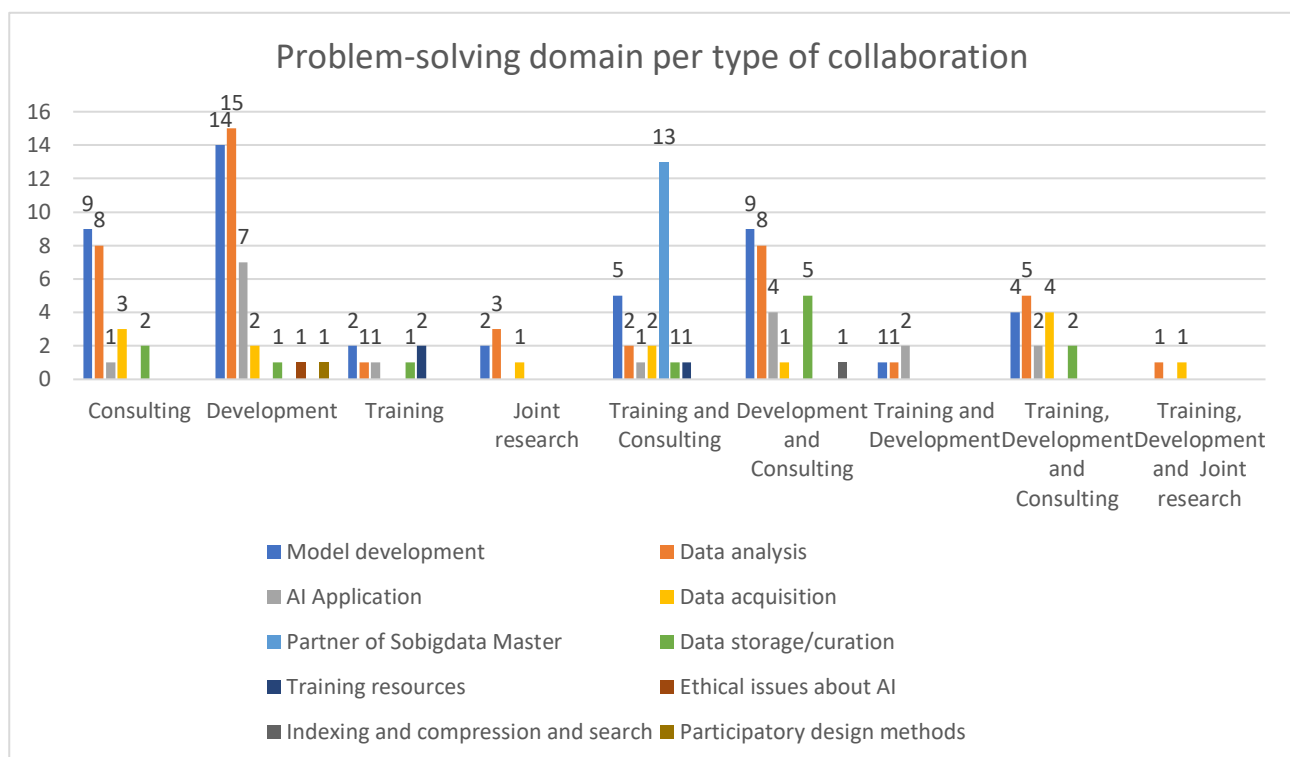


Figure 3.9.2.1 – Problem-solving domain per type of collaboration

3.10 SoBigData RI's research spaces

We can gain a deeper understanding of SoBigData RI's research spaces (formerly known as exploratories)¹ by defining them as the research spaces and topics in which scientific collaborations within the institute take place. This vital context helps elucidate the significance of these research spaces. Table 3.10.1 provides descriptions of each of SoBigData RI's research spaces, offering valuable insights into their respective scopes and objectives.

SoBigData's research spaces	Description
<i>Demography, Economy and Finance 2.0</i>	Explore changes and correlations in people's and companies' behavior after the economic crisis
<i>Societal Debates and Misinformation</i>	Analyze both discussions on social media and journalistic production to understand the most debated topics
<i>Sustainable Cities for Citizens</i>	Quantify traffic and the city's usage to take better decisions in mobility management and accessing real-time traffic information
<i>Social Impact of AI and Explainable Machine Learning</i>	Tracking changes from human coding to algorithms automatically learning to solve tasks
<i>Health Studies</i>	Research to address health-related issues, including medical, nutrition, and environmental health, is the focus of the multidisciplinary subject of the health science
<i>Pervasive Intelligence in Cyber-Physical Systems for Future Society</i>	Study AI models for a dynamic and automatic allocation of resources capable of reacting to changes in services, application requirements, sustainability, and user-generated traffic
<i>Disaster Response and Recovery</i>	Research of methods and tools to analyze, monitor, and improve post-disaster reconstruction processes in socio-economic areas, spatial planning, environmental health in cooperation with national and international institutions
<i>Societal and Industrial Impact of Next-Generation Internet & beyond 5G Networks</i>	Study of the capability of 5/6G networks and next-generation Internet to enable new services impacting society and manufacturing

Table 3.10.1. A description of SoBigData research spaces

¹ <http://sobigdata.eu/exploratories>

Having defined what are the SoBigData research spaces, we will now delve into a more detailed exploration of these themes in Figure 3.10.1. This figure shows the number of collaboration considering the specific research areas within SoBigData, providing a deeper understanding of their significance.

Among these research spaces, "Demography, Economy, and Finance 2.0" stands out as the most actively researched and applied topic among SoBigData stakeholders. It is followed closely by "Societal Debates and Misinformation," "Sustainable Cities for Citizens," and "Social Impact of AI and Explainable Machine Learning."

Subsequently, we find "Health Studies," "Pervasive Intelligence in Cyber-Physical Systems for Future Society," "Disaster Response and Recovery," and "Societal and Industrial Impact of Next-Generation Internet & beyond 5G Networks." These research spaces encompass a diverse range of critical research areas, underlining the broad and impactful scope of SoBigData's research aims.

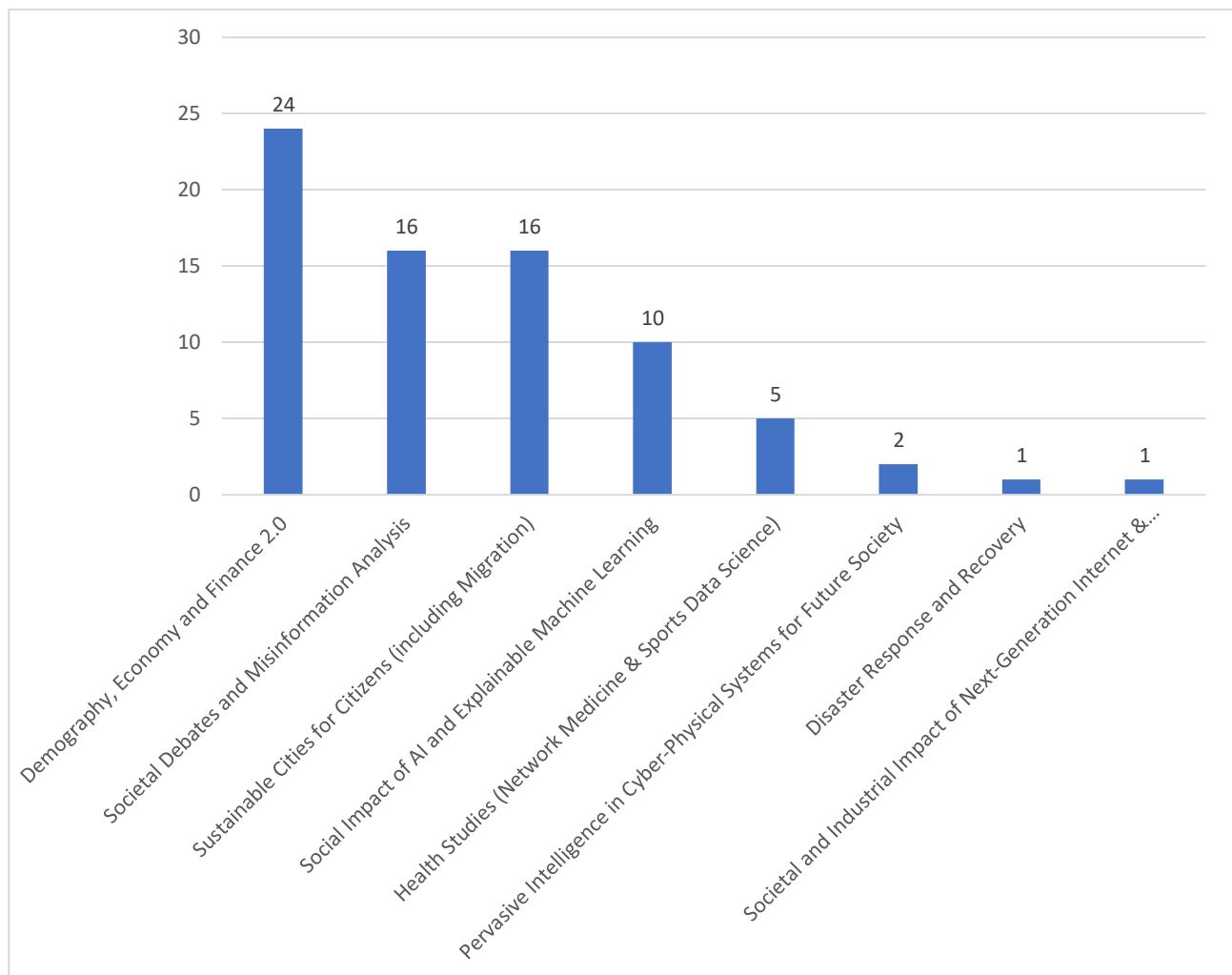


Figure 3.10.1 – SoBigData research spaces

3.10.1 SoBigData RI's research spaces per stakeholder types

In accordance with other analyses presented in this report, we enrich the discussion of these research themes per stakeholder type. Figure 3.10.1.1 provides a valuable perspective on the interplay between SoBigData RI research and the various types of stakeholders. This context allows us to appreciate the significance of each Exploratory in relation to the diverse stakeholder groups. We present the key insights as it follows:

- "Demography, Economy, and Finance 2.0" and "Sustainable Cities for Citizens" have found practical application and resonance among industry, researchers, public administrations, policymakers, and non-profit organizations. This broad engagement underscores the relevance of these research spaces in addressing societal and economic challenges;
- "Societal Debates and Misinformation" emerges as a pivotal topic, attracting active interest from researchers, public administrations, policymakers, non-profit organizations, journalists, students, and international organizations. This broad appeal highlights its importance in addressing misinformation and fostering informed societal debates;
- "Social Impact of AI and Explainable Machine Learning" holds significance for multiple stakeholders, including industry, researchers, public administrations, policymakers, teaching institutions, and students. This comprehensive engagement reflects the multifaceted impact of AI and machine learning on society;
- "Health Studies" has primarily been applied within industry, researchers, and student contexts, signifying its relevance to healthcare and research communities;
- "Pervasive Intelligence in Cyber-Physical Systems for Future Society" has primarily engaged with industry, emphasizing its importance in advancing intelligent systems and future technologies;
- On the other hand, "Disaster Response and Recovery" and "Societal and Industrial Impact of Next-Generation Internet & beyond 5G Networks" have been primarily implemented with researchers and non-profit organizations, respectively. These research spaces play a crucial role in addressing disaster management and emerging network technologies.

In summary, Figure 3.10.1.1 reveals the diverse and dynamic relationship between SoBigData RI research spaces and the different stakeholder groups, highlighting the ways in which the research intersects with the needs and interests of each respective audience.

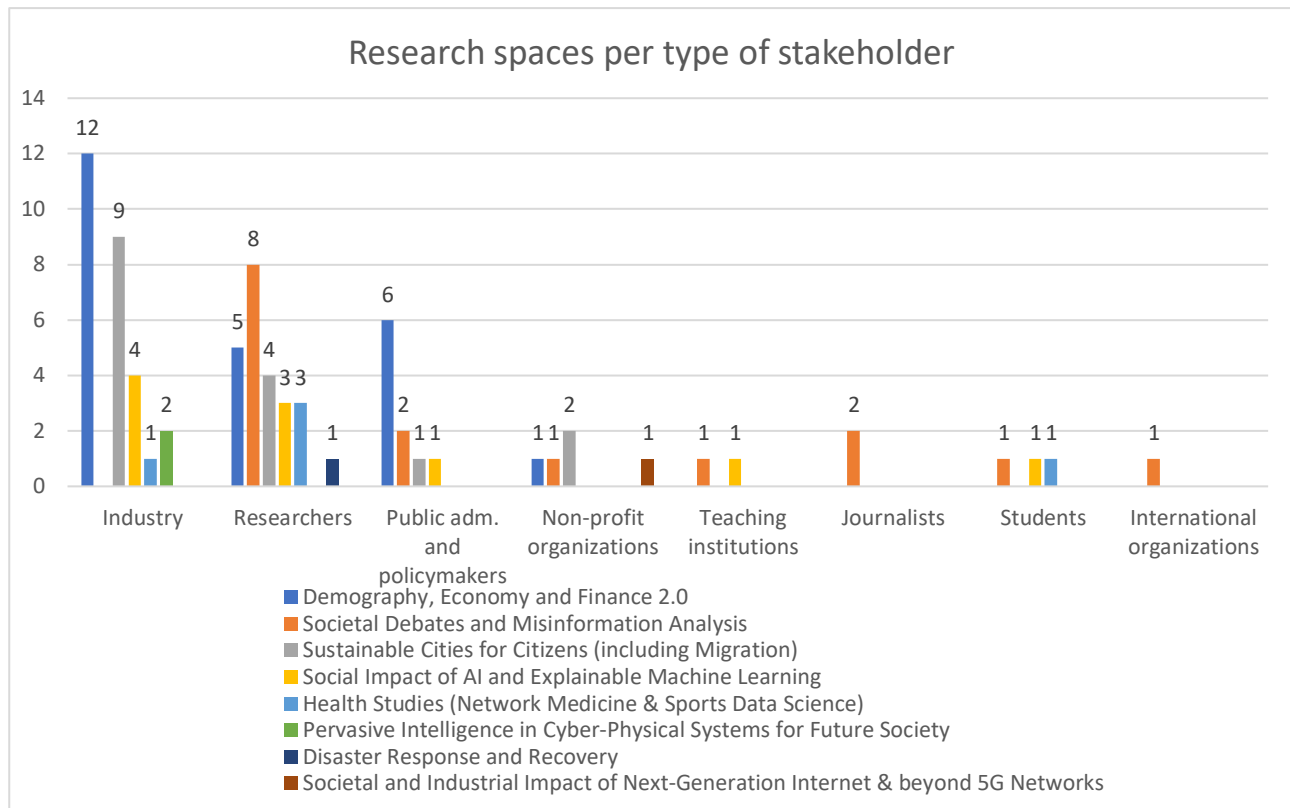


Figure 3.10.1.1 – SoBigData research spaces per stakeholder types

3.10.2 SoBigData RI's research spaces per collaboration types

We conclude this section by discussing a comprehensive view of SoBigData RI research in relation to the type of collaboration. As illustrated by Figure 3.10.2.1, this detailed breakdown underscores the varied applications and relevance of each research in different collaborative contexts. Below we present the key findings emerging from this analysis:

- The "*Demography, Economy, and Finance 2.0*" has been extensively applied in various collaborative settings, including "Development," "Consulting," "Joint Research," "Development and Consulting," "Training, Development, and Consulting," and "Training, Development, and Joint Research." This versatility highlights its adaptability and importance across different collaborative tasks;
- The "*Societal Debates and Misinformation*" exhibits a wide range of applications, spanning "Development," "Consulting," "Training," "Joint Research," "Development and Consulting," "Training and Development," and "Training, Development, and Consulting" activities. This broad utilization reflects its critical role in addressing societal debates and misinformation in diverse collaborative contexts;
- "*Sustainable Cities for Citizens*" has found relevance in "Consulting," "Training and Consulting," "Development and Consulting," and "Training, Development, and Consulting" activities. These

applications emphasize its significance in promoting sustainable urban development and citizen-centric cities;

- The "*Social Impact of AI and Explainable Machine Learning*" has been applied across "Development," "Consulting," "Training," "Joint Research," "Development and Consulting," "Training and Development," and "Training, Development, and Consulting" activities. This comprehensive engagement highlights its central role in understanding the societal implications of AI and machine learning;
- "*Health Studies*" has primarily been applied in "Development," "Training," and "Development and Consulting" activities, emphasizing its key role in advancing healthcare research and practice;
- "*Pervasive Intelligence in Cyber-Physical Systems for Future Society*" has found its primary application in "Development" activities, underscoring its relevance in shaping the future of intelligent systems;
- The "*Disaster Response and Recovery*" is primarily applied in "Training, Development, and Consulting" activities, highlighting its critical importance in disaster management and recovery efforts.
- The "*Societal and Industrial Impact of Next-Generation Internet & Beyond 5G Networks*" is predominantly utilized in "Consulting" activities, emphasizing its significance in shaping the industrial and societal impact of advanced network technologies.

Figure 3.10.2.1 offers a detailed perspective on the interplay between SoBigData RI research and the types of collaboration in which they are actively engaged, showcasing their adaptability and relevance in diverse collaborative contexts.

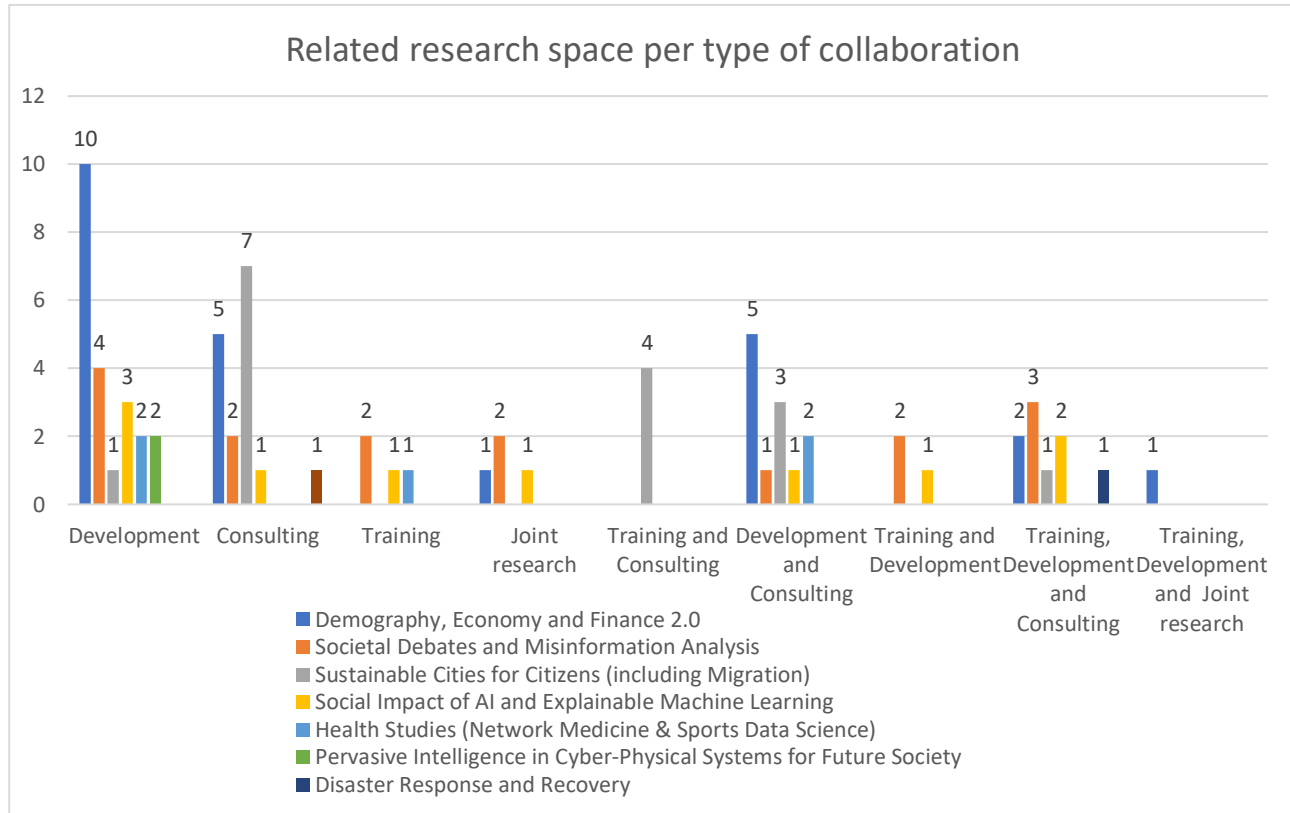


Figure 3.10.2.1– SoBigData research areas per collaboration types

4 SoBigData Partners' Stakeholders National nodes

In this section, we turn our attention to the analysis of stakeholders at the national nodes level within the SoBigData RI consortium. The consortium encompasses a total of 12 countries, with 9 of them actively participating in projects alongside multiple stakeholders. These countries include Belgium, Bulgaria, Estonia, France, Germany, Italy, Netherlands, Sweden, the United Kingdom joined by European Networks.

To account for the different number of institutions featured in each country, we scale the numbers of stakeholders for the number of national institutions participating in SoBigData RI.

As illustrated in Figure 4.1, Italy emerges within the SoBigData RI, engaging on average with 10 stakeholders. The United Kingdom, Belgium, Bulgaria, Estonia and France closely follow, underlining their substantial contributions to the consortium's initiatives. European Networks, Germany, Netherlands and Sweden, while fewer in number, are integral contributors with 1 average stakeholder each, showcasing the consortium's inclusive and diverse network of collaborators.

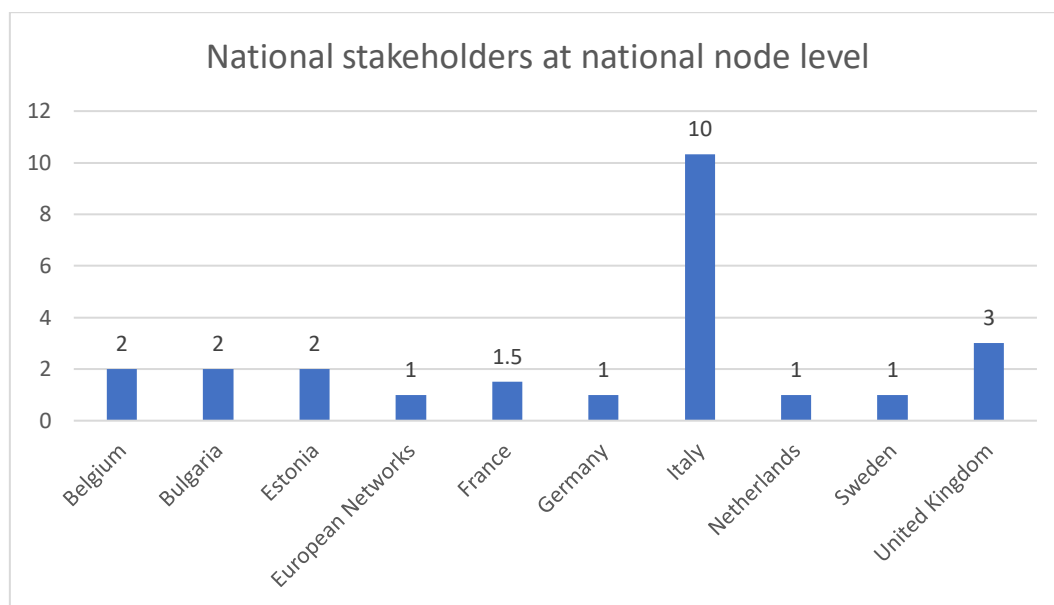


Figure 4.1. Stakeholders at national nodes level

4.1 Status of collaboration per each national node

Figure 4.1.1 offers an insightful view of the status of collaboration at each national node within the SoBigData RI. This data paints a picture of the progress and activity within the consortium. More details are presented below:

- Bulgaria, France, Italy, and Sweden have high levels of project completion, underscoring their dedication to collaborative efforts and the successful execution of projects. This demonstrates the significant contributions to achieve tangible outcomes within the SoBigData RI;

- Italy emerges as a notable leader in the planned projects. This highlights Italy's ongoing commitment to driving the research agenda forward;
- The United Kingdom, Germany, Netherlands and Bulgaria follow with the ongoing projects. This substantial engagement reflects the breadth of collaborative work being undertaken in these countries and by reinforcing the collaborative spirit of the consortium;
- It's worth noting that Belgium, Estonia, France, and Sweden do not currently have any ongoing projects. However, this does not diminish their role within the consortium, as they may have completed projects or have planned collaborations on the horizon;
- Speaking of planned collaboration, Belgium and Estonia are poised for future endeavours. Italy and the Netherlands also have planned projects, underscoring the forward-thinking approach to expand the collaborative landscape within the SoBigData RI.

Figure 4.1.1 offers a comprehensive overview of the status of collaboration at each national node, shedding light on both completed and ongoing projects, as well as the planned initiatives, within the SoBigData RI consortium.

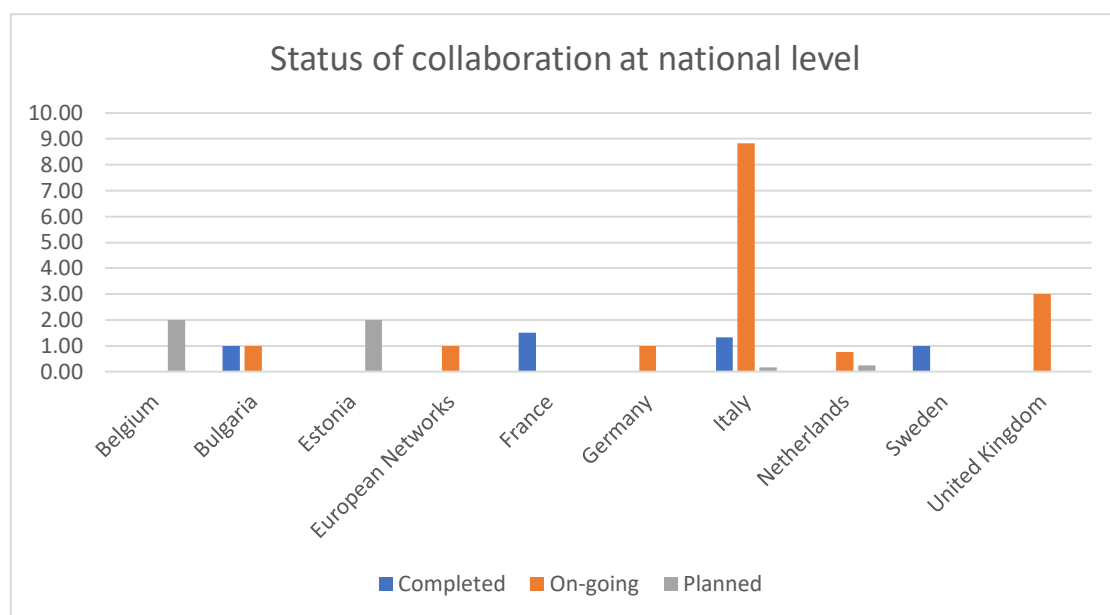


Figure 4.1.1 - Status of collaboration at national nodes level

4.2 Typologies of stakeholders per each national node

As depicted in Figure 3.3.1 (Section 3), the most numerous categories of stakeholders within the SoBigData RI are defined as Industry, Researchers, Public Administrations and Policymakers, with additional typologies including non-profit organizations. Furthermore, we categorize teaching institutions, journalists, students, and international organizations as residual typologies within this framework.

In Figure 4.2.1, we provide a detailed breakdown of the stakeholder typologies associated with each national node. This analysis helps us gain a clearer understanding of the various stakeholder engagements at the national level:

- Industry stakeholders are notably represented by Italy and Estonia, highlighting their active involvement in collaborative efforts that bridge the gap between research and industry applications;
- Collaborations with researchers are notably extensive in Italy, Netherlands, France, Germany, Belgium, and the United Kingdom, showcasing the significance of research partnerships within these national nodes;
- Public administrations and policymakers have played a central role in collaborative initiatives within Italy, Netherlands, Bulgaria, and Estonia, reflecting the importance of policy and governance considerations in these contexts;
- Non-profit organizations have been pivotal collaborators in projects conducted by Italy, the United Kingdom, France, Belgium, and by European Networks, underscoring their contributions to the consortium's mission.

This breakdown in Figure 18 offers valuable insights into the diverse stakeholder landscape at each national node, highlighting the unique strengths and foci of each node within the SoBigData RI.

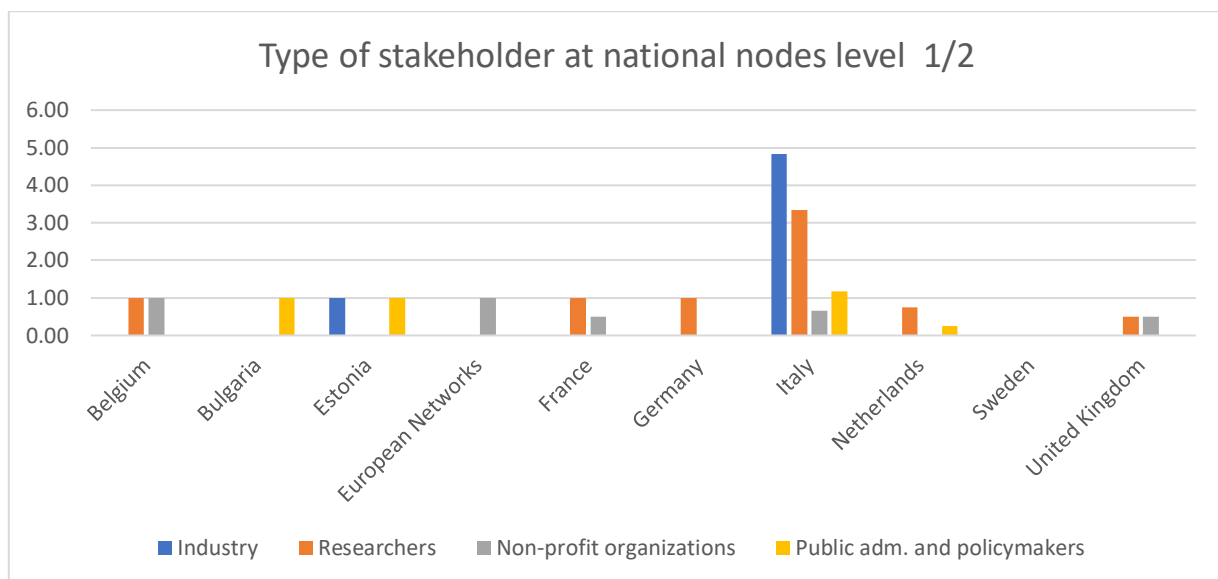


Figure 4.2.1. - Type of stakeholders per each national nodes level 1/2

Figure 4.2.2 provides a more nuanced view of the other typologies of stakeholders within each national node, underlining the diverse array of contributors and collaborators across the SoBigData RI consortium. This detailed analysis provides additional context on the various forms of collaboration at the national level:

- Teaching institutions have played a significant role in collaborative efforts within Italy and the United Kingdom. These institutions are instrumental in the dissemination of knowledge and skill development, contributing to the SoBigData RI's educational aims;

- Journalists have been specifically engaged in collaborative activities within the United Kingdom, reflecting the importance of media and communication professionals in communicating research outcomes and insights to the broader public;
- Collaborations with students are evident in Italy and Sweden, emphasizing the active involvement of student communities in research and learning activities;
- Bulgaria has been involved with international organizations, highlighting their global reach and engagement with entities beyond national boundaries.

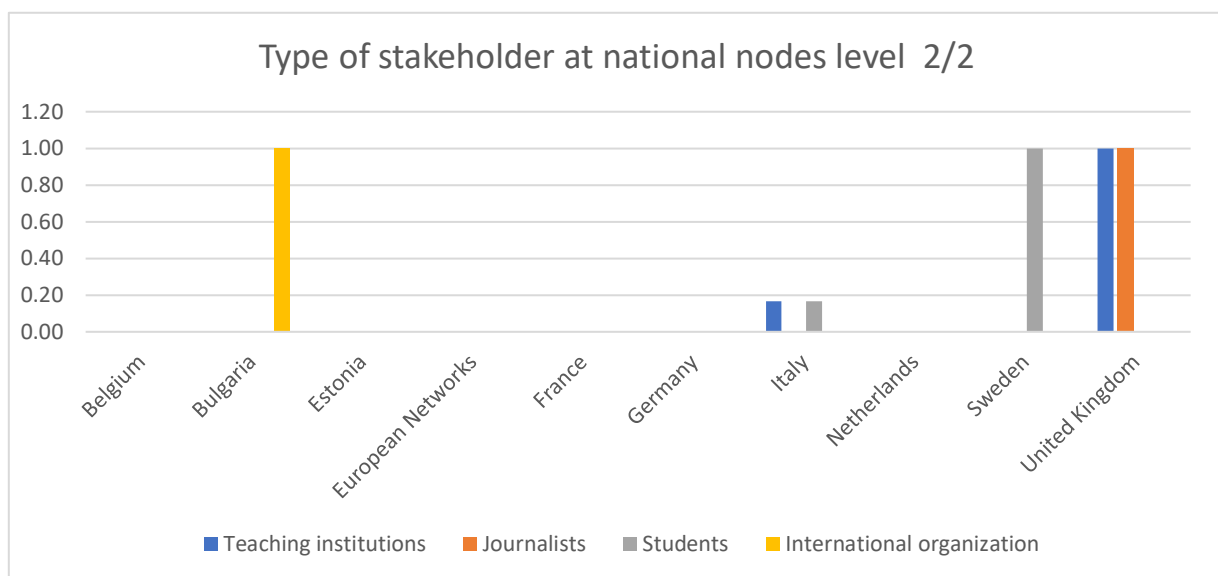


Figure 4.2.2. - Type of stakeholders per each national nodes level 2/2

4.3 Singular typologies of collaborations per each national node

Figure 4.3.1 offers an overview of the distinct types of collaborations within the SoBigData RI. This breakdown shows the specific roles and contributions of each national node:

- "Development" activities are notably distributed among Italy, the United Kingdom, the Netherlands, France, and Sweden. This shared focus on development underscores the collaborative efforts to advance research, technology, and solutions within the consortium;
- "Consulting" engagements are observed in Italy, the United Kingdom, the Netherlands, Belgium, and Bulgaria, showcasing the shared emphasis on offering expertise and advisory services in diverse contexts;
- "Training" activities have been actively implemented in Italy, the United Kingdom, and Bulgaria, underscoring the commitment to capacity-building and knowledge dissemination within these nodes;
- "Joint Research" initiatives have been carried out in Italy and Germany, emphasizing the significance of collaborative research projects that bring together expertise from different countries.

Figure 4.3.1 provides a perspective on the diverse types of collaborations, highlighting the specific roles and activities undertaken by each national node within the SoBigData RI.

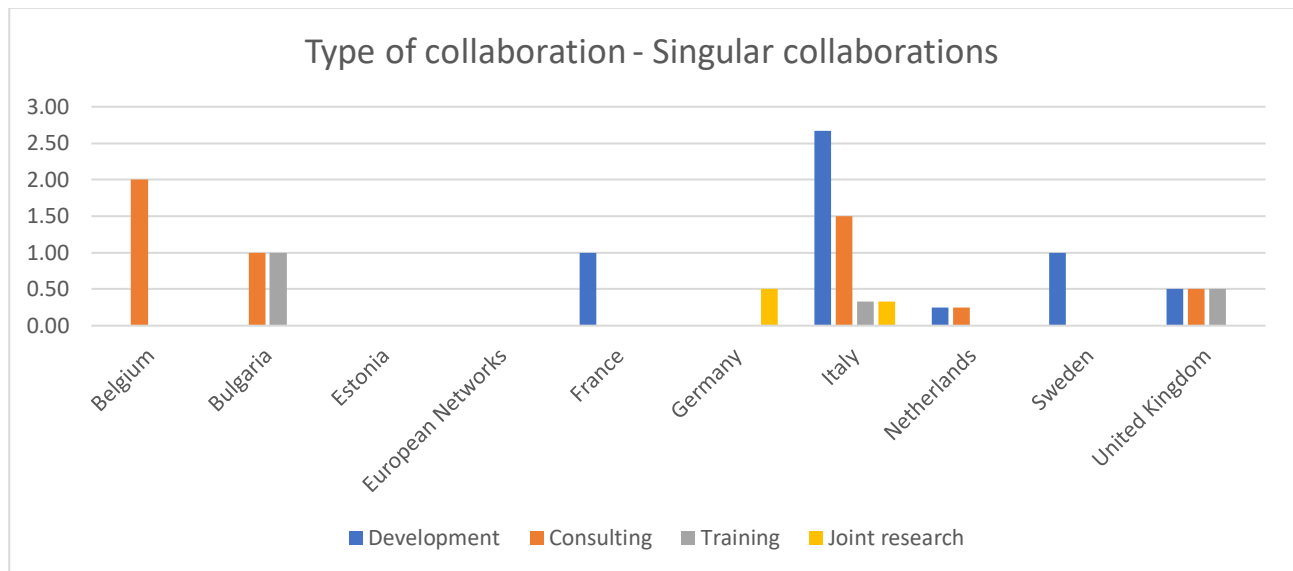


Figure 4.3.1 – Singular collaborations per each national node

4.3.1 Compound typologies of collaborations per each national node

Figure 4.3.1.1 provides a visual representation of the compound typologies of collaboration observed within the SoBigData network. A diverse array of collaboration types has been identified across various national nodes, reflecting the dynamic nature of the SoBigData Research Infrastructure. The following are the key collaborative typologies observed:

- **Development and Consulting:** This collaborative model has found implementation in multiple national nodes, including Italy, the United Kingdom, the Netherlands, and Estonia. It underscores the emphasis on collaborative development efforts coupled with advisory and consulting components.
- **Training and Consulting:** The combination of training and consulting activities has been noted in Italy, the United Kingdom, and within the European Networks. This collaborative structure highlights the sharing of knowledge through training initiatives, coupled with consultative support.
- **Training and Development:** The United Kingdom and Germany have engaged in collaborative efforts that involve training and development. This combination demonstrates a commitment to skill enhancement and capacity building, underpinning collaborative ventures.
- **Training, Development, and Consulting:** A multi-faceted approach to collaboration has been adopted in Italy, the Netherlands, Estonia, and European Networks, encompassing training, development, and consulting. This comprehensive model reflects a holistic commitment to knowledge exchange, skill enhancement, and advisory support.

- **Training, Development, and Joint Research:** France has implemented a collaborative model that includes training, development, and joint research. This structure underscores a multifaceted approach that encompasses educational aims, capacity building, and collaborative research initiatives.

These diverse compound typologies of collaboration reflect the nature of SoBigData RI partnerships and the adaptability of the infrastructure to accommodate a wide spectrum of collaborative models across various national nodes. These collaborations are instrumental in advancing the goals of SoBigData RI and fostering interdisciplinary exchanges.

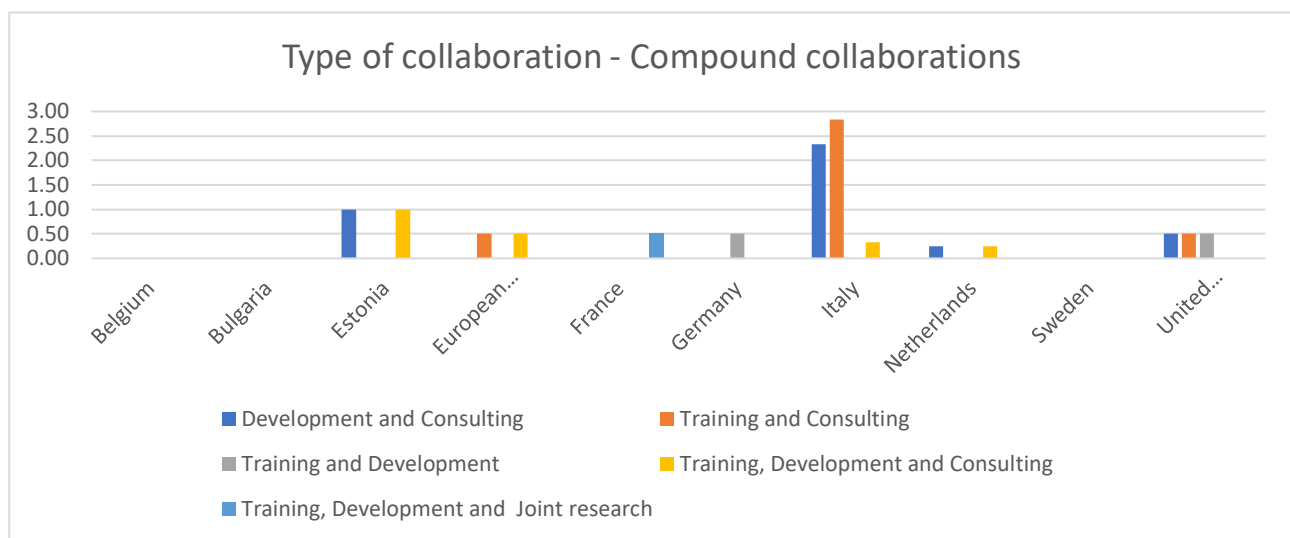


Figure 4.3.1.1 – compound collaborations per each national node

4.4 Stakeholders' sectors per each national node

Figure 4.4.1 provides a visual representation of the distribution of main sectors across the various national nodes within the SoBigData Research Infrastructure. The distribution of sectors across the national nodes reflects the diverse landscape of collaborative efforts. Each national node brings a unique set of sectors into focus, contributing to the multifaceted nature of SoBigData RI's research landscape. Here, we outline the key sectors and their respective presence in the national nodes:

- **Computer Science:** This sector enjoys active engagement across most SoBigData RI's partners, with the notable exceptions of Belgium, Estonia, and the United Kingdom. The ubiquity of computer science signifies its pivotal role in data-related research.
- **Economy and Finance:** Collaborations within the realms of economy and finance are prominent in France, Italy, and the Netherlands. These partnerships reflect a shared interest in the intersection of data science and economic domains.
- **Social Sciences:** The social sciences sector emerges as a central area of collaboration, involving partners from Belgium, Bulgaria, Estonia, France, Italy, the Netherlands, the United Kingdom, and

European Networks. This wide-ranging engagement underscores the importance of data-driven insights in addressing societal issues and dynamics.

- **Transport:** The transport sector is a focus of collaborative efforts primarily within the Italian node. This emphasis on transport-related research highlights the significance of data-driven solutions in addressing mobility and transportation challenges.
- **Humanities:** Collaborations within the humanities sector are distributed across Bulgaria, the Netherlands, Italy, and the United Kingdom. These partnerships signify a commitment to harnessing data science for humanities-focused research aims.
- **Media and Telecommunications:** Collaborations in the media and telecommunications sector are found within Estonia, Italy, and the United Kingdom. These partnerships underscore the intersection of data science with communication and media industries.
- **Energy:** Italy stands as the sole national node engaged in the energy sector, signifying a distinct focus on data-driven innovations in energy-related research.
- **Retail:** Italy is the exclusive national node to feature the retail sector within its collaborations, showcasing a dedicated effort to apply data science in the retail industry.

The distribution of these sectors across various national nodes exemplifies the interdisciplinary and cross-sectoral nature of SoBigData RI's collaborative efforts. This diversity highlights the adaptability and inclusivity of the infrastructure in accommodating a wide spectrum of research domains, promoting innovation, and fostering knowledge exchange.

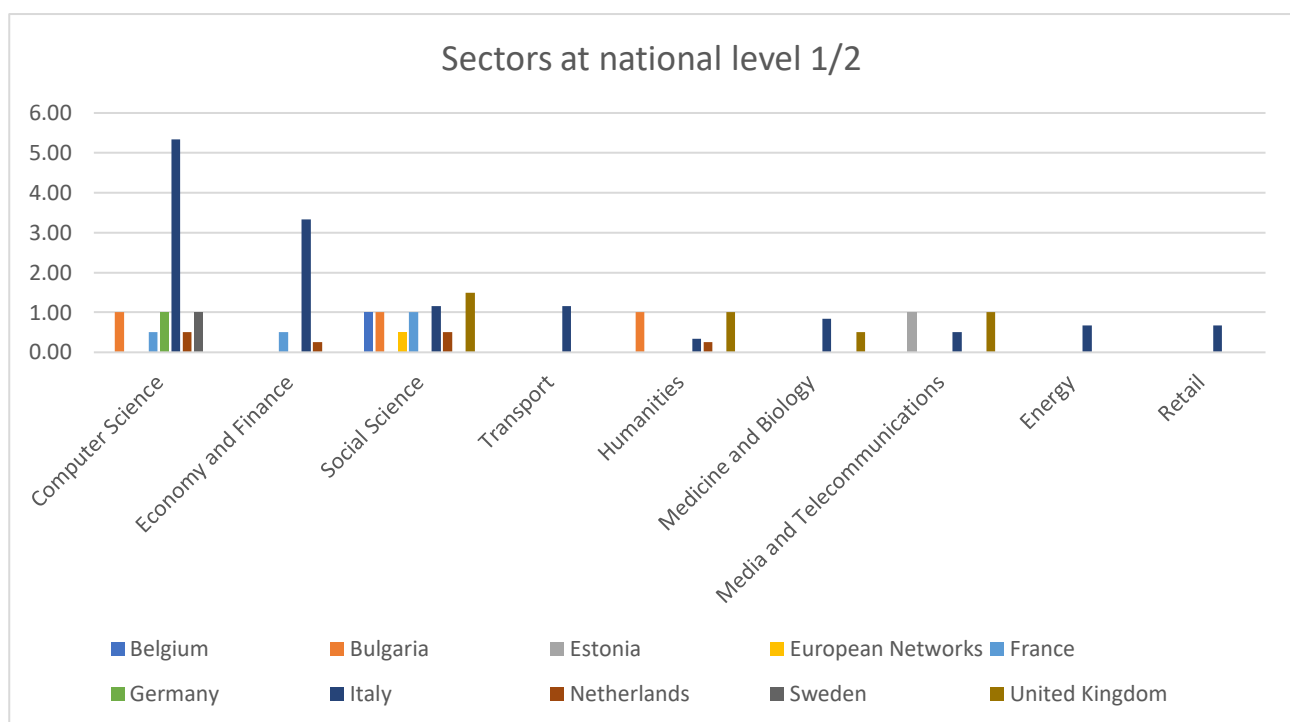


Figure 4.4.1 – Stakeholders' sectors per each national node part 1

Figure 4.4.2 provides an overview of the distribution of additional sectors among the various national nodes within the SoBigData Research Infrastructure. Diverse sectors not covered in the previous figure further accentuate the interdisciplinary and inclusive nature of SoBigData RI's collaborative landscape. Here, we outline the specific sectors and their respective presence in the national nodes:

- **Industry 4.0 and e-Infrastructures:** Collaboration in the domains of Industry 4.0 and e-Infrastructures is observed in Italy and within the purview of European Networks. This partnership underscores the critical role of data science in enhancing industrial processes and e-Infrastructures across borders.
- **Insurance Sector:** The insurance sector finds prominence in collaborations involving the Italian node. This emphasizes the application of data-driven insights in insurance-related research and decision-making.
- **Public Engagement and Government Services:** Collaborations centered around public engagement and government services are distributed across Bulgaria and France. This involvement underscores the use of data science in optimizing public sector operations and enhancing citizen engagement.
- **Statistics:** Italy and the Netherlands are the national nodes engaged in collaborations related to statistics. These collaborations underscore the importance of data-driven statistical analyses in various research domains.
- **Law and Ethics:** Italy stands as the exclusive national node where collaborations in the fields of law and ethics are reported. These collaborations underscore the intersection of data science with legal and ethical considerations.
- **Sport Analytics:** The sector of sport analytics features primarily in collaborations involving the Italian node. This highlights the use of data science in the analysis and enhancement of sports performance and strategies.
- **Open Science, Science, Technology, Innovation Analysis/Research:** European Networks exclusively house collaborations focused on open science, science, technology, and innovation analysis/research. These initiatives underline the role of data science in advancing scientific research, technology, and innovation on a broader scale.

The distribution of these diverse sectors across national nodes further underscores the adaptability and collaborative potential of SoBigData RI. These collaborations reflect the infrastructure's ability to engage in a wide array of research domains, fostering innovation, interdisciplinary exchanges, and knowledge sharing across borders.

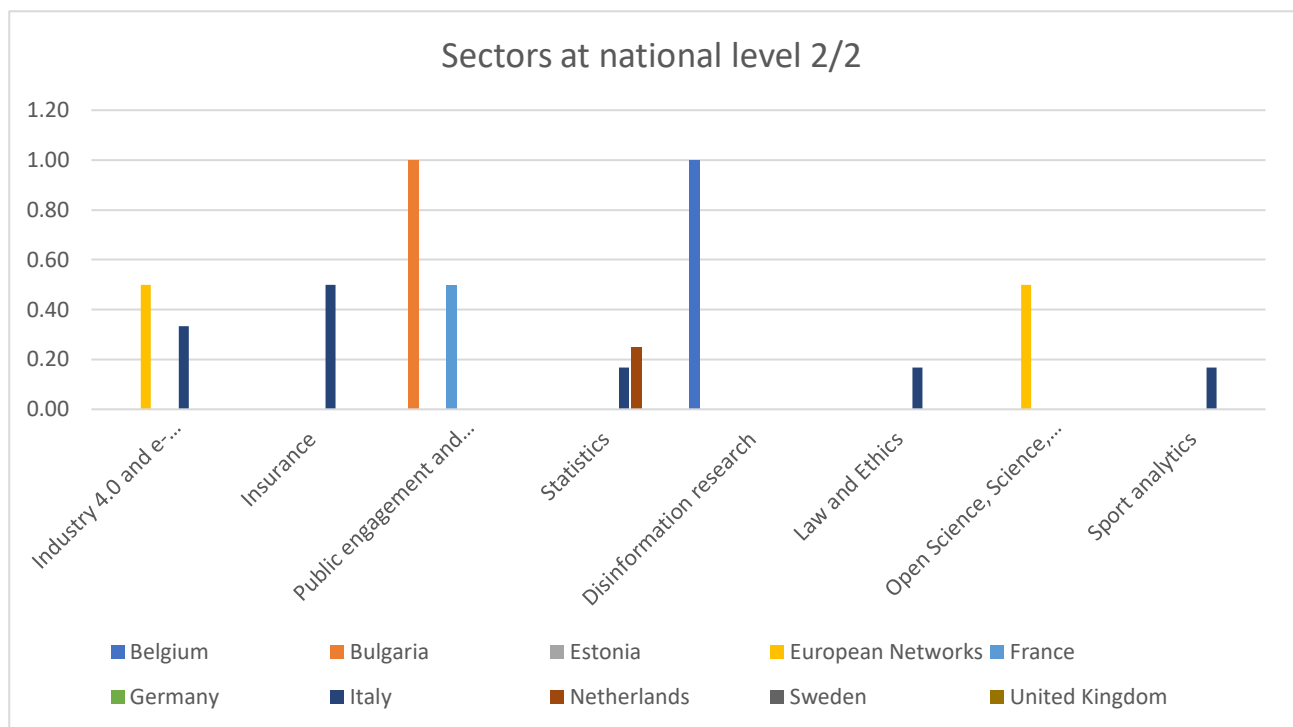


Figure 4.4.2 – Stakeholders' sectors per each national node part 2

4.5 Stakeholders' problem-solving domain for each national node

Figure 4.5.1 provides a visualization of the problem-solving domains represented by stakeholders within the national nodes of the SoBigData RI. A comprehensive understanding of the domains where stakeholders are actively engaged reveals the diverse landscape of problem-solving efforts within SoBigData RI. Here, we outline the specific domains and their respective presence within the national nodes:

- **Model Development:** Model development is a prevalent problem-solving domain and is widely applied by every SoBigData RI national partner. Notable exceptions include Estonia, France, and Germany, reflecting the varied research focus within the infrastructure.
- **Data Analysis:** Data analysis is a universally required domain, with an active presence in each national node. The prominence of data analysis underlines its central role in deriving insights and knowledge from data.
- **AI Applications:** AI applications are actively applied in multiple national nodes, including Italy, Germany, Sweden, the Netherlands, the United Kingdom, and European Networks. This reflects the significance of AI in addressing data-related challenges and opportunities.
- **Data Acquisition:** Data acquisition is another integral domain, and its application is observed in Belgium, Estonia, France, Italy, the Netherlands, and European Networks. This underscores the vital role of data collection in research and innovation efforts.

- **Partner of SoBigData Master:** The role of a "Partner of SoBigData Master" is primarily associated with Italy, in line with the national scope of SoBigData Master initiatives.
- **Data Storage/Curation:** Data storage and curation efforts are actively applied in Italy, the Netherlands, and within the purview of European Networks. These endeavors emphasize the importance of effective data management and preservation.
- **Training Resources:** Training resources are distributed among Bulgaria and the United Kingdom. This reflects the commitment to educational and capacity-building aspects within specific national nodes.
- **Ethical Issues about AI:** France is the national node where ethical issues related to AI are notably present. This signifies a dedicated focus on the ethical dimensions of AI applications.
- **Participatory Design Methods:** France is also home to participatory design methods, showcasing a commitment to inclusive and collaborative approaches in the research process.
- **Indexing and Compression and Search:** The domains of indexing and compression, as well as search, are primarily associated with Italy, underscoring the nation's expertise in these areas.

The distribution of these problem-solving domains illustrates the interdisciplinary nature of SoBigData RI's research aims. It highlights the diverse range of research challenges addressed by stakeholders across national nodes, underscoring the infrastructure's capacity to foster cross-disciplinary collaboration and innovation.

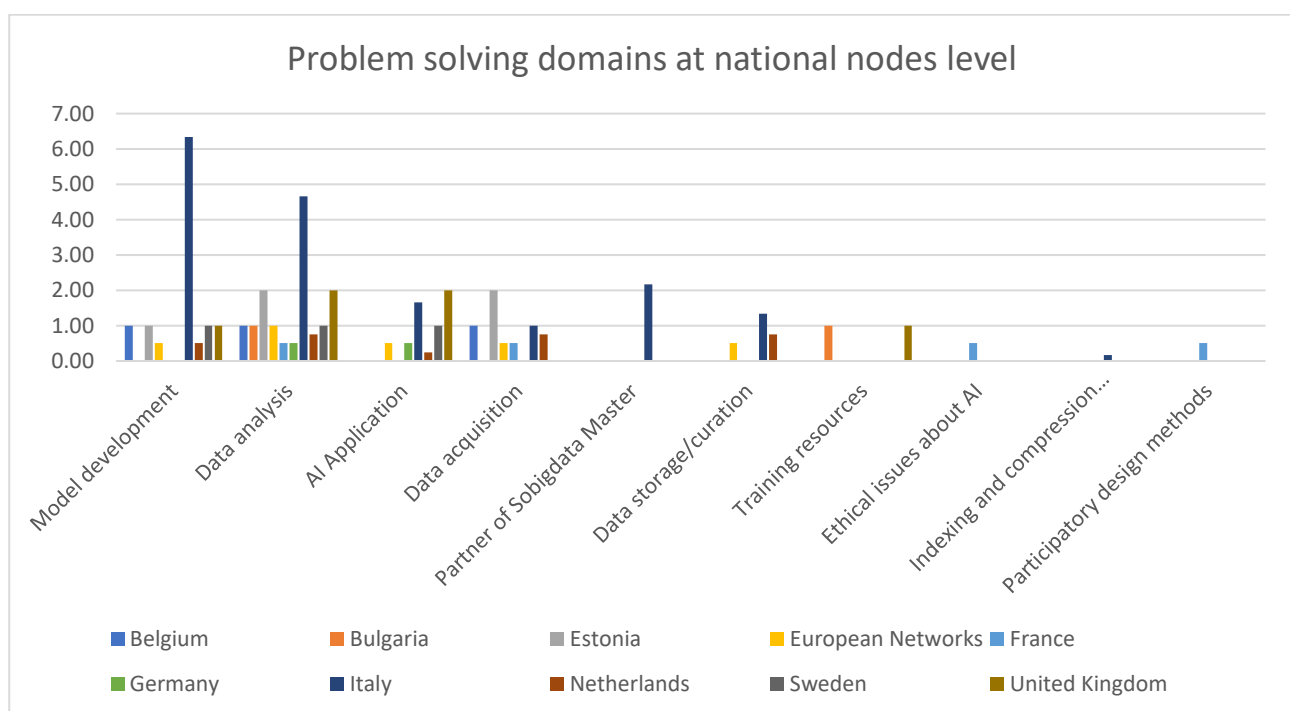


Figure 4.5.1. – Stakeholders' problem-solving domains per each national node

4.6 SoBigData RI's research spaces for each national node

Figure 4.6.1 offers a comprehensive view of the research spaces applied within the SoBigData RI across its various national nodes. The diversity and distribution of the research highlight the comprehensive approach

to research and innovation within SoBigData RI. Here, we outline the specific research and their respective prevalence within the national nodes:

- "*Demography, Economy, and Finance 2.0*" has witnessed extensive application in Belgium, France, Italy, and the Netherlands. Its prevalence underscores the importance of data-driven insights in the realms of demography, economy, and finance.
- "*Societal Debates and Misinformation*": The exploration of societal debates and misinformation is widespread and can be found in Belgium, Bulgaria, France, Germany, Italy, the Netherlands, and Sweden. This broad application reflects the infrastructure's commitment to addressing the societal implications of data-driven insights.
- "*Sustainable Cities for Citizens*": The focus on creating sustainable urban environments for citizens is primarily observed in Belgium and Italy, underscoring the commitment to data-driven urban planning and innovation.
- "*Social Impact of AI and Explainable Machine Learning*": This research space finds application across Germany, Italy, Sweden, and the United Kingdom. Its prevalence reflects the importance of understanding the social consequences and transparency in AI and machine learning.
- "*Health Studies*" Exploratory: Health studies explorations have a strong presence in Italy and the Netherlands, highlighting a focus on data-driven advancements in healthcare research and applications.
- "*Pervasive Intelligence in Cyber-Physical Systems for Future Society*": The exploration of pervasive intelligence in cyber-physical systems is primarily applied in Italy, showcasing a commitment to innovative solutions in the digital and physical convergence.
- "*Disaster Response and Recovery*": The focus on disaster response and recovery is predominantly applied in the Netherlands, underscoring the nation's commitment to data-driven strategies for managing and mitigating disasters.
- "*Societal and Industrial Impact of Next-Generation Internet & Beyond 5G Networks*": The exploration of societal and industrial impact within next-generation internet and beyond 5G networks is found in Belgium, .

The distribution of the research spaces reflects the multifaceted approach within SoBigData RI, addressing diverse societal, economic, and technological challenges. It underscores the infrastructure's capacity to foster cross-disciplinary collaboration and innovation, promoting data-driven solutions in various domains.

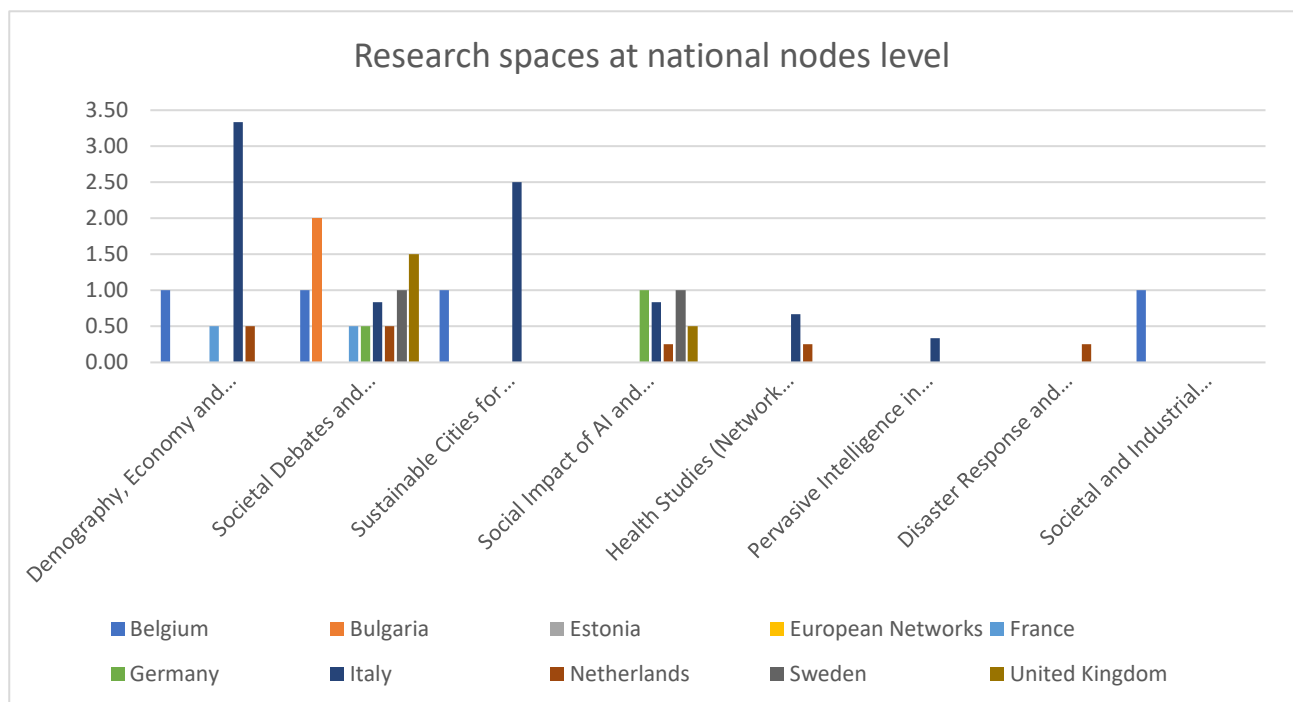


Figure 4.6.1. – SoBigData RI's research spaces for each national node

5 SoBigData Partners' Stakeholders National nodes in detail

In this section, we delve into a comprehensive analysis of the national nodes within the SoBigData Research Infrastructure, focusing on each partner's engagement and the number of stakeholders involved.

5.1 Number of stakeholders within each national node

Figure 5.1.1 provides an illustrative view of the number of stakeholders associated with each SoBigData RI partner across the national nodes. We highlight the most pertinent stakeholders per country below:

- **Belgium:** Within the context of SoBigData RI, Belgium collaborates closely with "Re-Imagine Europa," featuring 2 stakeholders engaged in research and innovation aims.
- **Bulgaria:** The "Centre for the Study of Democracy (CSD)" plays a pivotal role in Bulgaria's involvement, with 2 stakeholders contributing to the research ecosystem.
- **Estonia:** The "University of Tartu (UT)" is a key partner, with 2 stakeholders actively participating in collaborative efforts within Estonia.
- **European Networks:** In the domain of European Networks, two notable stakeholders, "EGI.eu" and "Openaire," each contribute 1 stakeholder, fostering cross-border collaboration and innovation.

- France: France is represented by two prominent entities - the "Centre National de la Recherche Scientifique (CNRS)" and the "Paris School of Economics (PSE)." These institutions boast 2 and 1 stakeholders, respectively, underscoring the significance of their contributions to the research landscape.
- Germany: Germany's engagement is enriched by the involvement of the "Fraunhofer Institute" and "Leibniz University Hannover (LUH)," each contributing 1 stakeholder to the collaborative efforts.
- Italy: Italy exhibits extensive involvement with stakeholders, primarily driven by key institutions such as the "University of Pisa (UNIPi)," "Consiglio Nazionale delle Ricerche (CNR)," and "Scuola IMT Alti Studi Lucca (IMT)." Additionally, "Scuola Normale Superiore (SNS)" and "Università degli Studi di Roma 'La Sapienza' (UNIROMA1)" also contribute to the rich research ecosystem within Italy.
- Netherlands: In the Netherlands, research is bolstered by institutions such as "Erasmus University of Rotterdam (EUR)," "TU Delft," "University of Utrecht (UU)," and "University of Amsterdam (UvA)," each having 1 stakeholder engaged in collaborative initiatives.
- Sweden: Sweden's contributions are represented by "KTH Royal Institute of Technology (KTH)," featuring 1 stakeholder that actively participates in research activities.
- United Kingdom: The United Kingdom is a dynamic research hub, with "University of Sheffield (USFD)" actively collaborating with 4 stakeholders. Additionally, "King's College of London (KCL)" contributes to the research landscape with 2 stakeholders, further enriching the collaborative efforts.

This comprehensive overview underscores the diverse network of stakeholders and their crucial roles within SoBigData RI's national nodes. The multi-dimensional collaborations and contributions from various partners and stakeholders illustrate the infrastructure's capacity to foster cross-disciplinary research, innovation, and knowledge exchange across international borders.

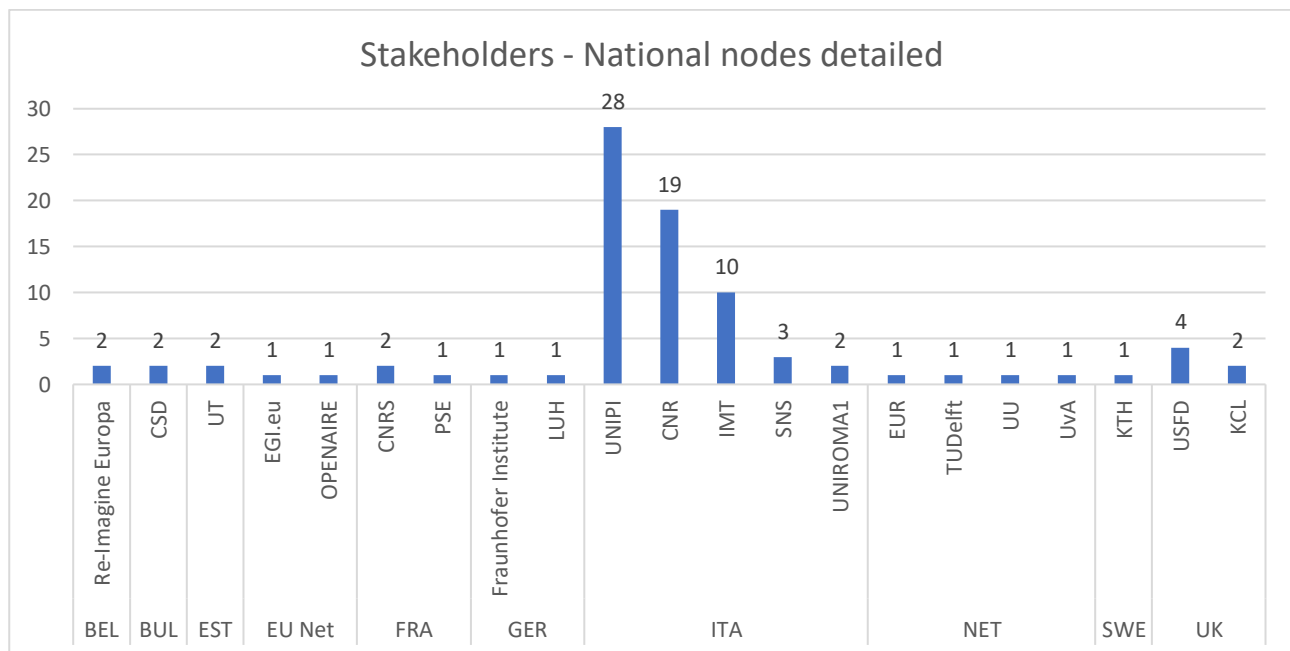


Figure 5.1.1. – Stakeholders - National nodes detailed

5.2 Status of collaboration between SoBigData RI's partners and the national nodes

Figure 5.2.1 offers an informative snapshot of the collaboration status between SoBigData RI's partners and the national nodes, detailing the completion, ongoing, and planned collaborations. We provide an overview of the collaboration status per country below:

- **Belgium:** The national node in Belgium collaborates actively with "Re-Imagine Europa," with 2 planned collaborations that signify the nation's ongoing commitment to research and innovation.
- **Bulgaria:** The "Centre for the Study of Democracy (CSD)" in Bulgaria has demonstrated active involvement, with 1 completed collaboration and an additional 1 ongoing collaboration. This exemplifies Bulgaria's dedication to research initiatives within the infrastructure.
- **Estonia:** The "University of Tartu (UT)" in Estonia is actively engaged, with 2 planned collaborations that reflect the nation's commitment to future research endeavors.
- **European Networks:** Within European Networks, both "EGI.eu" and "Openaire" are contributing through 1 ongoing project each, fostering cross-border collaboration and innovation.
- **France:** France is a dynamic contributor, with the "Centre National de la Recherche Scientifique (CNRS)" and the "Paris School of Economics (PSE)" actively participating in collaborative efforts. CNRS has completed 2 collaborations, while PSE has 1 completed collaboration, demonstrating the nation's rich history of research contributions.
- **Germany:** Germany showcases ongoing commitment to research with the "Fraunhofer Institute" and "Leibniz University Hannover (LUH)," each having 1 ongoing project, further enriching the collaborative landscape.
- **Italy:** Italy plays a central role within SoBigData RI, with several prominent partners actively involved in various stages of collaboration. The "University of Pisa (UNIPi)," "Consiglio Nazionale delle Ricerche (CNR)," and "Università degli Studi di Roma 'La Sapienza' (UNIROMA1)" have completed projects, while "UNIPi," "CNR," "Scuola IMT Alti Studi Lucca (IMT)," "Scuola Normale Superiore (SNS)," and "Università degli Studi di Roma 'La Sapienza' (UNIROMA1)" are engaged in ongoing projects. Additionally, CNR represents the sole Italian partner with a planned collaboration, reflecting Italy's substantial contributions.
- **Netherlands:** The Netherlands actively participates in research and innovation. "EGI.eu," "Erasmus University of Rotterdam (EUR)," "University of Utrecht (UU)," and "University of Amsterdam (UvA)" each have 1 ongoing collaboration, while "TU Delft" has 1 planned collaboration, signifying the nation's dedication to collaborative research.
- **Sweden:** In Sweden, "KTH Royal Institute of Technology (KTH)" has 1 completed project, showcasing the nation's active role in contributing to SoBigData RI's research initiatives.
- **United Kingdom:** The United Kingdom is a vibrant research hub, with "University of Sheffield (USFD)" involved in 4 ongoing collaborations and "King's College of London (KCL)" actively participating in 2 ongoing collaborations. These contributions highlight the nation's substantial commitment to research and innovation.

This detailed overview of collaboration status underscores the active and diverse engagement of national nodes and partners within SoBigData RI. The completion of projects, ongoing collaborations, and planned

initiatives highlight the infrastructure's capacity to foster cross-disciplinary research, innovation, and knowledge exchange, furthering the understanding of data science and big data analysis.

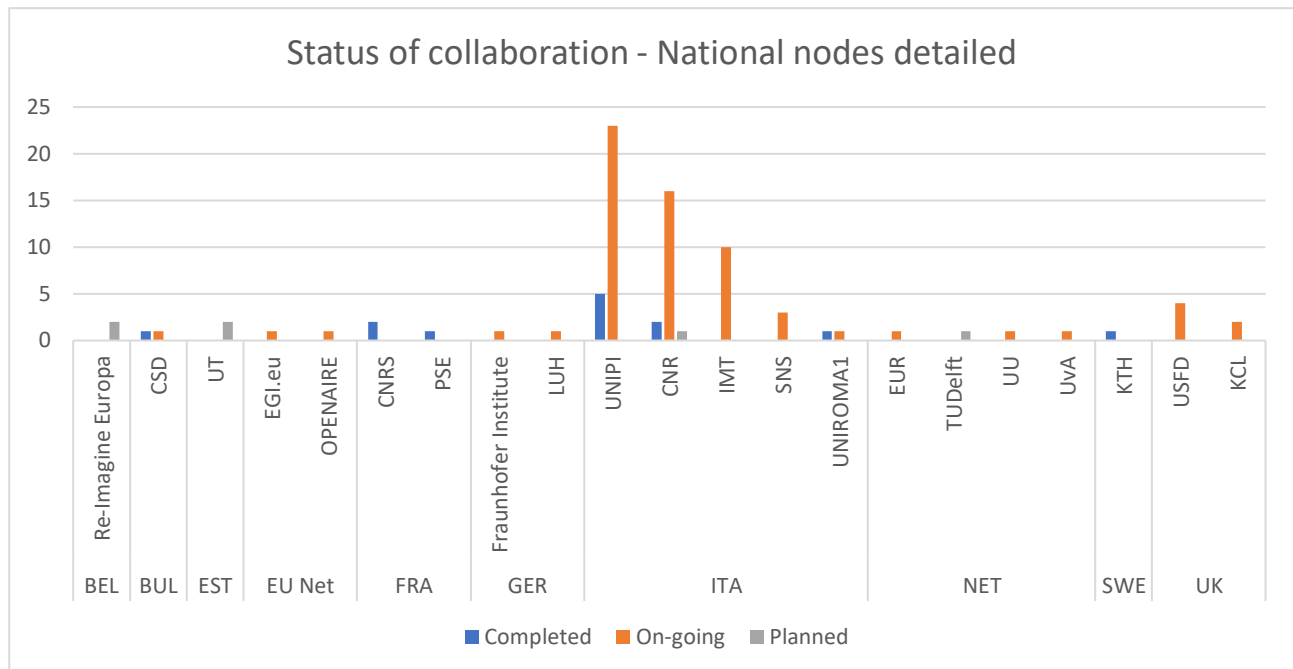


Figure 5.2.1 – Status of collaboration - national nodes detailed

5.3 Typologies of stakeholders engaged within each national node

Figure 5.3.1 provides a detailed overview of the main typologies of stakeholders engaged within the national nodes of the SoBigData Research Infrastructure. Each country exhibits unique and diverse stakeholder profiles, underscoring the rich tapestry of collaboration within the infrastructure. We discuss the stakeholder typologies for each country below:

- Belgium: "Re-Imagine Europa" in Belgium collaborates primarily with researchers, highlighting the significance of academic partnerships in advancing research and innovation.
- Bulgaria: The "Centre for the Study of Democracy (CSD)" is actively involved with public administrations and policymakers, emphasizing the importance of data-driven insights in the policy and governance realm.
- Estonia: The "University of Tartu (UT)" in Estonia engages with public administrations and policymakers, reflecting the institution's contributions to the public sector's data-related challenges and opportunities.
- European Networks: Within European Networks, "EGI.eu" and "OPENAIRE" establish partnerships primarily with non-profit organizations, fostering collaborative efforts aimed at addressing societal challenges.
- France: Both the "Centre National de la Recherche Scientifique (CNRS)" and the "Paris School of Economics (PSE)" in France have connections with researchers, highlighting the importance of

academic collaboration. Additionally, CNRS is also engaged with non-profit organizations, furthering the outreach and impact of research.

- Germany: The "Fraunhofer Institute" and "Leibniz University Hannover (LUH)" in Germany are actively involved with researchers, emphasizing the academic dimension of their contributions.
- Italy: Italy boasts a diverse and multi-dimensional stakeholder landscape. Industry stakeholders collaborate with key partners such as "University of Pisa (UNIPi)," "Consiglio Nazionale delle Ricerche (CNR)," "Scuola IMT Alti Studi Lucca (IMT)," and "Scuola Normale Superiore (SNS)," showcasing the importance of industry-academia collaboration. Researchers are involved with "UNIPi," "CNR," "IMT," and "Università degli Studi di Roma 'La Sapienza' (UNIROMA1)," highlighting the academic dimension of these partnerships. Furthermore, "UNIPi" and "CNR" are engaged with non-profit organizations, illustrating the multifaceted nature of their contributions. Public administrations and policymakers are actively involved with "CNR," "IMT," and "SNS," underlining the relevance of research insights in governance and policy.
- Netherlands: In the Netherlands, "Erasmus University of Rotterdam (EUR)," "University of Utrecht (UU)," and "University of Amsterdam (UvA)" engage in collaborative activities with researchers, emphasizing the academic dimension of their contributions. "TU Delft" collaborates with public administrations and policymakers, highlighting the role of data-driven insights in governance.
- United Kingdom: "University of Sheffield (USFD)" in the United Kingdom is actively involved with researchers and non-profit organizations, underscoring the collaborative nature of research and its societal impact.

This comprehensive overview of stakeholder typologies within the national nodes reveals the net of relationships, collaboration, and expertise that contribute to the advancement of data science and big data analysis within the SoBigData RI. The diverse engagement of researchers, industry stakeholders, non-profit organizations, and public administrations and policymakers highlights the interdisciplinary and cross-sectoral nature of the infrastructure's initiatives.

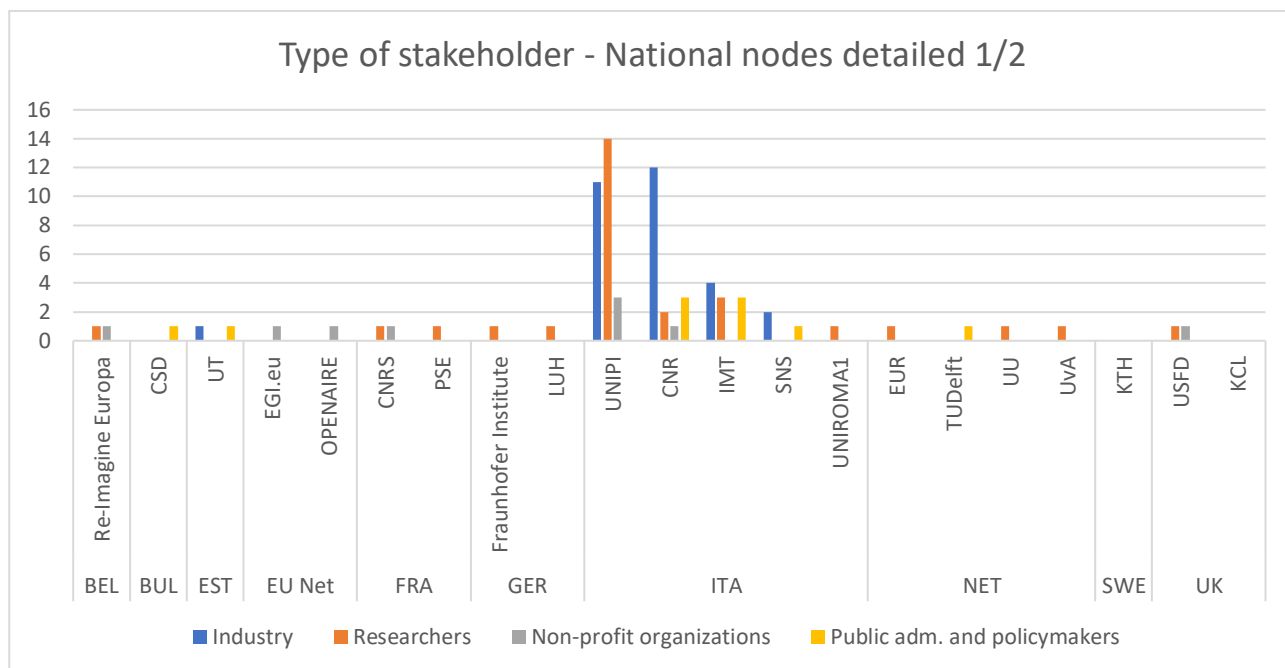


Figure 5.3.1. – Types of stakeholders - national nodes detailed part 1

Figure 5.3.2 provides insights into the residual typologies of stakeholders within specific national nodes of the SoBigData Research Infrastructure. These unique stakeholder profiles reveal additional dimensions of engagement, underscoring the diverse and multi-faceted nature of collaboration. We delve into the residual stakeholder typologies for each country below:

- Bulgaria: The "Centre for the Study of Democracy (CSD)" in Bulgaria establishes relations with international organizations, highlighting the institution's active engagement on the global stage and the importance of international partnerships in advancing research initiatives.
- Italy: Italy presents distinctive stakeholder engagement. "Consiglio Nazionale delle Ricerche (CNR)" collaborates with teaching institutions, emphasizing the academic dimension of their contributions to education and research. "Università degli Studi di Roma 'La Sapienza' (UNIROMA1)" has collaborations with students, showcasing the institution's commitment to student-centric initiatives and knowledge exchange.
- Sweden: In Sweden, "KTH Royal Institute of Technology (KTH)" engages with students, fostering a learning environment that actively involves the next generation of data scientists and researchers.
- United Kingdom: The United Kingdom's contributions encompass a diverse range of stakeholder engagement. "King's College of London (KCL)" collaborates with teaching institutions, emphasizing the institution's role in education and knowledge dissemination. "University of Sheffield (USFD)" is actively engaged with journalists, highlighting the institution's outreach to the media and the dissemination of research findings to a broader audience.

This detailed overview of residual stakeholder typologies within specific national nodes further enriches our understanding of the breadth and depth of collaboration within the SoBigData Research Infrastructure. The multi-dimensional engagement with international organizations, teaching institutions, students, and

journalists underscores the infrastructure's ability to foster diverse and impactful partnerships across various sectors and domains.

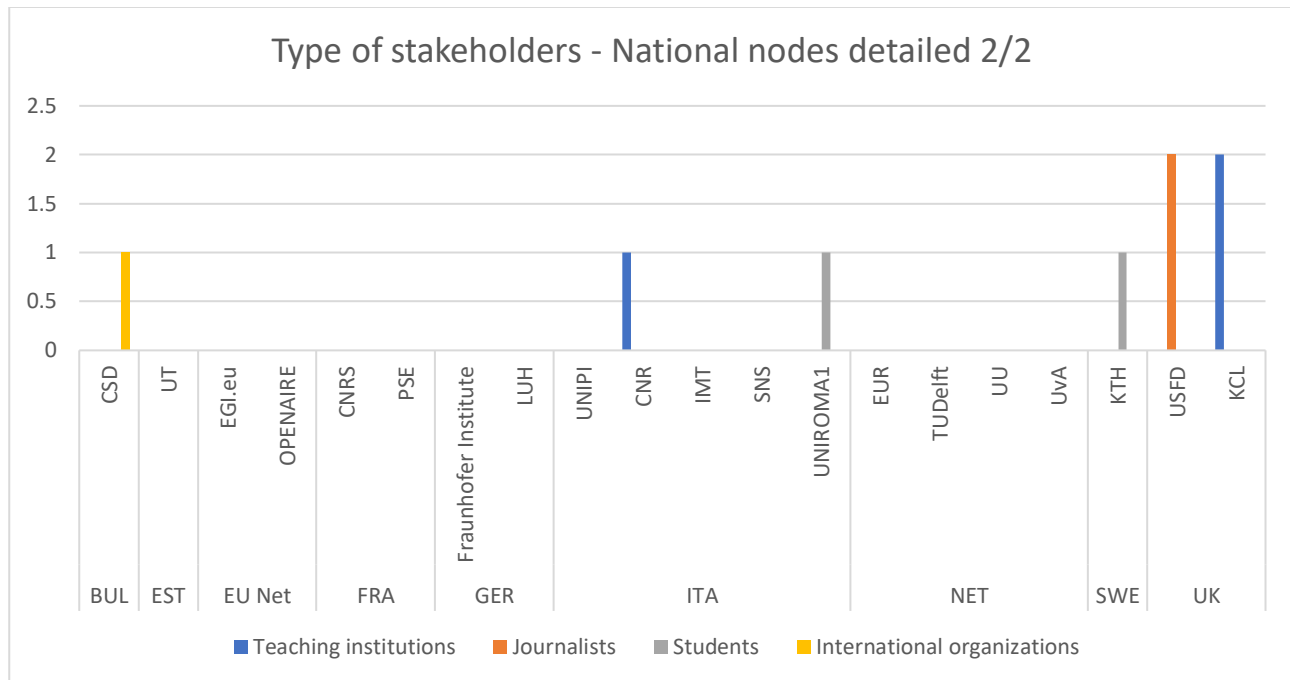


Figure 5.3.2. – Types of stakeholders - national nodes detailed part 2

5.4 Singular typologies of collaborations within each national node

Figure 5.4.1 provides an overview of the singular collaboration typologies within the national nodes of the SoBigData Research Infrastructure. These specific collaboration activities showcase the diverse range of contributions from each partner, enhancing our understanding of their unique roles. We delve into the singular collaboration typologies for each country below:

- Belgium: "Re-Imagine Europa" in Belgium actively engages in consulting activities, underscoring their role as valuable consultants in the research and data science domain.
- Bulgaria: The "Centre for the Study of Democracy (CSD)" in Bulgaria participates in both consulting and training activities, highlighting their dedication to knowledge sharing and capacity building.
- France: "Centre National de la Recherche Scientifique (CNRS)" in France contributes to development activities, emphasizing their role in advancing research and innovation in the data science field.
- Germany: The "Fraunhofer Institute" in Germany is actively involved in joint research activities, fostering collaborative research initiatives with a focus on shared outcomes.
- Italy: Italy showcases a diverse spectrum of singular collaboration typologies. "University of Pisa (UNIFI)," "Consiglio Nazionale delle Ricerche (CNR)," "Scuola IMT Alti Studi Lucca (IMT)," "Scuola Normale Superiore (SNS)," and "Università degli Studi di Roma 'La Sapienza' (UNIROMA1)" have been engaged in development activities, highlighting their substantial contributions to research advancements. Additionally, "UNIFI," "CNR," and "IMT" are involved in consulting activities, signifying

their expertise as consultants in the data science and big data analysis domain. "CNR" and "UNIROMA1" are actively contributing to training activities, emphasizing their commitment to knowledge dissemination and capacity building. "CNR" stands out as the sole partner involved in joint research activities, fostering collaborative research activities.

- Netherlands: In the Netherlands, "TU Delft" participates in consulting activities, showcasing their expertise as valuable consultants in data-related initiatives, while "University of Amsterdam (UvA)" is involved in development activities, highlighting their commitment to research advancements.
- Sweden: "KTH Royal Institute of Technology (KTH)" in Sweden actively contributes to development activities, emphasizing their role in advancing research and innovation in data science.
- United Kingdom: The United Kingdom demonstrates a dynamic range of singular collaboration typologies. "University of Sheffield (USFD)" actively engages in development and consulting activities, showcasing their expertise as consultants and their commitment to research advancements. "King's College of London (KCL)" is actively involved in joint research activities, fostering collaborative research initiatives.

This overview of singular collaboration typologies underscores the dynamic contributions and unique roles of each national node and partner within the SoBigData RI. The active engagement in consulting, training, development, and joint research activities highlights the diverse expertise and contributions that drive innovation and advancement in the field of data science and big data analysis.

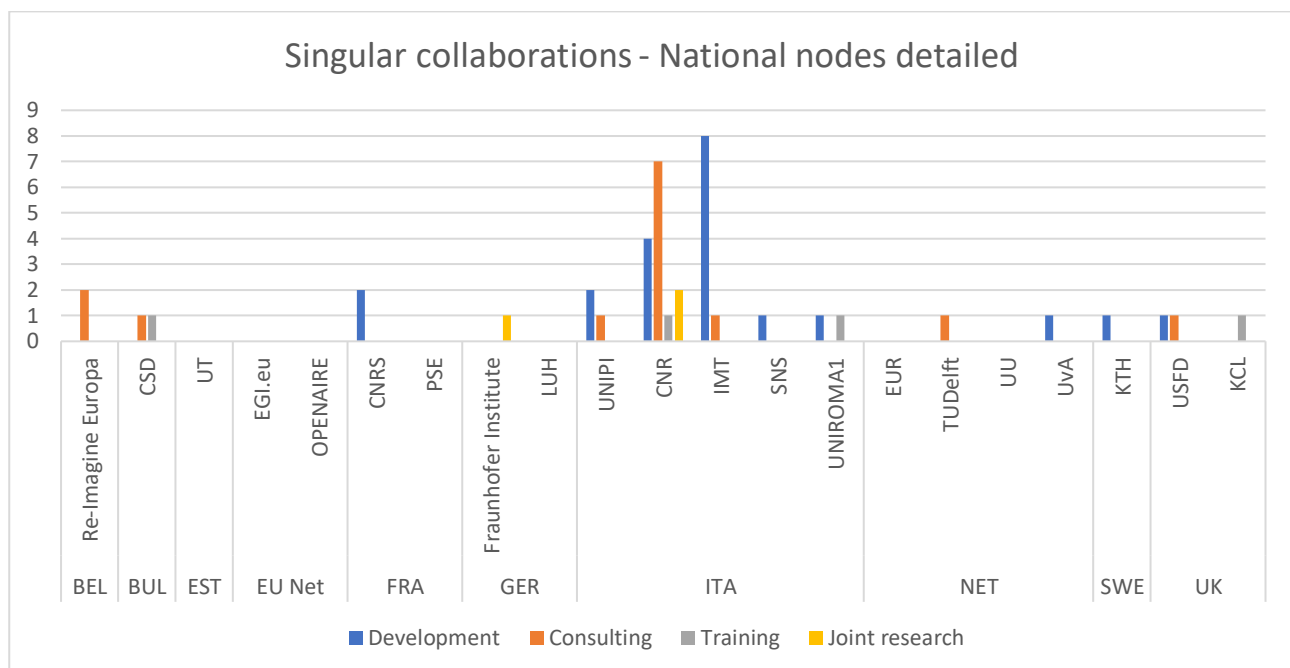


Figure 5.4.1. – Singular collaborations - national nodes detailed

5.5 Compound typologies of collaborations within each national node

Figure 5.5.1 provides an overview of the compound collaboration typologies within the national nodes of the SoBigData Research Infrastructure. These collaborative activities combine multiple dimensions, showcasing

the multifaceted nature of contributions from various partners. We delve into the compound collaboration typologies for each country below:

- Estonia: The "University of Tartu (UT)" in Estonia actively engages in two compound collaboration typologies, specifically "Development and Consulting" and "Training, Development and Consulting" activities. This multifaceted approach underscores the institution's commitment to fostering diverse and impactful collaborations.
- European Networks: Within European Networks, "EGI.eu" is involved in "Development and Consulting" activities, emphasizing their role in supporting research advancements and consultancy services. "OPENAIRE" is engaged in "Training and Consulting" activities, highlighting their commitment to capacity building and knowledge dissemination.
- France: The "Paris School of Economics (PSE)" in France contributes to "Training, Development, and Joint research" activities, showcasing their multifaceted approach to collaborative research and knowledge sharing.
- Germany: "Leibniz University Hannover (LUH)" in Germany actively participates in "Training and Development" activities, highlighting their commitment to education and skill development in the data science and big data analysis domain.
- Italy: Italy presents a range of compound collaboration typologies. "University of Pisa (UNIPi)," "Consiglio Nazionale delle Ricerche (CNR)," and "Scuola Normale Superiore (SNS)" are actively engaged in "Development and Consulting" activities, emphasizing their multifaceted contributions to research advancements and consultancy services. "UNIPi" and "Scuola IMT Alti Studi Lucca (IMT)" are involved in "Training and Consulting" activities, reflecting their dedication to knowledge dissemination and capacity building. "CNR" stands out by actively contributing to "Training, Development, and Consulting" activities, underscoring their multifaceted approach to collaborative endeavors.
- Netherlands: In the Netherlands, "Erasmus University of Rotterdam (EUR)" and "University of Utrecht (UU)" are involved in "Training, Development, and Consulting" activities, showcasing their multifaceted contributions to research advancements, education, and consultancy services.
- United Kingdom: The United Kingdom demonstrates a dynamic range of compound collaboration typologies. "University of Sheffield (USFD)" actively engages in "Development and Consulting" as well as "Training and Development" activities, highlighting their multifaceted contributions to research advancements and education. "King's College of London (KCL)" is actively involved in "Training and Consulting" activities, emphasizing their commitment to knowledge dissemination and consultancy services.

This comprehensive overview of compound collaboration typologies underscores the dynamic and multifaceted contributions of each national node and partner within the SoBigData Research Infrastructure. The active engagement in "Development and Consulting," "Training and Development," "Training and Consulting," and "Training, Development, and Joint research" activities highlights the diverse expertise and contributions that drive innovation and advancement in the field of data science and big data analysis.

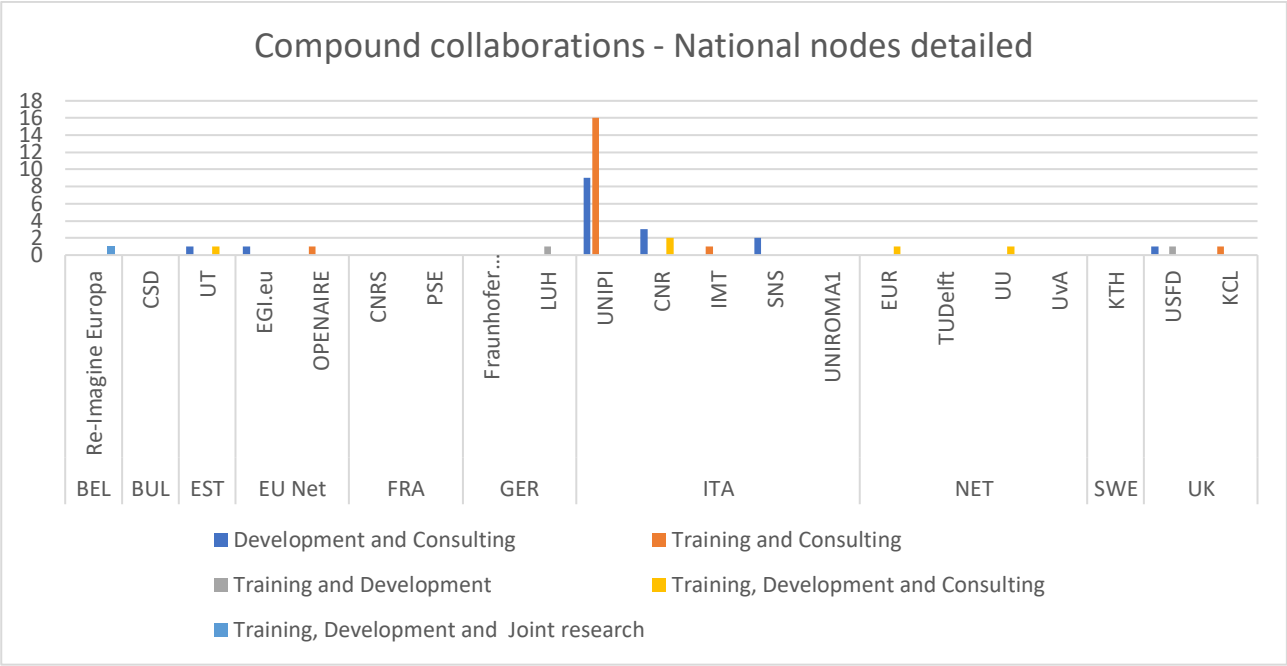


Figure 5.5.1. – Compound collaborations - national nodes detailed

6 SoBigData partners' interviews

Up to this point, this document has primarily presented an analysis of stakeholder data contributed by the SoBigData partners. In this section, we aim to further enrich our analysis by providing insights derived from interviews with selected stakeholders, conducted under the purview of WP3 leadership. This section is thoughtfully structured into three distinct subsections: 1) *preparation*; 2) *execution*, and 3) *analysis*, each serving as a critical building block in the process.

6.1 Preparation

The stakeholder interview process began in August 2023, starting with a thorough preparation phase. Our first task was to develop a well-thought-out interview protocol, a structured roadmap that would guide our interactions with stakeholders. The rationale behind our stakeholder selection criteria was rooted in the desire to engage individuals who had actively participated in completed collaborations within the SoBigData framework. This selection strategy aimed to tap into the unique perspective that such stakeholders possessed, having been involved in every facet of the collaboration. The inherent advantage of this approach was the stakeholders' ability to provide a holistic and nuanced evaluation of the entire collaborative process. Their insights and judgments on the quality and success of these collaborations were, therefore, of paramount importance.

In the subsequent sections, we provide a detailed account of the interview questions formulated for this purpose. These questions were designed to elicit rich insights from stakeholders, shedding light on their experiences, observations, and recommendations, thus contributing to a deeper understanding of SoBigData's collaborative dynamics.

1. How did you hear about SoBigData?
2. Could you tell us some details toward the collaborative relationship between your organization and the partner of SoBigData (in terms of goals, time, people involved, etc.)?
3. What are the results of this collaboration?
4. What is the added value of this collaboration for your organization?
5. What are the main difficulties that emerged during this collaboration?
6. Are you satisfied about the collaboration? If so, could you give us some information on how to improve the services? If not, what are the most important things that must be changed to improve the service offered by SoBigData and its partners?

The questions that we selected cover the entire process of the involvement of the stakeholder within the collaboration with SoBigData's partner, the focus of the collaboration, its results, and the benefits of the collaboration. Moreover, it reports a question toward the potential difficulties (in a broad sense) of the collaboration. Finally, the protocol leaves a space for satisfaction and suggestions, useful for improving the activities of the RI.

During September 2023, we contacted the SoBigData partners who have conducted completed collaborations with stakeholders. We asked to them for the email contacts of the external members who was involved in the collaborations. After getting the contacts, we wrote to every stakeholder to organize an online interview. In particular, the partners reported 8 completed collaborations. After contacting them 2 were available for the interview in time to be included in this deliverable. Therefore, the response rate was 25%.

6.2 Execution

In October 2023, we made a significant stride into conducting thorough online interviews with stakeholders. We selected this approach with care to uncover the various challenges and opportunities that stakeholders, representing diverse backgrounds and experiences, might reveal for the larger SoBigData RI.

These online interviews were conducted following a rigorous methodological approach in qualitative research (Gehman et al., 2018), with each session spanning a duration of 30 to 45 minutes. The extended time frame for these interactions allowed for thorough exploration and a deep dive into the thoughts, perspectives, and insights of the participants. It enabled us to engage in meaningful dialogues that transcended the surface and delved into the intricate details of their experiences.

In the subsequent sections, we will meticulously report on the results that emerged from the execution of these interviews, presenting a comprehensive overview of the process, the rich insights gathered, and the implications of these interactions for the SoBigData RI.

6.3 Analysis

As we analyze the interviews, a key recurring theme is the crucial role that each researcher or partner plays in introducing stakeholders to the project.

This introduction sets the stage for the collaboration, shaping its initial trajectory. For example, one stakeholder shared a personal experience, stating,

"I have known [name of the researcher, NR] when I was a PhD researcher at Palermo, and I did a seminar. Then, fast forward, I started the Lab in Boston, and I contacted people that I appreciate in Italy to keep in touch. [NR] was involved in many projects, including SoBigData, and we started to talk about potential collaborations."

This anecdote underscores the significance of pre-existing relationships and proactive outreach in sparking collaborations. Another aspect that emerges is the diverse and complex nature of the projects requested by stakeholders. Researchers and partners alike are often required to delve into unfamiliar subject matter to effectively address these requests. For instance, in the context of a project involving genomics and big data,

"[NR] learnt very soon all the basic notions of biology that were useful to formalize the problem." This demonstrates the adaptability and dedication of the SoBigData community in acquiring new knowledge domains to tackle complex challenges.

Furthermore, these collaborations have resulted in the generation of significant scientific knowledge. In one instance, a project related to mobility prompted collaborative efforts to develop a streamlined model that yielded similar results to the original, but with greater efficiency. This collaborative effort, which also involved a dedicated student, culminated in a high impact publication. An essential aspect of these collaborations is the substantial knowledge sharing among researchers. This exchange of expertise and insights amplifies the value derived from these collaborations, contributing to a broader knowledge base and fostering a culture of continuous learning. Importantly, stakeholders do not report any notable difficulties, be they technical or administrative/bureaucratic, in participating in these collaborations. This underscores the efficiency and effectiveness of the collaborative framework. Moreover, there is a consistently high degree of satisfaction expressed by stakeholders regarding their collaborative experiences. Each stakeholder affirms their willingness to engage in such collaborations again, underscoring the mutually beneficial nature of these partnerships.

In conclusion, it is worth highlighting a concern voiced by one researcher regarding potential limitations imposed by EU rules, particularly in terms of privacy for research purposes. This concern revolves around the balance between safeguarding citizen privacy and the need for detailed data for in-depth analysis. The researcher notes,

"We are interested in having detailed information that permits very deep analysis, which is not possible if data is aggregated. The higher the aggregation, the lower the detail, and the lower the possibility to understand what is really happening."

This perspective highlights the importance of having clear rules related to the ethical and legal aspects when data governance and data analysis are performed.

7 SoBigData RI future developments

This report serves as a repository of information and insights crucial for crafting the business model for SoBigData RI. Furthermore, our findings hold the potential to benefit the entire research infrastructure. In alignment with the project's leadership team, we propose three principal trajectories for the future engagement of SoBigData RI partners:

1. **Renewed Engagement and Involvement:** The first proposed trajectory, emphasizing renewed engagement and involvement, is not only rooted in the insights gained from stakeholder interviews but also underscores the necessity for proactive knowledge dissemination. This proactive approach is seen as a pivotal step in ensuring that the benefits of data science and big data analysis are widely accessible and democratically distributed. It involves encouraging all stakeholders, including researchers, industries, government entities, and not-for-profit organizations, to actively participate in and contribute to the ongoing success and evolution of SoBigData RI. This engagement is not limited to the existing scope of the project but extends to cultivating a more dynamic relationship with these stakeholders.

2. **Service Catalog Creation:** The second trajectory, which focuses on the creation of a service catalog, is a response to the multifaceted needs of SoBigData RI. A comprehensive understanding of the RI's core activities, resources, and costs is critical for strategic planning and efficient resource allocation. The service catalog serves as a roadmap that encapsulates the range of services offered by SoBigData partners. By categorizing these services into Key Exploitable Results (KERs), we gain a structured view of how the RI delivers value. These KERs encompass diverse areas, such as ethics and legal services, educational offerings through the SoBigData Academy, social mining and AI resources, infrastructure-as-a-service, and community building and networking. The service catalog not only provides clarity for internal management but also strengthens communication channels with governmental bodies and external stakeholders. It becomes a tool to showcase the wide spectrum of services that SoBigData RI can offer to support data-driven endeavors across sectors.
3. **Drafting a Business Plan:** The third trajectory pertains to the formulation of a preliminary business plan, particularly with a focus on government engagement. SoBigData RI's potential for generating substantial social impact necessitates a structured document that articulates a well-thought-out approach on the entrepreneurial aspects of the project. A well-constructed business plan highlights the financial sustainability of the RI and provides assurance to stakeholders, including policymakers. It serves as a compelling instrument for conveying the RI's commitment to effectively bridge the gap between research and practical applications. The plan is not merely a static document but a dynamic framework for encouraging data-driven entrepreneurial initiatives that align with the broader goals of societal progress and innovation, thus making a case for incentivizing and promoting such initiatives at the policy level.

In summary, these three proposed trajectories are integral to SoBigData RI's journey toward a more robust, dynamic, and sustainable future. They underscore the importance of active engagement, comprehensive service documentation, and a structured plan to harness the full potential of data science and big data analysis for the betterment of society. These strategic suggestions delineate a clear and resilient path forward for SoBigData RI, ensuring a more active role for partners, a comprehensive overview of services, and a solid foundation for the formulation of a business plan. This, in turn, can bolster the RI's prospects for long-term success and its enduring value as a driver of data science and big data analysis.

8 Conclusions

This report stands as an exploration of the network of stakeholders entwined within the SoBigData RI, a collaborative task led by the WP3 team at the Institute of Management at the Sant'Anna School of Advanced Studies. At its core, our aim with this stakeholder analysis was to painstakingly identify, categorize, and scrutinize the stakeholders collaborating with the SoBigData RI.

Our approach started with a comprehensive overview of the data collection process, laying bare the methodologies employed and the sources tapped into for the extraction of vital information about these stakeholders. Concerning the pertinence of these stakeholders, we delved into how distinct groups of stakeholders engage with and exert their presence over the services offered by the RI. The involved stakeholders encompass a mix, ranging from "Industry," "Researchers," "Non-profit organizations," to "Public administrations and policymakers." Meanwhile, the net of collaboration unfolds in the forms of "Development," "Consulting," "Training and Consulting," and "Development and Consulting." The list of primary sectors inside the RI include the domains of "Computer science," "Economy and Finance," and "Social sciences," while the bulk of the researcher activities gravitate towards "Public administrations and policymakers," "Students," and "Industry." The problem-solving domains featured in our analysis exhibit a rich diversity, encompassing "Model development," "Data analysis," "AI application," and "Data acquisition." It is important to observe that "Industry," "Researchers," and "Non-profit organizations" display a mosaic of heterogeneous problem-solving domains, while other stakeholders specialize in specific, well-defined domains. In addition, we studied the national nodes, unfurling the network of stakeholders linked to these entities. The sectors of "Computer science" and "Social Sciences" play pivotal roles and are strategically distributed among these nodes. The problem-solving domains of "Model development," "Data analysis," and "AI acquisition" similarly find their distribution and network of influence among these nodes. Additionally, the research paths include "Demography, Economy, and Finance 2.0," "Societal Debates and Misinformation analysis," and "Social impact of AI and Explainable Machine Learning," each having its footprint distributed among these nodes. This report also included a presentation of the invaluable insights garnered from interviews conducted with specific stakeholders. Key takeaways underscore the importance of each partner's proactive role in introducing third-party entities into the RI, the far-reaching influence of cross-disciplinary approaches on collaborations, the efficiency and effectiveness of administrative processes, and the remarkable degree of satisfaction from these partnerships.

Furthermore, we showed an array of prospective pathways for future development. These recommendations are interwoven with our understanding of the stakeholders and the pivotal roles they perform within the SoBigData RI PPP. In a nutshell, it appears judicious to endorse a renewed "involvement-role" among all partners, to craft a comprehensive service catalogue, and to formulate a strategic business plan. This report painted a panoramic view of the stakeholder landscape, underscoring the critical of their contributions to the enduring success and sustainability of the project. This report does not serve as a static document but as a dynamic and empowering tool for stakeholders, project managers, and decision-makers alike. It empowers

them with the knowledge and insight needed to make well-informed decisions, foster enhanced collaboration, and fortify the project's ongoing growth and prosperity.

In conclusion, the stakeholder analysis of SoBigData RI uncovers a complex, multifaceted, and dynamic network of collaborations. It entails diverse stakeholders, extensive projects, and a wide spectrum of collaboration types. These interactions show the pivotal role the infrastructure plays in the ever-evolving landscape of research, innovation, and societal progress.

References

Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge university press.

Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2018). Finding theory–method fit: A comparison of three qualitative approaches to theory building. *Journal of Management Inquiry*, 27(3), 284-300.

Grossi, V., Giannotti, F., Pedreschi, D. et al. Data science: a game changer for science and innovation. *Int J Data Sci Anal* 11, 263–278 (2021). <https://doi.org/10.1007/s41060-020-00240-2>

Appendix 1. Questionnaire “SoBigData PPP Stakeholder Analysis”

86 risposte

 Visualizza in Fogli

Accetta risposte ☒

Riepilogo

Domanda

Individuali

 **SOBIGDATA** ^{PPP}
PREPARATORY PHASE PROJECT

SoBigData PPP Stakeholder Analysis - deadline extended to September 8th

 [Cambia account](#)

 Non condiviso

* Indica una domanda obbligatoria

SoBigData Partner *

La tua risposta

Contact point e-mail *

La tua risposta

Stakeholder name *

La tua risposta

Is it a relevant stakeholder for SoBigData RI in terms of collaboration and development of the Research Infrastructure? *

☐ Yes

☐ No

Type of collaboration *

- **Training:** SoBigData partner organizes training materials or courses for the stakeholder

- **Development:** SoBigData partner develops a method, an experiment, or deploys a technological infrastructure for the stakeholder

- **Consulting:** SoBigData partner supports the stakeholder with experts in the field in order to improve/develop an internal project. This may also include the development of a proof of concept

☐ Training

☐ Development

☐ Consulting

☐ Altro: _____

Status of collaboration *

☐ Completed

☐ On-going

☐ Planned

Type of stakeholder *

- ☐ Researchers
- ☐ Industry
- ☐ Policy and law makers
- ☐ Public administrations
- ☐ Non-profit organizations
- ☐ Teaching institutions
- ☐ Funders
- ☐ Data analysts
- ☐ Journalists
- ☐ Students
- ☐ Altro: _____

Stakeholder Sector *

- ☐ Computer Science
- ☐ Economy and Finance
- ☐ Social Science
- ☐ Humanities
- ☐ Medicine
- ☐ Biology
- ☐ Law and Ethics
- ☐ Telecommunications
- ☐ Insurance
- ☐ Retail
- ☐ Transport
- ☐ Sport analytics
- ☐ Energy
- ☐ Statistics
- ☐ Altro: _____

Related Exploratory, if any *

- ☐ Demography, Economy and Finance 2.0
- ☐ Disaster Response and Recovery
- ☐ Health Studies (Network Medicine & Sports Data Science)
- ☐ Pervasive Intelligence in Cyber-Physical Systems for Future Society
- ☐ Societal and Industrial Impact of Next-Generation Internet & beyond 5G Networks
- ☐ Societal Debates and Misinformation Analysis
- ☐ Social Impact of AI and Explainable Machine Learning
- ☐ Sustainable Cities for Citizens (including Migration)
- ☐ None

Problem-solving domain *

☐ Data acquisition

☐ Data analysis

☐ Data storage/curation

☐ Model development

☐ AI Application

☐ Altro: _____

Number of SoBigData Partner's researchers involved *

La tua risposta _____

Developing and consolidating the European research infrastructures landscape, maintaining global leadership (2021)", Grant Agreement n.101079043, "SoBigData RI PPP: SoBigData RI Preparatory Phase Project"

Invia

Cancella modulo

Appendix 2. Interviews' questions

- How did you hear about SoBigData?
- Could you tell us about some details toward the collaborative relationship between your organization and the partner of SoBigData (in terms of goals, time, people involved, etc.)?
- What are the results of this collaboration?
- What is the added value of this collaboration for your organization?
- What are the main difficulties that emerged during this collaboration?
- Are you satisfied about the collaboration? If so, could you give us some information on how to improve the services? If not, what are the most important things that must be changed to improve the service offered by SoBigData and its partners?

Appendix 3. Interviews' transcripts

Interview n.1

Date: 3rd October, 2.27 pm (Rome time)

Location: the interview has been conducted and registered on Microsoft Teams

Duration: 30 minutes, 34 seconds

Interviewer

Could I ask you how you came to know about the SoBigData project? We know that you had the pleasure of collaborating with one of our core partners....

Interviewed n.1

I have known [name of the researcher, NR] when I was a PhD researcher at Palermo, and I did a seminar. Then, fast forward, I started the Lab in Boston, and I contacted people that I appreciate in Italy to keep in touch. [NR] was involved in many projects, including SoBigData, and we started to talk about potential collaborations. Then [NR] came to Boston, visited my Lab, and we developed them... There was a specific genomics problem, I don't know how much your background is on genomics, but, eh, there is a big finger problem, in the sense that we had a problem... We were trying to study DNA variances, mutations, that is, they are present, let's say in populations and there are methods that allow you to associate these variances with different diseases, so you have this Correlation Association however you don't know why let's say these variances have this impact and then you can make hypotheses about what these variances do... One of these hypotheses is that these variances change for certain proteins that are called transcription factors... Enrollment factors change the expression of genes, so you imagine a normal program where these factors can bind... if you have this variance maybe they can't bind anymore, or they bind in places where they're not supposed to bind and so your, your cellular program changes and if it changes, if you're unlucky, maybe you get let's say diseases.... We had this problem of scanning millions of variances with transcription factor models and we did kind of a prototype. But let's say our code was taking months... So we contact [NR]... we have this interesting problem however let's say waiting months is not a feasible thing, can you help us? He says yes. We explained the problem. [NR] learnt very soon all the basic notions of biology that were useful to formalize the problem and in about two weeks, a month or so, he had a prototype that he made. We integrated it into our pipeline and from months of computation we were up to three hours. So that was in my opinion a nice accomplishment... And then we wrote a paper describing this... we sent it to [Name of the journal, NJ] and published it. Then with [NR] we kept thinking about this problem and we're exploring some other models that are a little bit more sophisticated. And so, let's say, this is the collaboration that we started related to big data. Let's say, the nice thing is that then we actually published this paper, however we are still exploring other topics related to this initial collaboration.

Interviewer

Great, great thank you very much, very clear -- no, I don't do genomics. My background is all about economics and management however I thank you for the clarity. That's... let's say, the problem has been framed, and, if I understand correctly, thanks to the collaboration with Professor [NR], within three weeks more or less making an estimate, we went to decrease a lot those waiting times that otherwise required. And so now let's say from months that it used to take now it takes three hours... And as far as who was involved? Was there only Professor [NR] in addition to your internal team?

Interviewed n.1

Yes, yes, correct.

Interviewer

Perfect... if we wanted to look a little bit at the major difficulties... let's say, if there were difficulties in collaboration, something that maybe was more cumbersome... Let's talk broadly, in a bureaucratic rather than administrative sense...

Interviewed n.1

No, in the end no... The only thing maybe I was sorry... [NR] is very busy, isn't he? I would have liked to, maybe deepen the visit... however, in the end we didn't have any problems, everything went smoothly even in my department.

Interviewer

Perfect, great, great, so let's say if you were to go back would you still do it again?

Interviewed n.1

Yes!

Interviewer

If I can try to summarize... let's say that this kind of collaboration, with all the steps that were followed to get to this result... The results were seen both in terms of, precisely, data processing time, but also in terms of papers, so also scientific publications, that is, the added value...

Interviewed n.1

Yes... also, let's say from a point of view... being in the Lab, [NR] maybe gave us also other ideas, let's say, it's nice to talk... I don't know, maybe about the project that we worked on, but for example he did another seminar on other technical things that, let's say, enriched the whole Lab... so, let's say... so many benefits, not only related to this project. The fact that he is a person who spends time in your Lab and every day you can talk to him, in my opinion is a very very nice thing.

Interviewer

Excellent. In fact even this exchange of knowledge that goes beyond just technical, right? I mean, solving a problem, I solve it, but ... there is also then something else that in the collaboration you found that worked ... if I understand correctly ... an interchange, something that goes beyond the project itself ...

Interviewed n.1

Yes!

Interviewer

Perfect, so I look for the meantime thank you so much for these impressions, I see that there is a general satisfaction. Here, for what is precisely the collaboration with the research world. So. So here you go, thank you if there's any really freewheeling suggestions that you want to give, or maybe if there's, if you think, of... here we've said it all and then we'll bring back a little bit these impressions...

Interviewed n.1

As for me, I will try to stay updated on SoBigData project initiatives on the genomics side and services related to young researchers. I see that there are several opportunities... thanks for your time!

Interviewer

I notice a tendency, here's... a general satisfaction, so it's good to also report these feelings, so I also say thank you for the time that you have dedicated to me, that you have dedicated to us...

Interviewed n.1

Okay, have a good day! Bye.

Interviewer

Bye!

Interview n.2

Date: 3rd October, 3.26 pm (Rome time)

Location: the interview has been conducted and registered on Microsoft Teams

Duration: 45 minutes, 21 seconds

Interviewer

Could I ask you how you learned about the SoBigData project?

Interviewed n.2

Let's say an indirect knowledge, in short, it's not that I have ever participated in the activity of the project... but whatever clearly, having in Pisa a main base, being part of the CNR in Pisa, working on the big data sector, I know [Name of the researchers, NRs], clearly this project has not gone unnoticed. It had a big funding and so in short, I am aware of the project. I know several people from the CNR group. From the former group where [NR] worked. I mean, I know quite well, in short.

Interviewer

Perfect. We know precisely that you were fortunate enough to collaborate with Professor [NR], right? Through some projects. We wanted to ask you, here in a way also very much so, without possibly going into too much detail, if too technical, what was a little bit the object let's say of the collaboration. How long did it last? The subjects that got involved.

Interviewed n.2

Yes, so the collaboration came about because of a collaborative program that exists between [Name of the Institution, NI] and the University of [NI], which is called [Name of the Program, NM]. I don't know if it's always active, honestly though it was active a few years ago which is a program just focused in funding a with limited amounts. I mean, to make trips essentially in the two directions, mainly to get, let's say, professors from [Name of the city, NC] to [NI] to establish research collaborations. So, with [NR] and one of his thesis students, we have been working basically on improving some of the algorithms that we have been working on. So, in terms of shared mobility, we did one of the first projects that I worked on when I came to the [Name of the Lab] was just to try to understand the potential of ride sharing. Cab, for example. Using data from New York and other cities and being a computationally very complex problem, the data structures, that is, how to model, let's say the possibilities of sharing or not sharing a ride, needed to be handled as efficiently as possible. Together with [NR] we tried to introduce a simpler model than the original that produced similar results, but faster. We worked together with a student that focused the thesis on this topic. And then the result was a good publication. I know that now this thesis student is a doctoral student with Professor [NR]. So, I would say the collaboration has been a success.

Interviewer

Great and so let's say the academic results, if we want to summarize it a little bit were precisely the publication in an academic journal and let's say a kind of how to say... of knowledge distribution, of knowledge sharing that then allowed a subject to possibly grow in its training, if we want to summarize, if that collects a little bit what the added value let's say of this collaboration, which anyway came in a little bit under the hat for some aspects on big data, here. Do you agree?

Interviewed n. 2

Mhm, yes, yes, I absolutely agree on with what you say was clearly seen the limited extent of the collaboration, because precisely the funds covered visits of two or three weeks, so in short, we worked together on this idea. From this idea, then, thanks to the work developed by the thesis in collaboration also with other CNR colleagues in Pisa, we precisely finished developing the idea and then a publication came out in a very important journal, so... from a point of view let's say of the collaboration, then, also of this, of this thesis student who then, precisely, thanks also to this work became a doctoral student at the University of [NC]. I think, in short, the game has produced a good candle.

Interviewer

Ok. And in terms of the number of people involved, more or less, let's say in this project, the main ones were three of you, if I understand correctly, so Professor [NR], you and the thesis student.

Interviewed n.2

Yes, and then there was another colleague of mine from CNR in Pisa, from the [NI] who worked with me on the original idea, let's say, of this, this original model. And then there was another professor who worked in collaboration with [NR], I think from the University of [NC]. I mean, in the end this is what came out.

Interviewer

Excellent... were there any critical issues? From all points of view. I say administrative, bureaucratic rather than anything else and that they turned out to be a little bit cumbersome or possibly that could be changed in the future.

Interviewed n.2

I have to say that that kind of collaboration, which is precisely under this broad hat of this collaboration that is called [NP] because clearly it is a kind of collaboration that requires a proposal that has to be submitted to both the University of [NC] and [NI]. So, I took care of the [NI] side and Professor [NR] took care of the [NC] side, let's say, the proposal was exactly the same, simply sent to the two entities. Actually, then the selection phase was done by [NI]. And so, from that point of view it was a relatively quick process. I mean, the proposal itself, the technical part, I think it was 1500 characters, so a couple of pages essentially. Basically, it's about having the idea and more about getting the need across. How is it important to collaborate? The respective expertise, how are they complementary and after that within a few months we had a result, it was approved. And again, I see the [NI] side of the collaboration. The management of the funds was relatively simple, partly because they were travel funds; therefore, travel funds are quite simple to use. Essentially one makes the trip and ask for reimbursement. Here, it's not that there is to prepare a contract for a researcher or in short that would be a much more complicated thing, so from that point of view there it seems to me that it's a process that worked quite smoothly. Here, without any major hiccups, I don't know, I'd have to check whether it's always active this program because it depends on whether the partner universities continue to want to pursue the initiative or not, so I don't know... [NI] has a lot of collaborative programs with universities scattered all over the world. And I think there is always the one with the University of [NC], but I would have to check.

Interviewer

Ok, perfect, so let's say that here, if you could go back, if you could go back would you do this kind of collaboration again, I mean would you replicate it?

Interviewed n.2

Yes, yes, absolutely. Now, not with Professor [NR], I've had other proposals with other professors at the University of [NC] that have gone through, both myself personally and also other people here at the lab where I work, so it's definitely a program that works well, that has led to establishing collaboration. The most complicated thing I have to say, because these programs in theory would serve to establish more stable collaborations, even with more substantial funding, because, as I tell you again this is just travel funding. Effectively it is not easy, because as we know in short, the collaboration between a North American, U.S. entity and Italy is never easy because of course everyone tends to fund their own part. And if there are international collaboration programs, they are usually with third countries, I'm talking for example European programs, maybe they fund a partnership with a country, I don't know North African, but with the United States. No, because they are perceived as rich nations, with their own funds, and so it's a matter of possibly doing these matching. So, there are some programs where in theory one submits the same proposal and it has to be approved by one side and the other side. In my opinion they are very complex and, objectively, I have never seen them particularly efficient.

Interviewer

Ok, so if we bring back an example for a moment, the example precisely of SoBigData which is a research infrastructure, however with a European hat the difficulty might be in the future, that of being able to have more or less formalized collaborations, etc.

Interviewer n.2

Yes, in my experience it is not easy this kind of collaboration because then in the data world I have to say that the European Union is there... The GDPR talk that is a wall. I understand all the privacy requirements. From the researchers' point of view the GDPR certainly doesn't make life easier for researchers and that's in a competitive environment and so when one comes across I don't know... Simple data sharing between a European entity is a non-European entity, it's very complicated or almost impossible... and so collaboration, again you have to see what you mean, it's sort of everybody does their own thing and then somehow we come together, but a more substantial collaboration where one shares resources. I see it really very complicated in terms of issues, precisely talking about data, as precisely here we are talking about infrastructure on big data, national regulations can be an issue. European law is probably the strictest in the world. So it's not easy to be able to work and on the other hand, precisely, being people who work analyzing data, it's clear that the quality of the data is the fundamental starting point, i.e. one can be the most brilliant researcher in the universe, but if one has to work I don't know... with aggregated data at the level of a whole city rather than I don't know, a street rather than a neighborhood rather than the individual person... let's say the kind of research that can be done with this data is completely different and has a fundamentally different impact. I think this aspect here is not part of the public debate about on data... data is always seen as a battle between corporations and private individuals. I think it is yet another thing that puts Europe at a competitive disadvantage because let's say I am in a position where I am largely, as mentioned, working in a different reality. That is, I find myself collaborating working in a North American reality. I'm familiar with, obviously I collaborate and work in the European reality all the time through [NI]. We have so many collaborations with Chinese realities and really the kind of access to data that there is in China is not comparable to anything else. So then if one starts to say why are Chinese universities improving because in the journals, like Nature and in short in the most important ones you see more and more Chinese authors... Then, I mean, there is not, there is not only the big infrastructure, like the Cern infrastructure. The particle accelerator or the shared telescopes, I mean when one thinks of infrastructure one always thinks of these big, collaborations in the area of I don't know astrophysics or nuclear physics, but there are also digital technologies today. To not have that vision in the world of big data... In my opinion is not very, how to say... Looking very far ahead, here, I see it as a little bit short-sighted.

Interviewer

Ok, so let's say the issue, precisely, at the European level may be the risk of not having enough -- how should I put it? Enough flexibility, here in the use of data, especially, let's say for research more than for corporations...

Interviewed n.2

For research we often interact with telephone companies or the like that have for example data on in short on mobility of people, etc. And European companies are already from a company's point of view, giving data for research or research collaborations are mainly seen as an activity that unnecessary, i.e., as a cost. However, they are a cost item... just giving, preparing data and getting it externally accessed. They have to commit resources to it, beyond whether they then also pay the research organization or not, but even if they don't pay the research organization and make what is called a contribution, for them it is a line item in the budget that they have to, add. If you add to that the associated risk, I'm talking about the European ones to GDPR and the procedures required to have all the required infrastructure of GDPR etc. and the additional risks of sharing data... It's already difficult to forge partnerships, it's already difficult to get companies to share the data they have, and the more obstacles you put in the way, the harder it is, the harder it is to get to the bottom line. This, this is unfortunately a fact that exists that we experience on a daily basis. The consequences of these choices aren't that you see today and tomorrow you will probably see in a few years. And yet what I verify is that not only is it not talked about in the newspapers that that, whatever, but it's not even talked about, it seems to me, in the research community that much. Here, I don't see a big debate, a big movement to try to get access for research purposes more -- I don't know, more uniform, more global. Facilitating access for research purposes and this especially, I mean, as Italians, as Europeans... I mean, it has to make us think a little bit, because I think Europe from a certain point of view wants to have this ambition, to be the beacon as far as cutting-edge regulations are concerned, I think also of the forced transition to the electric car... And that in some respects is commendable! But I have the impression that sometimes one makes choices a little bit let's say, ideological or principled, regardless or without assessing or without taking into account the side effects of these choices or the broader implications and that can then have very noticeable undesirable effects on Europe... in short, on then what it manages to do, because anyway, beyond everything one has to remain a leading continent in the world. Even if one wants to promote certain initiatives or certain attentions to certain issues, if it then becomes insignificant on a global scale it's an end game in itself, that is this part of remaining a key

global player. Eh, I don't know how much it's perceived, that's it. And already there is the problem precisely of governance in Europe, which is much more complex than the giant Asian countries like India, China and the United States. Eh, I mean, so in my opinion it would take a little bit more sharing.

Interviewer

Clear... and so this, in addition to stopping research, let's say it spills over, if I understood your words a little bit correctly, to businesses, to what are the Corporations but not only... the world, let's say of for-profit or nonprofit organizations, but also to governments ... I mean, this can risk leaving Europe behind? Not just on a legislative level, on a regulatory level? Or is there less risk on that?

Interviewed n.2

So, at the level of regulations, I don't think Europe lacks... maybe too many? Or fine-tuned in a particular way... I give another example, because we also work a lot in the study of mobility, electric mobility... This so peremptory action... The principle is all very well, I think we all agree, but beyond the principle, that is the implementation... there are the consequences of what you do and I have the impression that I don't know... There, I when I see these decisions taken by Europe... boh, I have big doubts. Here, I think about how many people work and the supply chain that there is, for example, in the automotive industry... it's important for Europe and to proceed to these things, in such a, very peremptory way, without maybe trying to understand what are the large-scale implications is a little bit worrying. So, from a certain point of view, it's no wonder that the research world is not being listened to, because in my opinion the case of the electric transition is even more striking in terms of also the consequences that the world of privacy... of the extremely strict regulation and the impact can have. This, precisely, I repeat, is an impact that then has a spillover everywhere, because clearly if one cannot do research one cannot do innovation. Startups are not born, companies are not born, and they are born elsewhere. And it's hard to stay competitive if one say doesn't start, at least on a level playing field, right? So, if precisely we, working with data, in Europe certain ones you can't acquire... you lose. It's no use talking about brain return, it's no use talking or putting money on European programs if you don't put people in a position to compete, let's say. Again, it's clear that, I understand the motivation of the privacy of the citizen in relation with the corporation. But there is another world that is not interested in what I do, in what you do. We are interested to have detailed information, that permit to do very deep analysis, that are not possible if you aggregate data. The higher the aggregation, the lower the detail and the lower the possibility to understand what is really happening... So that is kind of the reality of the facts.

Interviewer

Okay, so let's say there's more ideology than strategy behind it at the moment... at least the perception that comes out externally especially...

Interviewed n.2

The perception is that, or at least there are strong sensitivities on certain issues, like precisely privacy, the need to make the transition to the electric car... which are principles, again, I think globally shared by the vast majority of European citizens. But in the implementation, as this is done, you kind of lose sight of the strategic vision... I think one has to reconcile the vision with an implementation that has the idea of providing benefits to Europe... You don't want to shoot yourself in the foot and precisely end up then relegated to not being able to compete. In research, to see the market invaded by non-European Electrics, just to name one, and to see people losing their jobs because there is no longer a need to make engines... There are a lot of things that drawn on paper look good... And then when you go and analyze them better, you realize that there are a number of second-, third- and fourth-level implications that I think the job of politicians would be to take into account... I don't know how much this then depends on the fact that public opinions are probably sensitive to these issues, at the level of precisely slogans... things like that... However, I think politicians should also lead possibly toward more in-depth thinking on the topics.

Interviewer

Yes, if I can try to summarize... Surely we should keep the purpose of research separate from the purpose of profit... The data we use often gets put in a cauldron, doesn't it? Everything has to be covered by privacy...however, be careful...we researchers don't care so much about the individual...however, we care about the detail. If we can get the most detailed data possible, clearly then we also have the ability to do better research.

Interviewed n.2

For sure!

Interviewer

Thank you, no further questions.

Interviewed n.2

I had this talk willingly.

Interviewer

If you have anything else to add that so you would like to share, we are very open, of course, to receive all this feedback. These perceptions... even especially from those who work outside because then we are kind of immersed here.

Interviewed n.2

These thoughts I gladly shared. Because I mean, my experience... I mean, I have a little bit two visions, let's say... being always, I mean clearly I'm based in Pisa, but now for many years I have a very good visibility also on what's going on at [NI] comparing myself a little bit... how to say... it puts your point of view a little bit in the broader context and that... It makes you think a little bit, so I'm... I'm glad I could share these thoughts with you. Then, just, they're thoughts in no particular order and I don't know, they're going to leave time for them, however, in my opinion it's okay to do them, here.

Interviewer

Thank you so much, so I now thank you and we hope you give to see us in Pisa in some occasion on big data.

Interviewed n.2

All right, it was nice to meet you. Bye Henry, have a good job.

Interviewer

Nice to meet you my pleasure. Bye!

Interviewed n.2

Bye!