Social Mining & Big Data Analytics

SoBigData

RESEARCH INFRASTRUCTURE ++

Deliverable D3.3

Dissemination and Impact Report and Plan for following period 2

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

DOCUMENT INFORMATION

PROJECT	
PROJECT ACRONYM	SoBigData-PlusPlus
PROJECT TITLE	SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics
STARTING DATE	01/01/2020 (60 months)
ENDING DATE	31/12/2024
PROJECT WEBSITE	http://www.sobigdata.eu
TOPIC	INFRAIA-01-2018-2019 Integrating Activities for Advanced Communities
GRANT AGREEMENT N.	871042
DELIVERABLE INFORMATION	
WORK PACKAGE	WP3 NA2 - Dissemination, Impact, and Sustainability
WORK PACKAGE LEADER	USFD
WORK PACKAGE PARTICIPANTS	USFD, CNR, UNIPI, IMT, KCL, SNS, PSE, RIE, CEU
DELIVERABLE NUMBER	D3.3
DELIVERABLE TITLE	Dissemination and impact report and plan for following period 2
AUTHOR(S)	Joanna Wright (USFD), Beatrice Rapisarda (CNR), Matteo Nicolosi (RIE), Roberto Pellungrini (SNS), Daniele Fadda (CNR)
CONTRIBUTOR(S)	Vittorio Romano (CNR), Erika Widegren (RIE), Saverio Barabuffi (SSSA)
EDITOR(S)	Beatrice Rapisarda (CNR), Valerio Grossi (CNR)
REVIEWER(S)	Katia Genovali (CNR), Saverio Barabuffi (SSSA), Michela Natilli (CNR)
CONTRACTUAL DELIVERY DATE	31/12/2022
ACTUAL DELIVERY DATE	26/04/2023
VERSION	V1.1
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	73
KEYWORDS	Dissemination activities, communication, events, stakeholder analysis

EXECUTIVE SUMMARY

This deliverable provides an overview of the dissemination activities that have taken place from M19 to M36 of the project.

During the first months of this period Covid-19 was still very much in existence meaning that all events were held virtually. Events started to return to face to face from around February 2022, however many events continue to offer a virtual presence alongside being an 'in person' event.

The number of events has increased from the first period which means the dissemination of the work and value of the SoBigData++ project is reaching further, and the project will continue to cement its position as a pillar of the data science community on a European level.

The collection of data from participants of the events continues to be a challenge. Throughout this deliverable the numbers and genders of participants at each event must be taken as a 'best estimate' and not confirmed figures. The same goes for the collection of data regarding the stakeholder type - these figures have been estimated by the local organisers.

The number of females involved in the various events has shown a slight increase to 38.74% which is approximately 2/5ths of the total numbers. During the SoBigData project female participation was approximately 1/3rd of total numbers and during the first period of SoBigData++ female participation was recorded as 34%.

The document also contains the plan for 2023 for dissemination and outreach activities and an update on the stakeholder analysis within the consortium.

During this timeframe, SoBigData initiated a redesign of the corporate image, which is still ongoing and includes the logo, website, and all media. In this document, a chapter is dedicated to the redesign of the SoBigData RI logo.

DISCLAIMER

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

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

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

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

GLOSSARY

AI	Artificial Intelligence
CSO	Civil Society Organisations
EU	European Union
EC	European Commission
GANs	Generative Adversarial Networks
GDPR	General Data Protection Regulation
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
ML	Machine Learning
NGO	Non-governmental Organisation
P2P	Peer-to-peer
RI	Research Infrastructure
SBD	SoBigData

TABLE OF CONTENTS

1 Relevance to SoBigData++ 7
1.1 Purpose of this document 7
1.2 Relevance to project objectives 7
1.3 Relation to other work packages 8
1.4 Structure of the document 8
2 Dissemination Activities from M19 to M36 9
2.1 Introduction 9
2.2 Timeline of Events 9
2.3 Periodic Training Planning Report Data 15
2.4 Detailed Report on Selected Training Activities 24
2.4.1 Webinars 28
2.4.2 Conferences 32
2.4.3 Awareness Panels 35
2.4.4 Workshops 37
2.4.5 Other events 44
2.5 Other Events where SoBigData++ has been presented 51
2.6 Overview of Future Events 54
2.7 Scientific Production 55
2.8 SoBigData++ Magazine 58
2.9 Dissemination Impact 58
2.9.1 Project dissemination indicators 58
2.9.2 Impact of social media 60
2.10 Future considerations 61
3 Brand Identity 62
3.1 Pictogram construction 62
3.2 Text composition 63
3.3 Colors 63
3.4 Logo declination 64
4 Impact of outreach towards policy makers and the public at large 66
4.1 Introduction 66
4.1.1 The three main events organized 66
4.2 Planned Activities for 2023 69
4.2.1 Other activities 70
5 Stakeholder Analysis Updates 71
6 Conclusions 73

1 Relevance to SoBigData++

Effective dissemination is vital to ensure the project is as impactful as possible. The further the project reaches and the more participants it attracts means that it will become further embedded into the fabric of the Data Science community.

SoBigData++ will continue to set the standard for open, fair and ethical data mining across the industry and the dissemination events that the project organises or co-organises will continue to demonstrate and reflect the project's ethos.

1.1 Purpose of this document

This document is to provide the consortium with an overview of the dissemination activities carried out in the second period of the project, from M19 to M36, together with a report on the project's impact so far.

The objectives can be summarised as follows:

- Communication of project results to the general public, scientific communities, and potential adopters through various dissemination channels.
- Dissemination of the project results to the partners' research and business associates, clients and public through media and existing and future contacts.
- Identification of user communities for potential academic and commercial impact.
- Exploration of initial ideas on generating impact from the project results on the target stakeholders.
- Establishment of the appropriate metrics and indicators in order to assess the success of the dissemination and impact of the project results.

1.2 Relevance to project objectives

The dissemination of the project results and impact generation are key activities to reach the project objectives.

The number and variety of activities that have taken place during this second period continue to demonstrate the project's commitment to disseminate its message of ethical and fair data mining, to as wide an audience as possible.

The project aims to reach further than the immediate data science community into the realms of industry and policy making whilst continuing to encourage new data scientists into the community with events designed to attract students and early career researchers.

1.3 Relation to other work packages

Dissemination and impact generation activities are horizontal activities that last for the entire duration of the project. Therefore, there is clear interaction with the other work packages on an ongoing basis. All work packages continue to provide feedback and support to each other via meetings, plenary meetings, emails and discussions - for both dissemination and impact purposes.

1.4 Structure of the document

The document is categorised into four different sections:

- Section 2 provides a report of the dissemination activities that have taken place in the reporting period. In more detail, it provides dates in which events took place, partners who organised the event and a short summary of the event including details of the format. It also includes details of events that have been planned but have not yet taken place. The last part of section 2 looks at the results of the project indicators and assesses the progress and success of the dissemination activities.
- Section 3 describes the new brand identity of SoBigData with the study for the design of the new logo and new palette of colors.
- Section 4 provides an overview of the project's impact of outreach towards policy makers and the public at large during the reporting period and the plan for next period.
- Section 5 gives an update about the stakeholder analysis already published in D3.2.
- Section 6 of the document concludes with consolidated findings so far.

2 Dissemination Activities from M19 to M36

2.1 Introduction

As Covid-19 was still a feature in the earlier months of this second period, all events from July 2021 to February 2022 were held virtually. Since February 2022, most have been able to return to a face-to-face format, although some have continued to offer online participation in addition to attendance in person.

The collection of data relating to the ages, stakeholder breakdown, genders and numbers of participants continues to create challenges. As there is no formal registration required for events (whether they are SoBigData++ or not) and in respect of GDPR and the protection of personal data, the collection of data for reporting purposes has its own inherent challenges. Local organisers have provided estimated figures for attendance numbers and the gender split and they have given their best estimates regarding the type of stakeholder, however all these figures must be taken taken as an estimate and not be considered as accurate data.

The mandatory forms used for the collection of data after an event has happened are continually assessed for their usefulness and are amended as necessary to ensure the data collected is fit for purpose.

2.2 Timeline of Events

The table below (Table 1) is a comprehensive list of the various events that have taken place from Month 19 to Month 36 of the project together with a brief description of the event. The first word in the description denotes what type of event took place.

Month	Date	Event	Partner	Description
Month 18	30 Jun 2021	**AI & Society Roundtable** https://www.humane-ai.eu/event/ai-society-roundtable/	UNIPI	**Other**: a collective intelligence exercise towards shaping the research questions of Social AI, driven by societal challenges. Talks and discussion.
Month 19	1 May – 1 July 2021	**HACK@EO L'Aquila 2021 – City Sustainability Indices for Citizens** https://hackat.it/hackat-eo-laquila-2021/	UNIVAQ	**Hackathon:** consisted of three challenges which aimed to assess air quality, walkability of roads and accessibility to services. The results formed the basis of a final challenge for the participants.
	1 July 2021	**CartoCafe: A Theory of data patterns and visual analytics of football** **No URL available**	FRH	**Webinar:** Using football data to find and analyse data patterns from the theoretical concepts to designing an analytical workflow to find tactical patterns in the behaviour of teams in a football match.

Month	Date	Event	Partner	Description
	6 July 2021	**5th SoBigData++ and LeADSjoint Awareness Panel:** **Legal Materials as Big Data: (algo)Rithms to Support Legal Interpretation.** **No URL available**	UNIPI	**Awareness Panel:** to share awareness and support networking with data scientists. Involved 6 talks from experts.
	13 July 2021	**Automated Methods of Urban Green Analysis** https://www.youtube.com/watch?v=C_6cAvYV93c	IMT	**Webinar**: The seminar aimed at providing a preliminary definition of the state-of-the-art automatic methods for systematised urban green data collection.
Month 21	13-17 Sept 2021	**XKDD 2021: 3rd International Workshop and Tutorial on eXplainable Knowledge Discovery in Data Mining as part of ECML PKDD 2021** https://kdd.isti.cnr.it/xkdd2021/	UNIPI	**Workshop:** to encourage principled research that will lead to the advancement of explainable, transparent, ethical and fair data mining and machine learning.
Month 22	6-9 Oct 2021	**DSAA 2021 – The 8th IEEE International Conference on Data Science and Advanced Analytics** https://dsaa2021.dcc.fc.up.pt/	CNR	**Conference:** features a strong interdisciplinary synergy between statistics, computing and information/intelligence sciences and cross domain interactions between academia and business/industry for data science and analytics.
	7-8 Oct 2021	**Ethics and Privacy of Big Data Use for Migration Research** https://hummingbird-h2020.eu/news/event-items/WS07102021	UNIPI	**Workshop** that will discuss the theoretical, philosophical, legal and ethical aspects of Big Data with a focus on migration.
	9 Oct 2021	**SoBigData++ Webinar: A Research Infrastructure to Empower Data Science Analysis** (as part of DSAA 2021) http://sobigdata.eu/events/sobigdataeu-research-infrastructure-empower-data-science-analysis	CNR	**Webinar**: demonstrated the services made available by the SoBigData++ Research Infrastructure and focused on the computational resources provided to the users in the virtual laboratory.
	26 Oct 2021	**SoBigData: A European Research Infrastructure for Big Data and Social Mining** at SciDataCon 2021 https://www.scidatacon.org/virtual-2021/programme/	CNR	**Webinar:** Data science is an opportunity for boosting social progress, and data analysis tools are triggering new services with a clear impact on our daily life. SoBigData RI is a multi-disciplinary research infrastructure aimed at using social mining and big data to understand the complexity of our contemporary, globally-interconnected society.

Month	Date	Event	Partner	Description
Month 23	1&5 Nov 2021	**LQ 2021: 1st International Workshop on Learning to Quantify: Methods and Applications** https://cikmlq2021.github.io/	CNR	**Workshop:** The goal of this workshop is to bring together all researchers interested in methods, algorithms, evaluation measures, evaluation protocols, and methodologies for LQ, as well as practitioners interested in the practical application of the above to managing large quantities of data.
	3 Nov 2021	**MozFest '21 session 'A Scientfic Experiment on Privacy and AI Transparency'** https://www.mozillafestival.org/en/	BSC	**Workshop:** Interactive session with participants where some basic insights into online transparency are presented, discussed and exercised in a life experiment.
	17 Nov 2021	**SoBigData++@ Humane-AI-Net** https://www.humane-ai.eu/event/the-sobigdata-research-infrastructure-a-hands-on-tutorial/	CNR	**Webinar:** to introduce the SoBigData++ project as an ecosystem for Ethical Social Mining. Also discussed the SoBigData RI Services and provided an overview of the Exploratories (Vertical research contexts), the resource catalogue, the training area and the SoBigData Lab.
	25 Nov 2021	**1st MobiDataLab Webinar: Fostering a data-sharing culture for better mobility in Europe** https://mobidatalab.eu/event/1st-mobidatalab-webinar/?is_past	CNR	**Webinar:** SoBigData will join the 1st MobiDataLab Webinar with a Panel discussion: the importance of a data-sharing culture in Europe – looking at the main challenges and opportunities.
Month 24	1-3 Dec 2021	**2nd Italian Workshop on Explainable AI** http://www.di.uniba.it/~swap/xai-it/	UNIPI	**Workshop:** looking at the dichotomy between the need for effective adaptive systems and the right to transparency and interpretability.
	6-7 Dec 2021	**10th International Symposium DataMod 2021** https://datamod2021.github.io/	UNIPI	**Satellite:** DataMod 2021 will look at the combined application of computational modelling methods with data-driven techniques from the areas of knowledge management, data mining and machine learning.
Month 25	18 Jan 2022	**6th SoBigData++ Awareness Panel: Recent perspectives on Dynamic Consent in Research: a Combined Legal & Technical Approach** https://www.lider-lab.it/2022/01/13/sobigdata-and-leads-joint-awareness-panel-3/	SSSA	**Awareness Panel:** Looking at Dynamic Consent from 3 different perspectives: ethical and legal, research participant's and technological points of view.

Month	Date	Event	Partner	Description
	21-23 Jan 2022	**Soccer & Data Cup – Expo Dubai 2020** https://www.youtube.com/watch?v=qtpzOLrkfjQ	UNIPI	**Datathon/Hackathon:** Soccer & Data Cup at EXPO 2020 Dubai was a 3-day international hybrid marathon of Sport Analytics combining fundamental techniques of data analysis and Artificial Intelligence.
Month 26	10-11 Feb 2022	**7th Edition of Energy Finance Italia** http://energyfinanceitalia.unicam.it/index.php/energy-finance-italia-7-workshop	IMT	**Workshop:** involved researchers from different disciplines to promote research, education in the energy sector and promoting the dissemination and practical application of research results.
Month 27	7 Mar – 22 Jun 2022	**Digital Epidemiology** https://sites.google.com/diag.uniroma1.it/phdprogramindatascience/home-page/digital-epidemiology	UNIROMA1	**Webinars:** A series of six seminars on various subjects of epidemiology.
	9th Mar 2022	**Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications** https://reimagine-europa.eu/events/towards-a-digital-ecosystem-of-trust-ethical-legal-and-societal-implications	RIE & CNR	**Conference:** the need for a strong transnational and safe data processing environment has emerged with great clarity during the covid-19 pandemic. The assessment of factors that shape the digital ecosystem of trust and dictate responsible data science is at the core of this first SoBigData++ White Paper.
	17 Mar 2022	**VA Lecture – GIScience MSc in Univ Brno, CZ** **No URL available**	FRH	**Lecture:** Invited lecture followed by discussion.
	28 Mar 2022	**AMLD EPFL 2022 – 'Advances of ML Approaches for Financial Decision Making & Time Series Analysis'**. https://appliedmldays.org/events/amld-epfl-2022/tracks/advances-of-ml-approaches-for-financial-decision-making-time-series-analysis	ETHZ	**Conference:** highlighting recent ML advancements like transformers, physics-informed neural networks, graph neural networks, and complexity tools and their impact on decision making, data-driven analysis, and time series predictions in finance.
Month 28	19 Apr 2022	**Visual network analysis and transdisciplinary approaches to digital methods** https://liss-dtp.ac.uk/event/methods-workshop-visual-network-analysis-and-transdisciplinary-approaches-to-digital-methods/	KCL	**Datathon:** designed to allow PGR students to explore and critically engage with innovative approaches to digital methods, in particular Visual Network Analysis.

Month	Date	Event	Partner	Description
	20-28 April 2022	**Deep Learning Seminars** https://sites.google.com/uniroma1.it/deep-learning-seminars/home	UNIROMA	**Webinar:** Seminars for the PhD in Data Science covering several topics including meta learning, continual learning and data engineering for deep learning.
Month 29	9-31 May 2022	**Cultural Analytics** https://sites.google.com/diag.uniroma1.it/phdprogramindatascience/home-page/cultural-analytics	UNIROMA	**Other:** a 6 day PhD course on Cultural Analytics with guest speakers.
	23-25 May 2022	**Workshop series Issues in XAI 4: 'between ethics and epistemology'** https://juanmduran.net/xai4/	TU Delft	**Workshop**: focuses on the normative and epistemic aspects of explainable AI (XAI). Providing knowledge or understanding of the inner workings of AI models.
Month 30	13 June 2022	**IAIL 2022 – Imagining the AI Landscape after the AI Act** http://iail2022.isti.cnr.it/	CNR	**Workshop:** to analyse how the new EU regulation proposal could shape AI technologies of the future.
	13-16 June 2022	**Theory and Practice of Deep Learning** https://sites.google.com/diag.uniroma1.it/phdprogramindatascience/home-page/theory-and-practice-of-deep-learning?authuser=0	UNIROMA1	**Summer School:** A series of lectures on topics at the cutting edge of research in Deep Learning and their applications ranging from geometry to optimization.
	28 June 2022	**Thinking like a moderation Algorithm** https://medialab.sciencespo.fr/actu/devenez-moderateur-devenez-moderatrice/	CNRS	**Datathon/Workshop:** on content moderation and online speech regulation focusing on the public's concerns around digital freedoms.
Month 31	4-8 July 2022	**AI & Society 2022 Summer School** https://phd-ai-society.di.unipi.it/summer-schools/ph-d-ai-society-summer-school-2022/	UNIPI, CNR	**School:** Five days of lectures, panel, poster sessions and proactive project work, to advance the frontier of AI research together with internationally renowned scientists.
Month 32	29 Aug – 2 Sept 2022	**Visual Network Analysis for Social Data** https://cis.cnrs.fr/gephi-week/	CNRS	**Datathon:** bring together research working on social network analysis and network science to contribute to the development and integration of a set of tools for the visual analysis of social graph.

Month	Date	Event	Partner	Description
Month 32	29 Aug – 2 Sept 2022	**eXplainable AI Summer School** https://xaiss.eu/	TU Delft	**School:** covering some of the most important topics in explainable AI (e.g. post-hoc interpretability in ML, interpretable representation learning in language and vision tasks, counterfactuals, human-centric explainability and others) and their applications in important subject areas (e.g. language, vision, search and recommendation systems).
Month 33	7 Sept 2022	**LeQua 2022: A lab on Learning to Quantify @ CLEF2022** https://lequa2022.github.io/	CNR	**Workshop:** to encourage the comparative evaluation of methods for 'learning to quantify' in textual datasets, i.e., methods for training predictors of the relative frequencies of the classes of interest in sets of unlabelled textual documents.
	5-8 Sept 2022	**Conference and Labs of the Evaluation Forum (CLEF 2022)** https://clef2022.clef-initiative.eu/	CNR	**Conference:** covering the broad range of issues in the fields of multilingual and multimodal information access evaluation, and a set of labs and workshops designed to test different aspects of mono and cross-language Information retrieval systems.
	19 Sept 2022	**4th International Workshop on eXplainable Knowledge Discovery in Data Mining (ECML PKDD)** https://kdd.isti.cnr.it/xkdd2022/	CNR	**Workshop:** to encourage principled research that will lead to the advancement of explainable, transparent, ethical and fair data mining and machine learning.
	21-23 Sept 2022	**Privacy in Statistical Databases (PSD 2022)** https://crises-deim.urv.cat/psd2022/	URV	**Conference:** exploring the tension between the increasing societal and economical demand for accurate information and the legal and ethical obligation to protect the privacy of individuals and enterprise.
	23 Sept 2022	**LQ 2022: 2nd International Workshop on Learning to Quantify: Methods and Applications** https://lq-2022.github.io/	CNR	**Workshop:** Learning to quantify is the task of training class prevalence estimators. Bringing together researchers interested in algorithms, evaluation measures, evaluation protocols and methodologies for managing big data.
Month 34	10-11 Oct 2022	**New data and methods for migration studies: going beyond traditional data sources** https://hummingbird-h2020.eu/news/event-items/SBDpapers	UNIPI	**Workshop:** aimed at bringing together migration scholars from various disciplines to study human migration from non-traditional data sources and methods.

Month	Date	Event	Partner	Description
Month 35	9-10 Nov 2022	**2nd Edition of the School 'Machine Learning of Dynamic Processes and Time Series Analysis** https://mldyn2022.wordpress.com/	SNS	**School:** A two-day school on the use of mathematical and computer science approaches to time series analysis, leveraging on machine learning and artificial intelligence.
	17-18 Nov 2022	**The International Forum on Digital and Democracy (IFDaD 2022)** https://www.ifdad.org/	RIE	**Conference:** will explore the interplay between digital technology and data, political participation, and governance, also looking at eGovernment and disinformation.
	25 Nov 2022	**Workshop organised in collaboration with WHO, UNIPI & CNR on Design for Values in medicine and healthcare** https://juanmduran.net/who-tudelft-workshop/	TUDelft	**Workshop:** The workshop's core topic is designing and operationalising values in AI systems for healthcare and medicine. Values such as moral, medical, epistemic and social values were investigated.
	28 Nov 2022	**7th SoBigData++ Awareness Panel** **Data portability: integrating the notion in the legal framework** https://www.youtube.com/watch?v=W4_aleTeDyI&list=PLEqfZUNgk2sQjKAbTptmxNYpeIIz6QePR&index=8	SSSA	**Webinar:** Data Portability Right under the GDPR: limits and opportunities. The interplay between the GDPR's Data Portability Right and the European strategy for data: the Data Governance Act and the Data Act. The interplay between the GDPR's Data Portability Right and other EU-specific data access frameworks: The Open Data directive and the free- flow of non-personal data.

Table 1: List of Events that have taken place between M19 and M36

2.3 Periodic Training Planning Report Data

In order for the project to assess its impact and to demonstrate its reach, the organisers of the above events were tasked with collecting data covering 4 areas:

1. Number of participants
2. Gender split of participants
3. Age ranges of participants
4. Stakeholder type of participants

Below, in Table 2, there are the data collected with the data which has been collected relating to the numbers of participants which also details the gender split. As previously noted, these figures should be taken as an estimate and not as accurate figures.

Event	Number of Female Participants	Number of Male Participants	Total number of participants
AI & Society Roundtable	35	75	110
HACK@EO L'Aquila 2021	10	20	30
CartoCafe: A theory of data patterns and visual analytics of football	20	30	50
5th SoBigData++ and LeADSjoint Awareness Panel - Legal Materials as Big Data: (algo)Rithms to Support Legal Interpretation	11	10	21
Automated methods of urban green analysis	12	13	25
XKDD 2021 Workshop within ECML PKDD 2021	25	25	50
Ethics and Privacy of Big Data for Migration Research	45	46	91
A Research Infrastructure to Empower Data Science Analysis within DSAA 2021	4	12	16
SoBigData++: A European Research Infrastructure for Big Data and Social Mining at Virtual SciDataCon 2021	Not provided	Not provided	18
LQ 2021 – 1st International Workshop on Learning to Quantify	5	15	20
MozFest'21 Session: 'A Scientific Experiment on Privacy and AI Transparency'3	15	15	30
SoBigData++ @Humane-AI-Net	10	16	26
6th SoBigData++ Awareness Panel: Recent Perspectives on Dynamic Consent in Research: a Combined Legal and Technical Approach	15	10	25
Soccer & Data Cup – Expo Dubai 2020	22	49	71
7th Edition of Energy Finance Italia	16	19	35
Digital Epidemiology	4	19	23
Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications	50	50	100
VA Lecture - GIScience MSc in Univ Brno, CZ	25	25	50

Event	Number of Female Participants	Number of Male Participants	Total number of participants
AMLD EPFL 2022 – 'Advances of ML Approaches for Financial Decision Making & Time Series Analysis'	30	40	70
Visual network analysis and transdisciplinary approaches to digital methods	7	11	18
Deep Learning Seminars	20	30	50
Cultural Analytics	15	15	30
Issues in XAI: Between ethics and epistemology	15	25	40
Theory and Practice of Deep Learning	40	80	120
Thinking like a moderation algorithm	4	9	13
AI & Society 2022 Summer School	21	62	83
eXplainable AI Summer School	13	26	39
Le Qua 2022	5	10	15
Conference and Labs of the Evaluation Forum (CLEF 2022)	110	140	250
Privacy in Statistical Databases (PSD2022)	20	55	75
LQ 2022: 2nd International Workshop on Learning to Quantify: Methods and Applications	2	18	20
New data and methods for migration studies: going beyond traditional data sources	12	11	23
2nd Edition of the School 'Machine Learning of Dynamic Processes and Time Series analysis'	12	38	50
International Forum on Digital and Democracy (IFDaD)	40	60	100
Workshop organized in collaboration with WHO, UNIPI & CNR on Design for Values in medicine and healthcare	15	20	35
7th SoBigData++ Awareness Panel – Data Portability: integrating the notion in the legal framework	4	4	8

Event	Number of Female Participants	Number of Male Participants	Total number of participants
Total	709	1103	1830

Table 2: The number and gender breakdown of participants at each event.

There is a discrepancy of 18 in the total above. This is due to the event '*A European Research Infrastructure for Big Data and Social Mining at Virtual SciDataCon 2021*' which gave a total of 18 participants but did not provide a breakdown between the genders.

The chart below (Figure 1) illustrates the gender breakdown more clearly. In previous deliverables the split was around one third female and two thirds male. The current split is 38.74% female and 60.27% male. This equates to approximately 2/5 female participants and 3/5 male participants – which shows an increase in female participation.

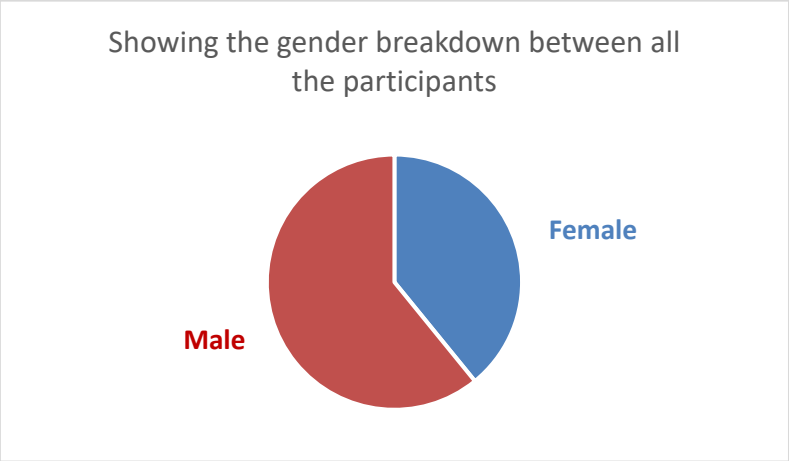

Figure 1: Showing the split between male and female participation in M19 – M36 of the project.

Table 3 below details the estimated age ranges of the participants for the events that were able to collect this data. A range of ages was provided in each of the four categories for the organisers to use:

- Under 18s
- between 18-30 years of age
- between 30-50 years of age
- over 50 years of age

As there is no requirement for a formal registration, organisers were tasked with collecting (to the best of their ability) estimates of how many people were in each of the above ranges. Obviously, this method has many flaws, but it is useful as a guide as to which age range the project reaches and appeals to most.

The ranges provided for ascertaining the number of participants in each range are:

- 1-5 participants
- 6-10 participants
- 11-15 participants
- 16-20 participants
- 21-30 participants
- 31-40 participants
- 41-50 participants
- Over 50 participants

Event	Under 18 years of age	Between 18-30 years	Between 30-50 years	Over 50 years of age
AI & Society Roundtable	0	21-30	31-40	31-40
XKDD 2021 Workshop within ECML PKDD 2021	0	21-30	31-40	21-30
Ethics and Privacy of Big Data for Migration Research	0	21-30	41-50	16-20
LQ 2021 – 1st International Workshop on Learning to Quantify	0	1-5	6-10	6-10
MozFest'21 Session: 'A Scientific Experiment on Privacy and AI Transparency'	0	6-10	11-15	6-10
6th SoBigData++ Awareness Panel: Recent Perspectives on Dynamic Consent in Research: a Combined Legal and Technical Approach	0	1-5	1-15	1-5
Soccer & Data Cup – Expo Dubai 2020	16-20	0	0	0
Digital Epidemiology	0	16-20	1-5	0
Visual network analysis and transdisciplinary approaches to digital methods	0	16-20	1-5	0
Deep Learning Seminars	0	41-50	1-5	1-5
Cultural Analytics	0	21-30	1-5	1-5
Issues in XAI: Between ethics and epistemology	0	11-15	16-20	1-5
Theory and Practice of Deep Learning	0	50+	6-10	6-10
Thinking like a moderation algorithm	0	6-10	1-5	1-5
AI & Society 2022 Summer School	0	50+	16-20	11-15

Event	Under 18 years of age	Between 18-30 years	Between 30-50 years	Over 50 years of age
Le Qua 2022	0	1-5	6-10	1-5
Conference and Labs of the Evaluation Forum (CLEF 2022)	0	50+	50+	50+
Privacy in Statistical Databases (PSD2022)	0	21-30	21-30	11-15
New data and methods for migration studies: going beyond traditional data sources	0	6-10	11-15	0
International Forum on Digital and Democracy (IFDaD)	0	6-10	21-30	41-50
Workshop organized in collaboration with WHO, UNIPI & CNR on Design for Values in medicine and healthcare	0	6-10	11-15	1-5
7th SoBigData++ Awareness Panel – Data Portability: integrating the notion in the legal framework	0	1-5	1-5	1-5
Cursory Totals	**16**	**373**	**285**	**207**

Table 3: Breakdown of age groups in selected events

As definitive totals cannot be achieved in each of the 4 age ranges, the lowest number in each range has been added together to provide cursory totals. These totals have been used to create the visual chart below (Figure 2) to give as clear a picture as possible of the division of the participants' age ranges.

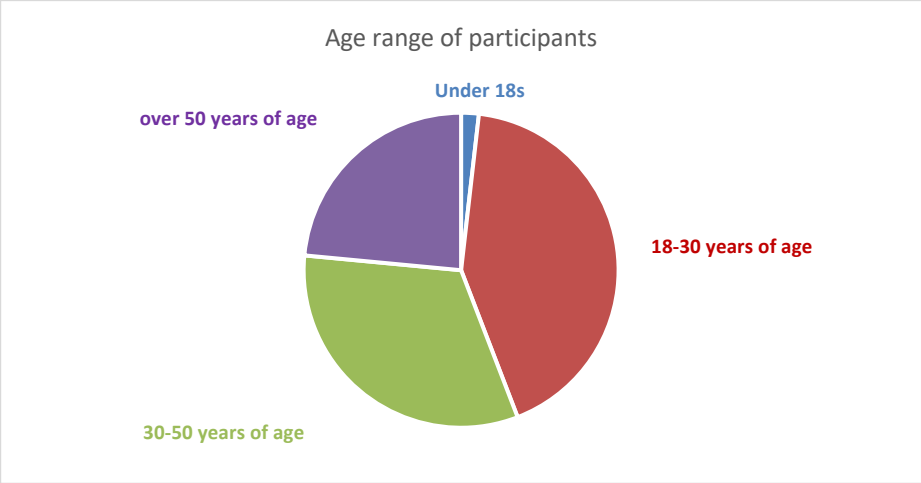

Figure 2: The split of the age ranges from M19 to M36 of the project.

As expected, most participants attending the events were between the ages of 18-30 followed by the 30-50 years and the over 50 years categories. As expected, the smallest group to participate were the under 18s.

The project is also interested in the type of individual that is attracted to the events. This is relevant to understand who participates in the events and what type of stakeholder the project reaches. The table below (Table 4) shows the stakeholder breakdown. Again, the collection of data proved challenging and as such only the events that managed to collect data are included in the table.

Event	Type of Stakeholder						
	Teaching/ Academic Institutions						
	Undergrads	PhD Researchers	Academics	Early-career Researchers	Data Analysts	Industry	Policy Makers
Ethics and Privacy of Big Data for Migration Research	1-5	11-15	11-15	16-20	1-5	1-5	1-5
LQ 2021 – 1st International Workshop on Learning to Quantify	1-5	1-5	6-10	1-5	0	1-5	0
6th SoBigData++ Awareness Panel: Recent Perspectives on Dynamic Consent in Research: a Combined Legal and Technical Approach	0	6-10	11-15	6-10	1-5	6-10	6-10
Soccer & Data Cup – Expo Dubai 2020	16-20	+ 16-20 high school age					
Digital Epidemiology	0	0	1-5	0	0	1-5	0
Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications	0	1-5	16-20	16-20	11-15	6-10	11-15
Visual network analysis and transdisciplinary approaches to digital methods	0	16-20	0	6-10	0	0	0
Deep Learning Seminars	0	21-30	0	16-20	0	0	0
Cultural Analytics	6-10	11-15	1-5	1-5	0	0	0
Issues in XAI: Between ethics and epistemology	1-5	6-10	11-15	11-15	0	0	0
Theory and Practice of Deep Learning	1-5	31-40	31-40	31-40	0	0	0
Thinking like a moderation algorithm	0	1-5	1-5	1-5	0	1-5	0
AI & Society 2022 Summer School	0	50+	16-20	6-10	0	0	0

Event	Type of Stakeholder						
	Teaching/ Academic Institutions						
	Undergrads	PhD Researchers	Academics	Early-career Researchers	Data Analysts	Industry	Policy Makers
eXplainable AI Summer School	1-5	21-30	31-40	1-5	0	1-5	0
Le Qua 2022	0	1-5	1-5	1-5	0	1-5	0
Conference and Labs of the Evaluation Forum (CLEF 2022)	21-30	50+	41-50	50+	6-10	31-40	1-5
Privacy in Statistical Databases (PSD2022)	0	16-20	16-20	11-15	0	6-10	6-10
New data and methods for migration studies: going beyond traditional data sources	1-5	6-10	16-20	1-5	0	1-5	1-5
Workshop organized in collaboration with WHO, UNIPI & CNR on Design for Values in medicine and healthcare	1-5	1-5	11-15	6-10	1-5	1-5	0
7th SoBigData++ Awareness Panel – Data Portability: integrating the notion in the legal framework	0	1-5	1-5	1-5	1-5	0	0
Total	**50**	**251**	**222**	**182**	**21**	**57**	**26**

Table 4: Stakeholder breakdown in selected events.

The events are focussed on encouraging participation from those within the academic realm – with PhD researchers, early-career researchers and academics making up the majority of the participants. However, the project also attracts a good number from industry and a smaller number of policy makers demonstrating the project's reach.

The figure below (Figure 3) provides a visual representation of the participants' stakeholder range. The numbers used to produce this chart have been taken from Table 4 above. As previously, the lowest numbers in each range were added together. The chart illustrates that the spread between the project's main target audience – PhD researchers, early career researchers and academics is fairly even making up 81% of the total participants. The other four stakeholders being undergraduates, those working in industry, data analysts and policy makers make up approximately 19% of total participation. It must be reiterated here that the numbers collected were taken from only half of the total number of events (21 events out of 43 events).

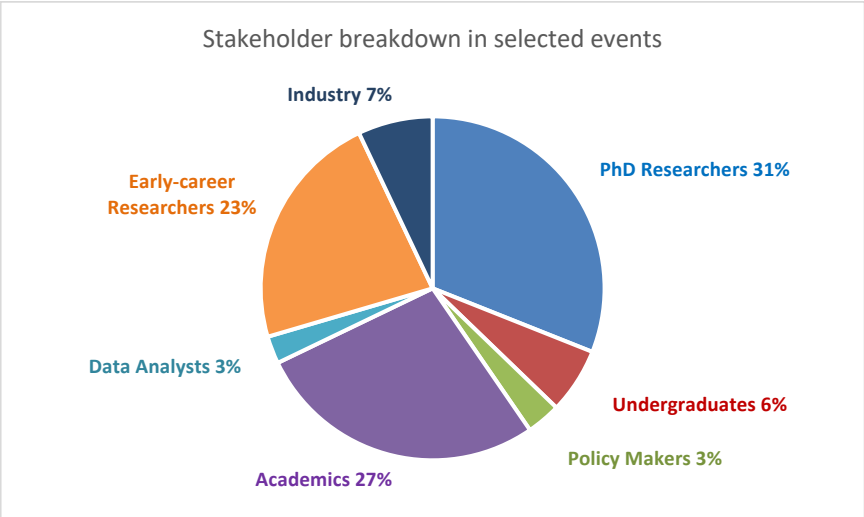

Figure 3: The split of the participants' stakeholder ranges from M19 – M36 of the project.

The project has added data analysts as a separate specific area during SoBigData++. This was because it seemed a useful tool to see how far the project reaches beyond the expected and target audience of data scientists within the academic realm. For data analysts, those working in industry and policy makers to make up approximately 12.53% of total participation is a strong indication that the project is relevant and attractive to those outside of academia.

Table 5 shows the approximate percentages of participation of each category. Although this shows the same information as the pie chart above (Figure 3), it is useful when grouping different stakeholder types together to see the actual figures.

Stakeholder Type	**% of total participation**	
Undergraduates	6.02%	**Academia: 84.94%**
PhD Researchers	30.24%	
Academics	26.75%	
Early-career Researchers	21.93%	
Data Analysts	2.53%	**Outside Academia: 12.53%**
Industry	6.87%	
Policy Makers	3.13%	
Total	**99.47%**	**99.47%**

Table 5: Showing the percentages of participation within each of the categories.

2.4 Detailed Report on Selected Training Activities

Below is a detailed report on selected training activities that have taken place from June 2021 to December 2022. It includes short descriptions of the events and where applicable there is a link to the event's website or the recording of the event.

The various events that have taken place during this reporting period are split into the following categories:

- Webinars
- Conferences
- Awareness Panels
- Workshops
- Other Events (Datathons, Hackathons, Schools, Satellites, Other)

Table 6 shows how many of each event there has been:

Type of Event	**Number of Events**
Webinars	8
Conferences	6
Awareness Panels	3
Workshops	13
Other Events (Datathons, Hackathons, Schools, Satellites)	13
Total	**43**

Table 6: Showing the number of events in each of the categories.

Then, we provide an analysis of the impact of each type of event to understand how to encourage and target the participation in specific categories of event by covering the same 4 areas presented in Chapter 2.3:

1. Number of participants
2. Gender split of participants
3. Age ranges of participants
4. Stakeholder type of participants

In Table 7 are shown the number and the percentage of participants in each type of event. The most attended type of events has been Other Events (Datathons, Hackathons, Schools, Satellites) and Conferences with 614 and 595 participants which account for 34% and 33% of the total participants in SoBigData events, respectively. However, participation in Webinars and Awareness Panels needs to be encouraged in the next period.

Type of Event	Total number of participants	Percentage of total participants
Webinars	208	11%
Conferences	595	33%
Awareness Panels	54	3%
Workshops	359	20%
Other Events (Datathons, Hackathons, Schools, Satellites)	614	34%
Total	**1830**	**100%**

Table 7: number and percentage of total participants in each category of the events.

The figure 4 provides a visual representation of the breakdown of total number of participants. In Table 8, we report the total number and percentage of the gender breakdown in each type of event. The most female participated type of event are Awareness Panels that accounts for 56% of the participants. On the one hand, female participation should be improved in Other Events (Datathons, Hackathons, Schools, Satellites) and Webinars since they account for 33% and 37% of female participants, respectively. On the other hand, Workshops and Conferences show a more balanced participation between female and male participants.

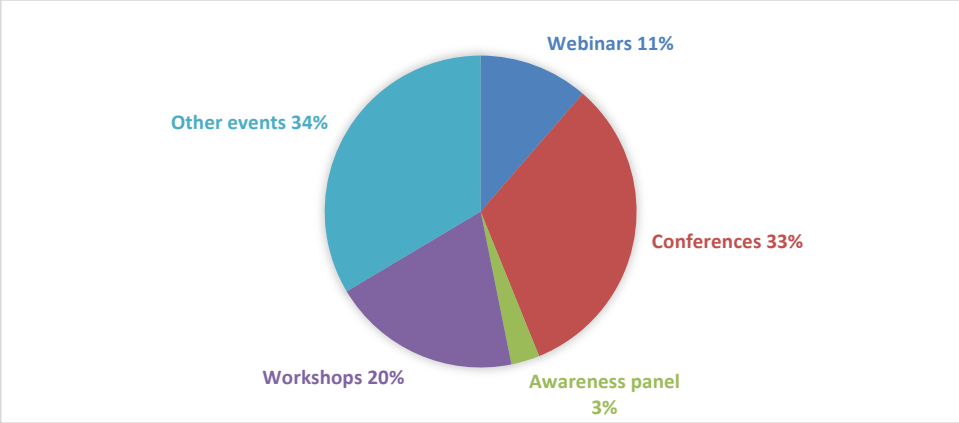

Figure 4: Breakdown of the total number of participants in each type of event from M19 – M36 of the project.

Type of Event	Total number of female participants	Percentage of female participants	Total number of male participants	Percentage of male participants
Webinars	70	37%	120	63%
Conferences	250	42%	345	58%
Awareness Panels	30	56%	24	44%
Workshops	155	43%	204	57%
Other Events (Datathons, Hackathons, Schools, Satellites)	204	33%	410	67%
Total	**709**	**100%**	**1103**	**100%**

Table 8: total number and percentage of female and male participants in each category of the events.

In Figure 5, we show a visual representation of the participants' gender breakdown.

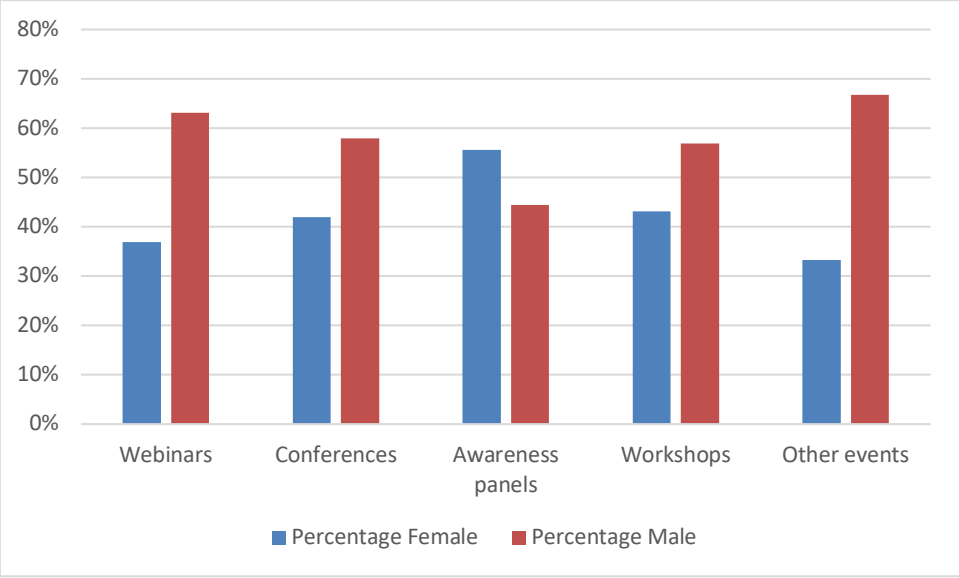

Figure 5: participants' gender breakdown for each type of event from M19 – M36 of the project.

In Table 9, we report the participants' age breakdown for each type of event. Participants under 18 years of age are only represented in Other Events (Datathons, Hackathons, Schools, Satellites) due to the specificities of the event "Soccer & Data Cup – Expo Dubai 2020". Webinars target mostly participants in the age range 18-30 since 93.44% of webinars participants are in the age range 18-30. In Conferences both the age range

18-30 and 30-50 are mostly represented, with 44.77% and 53.49% of Conferences participants respectively. In both Awareness Panels and Workshops the age range 30-50 is the most represented, accounting 75% and 62.15% of total participants of each type of event. Then, Workshops are attended mostly by participants in the age range 18-30, accounting for 34.11% of Workshops' total participants.

Type of Event	Percentage of Under 18 years of age	Percentage of 18-30 years of age	Percentage of 30-50 years of age	Percentage of over 50 years of age	Total
Webinars	0%	93.44%	3.28%	3.28%	**100%**
Conferences	0%	44.77%	53.49%	1.74%	**100%**
Awareness Panels	0%	12.50%	75%	12.50%	**100%**
Workshops	0%	34.11%	62.15%	3.74%	**100%**
Other Events (Datathons, Hackathons, Schools, Satellites)	6.58%	67.49%	23.05%	2.88%	**100%**

Table 9: participant's age breakdown for each type of event from M19 – M36 of the project.

In Table 10 is reported the stakeholders' breakdown for each type of event. Firstly, Webinars are mostly participated by PhD researchers and Early-career researchers, accounting respectively 53.85% and 41.03% of total participants in this this type of event. Secondly, Conferences participations is evenly balanced among Early-career researchers, academics and Phd researchers and Industry stakeholders (24.37%, 23.10%, 21.20% and 13,60% of total participants, respectively), however it still lacks the participation of Data analysts and Policy makers stakeholders (5.38% and 5.70%, respectively). Thirdly, Awareness panels are mostly attended by Academics (30.00% of total participants),but are evenly participated by PhD researchers, Early-career researchers, Industry and Policy makers, accounting for 17.50%, 17.50%, 15% and 15% of total participants of Awareness Panels, respectively. Finally, Workshops and Other Events are mostly attended the most by Academics, Early-career researchers and PhD researchers, which together account for 89.39% of total participants in Workshops and 88,79% of total participants in Other Events. This Table shows that the type of stakeholders that are mostly involved in participation of SoBigData events are PhD researchers, Early-career researchers and Academics. However, Industry stakeholders mainly participated in Conferences and Awareness Panels, but Data analysts and Policy makers participation has been poor, hence their participation should be increased and encouraged in the next period.

Type of Event	Undergraduates	PhD researchers	Early-career researchers	Academics	Data Analysts	Industry	Policy makers	Total
Webinars	0.00%	53.85%	41.03%	2.56%	0.00%	2.56%	0.00%	**100%**
Conferences	6.65%	21.20%	24.37%	23.10%	5.38%	13.60%	5.70%	**100%**
Awareness Panels	0.00%	17.50%	17.50%	30.00%	5.00%	15.00%	15.00%	**100%**
Workshops	3.78%	19.70%	27.27%	42.42%	1.52%	3.79%	1.52%	**100%**
Other Events	10,35%	34,48%	19,83%	34.48%	0.00%	0.86%	0.00%	**100%**

Table 10: stakeholders' breakdown for each type of event from M19 – M36 of the project.

2.4.1 Webinars

2.4.1.1 CARTOCAFE: A THEORY OF DATA PATTERNS AND VISUAL ANALYTICS OF FOOTBALL

Objectives: This webinar took place on 1 July 2021 and was organised by FRH in co-operation with DGfK (Gesellschaft für Kartographie und Geomatik). The event consisted of a lecture followed by an intensive and interesting discussion.

A fundamental goal of data analysis is to derive general knowledge at a high level from elementary data, for example in order to derive an understanding of the tactics of football teams from the paths of the players and the trajectories of the ball. Data patterns are combinations of relationships between connected elements of two or more data components that can be holistically viewed and represented as a single object, such as a cluster, trend, or correlation. Different types of relationships in forming data patterns were discussed and theoretical concepts, using examples of soccer data, were illustrated. In the second part of the session the process of designing an analytical workflow (to find tactical patterns in the behaviour of teams in a soccer game) was discussed.

Participants: There were 50 participants – 20 female and 30 male, however no data was collected on stakeholder or age range. The event was aimed at teaching/academic institutions, researchers and data analysts. The event was promoted via the SoBigData++ contact list and DGfK's own dissemination channels.

2.4.1.2 AUTOMATED METHODS OF URBAN GREEN ANALYSIS

Objective: This webinar was organised by IMT and took place virtually on 13 July 2021. Its aim was to provide a preliminary definition of state-of-the-art automatic methods for systematised urban green data collection. The talk focused on the methods and tools that are currently available for the analysis of urban green, considering their degree of accuracy (e.g. location, size, above ground volume, canopy cover, leaf area, species identity) in relation to the development of the urban green infrastructures.

Participants: Aimed at teaching/academic institutions, researchers and data analysts, this event attracted a total of 25 participants. Of these, 12 were female and 13 were male demonstrating an equal split between the genders. No data was collected regarding participants ages or stakeholder status. The event was successful and raised many interesting questions. A recording of the full webinar is available through the SoBigData++ YouTube channel.

The event was advertised via the usual SoBigData++ channels of the email contacts list, project website and Twitter account.

URL: https://www.youtube.com/watch?v=C_6cAvYV93c

2.4.1.3 SBD++ WEBINAR: A RESEARCH INFRASTRUCTURE TO EMPOWER DATA SCIENCE ANALYSIS

Objective: This webinar took place virtually on 9 October 2021 as part of the 8th IEEE International Conference on Data Science and Advanced Analytics (DSAA). The event showed the services made available by the Research Infrastructure and focused on the computational resources provided to the users in the SoBigData++ virtual laboratory. Examples of usage of the SoBigData++ libraries and its method engine were presented, and the users were able to follow and repeat the experience on a dedicated Virtual Research Environment built for DSAA 2021. The event was directed towards teaching/academic institutions, researchers, and data analysts.

Participants: Sixteen participants took part, with 4 being female and 12 being male. The webinar generated many questions and there was much interaction among the attendees. The speakers felt positive about the event despite it being virtual, with the limitation of using web-based communication tools such as zoom, instead of having a face-to-face presence.

URL: http://sobigdata.eu/events/sobigdataeu-research-infrastructure-empower-data-science-analysis

2.4.1.4 SOBIGDATA++: A EUROPEAN RESEARCH INFRASTRUCTURE FOR BIG DATA AND SOCIAL MINING AT SCIDATACON 2021

Objective: This webinar, which took place virtually on 26 October 2021, was a mix of research and practice presentations where the SoBigData++ project and the Research Infrastructure (RI) was presented. The RI was introduced to the participants, together with some of the newest research results supported by the SoBigData++ community. There was a practice session on how to access and use the resources provided by SoBigData++. There was also a short tutorial showing the participants how to use the e-infrastructure cloud services.

The event benefitted from good participation, as demonstrated by the questions raised during the presentations.

Participants: There were 18 participants to this event, although no breakdown between the genders was collected. The event was aimed at teaching/academic institutions, researchers, data analysts and those working in industry, although no data was collected on the stakeholder type.

URL: https://www.scidatacon.org/virtual-2021/programme/

2.4.1.5 SOBIGDATA++@HUMANE-AI-NET

Objective: This webinar took place on 17th November 2021 and was organised by CNR to promote the SoBigData++ project at Humane-AI-Net.

The event was a mixture of talks and practical sessions. The SoBigData++ project was introduced, and the Research Infrastructure was discussed. Next was a hands-on JupyterHub service session and an explanation of the SoBigData Libraries. Another hands-on session focused on the computational engine provided by SoBigData. After the practical sessions, there were two further talks: '*Legality Attentive data Science: it is needed and it is possible!'* and '*FAIR: an E-learning module for GDPR compliance and ethical aspects'*. These talks were followed by an open discussion with the attendees on the various aspects presented.

The event was lively and well-received. Several questions were asked, about how to use the platform and how much machine/deep learning is already integrated into the platform.

Participants: There was a total of 26 participants with 10 being females and 16 males, although no breakdown of ages was collected. The webinar was directed at researchers and was promoted through the usual SoBigData++ channels – being the project website, the email distribution list and Twitter.

URL: 1 - http://www.sobigdata.eu/events/sobigdata-humane-ai-net

2- https://www.humane-ai.eu/event/the-sobigdata-research-infrastructure-a-hands-on-tutorial/

2.4.1.6 1ST MOBIDATALAB WEBINAR: FOSTERING A DATA-SHARING CULTURE FOR BETTER MOBILITY IN EUROPE

Objective: This online event took place on 25 November 2021 and was organised by CNR and the MobiDataLAb project. The session focused on how worldwide statistics on tourism and air passenger traffic on the one hand and bibliometric data on the other can be used to measure migration. The webinar looked at how MobiDataLab is supporting the data sharing culture in Europe and to get some key insights into two main areas:

(i) who are the relevant actors for mobility data sharing, how the lack of data impacts them and what are their needs?
(ii) what is the current status of the mobility data sharing market – and what companies, products and services are already operating in Europe?

Participants: There were 37 participants, 17 females and 20 males. Between 6-10 participants were in the 18-30 years of age category and between 16-20 were aged between 30-50 years with between 1-5 being in the over 50 years of age category.

The webinar was aimed at teaching/academic institutions, researchers, data analysts, those in industry and public administration. No stakeholder data was collected. It was promoted via the usual channels of the project website, email contacts and Twitter.

URL: https://mobidatalab.eu/event/1st-mobidatalab-webinar/?is_past

2.4.1.7 DIGITAL EPIDEMIOLOGY

Objective: These were a series of six webinars held between 7 March and 22 June 2022. The first one was held both live and online, whilst the remaining 5 were available online. They were organised by Università Degli Studi di Roma la Sapienza (UNIROMA1).

The 6 webinars were as follows:

1. 'Study Design in Epidemiology'
2. 'Digital Disease Detection'
3. 'Digital Contact Tracking and Exposure Notification'
4. 'Human Proximity: From Measurement to Models and Interventions'
5. 'What We Talk About When We Talk About Vaccine Hesitancy'
6. 'Bayesian Reconstruction of Transmission Chains from Epidemiological Surveillance and Contact Tracing Data'

Participants: There were 23 participants of which 4 were female and 19 were male. Most participants were between 18-30 years of age, and a few were over 50 years of age. The event was aimed at teaching/academic institutions, researchers and data analysts and was promoted via the usual SoBigData++ channels of the project website, email distribution lists and Twitter.

URL:https://sites.google.com/diag.uniroma1.it/phdprogramindatascience/home-page/digital-epidemiology

2.4.1.8 DEEP LEARNING SEMINARS

Objective: This series of webinars was organised by UNIROMA1 and took place from 20-28 April 2022. They covered several advanced topics in deep learning: meta learning (i.e., 'learning to learn'), continual learning (i.e., learning from a continuous stream of tasks), and data engineering for deep learning (i.e., preparing data for being used in deep learning pipelines).

Participants: One event attracted approximately 50 participants, 20 being female and 30 being male. 41-50 participants were in the 18-30 years of age category and between 1-5 participants were in each of the categories 30-50 years of age and over 50 years of age.

The event was aimed at teaching and academic institutions and researchers. Most of the participants were PhD researchers, with 1-5 being early career researchers. The events were promoted via mailing lists.

URL: https://sites.google.com/uniroma1.it/deep-learning-seminars/home

2.4.2 Conferences

2.4.2.1 TOWARDS A DIGITAL ECOSYSTEM OF TRUST: ETHICAL, LEGAL AND SOCIETAL IMPLICATIONS

Objective: On 9 March 2022, SoBigData++ and its partner Re-Imagine Europa organised a virtual conference to present the white paper 'Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications', which was authored by Jeroen van de Hoven, Giovanni Comandè, Salvatore Ruggieri, Josep Domingo-Ferrer, Francesca Musiani, Fosca Giannotti, Francesca Pratesi and Marc Stauch as part of the SoBigData++ project. The event, chaired and moderated by Re-Imagine Europa's Media Research Director Luca De Biase, saw the participation of over eighty key representatives from academia, policy, industry, CSOs, NGOs and other stakeholders. The event was very interesting and reached a large audience.

Participants: There were approximately 100 participants with an estimated equal split between the genders. The conference was aimed at teaching/academic institutions, researchers, data analysts, law & policy makers, public administration and the general public.

There were between 1-5 PhD researchers, 6-10 individuals from industry, between 11-15 individuals were policy makers, between 11-15 were data analysts and between 16-20 participants were early career researchers and academics.

The event was advertised via the usual SoBigData++ channels and through the social media account of the organising institutions – RIE.

URL: https://reimagine-europa.eu/events/towards-a-digital-ecosystem-of-trust-ethical-legal-and-societal-implications

2.4.2.2 AMLD EPFL 2022 – 'ADVANCES OF ML APPROACHES FOR FINANCIAL DECISION MAKING & TIME SERIES ANALYSIS'

Objective: This was a conference organised by ETH Zurich and was held within AMLD 2022 on 28 March 2022.

The purpose of the event was to enable the exchange of recent research and insights amongst researchers interested in machine learning approaches for decision making and times series analysis of financial markets. There were world-class presenters from academia and industry working on topics such as:

- Deep learning for financial time series
- Reinforcement learning and data-driven optimal control for financial decision making
- Transformer-based and related NLP approaches for financial sentiment and event analysis
- Graph-based neural network techniques in finance.

It was a half-day event, that consisted of longer talks as well as lightning talks. These were followed by a panel discussion and a social event for the end of the day.

Feedback from one of the organisers stated, *"This track resulted in a most vibrant and fruitful exchange of ideas and information from researchers from different disciplines such as machine learning, complex systems, physics, mathematics and quantitative finance"*.

Participants: There were 70 participants with an approximate split of 30 female to 40 male. The event was aimed at teaching/academic institutions and researchers though no data was collected on age or stakeholder status. It was promoted via the usual channels of the organising body.

URL: https://appliedmldays.org/events/amld-epfl-2022/tracks/advances-of-ml-approaches-for-financial-decision-making-time-series-analysis

2.4.2.3 CONFERENCE AND LABS OF THE EVALUATION FORUM (CLEF 2022)

Objective: This was an 'in person' event organised by CNR and took place from 5-8 September 2022.

CLEF 2022 is the 13th CLEF conference continuing the popular CLEF campaigns which have run since 2000 contributing to the systematic evaluation of information access systems, primarily through experimentation on shared tasks. Building on the format first introduced in 2010, CLEF 2022 consists of an independent peer-reviewed conference on a broad range of issues in the fields of multilingual and multimodal information access evaluation, and a set of labs and workshops designed to test different aspects of mono and cross-language Information retrieval systems.

The event was held in person, but remote participation was possible. The event consisted of a main conference plus 14 'labs' running in parallel (i.e. sessions devoted to a specific topic on which a data challenge had been set up by the organisers). Within the sessions, talks were given by teams that participated in the data challenge, and by other keynote speakers.

Fabrizio Sebastiani of CNR (one of the organisers of Le Qua 2022 – held within CLEF 2022) provided some feedback after the event: *'The event went extremely well. The number of participants (250) equaled or exceeded the number of participants of previous editions of the conference, despite the uncertainties due to COVID. The number of labs also exceeded that of previous editions of the conference'*.

Participants: The event was aimed at teaching /academic institutions, researchers, data analysts, individuals from industry and law & policy makers. There were an estimated 250 participants, with approximately 110 being females and 140 males. More than half the participants attended in person. There was an even spread of ages from 18 to over 50s. The event was promoted via the usual SoBigSata++ channels, mailing lists and dedicated pages in social media.

URL: https://clef2022.clef-initiative.eu/

2.4.2.4 PRIVACY IN STATISTICAL DATABASES (PSD2022)

Objective: This conference took place on 21-23 September 2022. The purpose of the 'Privacy in Statistical Databases' (PSD) conference series is to attract world-wide, high-level research in statistical database privacy. During the 3 days, researchers, statisticians from national agencies and people from the industry, exchanged their ideas and discussed the latest topics related to privacy models, machine learning and privacy, microdata protection, disclosure risk assessment and record linkage, unstructured and mobility data, tabular data, synthetic data and case studies.

Participants: There were 75 participants of which 20 were female and 55 were male. Between 21-30 people were between 18-30 years of age and 30-50 years of age and between 11-15 people were over 50 years of age. Between 16-20 were PhD researchers, between 16-20 were academics and between 11-15 were early career researchers. There were between 6-10 policy makers and 6-10 participants from industry and between 1-5 data analysts. The event was aimed at teaching/academic institutions, researchers, law and policy makers and those in public administration.

The event was promoted via the SoBigData++ email contacts, project website and other channels of URV Universitat Rovira I Virgili in Spain.

URL: https://crises-deim.urv.cat/psd2022/

2.4.2.5 THE INTERNATIONAL FORUM ON DIGITAL AND DEMOCRACY (IFDAD 2022)

Objective: This conference was organised by RIE and took place in person on 17-18 November 2022. Virtual attendance was also available making the event more accessible and reaching a larger audience.

The conference brought together politics and academia and promoted international collaboration in the exchange of information, ideas, and good practices. The expected outcome of the Forum was to gain a greater understanding and shared view of how digital technologies can positively affect the electoral process and pluralism. As a follow-up to the debate of the first edition, the Forum explored and developed the interplay between digital technology and data, political participation, and governance. This edition also focused on two additional issues: the functioning of eGovernment and disinformation.

The event featured a full morning session focusing on the SoBigData ++ Project. All the participants were satisfied with the content of the session and made good use of the networking opportunities. The event was aimed at teaching/academic institutions, research and law and policy makers. Several prominent participants attended - both in person and remotely - from academia, think tanks, policy making and journalism fields.

Figure 6: SoBigData++ session at IFDaD Forum 2022, Rome, Italy.

Participants: There were approximately 100 participants, with 40 being female and 60 being male. Between 6-10 were between the ages of 18-30, between 21-30 participants were in the 30-50 years category and between 41-50 fell into the over 50 years of age category. No data was collected regarding stakeholder status.

The event was promoted via the SoBigData++ website, Twitter account and through the social media accounts of the various partners.

URL: https://www.ifdad.org/

2.4.3 Awareness Panels

2.4.3.1 5TH SOBIGDATA++ AND LEADSJOINT AWARENESS PANEL - LEGAL MATERIALS AS BIG DATA: (ALGO)RITHMS TO SUPPORT LEGAL INTERPRETATION.

Objective: This Awareness Panel took place on 6 July 2021 and is a regular event organised by the University of Pisa. It involved talks from 6 experts relating to legal reasoning, judicial decisions, legal knowledge modelling and predictive AI, text mining and case law, a predictive justice platform and a discussion around high-risk databases and new regulations on AI.

The event aims to provide a forum for discussion, networking, advice and support and to share awareness of current and topical issues surrounding big data and its legal issues.

Participants: There were 21 participants with 11 females and 10 males, demonstrating an equal split between the genders.

2.4.3.2 6TH SOBIGDATA++ AWARENESS PANEL – RECENT PERSPECTIVES ON DYNAMIC CONSENT IN RESEARCH: A COMBINED LEGAL AND TECHNICAL APPROACH

Objective: This awareness panel was organised by SSSA (and EURAC Research Centre) and was held virtually on 18 January 2022. Dynamic Consent (DC) uses information technology to enable continuous communication and interactive consent. It allows research participants to change their choices and preferences on participation and receive updated information on the research that is being conducted with their data and samples.

This awareness panel reflects on DC from 3 different perspectives:

- From an **ethical and legal point of view**, particularly on the balance between the protection of a participants' fundamental rights, including their right to data protection under the GDPR, and the promotion of data sharing and research, with particular regards to the case of secondary uses.
- From the **research participant's point of view,** particularly on how DC may help build a transparent trust relationship between participants and researchers.
- From a **technological point of view**, particularly on how DC technically works and the related technical opportunities and challenges.

Participants: There were approximately 25 participants of which it is estimated 15 were female and 10 were male. Between 1-5 people fell into each of the categories of 18-30 and over 50 years of age. Between 11-15 individuals were between 30-50 years of age.

This event was aimed at teaching/academic institutions, researchers, individuals from industry, law and policy makers and the general public. A fairly equal spread between PhD researchers, early career researchers, academics, those working in industry and policy makers has been reported.

This event was promoted through usual SoBigData++ channels – the project website, email distribution list and Twitter account.

URL: https://www.lider-lab.it/2022/01/13/sobigdata-and-leads-joint-awareness-panel-3/

2.4.3.3 7TH SOBIGDATA++ AWARENESS PANEL – DATA PORTABILITY: INTEGRATING THE NOTION IN THE LEGAL FRAMEWORK

Objective: This Awareness Panel was co-organised SSSA in conjunction with LeADS and took place virtually on 28 November 2022.

These events reunite experts to discuss specific ethical and legal issues to increase awareness among researchers, policymakers and local industries. Each panel has three speakers who are given a 15-minute slot to present their topic. Each presentation builds upon the previous one, providing a brief but complete overview of the proposed topic. The presentations were followed by discussions and interactions between the speakers and participants.

This specific panel aimed to raise awareness of different regulatory rules that come into play to maximise the re-use of data, evidencing their benefits and drawbacks. The Panel broadly discussed the European Strategy for Data (Data Governance Act, Data Act, Open Data Directive & Free Flow non-personal Data) and the data portability rules under the GDPR.

Participants: This event had 8 participants – with an equal split between the genders. There was between 1-5 in each of the categories: 18-30, 30-50 and over 50 years of age. It was aimed at teaching/academic institutions, researchers, and data analysts. There was between 1-5 participants in each of the following categories: PhD researchers, early career researchers, academics, and data analysts. The event was promoted via the project website, email contacts and the LiderLab website.

Link:
https://www.youtube.com/watch?v=W4_aleTeDyI&list=PLEqfZUNgk2sQjKAbTptmxNYpelIz6QePR&index=8

2.4.4 Workshops

2.4.4.1 XKDD 2021: 3RD INTERNATIONAL WORKSHOP AND TUTORIAL ON EXPLAINABLE KNOWLEDGE DISCOVERY IN DATA MINING

Objective: This event was organised by the University of Pisa and took place on 13th September 2021 as part of the ECML PKDD 2021 conference.

The purpose of this workshop and tutorial was to encourage principled research that will lead to the advancement of explainable, transparent, ethical and fair data mining and machine learning. The event was organised into two elements: a tutorial to introduce the audience to the topic, and a workshop to discuss recent advances in the research field. The tutorial provided a broad overview of state-of-the-art applications for explainable and transparent approaches and their relationship with fairness and privacy. Moreover, it presented Python/R libraries that demonstrated how explainability and fairness tasks can be addressed.

This year's edition of XKDD went very well. It was rich in content and contributions. There were 22 submissions received, of which only 7 could be accepted. This was due to the limitation of a half day event for the workshop. The event organisers are convinced that a live session would be more productive in terms of the circulation of ideas and are therefore hoping that next year's event will be in person.

Participants: There were approximately 50 participants with an estimated equal split between the genders. The event was aimed at teaching/academic institutions, researchers, data analysts and industry. Estimated age ranges of participants are as follows: there were between 21 and 30 people in each of the 18-30 and over 50 age ranges and there were between 31-40 participants in the 30–50-year age range.

URL: https://kdd.isti.cnr.it/xkdd2021/

2.4.4.2 ETHICS AND PRIVACY OF BIG DATA USE FOR MIGRATION RESEARCH

Objective: This was a virtual workshop that took place on 7-8 October 2021. It was organised by the University of Pisa with the HumMingBird and SoBigData++ consortiums and the IMISCOE Meth@Mig (Methodological Approaches and Tools in Migration Research) standing committee. The workshop aimed at bringing together researchers and practitioners from various disciplines and sectors who work on theoretical, philosophical, legal and ethical aspects of Big Data to discuss and present their experience, knowledge and research on the topic. Both quantitative and qualitative approaches, as well as work on tools and frameworks were welcomed.

There were two days of workshops, with two keynote speakers, eight contributed talks and a final brainstorming session on ethics screening. The event was very successful. It attracted many participants with varied interdisciplinary backgrounds where experiences were shared, and the participants gained valuable knowledge and awareness from the workshop.

Participants: There were 91 participants with an approximate equal split between the genders. Between 21-30 of the participants were between the ages 18-30 and between 16-20 were over 50 years of age. Between 41-50 were between the ages of 30-50.

The event attracted those from teaching/academic institutions, researchers, data analysts and individuals from industry and law & policy makers. Most of the participants were PhD researchers (between 11-15), early career researchers (between 16-20) and academics (between 11-15). There were between 1-5 participants in each of the following categories: undergraduate/postgraduates, those working in industry, policy makers and data analysts. Promotion of the event was through the usual SoBigData++ channels.

URL: https://hummingbird-h2020.eu/news/event-items/WS07102021

2.4.4.3 LQ 2021 – 1ST INTERNATIONAL WORKSHOP ON LEARNING TO QUANTIFY

Objective: This was a virtual event organised by CNR that took the form of 2 half days within the CIKM 2021 (30th ACM International Conference on Information and Knowledge Management) conference on 1-5 November 2021.

LQ 2021 featured a keynote speech by a leading scientist, six presentations of submitted papers, and a live discussion session. A full report has been published in SIGKDD EXPLORATIONS at https://bit.ly/3XFbbAk

One of the organisers stated, *'The event was successful. It was the first of its kind, and as such it elicited very lively discussions. it attracted (despite the difficulties due to participating in the event from faraway time zones) about 20 participants'*.

Participants: The event attracted 20 participants, 5 females and 15 males. There were approximately 1-5 in the 18-30 years of age category and between 6-10 in both the 30-50 and the over 50 years of age categories. It was aimed at teaching/academic institutions, researchers and those working in industry, and it was estimated that there were between 1-5 in each of the following categories: undergraduate/post graduates, PhD Researchers, early careers researchers and working in industry. There were between 6-10 academics.

The event was advertised via thematic mailing lists and personal contacts.

URL: https://cikmlq2021.github.io/

2.4.4.4 MOZFEST'21 SESSION 'A SCIENTIFIC EXPERIMENT ON PRIVACY AND AI TRANSPARENCY'

Objective: This festival/workshop was organised by BSC and was held virtually on 3 November 2021.

MozFest's goal is to make the internet a healthy place and promote trustworthy AI. During this year's interactive session participants were provided with basic insights into online transparency. Discussions were encouraged and exercised in a life experiment. Participants were active throughout the session, and several expressed their interest in continuing a collaboration.

Participants: There were approximately 30 participants with an estimated equal split between the genders. No data was collected but it was estimated that between 6-10 of the participants were between 18-30 years of age and between 11-15 were between 30-50 years of age. Between 6-10 were estimated to be in the over 50 years of age category.

The event was aimed at teaching/academic institutions, researchers, data analysts, industry, public administration and the general public. It was promoted by MozFest.

URL: https://www.mozillafestival.org/en/

2.4.4.5 2ND ITALIAN WORKSHOP ON EXPLAINABLE ARTIFICIAL INTELLIGENCE XAI.IT

Objective: This workshop was organised by UNIPI and The University of Bari, Italy and it took place virtually on 1-3 December 2021. The workshop addressed how the dichotomy between the need for effective adaptive systems and the right to transparency and interpretability can be dealt with and aimed to provide a forum for the Italian community to discuss problems, challenges and innovative approaches in the various sub-fields of AI. The event was reported to have gone very well.

Participants: There were 20 participants with an equal gender split. No ages or stakeholder type was collected. The event was aimed at researchers and was advertised via the project distribution contact list.

URL: http://www.di.uniba.it/~swap/xai-it/

2.4.4.6 7TH EDITION OF ENERGY FINANCE ITALIA

Objective: This was an in-person workshop that was held in Naples on 10-11 February 2022. The workshop involved over 30 researchers from different disciplines, with the aim of promoting research and education in the energy sector, and of promoting dissemination and the practical application of research results. The talks were characterised by strong interdisciplinarity across data science, hard sciences (physics, mathematics, engineering, material sciences) and economics and finance.

Feedback from Angelo Facchini of IMT stated that the event '*was very successful and SoBigData++ gained a fully satisfactory visibility in the above-mentioned research fields*'.

Participants: There were 35 attendees, 16 female and 19 male demonstrating a near even balance across the genders. However, no data on ages or stakeholders was collected. The event was aimed at teaching/academic institutions, researchers, data analysts and law and policy makers.

URL: http://energyfinanceitalia.unicam.it/index.php/energy-finance-italia-7-workshop

2.4.4.7 WORKSHOP SERIES ISSUES IN XAI 4: 'EXPLANATORYAI: BETWEEN ETHICS AND EPISTEMOLOGY'

Objective: This workshop was organised as a hybrid dissemination event by TU Delft ad took place on 23-25 May 2022.

It was a 3-day workshop with invited keynote speeches and accepted presentations. The workshop brought together junior and senior scholars working on XAI and willing to explore the intersection between epistemological and ethical aspects. There were 2 Best Paper awards.

Participants: There were 40 participants of which approximately 15 were female and 20 were male. Between 11-15 participants were aged between 18-30 and 16-20 participants were aged between 30-50 years of age. Between 1-5 were over 50 years of age.

The event was directed at teaching/academic institutions and researchers. Most were early career researchers (between 11-15) and academics (between 11-15) with a few being undergraduates/post graduate students (between 1-5) and there were also between 6-10 PhD Students.

The event was promoted via the dissemination list for SoBigData++ and through other associated mailing lists.

URL: https://juanmduran.net/xai4/

2.4.4.8 IAIL 2022 - IMAGINING THE AI LANDSCAPE AFTER THE AI ACT

Objective: This live event took place at Vrije Universiteit in Amsterdam on 13 June 2022. It was organised by CNR in conjunction with SNS and Vrije Universiteit.

The goal of the workshop was to analyse from a multidisciplinary point of view (i.e., law experts and technicians) how the new EU regulation proposal could shape the AI technologies of the future.

The event was structured into 2 invited talks from international experts (both female), 11 presented papers and one group activity. Many questions were posed to the authors at the end of each talk and the feedback received by the organisers was positive and confirmed that there was an enthusiasm for the subject area.

The 4 workshop chairs and many of the other organisers of this event were female – which is visually inspiring for all participants.

Participants: The event attracted approximately 30 participants with an estimated even split between the genders. The number of participants exceeded expectations proving that interest in this area is high.

There were between 6-10 were in the age category 18-30 years old, between 11-15 participants in the age bracket 30-50 years and approximately 1-5 were in the 50+ years category.

The event was aimed at teaching/academic institutions and researchers. Approximately 6-10 participants were PhD researchers, between 11-15 were early career researchers and between 6-10 were academics.

The workshop was promoted via the usual SoBigData++ channels of the project website, email contacts and Twitter. It was also promoted via the mailing lists of other projects and institutions as well as the event website and social media.

URL: http://iail2022.isti.cnr.it/

2.4.4.9 LEQUA 2022: A LAB ON LEARNING TO QUANTIFY AT CLEF2022

Objective: This workshop was organised by CNR and took place on 5-8 September 2022 within the CLEF 2022 conference.

The aim of 'LeQua 2022' was to allow the comparative evaluation of methods for 'learning to quantify' in textual datasets, i.e., methods for training predictors of the relative frequencies of the classes of interest in sets of unlabelled textual documents.

The event was hybrid in character - in that it was held in person, but remote participation was possible. The session was devoted to a specific topic (learning to quantify) on which a data challenge had been set by the organisers. Within the session, talks were given by the teams that had participated in the data challenge, and by a keynote speaker. Most participants attended in person.

One of the organisers, Fabrizio Sebastiani commented, *"Given that this lab was the first of its kind, the event went well: 6 different labs, from 5 different countries, participated in the data challenge and gave talks at the workshop. The discussion was lively and animated, and it was decided that a new edition of this lab will be proposed for 2024".*

Participants: No data on the participants was collected at the event but it was estimated that there were 15 participants, with 5 being female and 10 being male. Between 1 and 5 people were 18-30 years of age, between 6-10 were between 30-50 and between 1-5 were over 50 years of age.

The event was aimed at teaching/academic institutions and researchers. There were between 1-5 people in each of the following categories: PhD researchers, early career researchers, academics and those working in industry.

Promotion of the event was via a dedicated website and international mailing lists.

URL: https://lequa2022.github.io/

2.4.4.10 XKDD 2022 - 4TH INTERNATIONAL WORKSHOP & TUTORIAL ON EXPLAINABLE KNOWLEDGE DISCOVERY IN DATA MINING (PART OF ECML PKDD 2022)

Objective: The purpose of XKDD, eXplainable Knowledge Discovery in Data Mining, is to encourage principled research that will lead to the advancement of explainable, transparent, ethical and fair data mining and machine learning. This was a full day conference/workshop that took place on 19 September 2022. It included four sessions with two keynote speakers. There was also a final panel.

A comment from an organiser stated, *"This edition of XKDD went very well. It was rich in both content and contributions".*

Participants: The event attracted approximately 80 participants. It was estimated that there were 20 females and 60 males although no data regarding age was collected. The event was aimed at teaching/academic institutions, researchers and data analysts, although no stakeholder data was collected. It was promoted via the SoBigData++ email distribution lists.

URL: https://kdd.isti.cnr.it/xkdd2022/

2.4.4.11 LQ 2022: 2ND INTERNATIONAL WORKSHOP ON LEARNING TO QUANTIFY: METHODS AND APPLICATIONS (AS PART OF ECML PKDD 2022)

Objective: Organised by CNR, LQ 2022 was a half-day workshop co-located with the ECML/PKDD 2022 conference and took place on 23 September 2022.

Learning to Quantify (LQ - also known as 'quantification', or 'supervised prevalence estimation', or 'class prior estimation'), is the task of training class prevalence estimators via supervised learning. The goal of this workshop was to bring together all researchers interested in methods, algorithms, evaluation measures, evaluation protocols, and methodologies for LQ, as well as practitioners interested in the practical application of the above to managing large quantities of data.

The workshop had a keynote speaker and four presented papers. The presentations were followed by discussions and brainstorming.

Participants: The event was aimed at researchers and attracted 20 participants, 2 female and 18 male. No data was recorded on the type of stakeholder or the participant's ages.

URL: https://lq-2022.github.io/

2.4.4.12 NEW DATA AND METHODS FOR MIGRATION STUDIES: GOING BEYOND TRADITIONAL DATA SOURCES

Objective: The Paris School of Economics, the SoBigData++ consortium, HumMingBird consortium and Institut Convergences Migrations (ICM) jointly organised a two-day workshop aimed at bringing together migration scholars from various disciplines from these institutions and beyond (Figure 5). The workshop, held

on 10-11 October 2022, was devoted to investigating and showcasing new methods to study human migration based on non-traditional data sources and methods.

Feedback from an organiser stated the event was very successful and was well attended. Fifty-eight participants signed up and the presenters and organisers numbered approximately 23, with an even split between the genders of 12 females and 11 males. Since the event there has been a great deal of interest in the recordings of the presentations, especially those of the keynote speakers. There were 14 presenters and 4 keynote speakers from Harvard, Berkeley, PSE and York.

One of the organisers from PSE stated, "*The event was highly interdisciplinary, and it provided a lot of interesting grounds for discussions, ranging from ethical issues to policy relevance and technical method*".

Figure 7: The workshop *New data and methods for migration studies.*

Participants: The event involved 58 participants and 23 presenters and organisers. No data was collected as to the ages and stakeholder type of the participants, but numbers were collected for the presenters and organisers; therefore, the following figures relate to the presenters and organisers. Between 6-10 were in the 18-30 years of age category and there were 11-15 in the 30-50 years of age category. The event was aimed at teaching/academic institutions and researchers. Between 16-20 presenters and organisers were academics, between 6-10 were PhD researchers and there were between 1-5 in each of the following categories: undergraduates/postgraduates, early career researchers, working in industry and policy makers.

The event was circulated within the SoBigData network, Humming Bird, and ICM for the call for papers. It was promoted on Twitter and there will be a follow up blog available on the website.

URL: https://hummingbird-h2020.eu/news/event-items/SBDpapers

2.4.4.13 WHO WORKSHOP DESIGN FOR VALUES IN AI FOR MEDICINE AND HEALTHCARE

Objective: This workshop was organised by TUDelft and took place on 25 November 2022. The event was available in a hybrid capacity to enable a wider participation.

This workshop is a collaboration between the Delft Digital Ethics Centre and the World Health Organisation. The workshop's core topic is designing and operationalising values in AI systems for healthcare and medicine.

The workshop covered a myriad of values, including moral values (e.g., responsibility and accountability), medical values (e.g., beneficence and non-maleficence, compassion), epistemic values (e.g., explainability and trust), and social values (e.g., equal access to health services and transparency). The main challenge is to operationalise these values so that AI systems are aligned with the high standards of medical research, practice, and health care. To this end, the workshop brought together the three main disciplines involved, namely, computer and data scientists (represented by members of SoBigData++, ITU/WHO, and TUDelft), medical and healthcare organisations and practitioners (represented by WHO), and moral philosophers (represented by TUDelft).

The SoBigData Research Infrastructure was clearly presented during one of the main talks of the workshop.

Participants: The workshop was attended by 35 participants, with 15 being female and 20 being male. Between 6-10 were between the ages of 18-30 and between 11-15 were between the ages of 30-50 years of age. There was also between 1-5 in the over 50 years of age category. The event was aimed at teaching/academic institutions, researchers and those working in industry. There were between 1-5 in each of the following 4 categories: undergraduates/postgraduate, PhD researchers, working in industry and data analysts. Between 6-10 were early career researchers and between 11-15 were academics.

The event was promoted via the project email distribution list.

URL: https://juanmduran.net/who-tudelft-workshop/

2.4.5 Other events

2.4.5.1 AI & SOCIETY ROUNDTABLE

Objective: This virtual event was organised by the University of Pisa and took place on 30 June 2021. It was designed as a collective intelligence exercise towards shaping the research questions of Social AI, driven by societal challenges. The event started with four speeches which were followed by 4 breakout sessions on the subjects of bias, inequality, polarisation and social good. The event was finished with a final restitution session.

Letizia Milli from the University of Pisa stated, *"It was a refreshing, brilliant conversation, that we believe brought to a deeper understanding of the societal challenges ahead and how a human, social AI might become a part of the solution, instead of a part of the problem".*

Participants: The event attracted 110 participants with approximately 35 being female and 75 male. Most of the participants were between 30 and 50 years of age and over 50 years. There were a number between 18 and 30 years of age, however no data was recorded on type of stakeholder. The event was aimed at teaching/ academic institutions, researchers, data analysts and individuals from industry. It was promoted via the usual SoBigData++ channels.

URL: https://www.humane-ai.eu/event/ai-society-roundtable/

2.4.5.2 HACK@EO L'AQUILA 2021 – CITY SUSTAINABILITY INDICES FOR CITIZENS

Objective: HACK @ EO L'Aquila 2021 took place from 1 May – 9 July 2021 and consisted of three challenges with the aim of assessing air quality, walkability of roads and accessibility to services. The results of the three challenges were then the basis of a final challenge where the participants, after having defined a final indicator of social well-being, competed for the title by creating a communicative and useful interface for institutions and citizens.

To compete, participants had to form a team capable of guaranteeing gender equality and being able to balance technical-communicative skills. Each team required a minimum of 3 and a maximum of 6 experts, graduates or undergraduates in STEM subjects and humanities, with at least two females and two experts in humanities. The hackathon was open to all, university students and non-university students and is international with the only constraint that the results had to be communicated in Italian.

Participants: There were 10 registered teams for the Hackathon with an average of 3 members per team. Although around 30 participants started the process, due to the pandemic and other factors, 2 teams completed the full challenge. The gender breakdown of participants was approximately 69% male and 31% female.

The stipulation that each team had to be made up of both males and females and required a breadth of knowledge covering technical and communicative skills shows how this event was strongly promoting the ethos of SoBigData++ and encouraging females with their varied skills to participate in the challenge.

URL: https://hackat.it/hackat-eo-laquila-2021/

2.4.5.3 SOCCER & DATA CUP - EXPO DUBAI 2020

Objective: Soccer & Data Cup at EXPO 2020 Dubai was a 3-day international hybrid marathon of Sport Analytics combining fundamental techniques of data analysis and Artificial Intelligence. This Datathon/Hackathon was organised by UNIPI and took place on 21-23 January 2022.

The event covered the subject area of Data Science and aimed to raise young people's awareness to the new frontiers of the complex analysis of digital data in the sport area. The students, with the help of mentors and experts from the Sport Analytics discipline, found themselves immersed in a digital co-design competition using soccer data of games played in the top five European football leagues. They were asked to find key solutions to a specific challenge question using selected sports data sets. All teams were encouraged to present their projects in front of a selected jury of experts.

The event saw the participation of 10 teams: 2 international and 8 Italian. The event proved to be a great success. High-school students participated with great enthusiasm and presented very interesting projects. The winning team was from Palma di Montechiaro, in Sicilia. The event received a good level of visibility in several newspapers.

Participants: There were 71 participants, of which 22 were female and 49 were male. Between 16-20 were high school children and therefore under 18. Under graduates and post graduates also took part and the event was open to the general public. The usual SoBigData++ channels of promotion were utilised including the project website, Twitter and email distribution lists.

This is the only SoBigData++ event in this reporting period that actively encouraged and supported the participation of school age children. Providing a fun, safe and structured event to inspire and engage young minds is an effective way to promote the project beyond its usual audience and be a visible and tangible entity outside of the data science community.

URL: https://www.youtube.com/watch?v=qtpzOLrkfjQ

2.4.5.4 GISCIENCE MSC IN UNIV BRNO, CZ

Objective: This was a training webinar organised by FRH and took place on 17 March 2022. The lecture was followed by a discussion. No further details have been provided regarding the event, but it was reported that it went very well.

Participants: Approximately 50 students attended the event with an estimated equal split between the genders.

2.4.5.5 VISUAL NETWORK ANALYSIS AND TRANSDISCIPLINARY APPROACHES TO DIGITAL METHODS

Objective: On 19 April 2022, a one day workshop was held at King's College London (KCL) in collaboration with the London Interdisciplinary Social Science Doctoral Training Partnership (LISS DTP). Organised as an introductory, interactive seminar, it was designed to allow PhD students and junior researchers to explore and critically engage with innovative approaches to digital methods. The afternoon was dedicated to presenting the techniques of visual network analysis both in theory and then in practice, through the analysis of the network extracted in the morning from the participant's data sets.

The event went very well, especially the afternoon session, where PhD students brought their own datasets and tested on them the digital methods presented in the morning. The students' work on their data was also presented to everyone and discussed collectively as part of the pedagogy of the workshop.

Participants: There were approximately 18 participants; 11 were female and 7 were male. Between 1-5 participants were aged over 50 and between 16-20 were between 18-30 years of age. The event was aimed at teaching/academic institutions and researchers. The majority were PhD researchers and there were a few early career researchers.

The event was promoted via the SoBigData++ Twitter account and within KCL.

URL:https://liss-dtp.ac.uk/event/methods-workshop-visual-network-analysis-and-transdisciplinary-approaches-to-digital-methods/

2.4.5.6 CULTURAL ANALYTICS

Objective: A PhD Course in Cultural Analytics was organised by UNIROMA1 involving a series of talks with guest speakers from other institutions. The course was run from 9-31 May 2022. The topics were as follows:

- Wikipedia Beyond the Encyclopaedic Value
- Measuring the Happiness, Health, & Stories of Society Through the Sociotechnical Dynamics of Social Media and Fiction
- Goodreads: A Computational Study
- Media Content Analysis and Culturomics

The course was aimed at teaching/academic institutions, researchers, data analysts and the general public.

Participants: There were approximately 30 participants with an estimated equal split between the genders. Between 11-15 individuals were PhD Researchers, between 6-10 were under graduates or post graduate students and a few were early career researchers and academics. Most were between 18-30 years of age with a few falling into the 30-50 years of age category and the over 50 years of age category.

The event was promoted via the usual SoBigData++ channels.

URL: https://sites.google.com/diag.uniroma1.it/phdprogramindatascience/home-page/cultural-analytics

2.4.5.7 THEORY AND PRACTICE OF DEEP LEARNING

Objective: This 4-day event was organised by UNIROMA1 as an in-person Summer School which included lectures from 4 well known experts in the field. It took place from 13-16 June 2022.

Despite the extraordinary simplicity of Deep Neural Network structures, approaching machine learning tasks using this tool has proven extremely effective. Nowadays, Deep Learning is pervasively used in many applications and has many important usages in different fields such as Mathematics, Physics, Biology, and Engineering. The lectures were on topics at the cutting edge of research in Deep Learning and their applications ranging from geometry to optimisation.

Participants: There were approximately 120 participants with 40 being female and 80 being male. Over 50 of the participants fell into the 18-30 age category, and there were between 6-10 participants who were between 30-50 years of age. Between 6-10 were also in the over 50 years of age category. Between 1-5 were undergraduates/postgraduates, and there were between 31-40 in each of the 3 categories; PhD researchers, early career researchers and academics.

The event was considered to be a huge success. One of the organisers stated, *"During the four days more than 120 students attended the four organised courses. We covered a lot of topics on the fringe of Deep Learning research"*.

The event was promoted via the usual SoBigData Channels of the project website, email contacts and Twitter.

URL:https://sites.google.com/diag.uniroma1.it/phdprogramindatascience/home-page/theory-and-practice-of-deep-learning?authuser=0

2.4.5.8 THINKING LIKE A MODERATION ALGORITHM

Objective: This Datathon/Workshop was co-organised with CNRS and Sciences Po's médialab and took place on 28 June 2022. It was a hands-on workshop on content moderation and online speech regulation as part of a series of two-hour workshops organised from May to December 2022. The workshop investigated the relationships between technologies of content moderation and 'moderated publics' and facilitated discussions around free speech and content moderation on social media focusing on the public's concerns around digital freedoms.

One of the organisers stated, "*In the context of SoBigData++, it was interesting to mobilise members of the consortium because of either their practice of computational science, or their interest for computational methods. It went well, overall providing data for the CNRS microproject 'Workshopping on social big data and adversarial publics', and hopefully offering SBD++ partners an opportunity to engage differently with disagreements on content moderation"*.

Participants: The event attracted 13 participants. There were 4 female and 9 male. Between 6-10 were between 18-30 years of age, between 1-5 were between 30-50 years of age and there were also 1-5 participants in the over 50 years of age category. It was estimated that there were 1-5 participants in each of the following categories: PhD researchers, early career researchers, academics and those working in industry. The event was promoted via the SoBigData email distribution lists.

URL: https://medialab.sciencespo.fr/actu/devenez-moderateur-devenez-moderatrice/

2.4.5.9 AI & SOCIETY 2022 SUMMER SCHOOL

Objective: Organised by UNIPI, this summer school took place in person from 4-8 July 2022 at Aula Magna Area Bruno Pontecorvo, University of Pisa.

One of the organisers stated. '*This Summer School is dedicated to the PhD students of the "AI & Society" branch of PhD-AI.it, and open to PhD students of the other branches. It consisted of five thrilling days of lectures, panel, poster sessions and proactive project work, to advance the frontier of AI research together with internationally renowned scientists. And plenty of social activities to mix up and build the community of next-generation AI researchers, innovators and professionals.*

The event went very well, it was attended by many people involved in the National PhD programme in Artificial Intelligence - Society Area, especially the PhD students. It was a very important opportunity for all to meet

and exchange ideas and dialogue on artificial intelligence topics. This Summer School gave a fundamental contribution in consolidate this scientific community, which is also one of the objectives of the PhD programme itself'.

Participants: There were approximately 83 participants with 21 being female and 62 being male. Over 50 individuals were in the age bracket 18-30 and between 16-20 were between 30-50 years of age. There were between 11-15 in the over 50 years of age category. The event was aimed at teaching/academic institutions and researchers. To this end, over 50 participants were PhD researchers, between 6-10 were early careers researchers and between 16-20 were academics.

The event was promoted via the SoBigData++ website.

URL: https://phd-ai-society.di.unipi.it/summer-schools/ph-d-ai-society-summer-school-2022/

2.4.5.10 VISUAL NETWORK ANALYSIS FOR SOCIAL DATA

Objective: This event was organised as a one-week datathon on Visual Network Analysis for Social Data, also called 'the Gephi Week' (as it focused on Gephi, an open-source software for the visualisation and exploration of networks). The week was co-organised by CNRS, the Gephi Consortium, the University of Aalborg and Sciences Po's médialab, in Paris and took place from 29 August to 2 September 2022.

The event served several purposes that drew upon the various dimensions of open-source software: the code itself; the community that maintains it; the uses and knowledge practices that it supports.

The event was very productive as demonstrated by the following feedback:

"The outcomes of the workshop have been numerous and remarkable, first in terms of maintenance: the backlog of technical issues in the Github repo has been addressed and many tickets have been closed, including by the new developers who have taken onboard this task; New documentation pages have been set up; Preparatory work has been done to update the UI (SVG monochromatic icons and dark mode compatibility) ; GEXF file format has been updated to include data relative to graph drawing documentation (legend and title); The neo4j and Export to the Web plugins have been updated. Apart from this work of technical maintenance and improvement, during the week the participants focused on the topic of clustering and the detection of communities in network graphs, since this issue is of particular interest for the SoBigData++ infrastructure and visual network analysis in general".

Participants: The event attracted 20 participants with 4 being female and 16 male. Approximately 11-15 participants were between the ages of 18-30, between 6-10 were aged between 30-50 years of age and between 1-5 were in the over 50 years of age category.

The event was aimed at teaching/academic institutions, researchers, data analysts, those in industry and the general public. It is estimated there were between 1-5 participants in each of the following categories: undergraduates/postgraduates, PhD researchers, early career researchers, academics, working in industry and data analysts.

The event was promoted via Twitter, the *Gephi Consortium* mailing lists and social media and the *Centre National de la Recherche Scientifique* mailing lists and social media.

URL: https://cis.cnrs.fr/gephi-week/

2.4.5.11 EXPLAINABLE AI SUMMER SCHOOL - XAISS 2022

Objective: This was an in-person event organised by LUH and hosted by TU Delft in the Netherlands between 29 August -2 September 2022. It was a Summer School and Hackathon that covered some of the most important topics in explainable AI (e.g. post-hoc interpretability in ML, interpretable representation learning in language and vision tasks, counterfactuals, human-centric explainability and others) and their applications in important subject areas (e.g. language, vision, search and recommendation systems).

The event comprised of four days of lectures with each lecture being accompanied by a practical hands-on session. On the fifth day, there was a hackathon, where groups of students worked on a topic related to XAI. Each group then presented their results. In order to bond the participants and encourage networking multiple social events were also planned including a BBQ, city tour and evenings out.

The event went very well, and the participants were happy to provide feedback. The SoBigData++ platform was utilised for announcements, Q&A and discussions.

Participants: This event attracted 39 participants of which 13 were female and 26 were male. It was aimed at teaching/academic institutions, researchers, and those in industry. There were a few undergraduate/post graduate participants, early career researchers and those working in industry. Most were PhD researchers and academics.

Promotion of the event was via the usual SoBigData++ channels and DB World, L3S's own Twitter account and in person at preceding conferences. Student grants (free participation) were offered to students of underrepresented groups. In this instance, one was granted.

URL: https://xaiss.eu/

2.4.5.12 2ND EDITION OF THE SCHOOL 'MACHINE LEARNING OF DYNAMIC PROCESSES AND TIME SERIES ANALYSIS'

Objective: This school was organised by SNS and took place in person on 9-10 November 2022. Participation was also available virtually.

The School presented recent mathematical and data-driven approaches to Machine Learning and artificial intelligence techniques for time series. In particular, the focus was on the mathematical and computational aspects of Signatures, Generative Adversarial Networks (GANs), mean field games, reservoir computing, Continual Learning and Reinforcement Learning.

The school was structured into four lectures from keynote speakers of four hours each, plus two hours for contributed talks. One of the organisers stated, "*The school went well, with good participation of young researchers and interesting lectures and talks*".

Participants: The school was primarily aimed at PhD students and early-career researchers, however there were also scholars from industry in attendance. It was estimated that 50 people attended, with 12 being female and 38 being male, however no age or stakeholder data was collected.

The event was advertised via the SoBigData++ email distribution list.

URL: https://mldyn2022.wordpress.com/

2.5 Other Events where SoBigData++ has been presented

Apart from the events where SoBig Data++ was an organiser, or co-organsier, there have been a number of other events where SoBigData++ was presented or discussed. Again, this demonstrates how the project reaches out beyond its usual circle and widens participation by promoting and discussing SoBigData++ where there is an appropriate audience.

Table 11 below details the other events where SoBigData++ was presented.

Date	Event with link	Type of Event	How was SoBigData++ Presented?
16 Aug 2021	Workshop on Learned Algorithms, Data Structures, and Instance-Optimized Systems (LADSIOS) co-located with VLDB 202 https://www.ladsios.org	Workshop	Talk
22 Sept 2021	12th Workshop on Information Logistics and Digital Transformation https://bir2021.omilab.org/workshops#	Workshop	Presentation and Paper
22 Sept 2021	7th Italian Conference on ICT for Smart Cities and Communities https://icities2021.unisa.it/program.html#details	Workshop	Presentation
6 Oct 2021	Huawei Strategy and Technology Conference 2021	Conference	Talk
21-22 March 2022	Plenary meeting of the 'Cybersecurity and Privacy Research Network' https://crises-deim.urv.cat/redCiberseguridad/meeting.html	Workshop	Presentation
10 April 2022	Third International Workshop on Algorithmic Bias in Search and Recommendation (Bias 2022) https://biasinrecsys.github.io/ecir2022/	Workshop	Presentation and Paper

Date	Event with link	Type of Event	How was SoBigData++ Presented?
25 April 2022	WebConf'22 Tutorial on Opinion Formation in Social Networks: Models and Computational Problems https://sites.google.com/view/tutorial-opinion-formation/home	Tutorial	Presentation
6 May 2022	L'Aquila Reconstruction : from Urban and Territorial Perspective to Technological Solutions. https://territoriaperti.univaq.it/laquila-reconstruction-from-urban-and-territorial-perspective-to-technological-solutions/	Workshop	Presentation
10 May 2022	3rd European Meeting on Algorithmic Challenges of Big Data (ACBD 2022) https://ideas-ncbr.pl/en/acbd/	Workshop	Talk
6 June 2022	NAACL (SemEval 2022 Multilingual News Article Similarity) https://2021.naacl.org/	Conference	Presentation and Paper
9 June 2022	Final Meeting of the DFG Priority Programme "Algorithms for Big Data" https://big-data-spp.de/events/		Talk
27 June 2022	Regions4PerMed KA5 Conference https://www.regions4permed.eu/eventi/tackling-ethical-economical-legal-and-social-aspects-of-personalised-medicine-conference/	Conference	Presentation
16-22 July 2022	Lipari School on Computational Complex and Social Systems https://liparischool.it/	School	Presentation
17-23 July 2022	DATA SCIENCE: Models, Algorithms, AI and beyond https://complex22.liparischool.it/	Conference	Presentation
21 July 2022	IEEE World Congress on Computational Intelligence https://wcci2022.org/programs/	Conference	Presentation and Paper
22 Aug 2022	IGAFIT Workshop for Algorithms Postdocs in Europe (AlgPiE by IGAFIT 2022) https://www.algpie.com/algpie-2022	Workshop	Talk
9-25 Sept 2022	Visiting Fellowship at SOFI at Stockholm University: presented our project on the intergenerational transmission of culture and norms surrounding education and occupational choices. SoBigData was presented as an affiliation to Donia during the interactions with researchers and during project presentations	Seminar	Presentation and visiting/ networking meetings
15 Sept 2022	8th Italian Conference on ICT for Smart Cities and Communities https://icities2022.unicam.it/	Workshop	Presentation

Date	Event with link	Type of Event	How was SoBigData++ Presented?
21 Sept 2022	itaData 2022 - The 1st Italian Conference on Big Data and Data Science https://www.itadata.it/	Conference	Presentation and Paper
5 Oct 2022	First International Summer School on Data Science and Mobility http://www.master-project-h2020.eu/summer-school/	Tutorial	Presentation
17 Oct 2022	How to make sustainable development in fragile areas- Festival dello Sviluppo Sostenibile 2022 https://2022.festivalsvilupposostenibile.it/cal/990/come-fare-sviluppo-sostenibile-nelle-aree-fragili#.Y5IWKnbP2Uk	Festival	
19 Oct 2022	International Conference on Social Informatics (SocInfo 2022) http://www.dcs.gla.ac.uk/socinfo2022/index.html	Conference	Presentation and Paper
19 Oct 2022	ICRI 2022 International forum on Research Infrastructures https://www.icri2022.cz/	Conference	Presentation
20-21 Oct 2022	Presentation of continuation of micro-project- Academic Migration and Academic Networks: Evidence from the Iron Curtain https://www.cesifo.org/en/event/2022-10-20/cesifoifo-junior-workshop-big-data	Conference	Presentation and Paper
20 Oct 2022	International Conference of the Catalan Association of Artificial Intelligence https://ccia2022.cat/	Conference	Presentation and Paper
21 Oct 2022	XVII Reunión Española sobre Criptología y Seguridad de la Información https://recsi2022.unican.es/es/programa/		Talk
24 Oct 2022	Beacon Project 2022 Annual Conference	Conference	Presentation
25-26 Oct 2022	EU Disinfo Lab 2022 Annual Conference https://www.disinfo.eu/conference/	Conference	Presentation
Sept/Oct 2022	Bulgarian debate forum on proliferation of misinformation	Workshop	Presentation
Nov 2022	XAI.it Workshop @ AIxIA 2022	Workshop	Presentation

Table 11: Showing other events where SoBigData++ was presented.

2.6 Overview of Future Events

Below, in Table 12, there is the list of future events that have been planned and will be included in more detail in the next deliverable.

Date	Event	Partner
9-13 Jan 2023	Digital Methods Winter School & Data Sprint: What actually happened? The use and misuse of open source intelligence (OSINT)	UvA
26 Jan 2023	8th SoBigData++ Awareness Panel Further processing of health data for research purposes: the interplay between the GDPR and the MDR	SSSA
Feb 2023	NetSciX	CNR
Feb 2023	Energy & Finance 8	IMT
20-24 Mar 2023	MozFest Virtual	BSC
8-10 June 2023	MozFest House	BSC
3-14 July 2023	Digital Methods Summer School	UvA
16-22 July 2023	Lipari School on Computational Complex and Social Systems	UNIPI
July 2023	CLMPST 2023: 17th Congress on Logic, Methodology and Philosophy of Science and Technology	TUDelft
2023	4th Edition of the 1st level university Master in post-disaster technical-administrative management in local authorities	UAQ
2023	Outreach events targeting policymakers, civil society organisations, and the public-at-large, with a special focus on the results of T2.2 and T2.3	RIE
2023	MP on “Explanatory AI in medicine and healthcare: mitigating epistemic injustice through stakeholder’s involvement”	TUDelft
2023	Text and data mining and access to research data: enabler or obstacle?	SSSA
2023	Cyberspace, Digital Economy and Cybersecurity: towards an integrated legal framework?	SSSA
2023	XKDD 2023	UNIPI

Table 12: Detailing planned future events for the next reporting period of the project.

2.7 Scientific Production

Many scientific papers have been written during the second period of the project. Table 13 lists the papers acknowledged by the project as reported by the publications tab in the EU Sygma portal. The main categories are Journal Articles, Publication in Conference proceedings/Workshop, and Book Chapters. All the publications are open access, either green or gold.

Sygma portal No.	Type	Title	Authors
132	Article in Journal	A faster horse on a safer trail: generalized inference for the efficient reconstruction of weighted networks	Federica Parisi; Tiziano Squartini; Diego Garlaschelli
133	Publication in Conference proceedings/Workshop	Living in a pandemic: adaptation of individual mobility and social activity in the US	Lucchini L.; Centellegher S.; Pappalardo L.; Gallotti R.; Privitera F.; Lepri B.; De Nadai M.
134	Article in Journal	Explainable AI for clinical and remote health applications: a survey on tabular and time series data	Di Martino, Flavio; Delmastro, Franca
135	Article in Journal	Privacy protection of user profiles in online search via semantic randomization	Mercedes Rodriguez-Garcia; Montserrat Batet; David Sánchez; Alexandre Viejo
136	Publication in Conference proceedings/Workshop	Recent results and open problems in spectral algorithms for signed graphs	Ordozgoiti Rubio, Bruno
137	Article in Journal	Cycle analysis of Directed Acyclic Graphs	Tim Evans; Paul Expert; Vaiva Vasiliauskaite
138	Other	Healthy Twitter discussions? Time will tell	Gnatyshak, Dmitry; Garcia Gasulla, Dario; Álvarez Napagao, Sergio; Arjona Martínez, Jamie; Venturini, Tommaso
139	Publication in Conference proceedings/Workshop	SismaDL an ontology to represent post-disaster regulation	Caroccia, F.; D'agostino, D.; Giordano d'Aloisio; Di Marco, A.; Stilo, G.
140	Article in Journal	Supplementary Information (SI) of the Manuscript "On Some Fundamental Challenges in Monitoring Epidemics" from On some fundamental challenges in monitoring epidemics	Vasiliauskaite, Vaiva; Antulov-Fantulin, Nino; Helbing, Dirk
141	Publication in Conference proceedings/Workshop	Focus! Rating XAI Methods and Finding Biases	Anna Arias-Duart; Ferran Pares; Dario Garcia-Gasulla; Victor Gimenez-Abalos
142	Article in Journal	On the impact of publicly available news and information transfer to financial markets.	Metod Jazbec; Barna Pasztor; Felix Faltings; Nino Antulov-Fantulin; Nino Antulov-Fantulin; Petter N. Kolm
143	Article in Journal	Human-agent coordination in a group formation game	Takko, Tuomas; Bhattacharya, Kunal; Monsivais, Daniel; Kaski, Kimmo
144	Publication in Conference proceedings/Workshop	a software engineering approach	Giordano D'Aloisio
145	Publication in Conference proceedings/Workshop	On-device modeling of user's social context and familiar places from smartphone-embedded sensor data	Mattia G. Campana; Franca Delmastro
146	Publication in Conference proceedings/Workshop	Information dynamics of price and liquidity around the 2017 Bitcoin markets crash	Vaiva Vasiliauskaite; Fabrizio Lillo; Nino Antulov-Fantulin
147	Article in Journal	A narrative review for a machine learning application in sports: an example based on injury forecasting in soccer	Alessio Rossi; Luca Pappalardo; Paolo Cintia
148	Article in Journal	Living in a pandemic: changes in mobility routines, social activity and adherence to COVID-19 protective measures	Lorenzo Lucchini; Simone Centellegher; Luca Pappalardo; Riccardo Gallotti; Filippo Privitera; Bruno Lepri; Marco De Nadai
149	Article in Journal	Analysing Twitter semantic networks: the case of 2018 Italian elections	Tommaso Radicioni; Fabio Saracco; Elena Pavan; Tiziano Squartini
150	Article in Journal	Supplementary information.pdf from On the impact of publicly available news and information transfer to financial markets	Jazbec, Metod; Pàsztor, Barna; Faltings, Felix; Antulov-Fantulin, Nino; Kolm, Petter N.
151	Article in Journal	Morningness–eveningness assessment from mobile phone communication analysis	Roy, Chandreyee; Monsivais, Daniel; Bhattacharya, Kunal; Dunbar, Robin I. M.; Kaski, Kimmo

152	Chapter in a Book	From mean-field to complex topologies: network effects on the algorithmic bias model	Valentina Pansanella; Giulio Rossetti; Letizia Milli
153	Publication in Conference proceedings/Workshop	Lightweight Blockchain-based Platform for GDPR-Compliant Personal Data Management	Cristofol Dauden-Esmel; Jordi Castellà-Roca; Alexandre Viejo; Josep Domingo-Ferrer
154	Other	A dataset to assess mobility changes in Chile following local quarantines	Pappalardo L.; Cornacchia G.; Navarro V.; Bravo L.; Ferres L.
155	Article in Journal	Towards a generalization of information theory for hierarchical partitions	Juan Ignacio Perotti; Juan Ignacio Perotti; Nahuel Almeira; Nahuel Almeira; Fabio Saracco
156	Publication in Conference proceedings/Workshop	SoBigData RI: european integrated infrastructure for social mining and big data analytics	Trasarti R.; Grossi V.; Natilli M.; Rapisarda B.
157	Chapter in a Book	Focus and bias: will it blend?	Anna Arias-Duart; Ferran Parés; Víctor Giménez-Ábalos; Dario Garcia-Gasulla
158	Article in Journal	X-Mark: a benchmark for node-attributed community discovery algorithms	Salvatore Citraro; Salvatore Citraro; Giulio Rossetti
160	Article in Journal	Blood sample profile helps to injury forecasting in elite soccer players	Alessio Rossi; Luca Pappalardo; Cristoforo Filetti; Paolo Cintia
161	Article in Journal	A Survey on Deep Learning for Human Mobility	LucaMassimiliano; BarlacchiGianni; LepriBruno; PappalardoLuca
162	Article in Journal	Influence spreading model in analysing ego-centric social networks	Vesa Kuikka; Daniel Monsivais; Kimmo Kaski; Kimmo Kaski
163	Article in Journal	Regional Lean Soft Tissue and Intracellular Water Are Associated with Changes in Lower-Body Neuromuscular Performance: A Pilot Study in Elite Soccer Players	Tindaro Bongiovanni; Grant Tinsley; Giulia Martera; Carmine Orlandi; Federico Genovesi; Giuseppe Puleo; Alessio Rossi; Athos Trecroci
164	Article in Journal	Provable randomized rounding for minimum-similarity diversification	Bruno Ordozgoiti; Ananth Mahadevan; Antonis Matakos; Aristides Gionis
165	Article in Journal	A Mechanistic Data-Driven Approach to Synthesize Human Mobility Considering the Spatial, Temporal, and Social Dimensions Together	Giuliano Cornacchia; Luca Pappalardo
166	Publication in Conference proceedings/Workshop	Evaluating community detection algorithms for progressively evolving graphs	Cazabet R.; Boudebza S.; Rossetti G.
167	Publication in Conference proceedings/Workshop	Deep Gravity: enhancing mobility flows generation with deep neural networks and geographic information	Simini F.; Barlacchi G.; Luca M.; Pappalardo L.
168	Article in Journal	A Deep Gravity model for mobility flows generation.	Filippo Simini; Gianni Barlacchi; Massimilano Luca; Luca Pappalardo
169	Publication in Conference proceedings/Workshop	Healthy Twitter discussions? Time will tell	Gnatyshak, Dmitry; Garcia Gasulla, Dario; Álvarez Napagao, Sergio; Arjona Martínez, Jamie; Venturini, Tommaso
170	Article in Journal	Understanding peacefulness through the world news	Voukelatou V.; Miliou I.; Giannotti F.; Pappalardo L.
172	Article in Journal	Unifying continuous, discrete, and hybrid susceptible-infected-recovered processes on networks	Lucas Böttcher; Lucas Böttcher; Nino Antulov-Fantulin
173	Article in Journal	The Weighted Bitcoin Lightning Network	Jian-Hong Lin; Emiliano Marchese; Claudio J. Tessone; Tiziano Squartini
174	Article in Journal	Opinion Dynamic Modeling of News Perception	Letizia Milli
175	Article in Journal	Conformity: A Path-Aware Homophily Measure for Node-Attributed Networks	Rossetti G.; Citraro S.; Milli L.
176	Article in Journal	A model for the Twitter sentiment curve.	Giacomo Aletti; Irene Crimaldi; Fabio Saracco
177	Publication in Conference proceedings/Workshop	Quantifying the presence of air pollutants over a road network in high spatio-temporal resolution	Böhm M.; Nanni M.; Pappalardo L.
178	Article in Journal	Detecting shadow lobbying	Ivan Slobozhan; Peter Ormosi; Rajesh Sharma
179	Article in Journal	Visual media analysis for Instagram and other online platforms	Richard Rogers
180	Article in Journal	The role of geography in the complex diffusion of innovations.	Balázs Lengyel; Eszter Bokányi; Riccardo Di Clemente; Riccardo Di Clemente; Riccardo Di Clemente; János Kertész; Marta C. González; Marta C. González; Marta C. González

181	Article in Journal	Modeling Adversarial Behavior Against Mobility Data Privacy	Roberto Pellungrini; Luca Pappalardo; Filippo Simini; Anna Monreale
182	Article in Journal	Temporal mixture ensemble models for probabilistic forecasting of intraday cryptocurrency volume	Nino Antulov-Fantulin; Tian Guo; Fabrizio Lillo
183	Article in Journal	Reconstructing firm-level interactions in the Dutch input-output network from production constraints	Leonardo Niccolò Ialongo; Camille de Valk; Emiliano Marchese; Fabian Jansen; Hicham Zmarrou; Tiziano Squartini; Diego Garlaschelli
185	Article in Journal	On some fundamental challenges in monitoring epidemics	Vasiliauskaite, Vaiva; Antulov-Fantulin, Nino; Helbing, Dirk
186	Publication in Conference proceedings/Workshop	Capturing Political Polarization of Reddit Submissions in the Trump Era	Morini V.; Pollacci L.; Rossetti G.
187	Article in Journal	BioTAGME: A Comprehensive Platform for Biological Knowledge Network Analysis	Antonio Di Maria; Salvatore Alaimo; Lorenzo Bellomo; Fabrizio Billeci; Paolo Ferragina; Alfredo Ferro; Alfredo Pulvirenti
188	Publication in Conference proceedings/Workshop	Sockpuppet detection: a Telegram case study	Pisciotta, Gabriele; Somenzi, Miriana; Barisani, Elisa; Rossetti, Giulio
189	Article in Journal	On the accuracy of short-term COVID-19 fatality forecasts	Nino Antulov-Fantulin; Lucas Böttcher
190	Article in Journal	Venture capital investments through the lens of network and functional data analysis	Christian Esposito; Marco Gortan; Lorenzo Testa; Francesca Chiaromonte; Giorgio Fagiolo; Andrea Mina; Giulio Rossetti
191	Article in Journal	Strong ensemble nonequivalence in systems with local constraints	Diego Garlaschelli; Qi Zhang
192	Article in Journal	Non-Markovian temporal networks with auto- and cross-correlated link dynamics	Oliver E. Williams; Piero Mazzarisi; Fabrizio Lillo; Vito Latora
193	Article in Journal	Question routing via activity-weighted modularity-enhanced factorization	Vaibhav Krishna; Vaiva Vasiliauskaite; Nino Antulov-Fantulin
194	Article in Journal	Ecology in the digital world of Wikipedia	Ogushi, Fumiko; Kert□□sz, J□□nos; Kaski, Kimmo; Shimada, Takashi
195	Article in Journal	Achieving security and privacy in federated learning systems: Survey, research challenges and future directions	Alberto Blanco-Justicia; Josep Domingo-Ferrer; Sergio Martínez; David Sánchez; Adrian Flanagan; Kuan Eeik Tan
196	Article in Journal	Gravity models of networks: integrating maximum-entropy and econometric approaches	Marzio Di Vece; Diego Garlaschelli; Tiziano Squartini
197	Article in Journal	Comparative analysis of geolocation information through mobile-devices under different COVID-19 mobility restriction patterns in Spain	Raquel Pérez-Arnal; David V. Conesa; Sergio Alvarez-Napagao; Toyotaro Suzumura; Martí Català; Enrique Alvarez-Lacalle; Dario Garcia-Gasulla
198	Publication in Conference proceedings/Workshop	Diversity-Aware k-median	Thejaswi, Suhas; Ordozgoiti, Bruno; Gionis, Aristides
199	Article in Journal	Maximizing diversity over clustered data	Zhang, Guangyi; Gionis, Aristides
200	Publication in Conference proceedings/Workshop	An individual-level ground truth dataset for home location detection	Pappalardo, Luca; Ferres, Leo; Sacasa, Manuel; Cattuto, Ciro; Bravo, Loreto
201	Publication in Conference proceedings/Workshop	Opinion Dynamic Modeling of Fake News Perception.	Cecilia Toccaceli; Letizia Milli; Giulio Rossetti
202	Article in Journal	Bitcoin Transaction Networks: an overview of recent results	Nicoló Vallarano; Claudio J. Tessone; Claudio J. Tessone; Tiziano Squartini
203	Article in Journal	On the equivalence between the kinetic Ising model and discrete autoregressive processes	Carlo Campajola; Fabrizio Lillo; Piero Mazzarisi; Daniele Tantari
205	Article in Journal	Fluctuating ecological networks: A synthesis of maximum-entropy approaches for pattern detection and process inference	Tancredi Caruso; Giulio Virginio Clemente; Matthias C. Rillig; Diego Garlaschelli
206	Article in Journal	Are Social Networks Watermarking Us or Are We (Unawarely) Watermarking Ourself?	Flavio Bertini; Rajesh Sharma; Danilo Montesi

Table 13: List of all publication from M19 to M36 as listed in the EU portal Sygma.

2.8 SoBigData++ Magazine

The project also publishes a bi-annual newsletter, called SoBigData Magazine, with the latest news on the project, a selection of the scientific production, proposed here in a popularised form, news on project events and reports on past events. The magazine also contains a selection of articles extracted from the blog. The magazine is accessible in both a virtual and printable version and is published on social media channels and on the project website in pdf format and disseminated through specific mailing lists. A page for subscribing/unsubscribing to the newsletter has been dedicated on the project website to make the management of subscriptions faster and more efficient (see http://www.sobigdata.eu/newsletter).

The magazine has an Editorial Board composed by Roberto Trasarti (CNR), Beatrice Rapisarda (CNR), Marco Braghieri (KCL) and Valerio Grossi (CNR) and an Editorial Secretariat composed by Beatrice Rapisarda and Marco Braghieri to manage the content and the publication of the magazine.

2.9 Dissemination Impact

In Deliverable 3.1 - "Initial Dissemination and Impact Plan" we have defined a set of indicators in order to monitor the success of the communication activities. The success of the communication activities is closely monitored and reported in the periodic activity deliverables. We will be following the dissemination map to ensure we reach the intended stakeholder targets and keep a comprehensive record of the categories and numbers of stakeholders approached.

2.9.1 Project dissemination indicators

Table 14 reports the indicators defined in Deliverable 3.1. The following indicators represent some metrics in order to evaluate the dissemination activities.

The extend of audience reached includes all the audience reached through the activities done within the various communication channels of the project, in particular social media, website, press coverage and events, as shown in Table 15.

The Number of SMEs, big companies and other institutional stakeholders includes all the various stakeholders engaged through the dissemination activities of the project. A detailed description of this success indicator can be found in Section 5 "Stakeholder Analysis".

The Papers published on OpenAIRE portal includes all the paper on the portal, distinguished between Open access and Not Open Access, as shown in Table 16.

Success Indicators	Target Reporting Periods		
	1	2	3
Extent of audience reached	2,500	5,000	6,500
Number of SMEs, big companies and other institutional stakeholders engaged through all dissemination activities (Companies & Institutions	100	150	200
Papers published on OpenAIRE portal	30	60	90

Table 14: Success Indicators for Reporting Period 1, 2 and 3

Extent of audience reached	Metrics	Period 1 (M1-18)	Results Period 2 (M19-36)
Web site	-Blog posts: 35 -Visits: 26217 -Page Views: 38511 -Page/Visits: 1.7	19868	26217
Twitter	-Tweets: 188 -Visualizations: 67125 -Interactions: 2796 -Retweets: 323 -Likes: 481 **-Followers: 1804**	1508	1804
Facebook	-Posts: 141 -Interactions: 306 **-Followers: 528**	495	528
YouTube	-YouTube channel subscribers: 81 **-Video views: 2811**	917	2811
LinkedIn	**-Followers: 334**	N.A.	334
Events	**-Total number of participants: 1830** -Male participants: 1103 -Female participants: 709	3416	1830
Total		26217	**33524**

Table 15: Extent of audience reached for reporting period 2.

As we can see in Table 15, we have largely achieved the Success Indicators established in D3.1. The results achieved show that the SoBigData community is increasingly growing, thanks to the variety of activities done and a very active consortium on the dissemination activities.

Below, in Table 16, we can see that the success Indicator is largely reached also for the papers published on the OpenAire portal. Furthermore, the majority of the publications within the project is in Open Access, showing that the commission's indications to move towards open science are increasingly accepted by the consortium.

Papers published on OpenAIRE portal	Period 1 (M1-18)	Results Period 2 (M19-36)
Open Access	67	243
Restricted	6	6
Closed Access	2	15
Embargo		1
Total	75	**266**

Table 16: Papers published on OpenAIRE portal for reporting period 2.

The SoBigData++ page on the OpenAIRE portal is available for consultation at:

https://explore.openaire.eu/search/project?projectId=corda__h2020::a45280b6e42e263da0c4e85e5bf6845b

2.9.2 Impact of social media

The team in charge of the project's communication is doing an intensive communication work through the various established channels to enlarge the SBD community. In particular, the project has a blog in which popular articles are published on the latest scientific research of the SBD community, results of projects and initiatives, such as the micro-projects, stories of TNA experiences and other relevant events within the SBD community. The project website also has a section that collects all events organized by SBD or in which researchers from the consortium participate. All this material is used by the communication team to promote the project on social channels.

In this regard, it is worth mentioning that SBD is present on Twitter, Facebook, YouTube, and LinkedIn (Table 15).

While the use of Twitter and Facebook have been the project's preferred channels from the outset, a remark must be made about LinkedIn. If until a couple of years ago, LinkedIn was a social network used more for job search/offer and for promoting actions related to the world of work, now its use has changed and it is increasingly used, not only in the corporate world but also in the world of public research for P2P communication and is highly regarded for the seriousness of the profiles that participate in it and for the quality of the information that circulates. For these reasons, it was decided to invest more energy in this platform, starting to use it not only for the publication of news, but also for the publication of in-depth articles on topics dear to the project, as well as on the research carried out by the SBD community.

In Table 15 there is an overview of the extent of audience reached through the actions done within social media and website.

2.10 Future considerations

Although the dissemination activities have been participated in by an impressive number of people, the use of bursaries to encourage those from under-represented groups could be promoted more and used as a device to encourage even wider participation.

The project will continue to use what it has learned and implemented throughout the pandemic to ensure appropriate events can be delivered in a hybrid format welcoming both participants in person as well as virtually.

One positive theme that SoBigData++ promotes is the appreciation of interdisciplinarity which is both encouraged and respected. As the traditionally soft subjects on the humanities side of academia often have a stronger and more fluent manner of expression, mixing these strong communication skills with the technical and numerical side of data science should produce a more engaging and inspiring platform involving participants from varied academic backgrounds. SoBigData++ will continue to encourage interdisciplinarity in its events from both the organisation side and participation. This, in turn, serves to promote the project into more fields and eventually into the consciousness of society. It also brings in individuals who may not previously have thought their skill set appropriate for a career in data science. A clear example of this would be the 'HACK@EO L'Aquila 2021' event which took place from May to July 2021.

3 Brand Identity

SoBigData is an infrastructure to strengthen and connect clusters of research excellence in big data, social mining, and artificial intelligence, creating a pan-European and interdisciplinary community. To best express, these values in redesigning the infrastructure's logo, the square, a graphic symbol of excellence of solidity, and stability, was chosen as the archetypal shape. The matrix of squares represents the dynamism and multiplicity of viewpoints present in SoBigData. Each square differs from the adjacent one by a slight variation in color, opacity, and scale, creating a significant vibration and enhancing the uniqueness of each element. The variation is not random but follows a precise and reproducible pattern. By interpreting this pattern, multiple declinations of the logo can be generated to identify SoBigData RI-related projects. The logo and its declinations were completely generated through an algorithm.

3.1 Pictogram construction

The construction of the pictogram is based on a 5 x 5 square matrix. The size of the single square of the matrix is taken as the reference unit for the construction of the entire logo. The composition begins in the central quadrant of the matrix, which has the maximum opacity and scale values (100% and 90%).

The composition proceeds towards the outside of the grid, decreasing the opacity and scale values of the elements according to the scheme shown in Figure 8. The two chosen colors have been applied in a chequered pattern. The final effect is very dynamic and strongly structured.

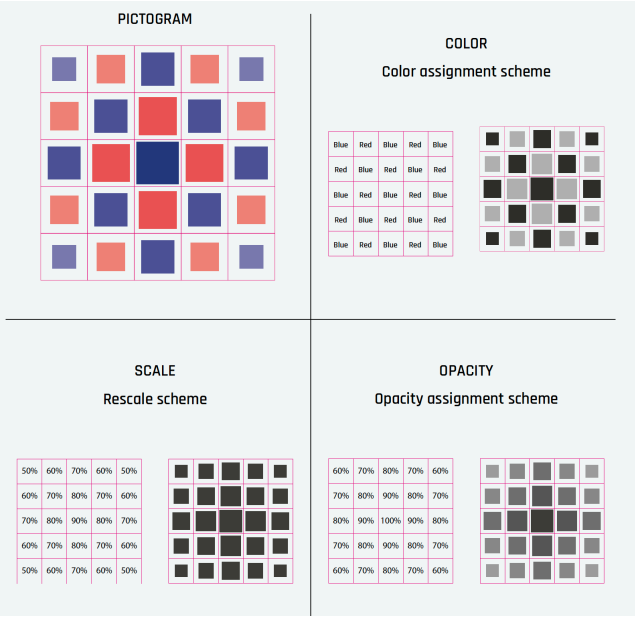

Figure 8: Pictogram construction.

3.2 Text composition

Consider x the size of a 5 x 5 matrix square needed to construct the pictogram. Each time the logo is applied, it needs a buffer area of 2x, while the distance between the pictogram and the text must be x. The height of the SoBigData text is 3x, while the underlying text must be spaced x apart and x high.

For the realization of the logo, the Rajdhani font family was used.

- Rajdhani Medium used for “SO”
- Rajdhani Semibold used for “BIG”
- Rajdhani Bold used for “DATA”
- Rajdhani Semibold used for “RESEARCH INFRASTRUCTURE” text

The weight of the text on the right side is balanced by the presence of the pictogram on the left side, creating a well-balanced logo.

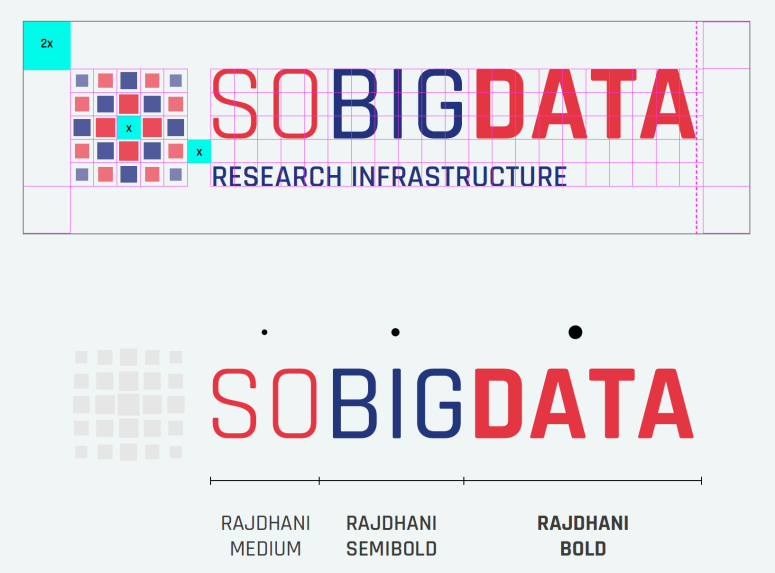

Figure 9: Text composition and font used.

3.3 Colors

Since the consortium's partners are many, and it is not easy to completely control the production of images and graphic visuals, the proposal of a strong visual identity is useful to increase the sense of coherence, regardless of the tool used to create them. The two chosen colors have strong tones, not primary, but very saturated and distant from each other, creating a strong contrast. The pure white background also contributes to the difference. The aim is to convey a sense of balance and, at the same time, vibrant energy. Furthermore, by using these two solid colors, it will be easy to create images consistent with SoBigData's brand identity, even for people with no specific training in visual design, making it easier for partners to create consistent flyers or images for communication activities.

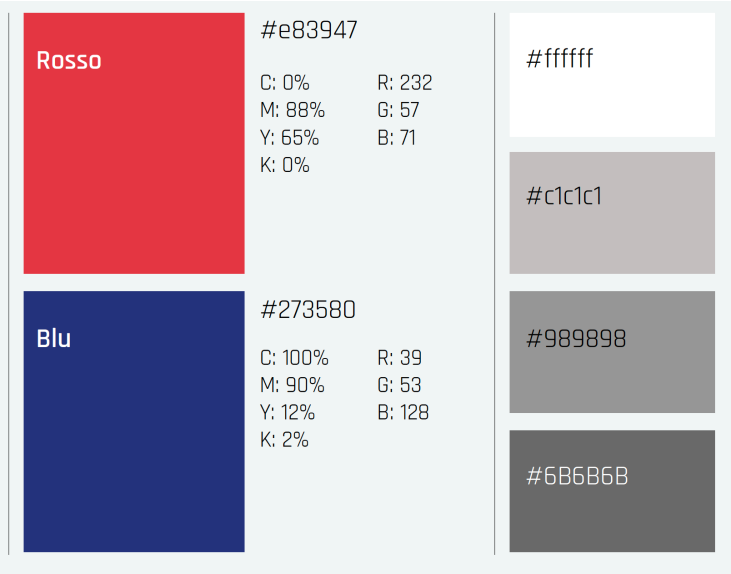

Figure 10: The new palette of the SoBigData logo pack.

3.4 Logo declination

Considering the growth of the SoBigData RI and the need to have multiple logos for each RI-related project (SoBigData++, SoBigData PPP, SoBigData.it and other national nodes), instead of creating a single logo, we created a system that allows us to create 25 unique but related logos by moving them around a matrix.The central position in the matrix is the primary logo of the Research Infrastructure. Projects related to the Research Infrastructure, as already mentioned, will be able to have their customized logos.

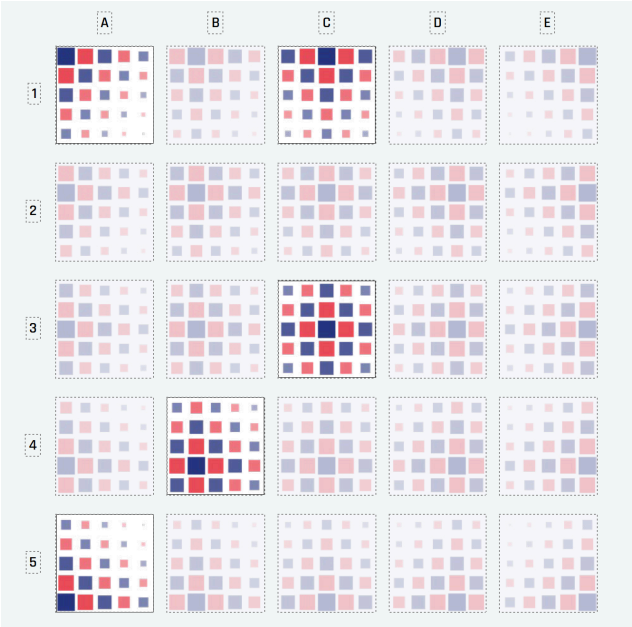

Figure 11: Examples of logo declination.

The logo is based on a visual system encoded through an algorithm that can reproduce a symbol for each project. The developed algorithm will be released as open-source code to allow each person in SoBigData to obtain their logo for each project. Finally, a script was created to create customized animations to generate video content from the logos. The SoBigData++ project logo is shown in the figure below (Figure 12).

Figure 12: The new SoBigData++ logo.

4 Impact of outreach towards policy makers and the public at large

4.1 Introduction

During months 18-36 of the project's timeline, the consortium member Re-Imagine Europa organised multiples events and activities towards policy makers and public at large, which are reported and described below. In addition, several dissemination and outreach activities are planned for 2023.

In particular, three main events were organised to disseminate and raise awareness about the Consortium's activities: a virtual roundtable on March 9th, 2022 (Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications); an in-person roundtable on "Novel Systemic Solutions for the European Information and Media Ecosystem ", organised in the context of the 2022 Re-Imagine Europa Forum on May 25th, 2022; a full morning session, including a presentation of the project and a high-level Roundtable ("Can Data Governance and Trustworthy Artificial Intelligence Help Strengthen our Democracies?"), in the context of the International Forum on Digital and Democracy, held in Rome and virtually on November 17th and 18th, 2022.

Other activities organised by Re-Imagine Europa include several bilateral conversations with the member of European Parliament Brando Benifei, particularly in relation to his task as the rapporteur of the Special Committee on Artificial Intelligence in a Digital Age, and with Mr Roberto Viola, the European Commission's Director General for Communications Networks, Content and Technology. In addition, Re-Imagine Europa performed monitoring activities for all policy news that could be of interest for the Consortium, with a particular focus on data and artificial intelligence policy, providing insights and novelties about the European Commission's activities on legislative proposals issued on those topics (e.g., Digital Services Act, Data Act, Digital Market Act, AI Act).

The three main events organised within the considered timespan are briefly described below.

4.1.1 The three main events organized

4.1.1.1 TOWARDS A DIGITAL ECOSYSTEM OF TRUST: ETHICAL, LEGAL AND SOCIETAL IMPLICATIONS

On March 9th 2022, Re-Imagine Europa and SoBigData++ organised an online event to present the White Paper "Towards a Digital Ecosystem of Trust: Ethical, Legal and Societal Implications" authored by Jeroen van de Hoven, Giovanni Comandè, Salvatore Ruggieri, Josep Domingo-Ferrer, Francesca Musiani, Fosca Giannotti, Francesca Pratesi and Marc Stauch as part of the SoBigData++ project.

Establishing a framework where data subjects can feel safe sharing their data is crucial to facilitating the development of data science in Europe. The white paper's aim is precisely to discuss the methods for promoting data sharing, building privacy-preserving technologies, and fostering data decentralisation and data altruism to develop a model for responsible digital innovation in Europe and beyond. Among the event participants were Brando Benifei, member of the European Parliament, Yvo Volman, Head of the Data Policy and Innovation Unit of DG Connect at the European Commission, Jeroen van de Hoven, Full Professor of

Technology Ethics at the Technische Universiteit Delft, and Gry Hasselbalch, Senior Key AI Ethics Experts and Research Lead at INTouchAI.eu.

Attendance was around 50 people.

URL: https://re-imagine.eu/events/towards-a-digital-ecosystem-of-trust-ethical-legal-and-societal-implications

4.1.1.2 SOBIGDATA EVENT DURING THE 2022 RIE FORUM - NOVEL SYSTEMIC SOLUTIONS FOR THE EUROPEAN INFORMATION AND MEDIA ECOSYSTEM

The impact of media in shaping opinions and public discourse is nothing new. Yet, the shift to digital technologies has transformed the media ecosystem dramatically. The disintermediation that dismantled the old power structure in information turned into novel intermediation, now much more concentrated in a few large platforms. This has drastically changed the way people receive and share information about local and global events.

Held in the context of the 2022 Re-Imagine Europa Forum, the event was organised in close collaboration with several members of the Consortium to raise awareness about the growing importance of the SoBigData Research Infrastructure. This is due to the uncertainty and disruption that the development of digital technology is causing in our era, with a particular focus on trust for democracy and democratic institutions. Since several attendees held high-level positions within the European Institutions or related organisations, Chatham House Rule was established for the meeting to allow all participants to express their views and perspectives without the seal of officialdom.

The event was held on May 25th, 2022. The participation was about 25 people.

URL: https://re-imagine.eu/wp-content/uploads/2022/05/Re-Imagine-Europa-Forum-2022-PROGRAMME.pdf

4.1.1.3 SOBIGDATA'S INVITATIONAL EVENT AT THE INTERNATIONAL FORUM ON DIGITAL AND DEMOCRACY - MEDIA ECOLOGY AND TECH DIPLOMACY: AN INTERTWINED AND INTERDEPENDENT PATH

The last decade has seen the emergence of new habits and behaviours in humans, driven by the increasingly rapid digital transformation. The Digital Media Ecosystem is a global arena in which many distinct battles are played, from civil rights and the rule of law to international relations and diplomacy. Facing global, regional and even local uncertainties could be made easier. What if we could combine a rational and ethical use of various digital technologies - especially what is now commonly defined as Artificial Intelligence systems - with more traditional decision-making mechanisms?

The event was organised at the John Cabot University on November 17th as a follow-up to the previous Roundtable on May 25th and held according to the Chatham House Rule. During the event, the attendees agreed to issue a joint declaration on technology diplomacy and media ecology (the Declaration of Rome)

that is currently being redacted and should be presented through a third event that will be organised during Q1 2023.

The event was held on November 17th, 2022. The participation was about 22 people.

URLs:

https://www.ifdad.org/wp-content/uploads/2022/10/IFDaD2022_Conference_Program.pdf

https://re-imagine.eu/the-rome-declaration-on-media-ecology-and-technology-diplomacy

4.1.1.4 SOBIGDATA'S SPECIAL SESSION AT THE INTERNATIONAL FORUM ON DIGITAL AND DEMOCRACY - CAN DATA GOVERNANCE AND TRUSTWORTHY ARTIFICIAL INTELLIGENCE HELP STRENGTHEN OUR DEMOCRACIES?

The 2022 edition of the International Forum on Digital and Democracy (IFDaD) returned to Rome on November 17th and 18th, 2022. It was organised in close cooperation between Associazione Copernicani, Fondazione Luigi Einaudi, Re-Imagine Europa, and SoBigData++. The IFDaD is a leading international event that brings together prominent thinkers from academia, policy, non-governmental organisations, and other stakeholders to debate and exchange information, ideas and good practices about Digitalization and Democracy. Among the dozens of leading thinkers from academia, policy, and non-governmental organisations that participated in this year's hybrid format event were Věra Jourová, Vice-President of the European Commission, Romano Prodi, former President of the European Commission and co-founder of the IFDaD, Gabriela Ramos, Assistant Director-General for the Social and Human Sciences of UNESCO, Jeffrey Sachs, Director of the Center for Sustainable Development at Columbia University, Roberto Viola, Director-General for Communications Networks, Content and Technology (DG CONNECT) at the European Commission, and Vincenzo Aquaro, Chief of Digital Government at the United Nations.

A special SoBigData++ session was organised to present the project to various policymakers, experts and citizens attending the forum (as described in Chapter 2.4.2.6 of this deliverable). The session was available in person (roughly 30 guests attended in the Aula Magna of Fondazione Luigi Einaudi) and virtually and broadcasted through the forum's specifically built streaming channel. The session included presentations by Roberto Trasarti, SBD++'s Project Coordinator; Fosca Giannotti, Professor at Scuola Normale Superiore; Kalina Bontcheva, Professor at the University of Sheffield; and Jeroen van den Hoven, Professor at Technische Universiteit Delft.

The Expert Roundtable "Can Data Governance and Trustworthy Artificial Intelligence Help Strengthen our Democracies?" followed the presentations and was moderated by Luca De Biase, Chief Editor at Nova and Media Research Director at Re-Imagine Europa, who presented the main questions to the speakers (Can civil and human rights be protected by implementing a different design for the current digital ecosystem? Is data transparency a key factor in protecting our democracies?).

The Roundtable was introduced by a special message from Guido Scorza, Member of the Garante per la protezione dei dati for Italy. The discussants were Fosca Giannotti, Jeroen van den Hoven, Paul Lukowicz, Professor at DFKI and the University of Kaiserslautern, and Gianluca Misuraca, Team Leader and Senior Expert in Technology Diplomacy at Intouchai EU. A special Highlight video summarising the main points touched during the discussion is currently in preparation.

The event was held on November 18th 2022. The participation was around 130 people, 80 males 50 females, between virtual and in-person.

4.2 Planned Activities for 2023

In 2023, dissemination and outreach activities to policymakers and the public at large will continue as planned according to what was outlined in the grant agreement. The main partner attending the task, Re-Imagine Europa, has recently won two separate grants issued by European Institutions to implement projects on topics closely related to those on which the Consortium is working. In addition, a third project developed in cooperation with a few prominent European Foundations will start in Q2, 2023. Several points of contact and collaboration could be established to improve the scope and outreach of the different projects. Below, you can find a summary of both:

ORBIS - Augmenting participation, co-creation, trust and transparency in Deliberative Democracy at all scales (Starting in February 2022)

ORBIS envisages addressing the disconnects between ambitious ideas and collective actions at a large socio-technical scale. It will respond to the profound lack of dialogue between citizenship and policy-making institutions by providing a theoretically sound and highly pragmatic socio-technical solution to enable the transition to a more inclusive, transparent and trustful Deliberative Democracy in Europe. ORBIS will shape and support new democratic models imagined and developed through deliberative democracy processes. These are processes of structured, inclusive and large-scale societal dialogue, which are inherently constructive, rather than transactional and are built through strong civic engagement. ORBIS follows a socio-constructive approach to European Democracy, in which deliberative democracy is not a theory which prescribes new democratic practices and models but rather the process through which we can collectively imagine and realise them.

NODES - Narratives Observatory combatting Disinformation in Europe Systemically (Starting in February 2022)

Recognising both the successes and failures of the methodologies put in place across the globe starting from 2016, NODES has developed an innovative and evidence-based approach to combat disinformation through a narrative methodology that aims to address some of the key critical points that are conducive to disinformation. To achieve this, NODES will develop and critically assess a novel large-scale mechanism to analyse and monitor the origin, development and spread of narratives classified as "disinformation", to

effectively provide quantitative and qualitative data on the narrative approach to disinformation, including the implementation of the Narrative Methodology, innovative media business models recommendations for future policy action and novel practices for practitioners in the field.

FUTURE4CITIZENS – The Future of Citizen's Engagement

Polarisation, social fragmentation, disaffection and a crisis of trust are some of the words that characterise the past years. This is not a European but a global trend. Since its launch in 2017, Re-Imagine Europa has worked to understand better why this is happening and to develop a systemic approach to reverse the existing trends. With the European elections looming and predictions of real-world challenges for 2023 (increasing energy and food prices), for the first time in history, there is a real risk of destructive political forces taking over. We need to develop a "vaccine against anger", move away from binary 'yes/no positions that divide, and find a constructive way forward that can start addressing the current crisis of trust.

4.2.1 Other activities

Besides these exciting new additions to Re-Imagine Europa's activities, which could prove very useful to develop further the scope and range of the Consortium's activities, the bilateral meetings with relevant members of the European Parliament and European Commission will certainly increase.

Moreover, RIEis currently organising a virtual event to present to the public the Declaration of Rome, which is currently (as of December 2022) being redacted as a follow-up to the two invitational RoundTables (*Media Ecology and Tech Diplomacy: An Intertwined and Interdependent Path and Can Data Governance and Trustworthy Artificial Intelligence Help Strengthen our Democracies?).* Such an event will feature several speakers from the Consortium and build on the principles included in the declarations to push the Consortium's agenda towards being the main testing ground for the EU's policy and regulation on data, artificial intelligence, and other related digital technologies. More virtual and in-person events will be developed according to the Consortium's strategy and needs during the next reporting period.

5 Stakeholder Analysis Updates

In this last year, and with the last remnants of the pandemic, the SoBigData++ consortium aim has shifted from opening new collaboration to strengthening and improving ongoing ones. Nevertheless, new stakeholders have been added to the project, with whom the partners have started collaborations. Most of the new stakeholders came into contact with SoBigData through participation in dissemination events of the project, or participation in consortia of new projects.

Overall, 10 of the partners of the consortium indicated a total of 26 new stakeholders, while the other partners reported no new stakeholder. In Figure 13, we report the macro-areas of activities of the new stakeholders, in Figure 14 we report the application fields, in Figure 15 we report the contribution of consortium partners to the acquisition of new stakeholders.

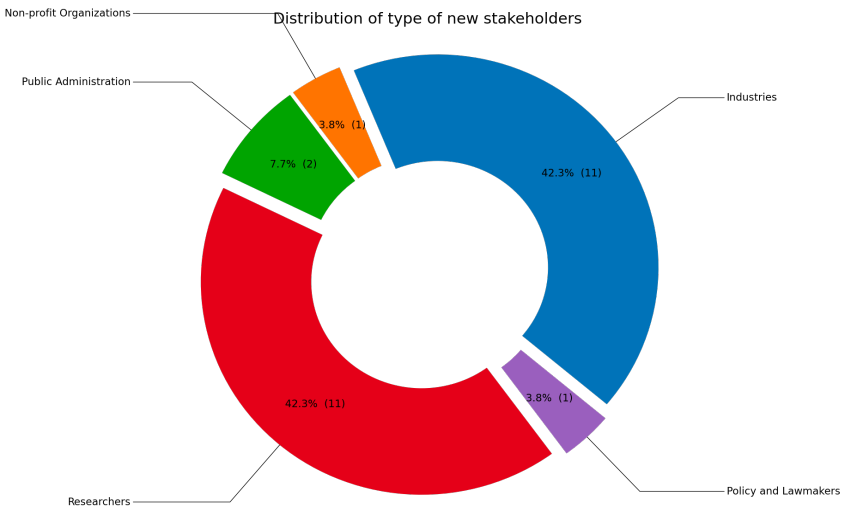

Figure 13: Distribution of macro-areas of activities of newly acquired stakeholders.

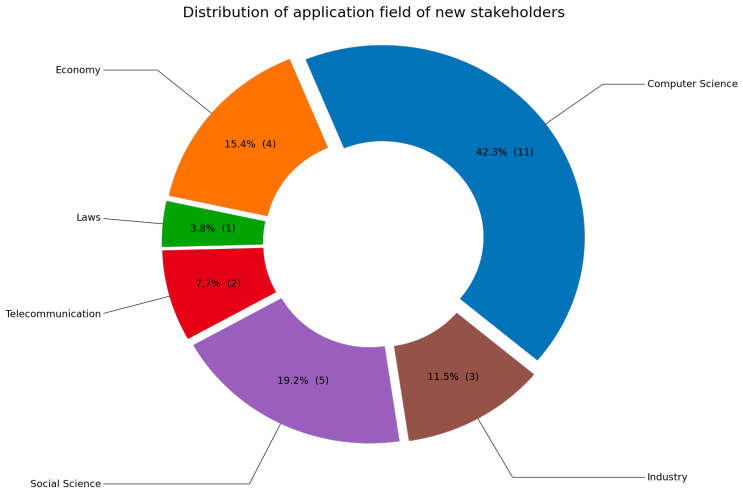

Figure 14: Distribution of application fields of newly acquired stakeholders.

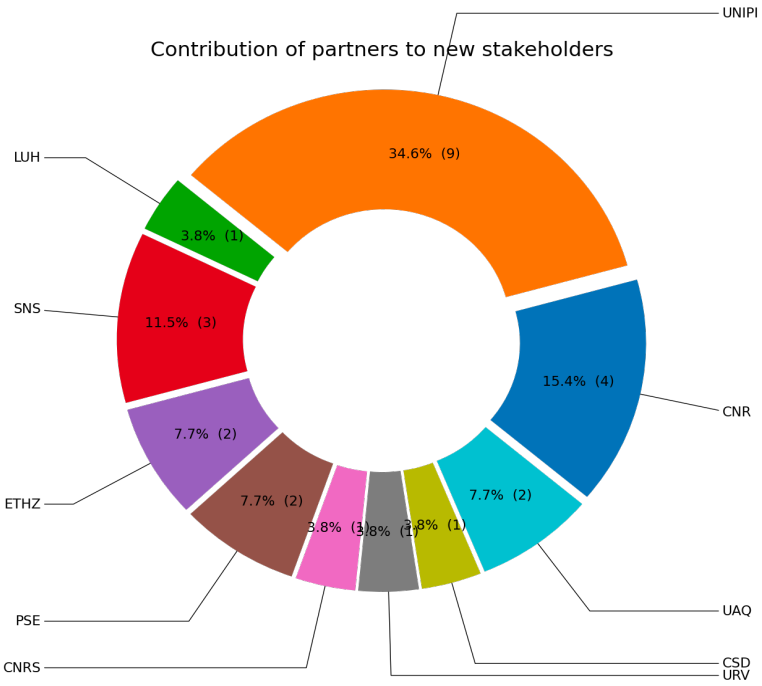

Figure 15: Contribution of partners to the acquisition of new stakeholders.

As expected from a wide research project like SoBigData++, the newly acquired stakeholders are mainly Industry or research focused. Interestingly, application fields on new stakeholders are much more varied: Computer Science stakeholders form the most numerous groups, but there are several new stakeholders in the fields of Social Science, Industry and Economy as well. Some partners of the consortium also continue collaboration with Policy Makers and local political entities, contributing to the broad vision of the SoBigData++ project.

The most prolific partner in terms of new stakeholders' acquisition has been UNIPI, with major contributions from SNS and CNR as well. In general, stakeholders have been acquired in a wide geographical range, attesting the good reach of the project and to the ability of most of the major partners to strengthen the position of the project.

Overall, the SoBigData++ project presents major point of attraction for stakeholders with a variety of interests, confirming the wide appeal of the research activities of all partners and the high potential for future development and collaborations.

6 Conclusions

This deliverable provided an overview of the dissemination activities carried out in the second period of the project (01 July 2021 - 31 December 2022), together with a report on the project's impact so far.

As we can appreciate, reading the document, all the actions done to reach general public, scientific communities, and potential adopters through various dissemination channels have achieved and even exceeded the objectives and KPI established in D3.1.

We also provided a comprehensive list and short description of all the events performed, where a good gender balance was reached. In this regard, we still need to take steps to improve, but we think we are in the right direction. Moreover, we should encourage Data Analysts, Industry and Policy makers stakeholders to be involved in SoBigData events.

In this deliverable, we presented also the new SoBigData brand identity. We recall that from now on, the logo presented here replaces the old one, thus this is the one that will be used in every mean of communication as well as official document. The templates for documents and presentation will be presented in the next deliverable.

Finally, a list of next year's planned events towards policy makers and a short update of the stakeholders' analyses are provided.