Social Mining & Big Data Analytics

SoBigData

RESEARCH INFRASTRUCTURE ++

Deliverable D4.1

Training Planning and Reporting

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 (48 months)
ENDING DATE	31/12/2023
PROJECT WEBSITE	http://www.sobigdata.eu
TOPIC	INFRAIA-01-2018-2019 Integrating Activities for Advanced Communities
GRANT AGREEMENT N.	871042

DELIVERABLE INFORMATION	
WORK PACKAGE	WP4 NA3 - Training
WORK PACKAGE LEADER	KCL
WORK PACKAGE PARTICIPANTS	CNR, USFD, UNIPI, FRH, UT, IMT, LUH, SNS, ETHZ, UNIROMA1, CNRS, URV, KTH, SSSA
DELIVERABLE NUMBER	D4.1
DELIVERABLE TITLE	Initial Planning for Training Programme
AUTHOR(S)	Mark Coté (KCL), Marco Braghieri (KCL), Joanna Wright (USFD), Beatrice Rapisarda (CNR)
CONTRIBUTOR(S)	--
EDITOR(S)	Beatrice Rapisarda (CNR), Valerio Grossi (CNR)
REVIEWER(S)	Roberto Trasarti (CNR)
CONTRACTUAL DELIVERY DATE	30/06/2020
ACTUAL DELIVERY DATE	13/10/2020
VERSION	1.1
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	23
KEYWORDS	Training, Coronavirus, Online Training, Datathons, Summer Schools, Events

EXECUTIVE SUMMARY

This deliverable provides an overview of the activities of Work Package 4 (Training) within the SoBigData++ project.

In section 1, the deliverable describes the purpose of the document, the relevance of the WP regarding project objectives and a description of the SoBigData++ project in regard to WP4 activities, which are then assessed in their relation with other Work Packages.

Section 2 focuses on training activities that have been performed in the reporting period, including Summer Schools (T4.2), Datathons (T4.3). Moreover, it provides an outlook on virtual events such as Webinars, Online Training Modules and Tutorials in accordance with Task 4.1.

Section 3 is devoted to describing the survey design that will be implemented to streamline data collection and harvesting regarding events, which will then feed into the reporting required by the SoBigData++ project indicators.

Section 4 describes the working group that WP4 has set up in order to address the challenges rising from the current coronavirus pandemic and how to address them in relation to the type of events that are more centred on in-person presence, such as datathons and workshops.

Section 5 updates on Cultivating diversity in data science through training (T4.4).

Section 6 provides an overview of possible future events, which necessarily have to take into account the current coronavirus pandemic as an influencing factor.

Finally, section 7 gives a summary of planned and reported training activities.

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

EU	European Union
EC	European Commission
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
EOSC	European Open Science Cloud
FAIR	First Aid for Responsible Data Scientists
ESFRI	European Strategy Forum on Research Infrastructures

TABLE OF CONTENTS

1 Relevance to SoBigData++

Work Package 4, entitled Training, aims to establish a joint training and education resource on big social data promoting the education of the next generation of data science researchers. The Work Package explores and develops both conventional and unconventional training experiences for master students, PhD students and early career post-doctoral researchers as well as an academically interested general public. Likewise, Work Package 4 proposes campaigns aimed promoting interest and participation of under-represented communities in data science with special emphasis on gender issues.

1.1 Purpose of This Document

This document aims to provide an overview of the planned activities for the reporting period and offer an overview of the activities that have already taken place, organised and performed by Work Package 4. Thus, the document is divided into sections, each detailing a different aspect of training within the SoBigData++ Project. By beginning with a report on planning and events, this document aims to provide a broad framework of training events that have been planned and organised by SoBigData partners. This section is followed by a description of planned training events for the reporting period and planned training events. Finally, this document describes relevant activities in the development of training materials.

This deliverable has been moved to M9 (originally was planned for M6 and then moved by the amendment currently under revision) due to the disruption related to Covid-19, which has been officially defined as a pandemic on 11 March 2020 by the World Health Organisation. This deliverable focuses on planning and on an overview of events that have taken place considering the sizeable impact of the Covid-19 pandemic. Further work package planning will be featured in Deliverable D4.2 due on M18.

1.2 Relevance to Project Objectives

The training activity within the SoBigData++ project is aimed at developing a unique, joint training and education resource centre on big social data. Building on the experience of the first iteration of the SoBigData project, Work Package 4 explores and develops conventional and unconventional training experience for master and PhD students and post-doctoral trainees. These experiences include the organisation of a number of different events and the development of the e-learning Area which has been created and integrated into the SoBigData Research Infrastructure. Among events, project-oriented summer schools and datathons have been planned and organised in order to match research (and industrial) needs and people skills. Moreover, activities will aim to address gender and diversity issues in data science through training.

This Work Package is organised around four different tasks.

- Task T4.1, named ‘Online Training Modules’ is centred on creating open-source training materials that are integrated into the SoBigData Research Infrastructure within the e-Learning Area. This part of the Research Infrastructure was designed, created and integrated into the SoBigData catalogue during the first iteration of the SoBigData project.

- Task 4.2 is centred on the organisation of a yearly SoBigData summer school, introducing participants to techniques and methodologies for analysing big data, in order to provide them with a solid background in the computational and mathematical theories behind algorithmic tools for empowering their future research. The summer schools will be strongly interdisciplinary and include experts across arts and sciences.

- Task T4.3 is centred around the organisation of Datathons, with a minimum of one per year, whose aim is to bring together young and bright minds in smaller dedicated groups, providing complementary theoretical and practical skills to visualise and analyse social big data questions addressing important societal problems. All the Datathon will be supported by the Operational Ethics and Law Board (operated by Work Package 2) in order to include Ethical and Law aspects in the Datathon activities.

- Task 4.4 Computer Science and Data Science currently fail to adequately embody staff equality and diversity issues. For instance, not only females but also minority groups, etc are still woefully underrepresented in data science. The aim is to leverage existing networks in order to raise awareness regarding the opportunities provided by employment in the field of data science. SoBigData++ will support specific events and provide travel grants for young female and minority group researchers, continuing an experience started in the first iteration of the SoBigData project.

1.3 SoBigData++ Project Description

SoBigData++ is an advanced community stemming from the experience of the SoBigData project (2015-2019). Its aim is 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.

This evolution of the SoBigData project is centered on strengthening tools and services to empower researchers and innovators through a platform for the design and execution of large-scale social mining experiments. It will be open to users with diverse background, accessible on project cloud, aligned with the European Open Science Cloud (EOSC), and exploiting supercomputing facilities. Further promoting First Aid for Responsible Data Scientists (FAIR) principles, SoBigData++ will render social mining experiments more easily designed, adjusted and repeatable by domain experts that are not data scientists. SoBigData++ will move forward from a starting community of pioneers to a wide and diverse scientific movement, capable of empowering the next generation of responsible social data scientists, engaged in the grand societal challenges laid out in its exploratories: Societal Debates and Online Misinformation, Sustainable Cities for Citizens, Demography, Economics & Finance 2.0, Migration Studies, Sport Data Science, Social Impact of Artificial Intelligence and Explainable Machine Learning. Moreover, an interest group centred on Network Medicine will be developed alongside exploratory activities.

SoBigData++ will advance from the awareness of ethical and legal challenges to concrete tools that operationalize ethics with value-sensitive design, incorporating values and norms for privacy protection, fairness, transparency and pluralism. SoBigData++ will deliver an accelerator of data-driven innovation that facilitates the collaboration with industry to develop joint pilot projects, and will consolidate an Research Infrastructure which has started its participation into the European Strategy Forum on Research Infrastructures (ESFRI) Roadmap and sustained by the creation of a new subject, the SoBigData Association.

1.4 Relation to Other Work Packages

The SoBigData++ project is organised around work packages which are combined in order to follow three main axes:

- Community building (including innovation and networking activities)
- Social mining research infrastructure building
- User accessibility (granted by virtual and trans-national access

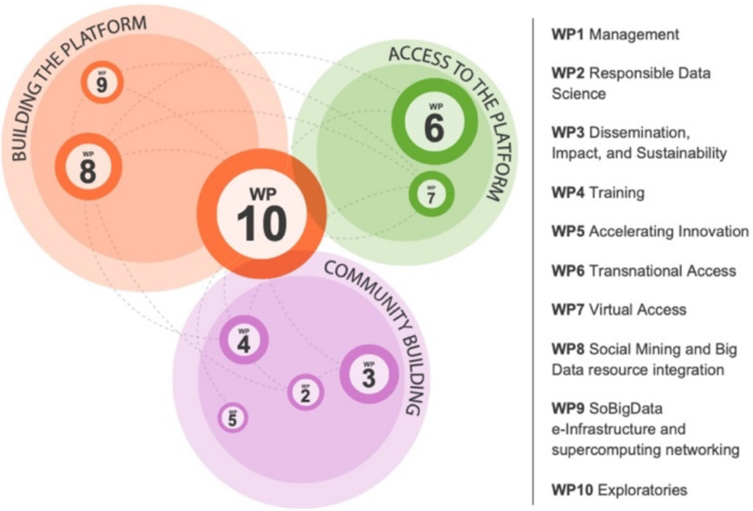

BUILDING THE PLATFORM
ACCESS TO THE PLATFORM
COMMUNITY BUILDING
WP 9
WP 8
WP 10
WP 6
WP 7
WP 4
WP 3
WP 2
WP 5
WP1 Management
WP2 Responsible Data Science
WP3 Dissemination, Impact, and Sustainability
WP4 Training
WP5 Accelerating Innovation
WP6 Transnational Access
WP7 Virtual Access
WP8 Social Mining and Big Data resource integration
WP9 SoBigData e-Infrastructure and supercomputing networking
WP10 Exploratories

Figure 1 ***SoBigData++ Work Packages organization***

Among all work packages, WP2 (Responsible Data Science), WP3 (Dissemination, Impact and Sustainability), WP4 (Training) and WP5 (Accelerating Innovation) are aimed at community building between excellence centres, other academic and industrial users and trainee data scientists. Thus, Work Package 4 works closely with:

WP2 – Responsible Data Science

This work package is mainly tasked with operationalising a legal and ethical framework for the whole SoBigData++ Research Infrastructure.

WP3 – Dissemination, Impact and Sustainability

This work package is mainly tasked with developing dissemination and impact strategies for the entire SoBigData++ project.

WP5 – Accelerating Innovation

This work package is tasked with widening the project's impact through innovation activities aimed at industry and other stakeholders, such as government bodies, non-profit organisations, funders and policy makers.

Aside from these work packages, WP4 will also work in collaboration with WP7 (Virtual Access) in order to design and integrate training modules into the SoBigData++ Research Infrastructure. Moreover, WP4 will work alongside WP9 (JRA2 - E-Infrastructure and Supercomputing Network) to create operation manuals for facilitating platform exploitation in all the aspects will be made accessible through a specialised operation portal dedicated to developers, ICT managers, and service providers. Finally, WP4 will also work alongside WP10 (JRA3 – Exploratories) in order to explore possible integration of activities performed by User Community Activists, which will be one of the two key figures alongside with Scientific Leaders, within the e-Learning area of the SoBigData++ Research Infrastructure.

2 Training Activities

Due to their in-person nature, both summer schools and datathons have had to a face severe disruption due to the coronavirus pandemic. Due to the health situation, both national lockdowns and travel restrictions have been implemented, rendering impossible to safely hold in-person events. The temporal distribution of events, as per Grant Agreement n°871042, originally planned for a yearly SoBigData++ summer school (M7; M19; M31; M43) and a yearly datathon (M10; M22; M34; M36).

In order to adjust to the present – and for the foreseeable future – situation, WP4 has adopted a series of strategies aimed at mitigating the impact of the impossibility of live, in-person events. Strategies have included transforming live events into virtual events or anticipating scheduled events in order to provide a real-time response to rising challenges, such as the coronavirus pandemic.

Even with the unfortunate pandemic situation SoBigData exceeded the number of events organized and planned in the first 18 months of the project thanks to the synergies with other European projects and to the collaborations with institutions (inside and outside the consortium).

2.1 Summer Schools and Datathons

2.1.1 School in Machine Learning of Dynamic Processes and Time Series

https://www.mldyn2020sns.com/

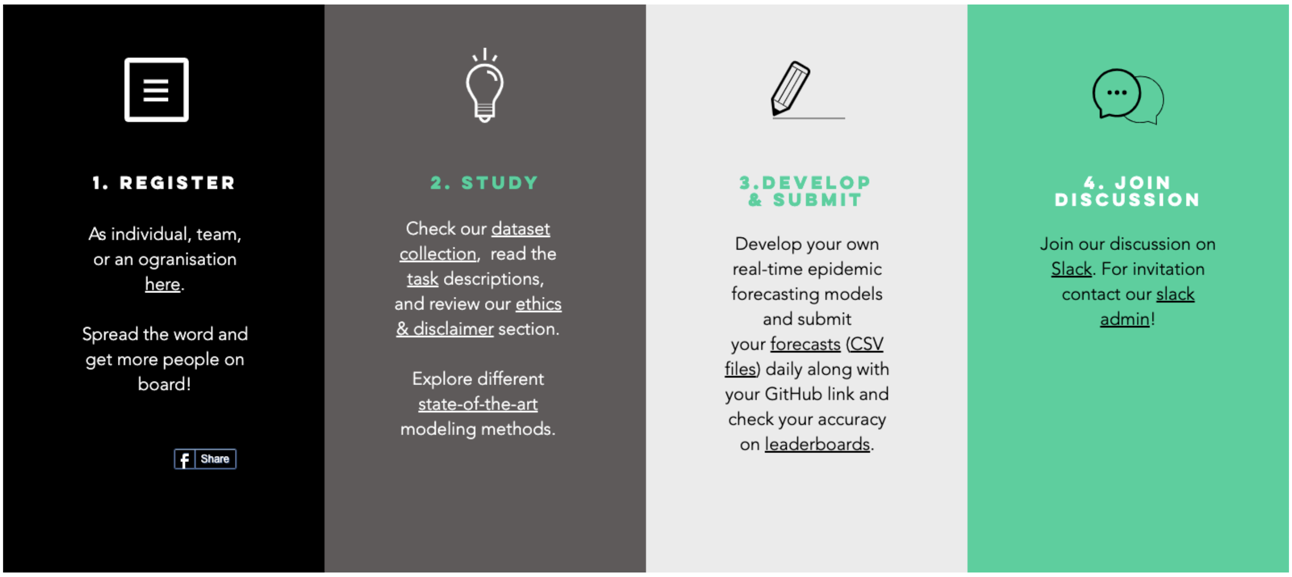
The Winter School will take place on 26-27 November 2020 in Pisa, Italy. Due to the current worldwide situation of the Covid-19 pandemic, the School will be in a hybrid format: can be attended either in presence or online. The aim of the School is to present recent developments in Machine Learning focusing on data-driven approaches to statistical learning and dynamical systems. Applications will also be discussed, such as the forecasting of financial time series.

Due to the coronavirus pandemic, the School which was originally scheduled to take place on M7 has been moved to M11.

2.1.2 Real-time epidemic datathon

https://www.epidemicdatathon.com

ETH Zurich seized the opportunity to work with the real time data of the unfolding Covid-19 pandemic and provide an opportunity for data science researchers to work on a collective open-source real-time forecasting challenge. The Datathon was open to everyone (individuals or teams) and ran from April 2020 to July 2020. The Datathon used publicly available data and encouraged data scientists to contribute to the global open-source scene by releasing real-time epidemic forecasting models.



1. REGISTER

As individual, team, or an ogranisation here.

Spread the word and get more people on board!

Share

2. STUDY

Check our dataset collection, read the task descriptions, and review our ethics & disclaimer section.

Explore different state-of-the-art modeling methods.

3.DEVELOP & SUBMIT

Develop your own real-time epidemic forecasting models and submit your forecasts (CSV files) daily along with your GitHub link and check your accuracy on leaderboards.

4. JOIN DISCUSSION

Join our discussion on Slack. For invitation contact our slack admin!

Figure 2 Epidemic Datathon Workflow

In order to enhance participation, alongside with the creation of an ad-hoc website, the Epidemic Datathon succeeded in creating a series of community tools in order to foster exchange between participants, and attracted 37 individuals and in line with the project's aim to encourage more female participation in data science, approximately one third of participants were female.

Due to the coronavirus pandemic, we decided to anticipate the datathon to give a real-time response to rising challenges, this was originally scheduled to take place on M10 has been moved to M4.

2.2 Webinars and Online Training Modules

In accordance with the community building pillar of the SoBigData++ work package organisation, a series of webinars aimed to substitute live, in-presence events have been created and are aimed to be integrated into the SoBigData++ Research Infrastructure. These live events were broadcasted via YouTube and their integration within the SoBigData++ Research Infrastructure is currently underway. The aim was to provide a flexible, timely resource regarding issues that were particularly pressing during the first months of the project. Moreover, a number of tutorials was organised within the Social Informatics Conference which took place in Pisa from 6 to 9 October. The conference is an interdisciplinary venue for researchers from Computer Science, Informatics, Social Sciences and Management Sciences to share ideas and opinions, and present original research work on studying the interplay between socially-centric platforms and social phenomena. It was an event co-organised by SoBigData++ within its WP3 dissemination, impact and sustainability activities.

2.2.1 Epidemics and the city: how human mobility and well-being changed during the COVID-19 era

https://www.youtube.com/watch?v=8Cc-vPeTACk

This was a dissemination **webinar** that was promptly organised to capture the threads of the topical Covid-19 crisis and the impact it has had on society from the perspective of Data Science and Environmental Epidemiology. It took place on 3 July 2020, and featured professor Dino Pedreschi from University of Pisa and professor Paolo Vineis from Imperial College London. Experts discussed the effect of Covid-19 on mobility, the impact on people’s well-being and on virus transmissibility as well as the quality of life of citizens. It also looked at co-benefits in relation to lockdown.

Epidemics and the city: how human mobility and well-being changed during the COVID-19 era

Dino Pedreschi (University of Pisa) and Paolo Vineis (Imperial College, London)

3rd July 2020 11:00am CEST

http://www.sobigdata.eu/events/webinar-epidemics-city-mobility-covid-19

Figure 3 Epidemic and the City: How Human Mobility and Well-being changed during the Covid-19 era

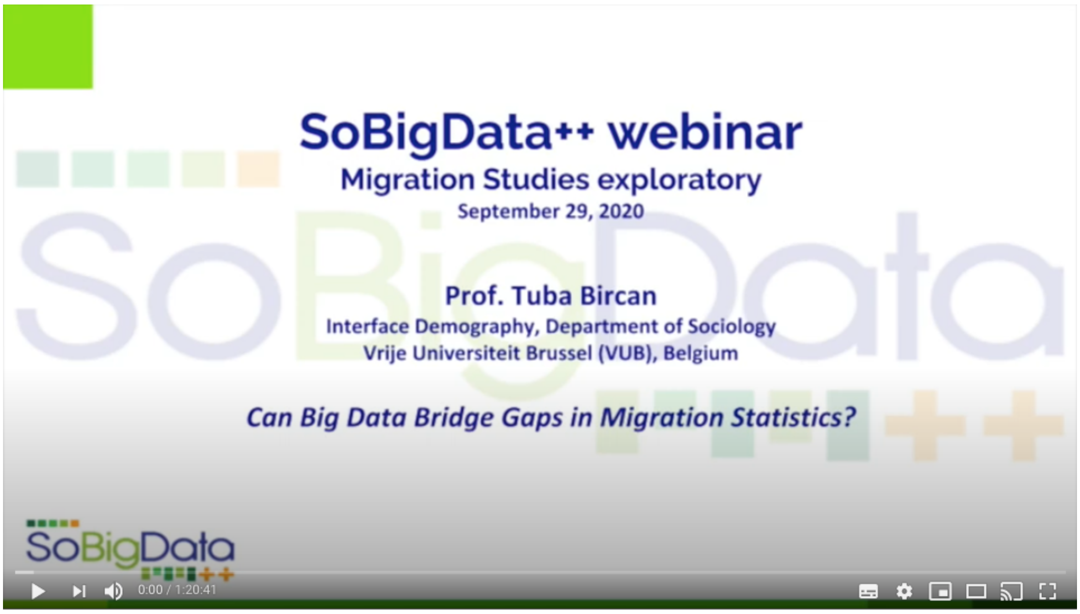
The webinar was aimed at experts, stakeholders of the SoBigData++ project and it was also open to the general public. Approximately 70 people were involved and the event was well received.

2.2.2 Can Big Data Bridge Gaps in Migration Statistics?

https://www.youtube.com/watch?v=ROQzO33KSnA

This **webinar** was organised by UNIPI and CNR and was born from the HumMingBird project – a Horizon 2020 project that aims at responding to these needs by improving understandings of changing nature of migration flows and the drivers of migration, by analysing patterns, motivations and new geographies, forecasting emerging and future trends. It took place on 29 September 2020 and featured professor Tuba Bircan of the Interface Demography, Department of Sociology of the Vrije Universiteit Brussels (VUB). Traditional statistical data on international migration suffers from the problems (gaps) of inconsistency in definitions, differences in geographical coverages, absence of reasons for migration, timeliness and limitations in demographic characteristics.

Although there is a novel list of potential data sources that could provide valuable, real-time insights, these remain largely untapped for the time being. There are sensitivity obstacles, legal issues, availability, accessibility, purpose of the data, etc. However, improving migration data is a crucial step to improving migration governance since better data is needed in order to bring about sustainable social and economic development and national migrant data strategies are needed to inform good policies. This talk discussed the existing gaps and shortcomings of the migration statistics and the potential utilization of Big Data analytics for bridging these gaps.



SoBigData++ webinar

Migration Studies exploratory

September 29, 2020

Prof. Tuba Bircan

Interface Demography, Department of Sociology

Vrije Universiteit Brussel (VUB), Belgium

Can Big Data Bridge Gaps in Migration Statistics?

Figure 4 Can Big Data Bridge Gaps in Migrations Statistics? Webinar

It was expected that this event would attract participation of around 30 people; however, 45 individuals accessed the webinar – an increase of 50% on plan, demonstrating the success of the project's outreach. Furthermore, over a third of the participants were female demonstrating an improving ratio between the genders. The event was aimed at experienced and non-expert researchers and academics interested in studying human migration, but coming from different research fields, (i.e., sociology, computer science, and demography). The organizers were very satisfied with the level of participation and there were various interesting questions and discussions that arose at the end of the presentation.

2.2.3 Data Protection for Research and Statistical Purposes: Towards Legally Attentive Datathons

This **webinar** took place on 22 July 2020 and was organised by Scuola Superiore Sant'Anna in Pisa, Italy as an open event. The program included 3 short talks from experts in the field followed by a Q&A session. The topics covered were 'Data Processing for Scientific Research and Statistics and the SoBigData++ Framework', 'Datathons: Risks & Opportunities' and 'Legally Attentive Datathons: Ready for the Check list'. Two of the three speakers were female following one of the main aims of the SoBigData++ project to showcase females in Data Science to inspire and encourage more females into this research sector. The event attracted 34 participants.

2.2.4 The SoBigData E-Infrastructure

https://www.youtube.com/watch?v=DAUVkCufX6o&t=1s

This **webinar** was conceived to explain how the SoBigData gateway works, what kind of resources and services are available, how to search for resources, how to upload new resources in the Catalogue.

SoBigData++ Infrastructure Webinar

Thursday, May 21, 2020

Figure 5 SoBigData++ Infrastructure Webinar

This webinar featured presentations from different members of the SoBigData++ project in order to illustrate the capabilities of the SoBigData++ research infrastructure and provide a step-by-step guide to its different functionalities. Valerio Grossi (CNR) introduced the e-infrastructure followed by Massimiliano Assante (CNR) who provided an introduction to the SoBigData++ Method Engine and a tutorial on Method Integration. Roberto Trasarti (CNR) provided a review of the interfaces and tools featured in the SoBigData++ platform. Finally, by Pasquale Pagano (CNR) who described the SoBigData++ Infrastructure workspace and catalogue. All slides are available on the SoBigData++ research infrastructure at https://data.d4science.net/PLsN.

2.2.5 Discovering Gender Bias and Discrimination in Language

https://kdd.isti.cnr.it/socinfo2020/tutorials2.html

The **tutorial** focused on the issue of digital discrimination, particularly towards gender. Its main goal is to help participants improve their digital literacy by understanding the social issues at stake in digital (gender) discrimination, and learning about technical applications and solutions.

The tutorial, held by Dr. Mark Coté, (KCL), Dr. Xavier Ferrer Aran (KCL) and Dr. Tom van Nuenen, (KCL) is based on their research in language modelling and Word Embeddings in order to clarify how human gender biases may be incorporated into AI/ML models. The tutorial is divided in four parts: it basically iterates twice through the social and technical dimensions.

1. We first offer a short introduction into digital discrimination and (gender) bias.
2. We give examples of gender discrimination in the field of AI/ML, and discuss the clear gender binary (M/F) that is presupposed when dealing with computational bias towards gender.
3. We then move to a technical perspective, introducing the DADD Language Bias Visualiser which allows us to discover and analyse gender bias using Word Embeddings.
4. Finally, we show how computational models of bias and discrimination are built on implicit binaries, and discuss with participants the difficulties pertaining to these assumptions in times of post-binary gender attribution.

The tutorial also included pre- and post-tutorial questionnaires, which are intended to track participants' digital discrimination literacy, as well as the explanatory value of the tutorial. The tutorial comes from our UK EPSRC-funded project Discovering and Attesting Digital Discrimination (DADD), a cross-disciplinary project at King's College London addressing research questions on digital discrimination involving academic (Computer Science, Digital Humanities, Law and Ethics) and non-academic partners (Google, AI Club), and the general public, including technical and non-technical users.

2.2.6 Learning to Quantify: Supervised Prevalence Estimation for Computational Social Science

https://kdd.isti.cnr.it/socinfo2020/tutorials2.html

This **tutorial** was aimed to raise the awareness of computational social science researchers on the fact that, when they are using classification technology, their research would almost always benefit from using quantification technology instead. The tutorial introduced the attendees to the main supervised learning techniques that have been proposed for solving quantification, to the metrics used to evaluate these techniques, and to some off-the-shelf, publicly available software packages that implement them.

2.2.7 Social Information for Emergency Response

The **tutorial** "Social Information for Emergency Response: Actionable Information from Unconventional Data Sources" introduced the basics for extracting and analyzing information from social media, with a specific focus of extracting images in an emergency after a natural disaster. The tutorial provided the basics about crawling and tweet analysis. A specific focus has been given to fine-grained geo-localization of tweets and crowdsourcing to filter relevant images and confirm geolocations, which are needed to provide high-quality information. The objective of the tutorial was to provide an introduction and hands-on experience in some of the tools available in the field of emergency information system and in particular on tools enabling the search of posts, focusing on Twitter and considering also alternative social media, post analysis, with text analysis techniques based on NLP and relevant image analysis approaches to filter images according to different criteria, and tools for setting up a crowdsourcing environment, based on the PyBossa open source tool, and for evaluating the quality of results from crowdsourcing.

2.2.8 Data Science Colloquium

https://datasciencephd.eu/events/data-science-colloquium-2020

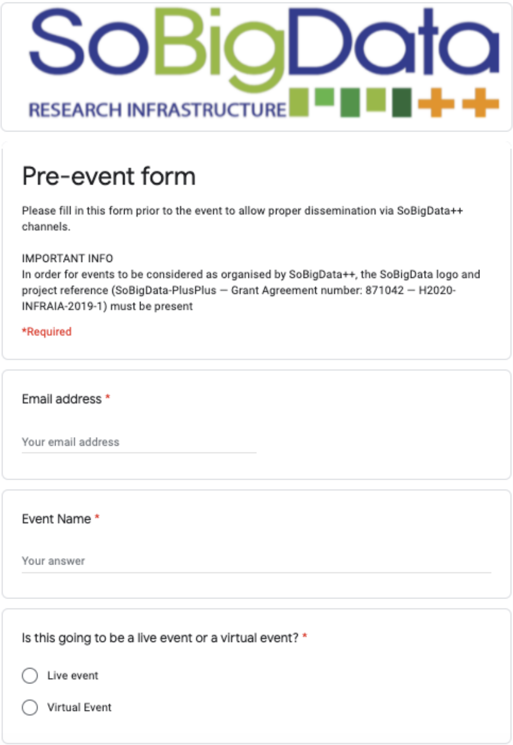
This event was organised by SNS, CNR, IMT and UNIPI, the Data Science Colloquium consisted of 2 hour sessions, 3 times a week from 20 May to 8 June. It included seminars held by professors and 3rd year PhD students to support and guide 1st year students through their research projects. The sessions were also open to any interested party. This event was both live and remotely accessible.

2.3 PhD in Data Science

The Ph.D. in Data Science is aimed at educating the new generation of researchers that combine their disciplinary competences with those of a "data scientist", able to exploit data and models for advancing knowledge in their own disciplines, or across diverse disciplines. To this purpose, the Ph.D. in Data Science develops a mix of knowledge and skills on the methods and technologies for the management of large, heterogeneous and complex data, for data sensing (how to harvest data), for data analysis and mining (how to make sense of data), for data visualization and storytelling (how to narrate data), for understanding the ethical issues and the social impact of Data Science.

3 Design of Survey for Training Activities

Quantifiable data collection is a key factor in providing sound indicators in participants to training activities which are a core part of Work Package 4. In order to further enhance this aspect and provide an evolution of the tools created for the SoBigData project, Work Package 4 decided to develop an ad hoc suite of data collection tools for SoBigData++. In order to streamline data collection and enhance cohesive data input from partners WP4 designed three different forms, hosted on Google Forms.



Pre-event form

Please fill in this form prior to the event to allow proper dissemination via SoBigData++ channels.

IMPORTANT INFO
In order for events to be considered as organised by SoBigData++, the SoBigData logo and project reference (SoBigData-PlusPlus – Grant Agreement number: 871042 – H2020-INFRAIA-2019-1) must be present

*Required

Email address *

Your email address

Event Name *

Your answer

Is this going to be a live event or a virtual event? *

○ Live event

○ Virtual Event

Figure 6 One of the three forms

3.1 Pre-Event Form

This form was created in order for event organisers to provide as much information as possible to other interested parties within the consortium and especially WP3 which is tasked with dissemination, impact and sustainability. The creation of an up-to-date calendar of training events will thus be streamlined by the incoming information provided by organisers.

3.2 Attendee Registration Form

This form was created to aid event organisers in data collection. The main objective was to provide organisers with a ready-made, collection tool on participants due to be collected when organising a SoBigData++ event.

3.3 Event Report Form

This form was created in order for organisers to collect all needed information around the event, in order to provide a precise snapshot of the event itself, and all relative data points. Moreover, by implementing the same forms for all training events, we aim to consolidate our dataset with three main aims: streamline event promotion, aid event organisers in collecting information from attendees and provide a full report. Thus, we shall be able to collate data from events in a more efficient manner.

Moreover, an event guideline and frequently asked question presentation has been created in order to provide guidance to event organisers on how to use these tools and facilitate data collection and harvesting.

4 Cultivating diversity in data science through training

The main goal is to set up a version of https://biaswatchneuro.com/ for SoBigData++ relevant fields. Bias Watch Neuro tracks the speaker composition of conferences in neuroscience, particularly with respect to gender representation. Mark Coté (KCL) has been in contact with project leader Fosca Giannotti and ideally WP4 will be able to develop an ad hoc form of speaker composition tracking in the SoBigData++ relevant fields, in order to provide an up to date and novel understanding of the issues that regard diversity in data science and the other fields that are embraced by SoBigData++.

5 Working group on the evolution of virtual events

During informal and formal WP4 meetings, a new need has risen due to the coronavirus pandemic. Seeing that the current situation will not be changing in the foreseeable future and wanting to proactively address some of the main issues regarding live events, WP4 has resolved to the creation of a working group dedicated to finding possible new avenues to introduce changes for those events that are based on a core in-person interaction, such as datathons and workshops. In such regard, WP4 has established a working group which is currently investigating best-practices and novel approaches to optimise virtual events to partially mitigate the absence of in-person events.

6 Future events

Planning of future events has proven very challenging due to the coronavirus pandemic, which due to the nature of national decisions, has influenced freedom of movement among countries with the implementation of quarantine periods and, more broadly, fundamental changes in the travel possibilities of many organisers and participants to live events. Seeing that this situation will last for the foreseeable future we are currently working on a revised schedule of events, which will be centred on virtual events. Hence, where are looking at different possibilities with a flexible approach, in order to address possible future modification in the influencing factors of the coronavirus pandemic. What follows is a tentative list of events that are currently being explored as feasible in the forthcoming future.

6.1 Datathon on “Social Good” (UNIPI and CNR)

This datathon will be organised in November 2020 among the UNIPI MA in Data Science and PhD Students and will involve a month-long series of virtual webinars.

6.2 Narratives and infrastructures of ephemeral attention Datathon (CNRS)

This is a datathon planned by CNRS in order to address part of the novel challenges connected to the coronavirus pandemic. A date is yet to be determined, but it would tentatively take place in the first part of 2021.

6.3 YouTube and Misinformation Datathon (CNRS)

This datathon, which is in its early planning stages, would ideally take place in 2021 and be focused on YouTube and misinformation and will be organised by CNRS.

6.4 Data Science Summer School (UNIROMA1 and UNIPI)

UNIROMA1 are planning for this summer school which would be a SoBigData++ summer school organised in collaboration with UNIPI.

6.5 Soccer Data Challenge Datathon (CNR)

This ‘Soccer Data Challenge’ is being planned for 2021 by CNR in collaboration with Futbol Club Barçelona.

6.6 Misinformation Analysis Summer School (USFD)

The University of Sheffield would like to organise a summer school on Misinformation Analysis. This would be the second edition after the first, which was held during the SoBigData project. However, feasibility has to be addressed.

7 Summary of Training Planning and Reporting

7.1 Training Planning

This deliverable has tried to account for the planning activities that have been implemented by Work Package 4. While training planning has been severely disrupted by the coronavirus pandemic, WP4 has managed to address some of the challenges of the current situation. Clearly, some of the sanitary and logistic issues deriving from the coronavirus pandemic will be present for the foreseeable future. However, WP4 has managed to maintain a proactive approach on future events planning, taking into consideration all present and possible future restrictions. Moreover, WP4 has developed a new streamlined approach to data collecting that will aid organisers and the project as a whole to keep track of participation to virtual events. This approach is intentionally being developed in a flexible manner, in order to remain in use once the coronavirus pandemic will allow in-person events. Finally, a working group has been set-up within WP4 in order to address the long-term consequences of virtual-only events, especially focusing on the event types that are more tightly connected with in-person presence, such as datathons and workshops.

7.2 Training Reporting

Section 2 of this deliverable is entirely dedicated to the events that have taken place despite the coronavirus pandemic. Moreover, WP4 has shown a proactive approach in addressing the impact of the pandemic with datathons, webinars and other activities. This demonstrates that WP4 has reacted promptly to a completely unprecedented situation and has allowed participants to its events to gather further insight into a globally-impacting phenomenon such as the coronavirus pandemic.