Social Mining & Big Data Ecosystem

SoBigData

RESEARCH INFRASTRUCTURE

Project Acronym	***SoBigData***
Project Title	***SoBigData Research Infrastructure*** ***Social Mining & Big Data Ecosystem***
Project Number	***654024***
Deliverable Title	**Periodic reports on TA activities 3**
Deliverable No.	**D6.3**
Delivery Date	***29 January 2020***
Authors	**Joanna Wright (USFD)**

HORIZON 2020

DOCUMENT INFORMATION

PROJECT	
Project Acronym	SoBigData
Project Title	SoBigData Research Infrastructure Social Mining & Big Data Ecosystem
Project Start	1st September 2015
Project Duration	52 months
Funding	H2020-INFRAIA-2014-2015
Grant Agreement No.	654024
DOCUMENT	
Deliverable No.	D6.3
Deliverable Title	Periodic reports on TA activities 3
Contractual Delivery Date	31 December 2019
Actual Delivery Date	29 January 2020
Author(s)	Joanna Wright (USFD)
Editor(s)	Beatrice Rapisarda (CNR)
Reviewer(s)	Cristina Muntean (CNR), Beatrice Rapisarda (CNR), Valerio Grossi (CNR)
Contributor(s)	
Work Package No.	WP6
Work Package Title	TA1 Transnational Access
Work Package Leader	USFD
Work Package Participants	CNR, FRH, UT, LUH, AALTO, ETH
Dissemination	PU
Nature	Report
Version / Revision	1.2
Draft / Final	Final
Total No. Pages (including cover)	36
Keywords	transnational access, short term scientific mission,

DISCLAIMER

SoBigData (654024) is a Research and Innovation Action (RIA) funded by the European Commission under the Horizon 2020 research and innovation programme.

SoBigData proposes to create the Social Mining & Big Data Ecosystem: a research infrastructure (RI) providing an integrated ecosystem for ethic-sensitive scientific discoveries and advanced applications of social data mining on the various dimensions of social life, as recorded by "big data". Building on several established national infrastructures, SoBigData will open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research.

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 document has been produced with the funding of the European Commission. 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.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 27 member states of the European Union. It is based on the European Communities and the member states' cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (http://europa.eu.int/).

GLOSSARY

ABBREVIATION	DEFINITION
TNA	Trans-National Access
STSM	Short Term Scientific Missions

TABLE OF CONTENTS

DELIVERABLE SUMMARY

This report details the calls initiated in the reporting period including the user selection process and outcomes, TA visit reports and self-evaluation.

It also includes summaries of the previous calls in addition to the details of the final call.

Statistics and a full overview has also been provided covering all 4 calls.

1 RELEVANCE TO SOBIGDATA

1.1 PURPOSE OF THIS DOCUMENT

This document describes the trans-national access activities carried out in the final months of the SoBigData project. More specifically, the preparation of the final call for TNA access is detailed, followed by applicant statistics, review process description, and outcomes.

We further developed the framework agreement and ethical approval processes, to ensure data protection, ethical, and legal aspects are adhered to.

1.2 RELEVANCE TO PROJECT OBJECTIVES

As detailed below, Transnational access is core to the SoBigData objectives, specifically addressing community widening and adoption.

1.3 SOBIGDATA PROJECT DESCRIPTION

SoBigData serves the wide cross-disciplinary community of data scientists, i.e., researchers studying all aspects of societal complexity from a data- and model-driven perspective, including data and text miners, visual analytics researchers, socio-economic scientists, network scientists, political scientists, humanities researchers, and more.

The SoBigData RI defines two ways for accessing its services:

Online access offers cutting-edge services for big data and social mining research.

● The Virtual Access is realized thanks to the e-Infrastructure which offers a web front-end comprising of a catalogue of SoBigData resources (both data and services) and a set of SoBigData Virtual Research Environments (VREs) dynamically created to address the requirements of specific research Experiment

  o Virtual Research Environments (VREs) are web-based working environments equipped with a number of applications, enabling scientists to have access to the set of data, services, models and algorithms needed to perform their investigation in a collaborative way.

● Transnational/ On-site access that offers world-leading research expertise from multiple disciplines, as well as big data computing platforms, big social data resources, and cutting-edge computational methods. Transnational access is granted by seven national infrastructures (see section 2 for details). The access will be granted through project calls in one of the following forms:
  o Exploratory projects are research experiments bound to multi-disciplinary themes specified in the calls and are expected to target data scientists from multiple disciplines, and address the definition of needed infrastructure resources and knowledge offered by specific hosting nodes
  o Open call, blue-sky projects are targeted at researchers wishing to explore the infrastructure for their own research topics.

The way these services are provided depends on the needs of the Final User. In Section 2.4. Actors and Use Cases for Virtual and Transnational Access, we report a set of use cases; starting from the simplest one, where the user can download data and/or software, to a complex one when both software and data are copied from a node to another of the infrastructure. It is important to highlight, that the envisaged data processing steps will be in compliance with ethical and legal requirements.

Currently, in our catalogue we have heterogeneous 63 datasets, and more than 70% contains personal data. We can find trajectory data, mobile call data, Twitter data, or retail data. In all these cases, the data sets contain personal data.

1.4   RELATION TO OTHER WORKPACKAGES

Transnational access draws on all technical SoBigData WPs, as they are developing the infrastructure and integrating the tools needed for the completion of the transnational access visits.

1.5   STRUCTURE OF THE DOCUMENT

Section 2 details the transnational access calls in the final months of SoBigData project. Appendix A details the call text itself. Appendix B shows the application form template. Appendix C is the proposal evaluation template, showing the selection and review criteria. Appendix D is the agreement template, which needs to be filled in and signed by each approved candidate.

2 TRANSNATIONAL ACCESS CALLS DURING THE REPORTING PERIOD

Call 3 ended on 20 April 2018 and the final call was released as an Open Call in September 2018 with a submission deadline of 1 October 2019. The visits period started in November 2018 with an end date of November 2019 which has been extended to 31 December 2019.

As with the previous calls this final call requested applications for Short Term Scientific Missions (STSMs) to carry forward big data analysis projects of relevance to SoBigData.

We welcomed applications from individuals with a scientific interest, professionals, start-ups and innovators that may benefit from training in data science and social media analytics.

Funding for a short-term scientific mission (2 weeks to 2 months) was made available up to €2,500 per participant (to cover the cost of daily subsistence, accommodation, and European flights) in call 2 and increased to €3,500 in call 3. The maximum of €3,500 continued during the final call. STSM bursaries were awarded on a competitive basis, according to the procedure described in section 2.1, and based upon the quality of the applicant, the scientific merit of the proposed project, and their personal statement.

Applications were offered to undertake STSM's at the following institutions:

- GATE (Societal Debates, Migration Studies) – The University of Sheffield, Sheffield, UK
- SoBigData.it (City of Citizens, Well-being and Economy, Societal Debates, Migration Studies, Sports Analytics[1]) – The European Laboratory on Big Data Analytics and Social Mining, Pisa, Italy
- Fraunhofer IGD (Societal Debates, Sports Analytics[1]) – The Competence Center for Information Visualization and Visual Analytics, Darmstadt, Germany
- UT (Well-being and Economy) – Tallin, Estonia
- LUH: L3S Research Center / Leibniz University Hannover (Societal Debates) – Hannover, Germany
- Aalto University (City of Citizens, Societal Debates) – The Data Mining Group, Aalto, Finland
- ETH Zurich (Migration Studies) – The Computational Social Science Group, Zurich, Switzerland

2.1 APPLICATION PROCEDURE

The application process was made up of a short biography of the applicant including qualification and experience in social media and big data research, the requested project objectives as well as a description of the work planned, a personal statement including their expected achievements and impact as well as dissemination plans. See Appendix B.1 Application form.

---

[1] Call 3 only

2.2 SELECTION PROCESS

The selection panel comprises the following members:

Internal Members		External Members	
Review Member	**Organisation**	**Review member**	**Organisation**
Kalina Bontcheva	USFD	Gerhard Lauer	University of Basel
Cristina Muntean	SoBigData.it	Arkaitz Zubiaga	University of Warwick
Thorsten May	Fraunhofer	Chedy Raïssi	INRIA
Anna Leontieva	UT	Matteo Magnani	Uppsala University
Avishek Anand	L3S	Marc Plantevit	Université Claude Bernard Lyon 1
Aristidis Gionis	AALTO	Barbara Plank	University of Copenhagen
Nino Antulov	ETZH	Celine Robardet	National Institute of Applied Science in Lyon
		Giorgina Ifrim	University College Dublin
		Nikos Pelekis	Pireus University

Table 1: The selection panel

For evaluation purposes each section of the application was ranked on a scale of 1-5 (5 being the highest). The areas comprised of: Quality of the applicant; Quality of the proposed research project; Originality of the proposed research project; Feasibility and Impact on SoBigData. We received 34 applications in the final call; 24 of these were accepted and visits were arranged. See the Table below detailing the applications we received and their outcomes.

2.3 FIRST CALL: APPLICANT SUMMARY

The first call received 22 applications with 18 being EC applicants and 4 non EU applicants.

Of the 18 EU applications there were 12 successful applicants. Of the 4 non EU applications there was 1 successful applicant.

There were 14 male applications with 7 being successful (all EU). There were 8 female applicants with 6 being successful (5 EU and 1 non EU). This ensured a good gender balance, in line with SoBigData policies.

USFD received 8 applications and 5 were successful; PISA received 9 applications with 6 being successful, Aalto received 2 applications with 1 successful, Zurich received 1 successful application and Fraunhofer received one application which was not successful.

2.4   SECOND CALL: APPLICANT SUMMARY

The second call received 19 applications with 15 being EC applicants and 4 non EU applicants.

Of the 15 EU applications there were 11 successful applicants. Of the 4 non EU applications there were 3 successful applicants.

There were 15 male applications with 10 being successful (7 EU and 3 non EU). There were 4 female applicants with all of them being successful (all EU).

GATE received 6 applications and 3 were successful; SoBigData.it received 9 applications with 7 being successful; Aalto received 3 applications with 3 successful, and Fraunhofer received one application, which was successful; L3S & ETH didn't receive any applications during Call 2.

2.5   THIRD CALL: APPLICANT SUMMARY

The third call received 19 applications with 12 being EC applicants and 7 non EU applicants.

Of the 12 EU applications there were 11 successful applicants. Of the 7 non EU applications there were 6 successful applicants.

There were 12 male applications with 11 being successful (8 EU and 3 non EU). There were 7 female applicants with 6 being successful (3 EU and 3 non EU).

GATE received 3 applications and all were successful; SoBigData.it received 9 applications with 8 being successful and one being deferred until some ethical issues have been addressed; L3S received 6 applications of which 5 were successful; Aalto received 1 application which was successful; ETH and Fraunhofer didn't receive any applications during this call.

In both the second and third calls we have tried to ensure a good gender balance, in line with SoBigData policies.

2.6   FINAL CALL: APPLICANT SUMMARY

The final call (an Open Call) received 34 applications with 18 being EC applicants and 16 being non EU applicants.

Of the 18 EU applications there were 17 successful applicants. Of the 16 non EU applications there were 7 successful applicants.

There were 22 male applications with 17 being successful (13 EU and 4 non EU). There were 12 female applicants with 7 being successful (4 EU and 3 non EU).

GATE received 13 applications and 7 were successful; SoBigData.it received 11 applications with 9 being successful; L3S received 5 applications of which all were successful; ETH received 3 applications, all of which were successful. Aalto and Fraunhofer did not receive any applications during this final call.

In line with SoBigData policy, we tried to ensure there was a good gender balance. Only 5 female applications were rejected, all of which were from outside the EU.

2.6.1 FINAL CALL - DATA

The host statistics for the final call can be found below, as you can see, the highest amount of applications were from Sheffield, UK (GATE) with 13 applications and Italy (SoBigData.it) with 11 applications. Of the GATE applications 7 were successful and of the SoBigData.it applications, 9 were successful. Table 3 also shows the total number of applications during the final call as well as the successful applications by country.

Host	No. of Applications in the Final Call	Percentage of All Applications in the Final Call
Fraunhofer	0	0%
GATE	13	38%
L3S	5	15%
SoBigData.it	11	32%
ETH	3	9%
UT	2	6%

Table 2: Total application figures by host

Host	Successful Applications in the Final Call	Percentage of All Successful Applications in the Final Call
Fraunhofer	0	0%
GATE	7	29%
L3S	5	21%
SoBigData.it	9	37.5%
ETH	3	12.5%
UT	0	0%

Table 3: Total successful application figures by host

Country Code	No. of Applications in the Final Call	No. of Successful Applications in the Final Call
DK	3	3
ES	1	1
HR	1	1
HU	1	1
IT	2	1
Non-European	16	7
NL	2	2
NO	2	2
PT	1	1
SI	1	1
SR	1	1
UK	3	3
Total	34	24

Table 4: Total Applications by Country

Country Code	No. of Female Applicants in the Final Call	No. of Successful Female Applicants in the Final Call
DK	1	1
ES	0	0
HR	1	1
HU	0	0
IT	0	0
Non-European	8	3
NL	0	0
NO	0	0
PT	0	0
SI	1	1
SR	0	0
UK	1	1
Total	12	7

Table 5: Total Female Applications by Country

The table below details all applications received during the final call. The successful applications are highlighted in green whilst the rejected ones are in grey.

	Project Name	Reviewer	Decision	Applicant Quality	Research Quality	Originality	Feasibility	Overall score
1	Using big data to measure the impact of Brexit on the mobility of researchers in the UK	Cristina Muntean	Accepted	4	4	3	4	4
2	High-order networks in sport analytics	Cristina Muntean	Accepted	5	5	4	5	5
3	Improved Explainable AI with Association Rule Mining	Cristina Muntean	Accepted	4	4	3	4	4
4	Contemporary Political Frontier: Analysing the Impact of Social media on General Elections of India, 2019 using Twitter Data	Rajesh Sharma	Rejected	3	1	2	?	1
5	Elections and Referendums: A network analysis approach	Cristina Muntean	Accepted	5	4	3	4	4
6	Gate (Text and Social Media Mining)	Kalina Bontcheva	Rejected					
7	Cross-lingual tweet content analysis in the Twitter Elections Integrity dataset	Kalina Bontcheva	Accepted					
8	Crowdsourcing Experiments on Toloka	Avishek Anand	Accepted	?	5	5	5	5
9	Keyword-Based Sentiment Analysis of Twitter Data on Violent Attacks in Nigeria using Recurrent Neural Networks.	Kalina Bontcheva	Rejected					
10	Quantifying the interaction of news and social media with the cryptocurrency market	Nino Antulov	Accepted	5	4	4	5	5
11	Gate (Text and Social Media Mining)	Kalina Bontcheva	Rejected					
12	Predicting bankruptcy of local government: a machine learning approach	Nino Antulov	Accepted	5	4	5	4	5
13	The relative influence of nodes with a low centrality in influence models	Cristina Muntean	Rejected	3	4	4	4	4
14	CitySensing - Big mobility and IoT data stream processing and analytics in Smart cities	Cristina Muntean	Accepted	4	3	3	4	4
15	Learning user specific network embedding for social media	Avishek Anand	Accepted	5	5	4	4	4
16	Predicting football transfers from player-club-agent collaboration networks using Supervised Dynamic Network Formation	Cristina Muntean	Accepted	4	4	3	3	4
17	Summarizing Social Media	Kalina Bontcheva	Rejected					
18	Sentiment Analysis for Twitter posts	Diana Maynard	Accepted	3	2	2	1	2
19	Scalable semi-supervised inductive network embeddings	Avishek Anand	Accepted	5	4	4	4	4
20	Mapping and Previewing trends in Regulatory Technology's function on social Media	Kalina Bontcheva	Rejected	1	1	1	1	1

#	Title	Name	Outcome					
21	Evaluation the human activity patterns in European cities	Cristina Muntean	Accepted	4	3	3	4	3
22	Designing a Model for automatic classical Arabic Part of Speech tagging based on hidden Markov model	Kalina Bontcheva	Rejected					
23	Mobility Assisted Community Service: a Data-Driven approach to Smart Service Design	Cristina Muntean	Accepted	5	4	4	4	4
24	Feedback effects in the experimental double auction with private information	Nino Antulov	Accepted	5	5	4	5	5
25	Content Verification and Linguistic-Based Approaches for Disinformation and Misinformation Distinction	Carolina Scarton	Accepted	4	4	3	3	4
26	On constructing corpus and machine learning models for classifying comments in social networks	Carolina Scarton	Accepted					
27	Leveraging Big Data for Improving Conversational Quality in Crowd-Robotics	Aviskek Anand	Accepted	3	3	4	3	3
28	Linguistic description for differentiation of disinformation and misinformation	Carolina Scarton	Accepted	3	4	4	4	4
29	Mul-Stance – Mul-Lingual Stance Classification	Kalina Bontcheva	Accepted					
30	The Impact of Misleading Information on the Multimodal-based Time Series Forecasting	Carolina Scarton	Accepted	4	4	5	3	4
31	Financial performance of luxury brands in the stock market: Do brand actions on digital social networks really care?	Anna Leontieva	Rejected					
32	Football transfer predicition using Ensemble Learning Methods	Cristina Muntean	Rejected	4	3	3	3	3
33	Understanding the Dynamics and Evolution of News Story Chains over Time using Natural Language Processing Techniques	Aviskek Anand	Accepted					
34	Mobility patterns from data	Cristina Muntean	Accepted	3	3	3	3	3

Table 6: *List of project applications for the final Call and their outcomes.*

2.6.2   CALL STATISTICS

The following statistics cover the whole period of the SoBigData project over the 52 months.

Overall the average number of successful applications is 72%, of these 34% were female candidates. Although this figure appears low, the total number of female applications throughout the project has in total been 31, this means that the percentage of female applications that were successful is actually 74% - higher than the average number of successful applications detailed above and also the percentage of successful male applications (71%).

The highest volume of applications within Europe came from Italy (38 applications with 30 being successful) and GATE (31 applications with 18 successful). The successful applications from Italy and GATE made up for 51% of the total successful applications. The total non-European applications accounted for 33% of the total applications and the success rate of these was 55%. The next highest number of applications from European countries came from the LS3 (11) and Aalto (6).

	Call 1	Call 2	Call 3	Final Call	Totals
Total no. of applications	22	19	19	34	94
EU applications	18	15	12	18	63
Non-EU applications	4	4	7	16	31
Male applications	14	15	12	22	63
Successful male applications	7	10	11	17	45
Female applications	8	4	7	12	31
Successful female applications	6	4	6	7	23

Table 7: *Number of applications received during all 4 calls and how many were successful*

	No. of Apps	No. of Successful Apps	No. of Apps	No. of Successful Apps	No. of Apps	No. of Successful Apps	No. of Apps	No. of Successful Apps	Total no. of Apps	Total no. of Successful Apps
GATE	9	5	6	3	3	3	13	7	31	18
SoBigData.it	9	6	9	7	9	8 (+1deferred)	11	9	38	30
Aalto	2	1	3	3	1	1	0	0	6	5
ETH Zurich	1	1	0	0	0	0	3	3	4	4
Frauhofer	1	0	1	1	0	0	0	0	2	1
L3S	0	0	0	0	6	5	5	5	11	10
UT	0	0	0	0	0	0	2	0	2	0
Totals	22	13	19	14	19	17	34	24	94	68

Table 8: *Number of applications in all 4 calls by host and how many were successful*

There were 24 visits hostedduring the final call. The average number of days that our applicants visited a host during this call was 25. In total, these visits made up 601 days which equates to 86 weeks of hosting in Call 4.

There were 96 weeks of hosting during the previous 3 calls so in total over the whole project there have been 182 weeks of hosted visits. The estimated number of weeks for the TNA visits was 122 – so this has been exceeded by 60 weeks – an increase of 67%.

The costs for Call 3 and the Final Call by partner so far are:

LUH	SoBigData.it	AALTO	USFD	Total
€ 8,613.15	€ 12,038.66	€ 3,507.07	€ 12,016.33	€ 36,175.20

CALL 4 EXPENSES BY PARTNER[2]				
LUH	SoBigData.it	AALTO	USFD	Total
€ 12,542.20	€ 16,018.99	€ 0	€ 8,821.10	€ 37,382.29

The budget to cover all TNA visits was €200,000.

The table below shows the costs incurred so far for hosting all visits. Not all expenses have been finalised so some figures will change. The revised figures will be delivered within the 45 days after finalisation of the project.

	Call 1	Call 2	Call 3	Final Call - provisional expenses	Provisional Totals
LUH	€ 0	€ 0	€ 8,613.15	€ 12,542.20	**€ 21,155.35**
SoBigData.it	€ 11,973.43	€ 6,971.80	€ 12,038.66	€ 16,018.99	**€ 47,002.88**
AALTO	€ 2,434.27	€ 6,008.75	€ 3,507.07	€ 0	**€ 11,950.09**
USFD - GATE	€ 7,845.28	€ 3,350.15	€ 12,016.33	€ 8,821.10	**€ 32,032.86**
Fraunhofer	€ 0	€ 2,227.16	€ 0	€ 0	**€ 2,227.16**
Provisional Totals	**€ 22,252.98**	**€ 18,557.87**	**€ 36,175.20**	**€ 37,382.29**	**€ 114,368.34**

Table 9: The costs incurred for hosting all visits

---

[2] Call 4 visitors figures are for the twenty four visitors – some expenses are not yet finalised

2.6.3 PARTICIPANT FEEDBACK

We have received positive feedback from the individuals who have undertaken visits. Below is a sample:

"This visit provides me deep insights into Big mobility data processing and analytics, as well as state-of-the-art methods, techniques and tools for mobility data analysis and visualization, particularly trajectory data clustering and outlier detection algorithms implemented on Edge and cloud infrastructure. I have acquired significant knowledge to identify different challenges in Big mobility data analytics, over both offline and streaming data, e.g. what are the main properties of a good real-life dataset, what are the problems with collecting the data, what features are important for various analytics scenarios, and what analytics algorithms can be successfully applied on Big mobility data in Smart City traffic control and m-health scenarios.

The participation in SoBigData project and access to its research infrastructure have improved my research expertise and experience in domains of mobile crowd sensing and Big mobility and IoT data processing, analysis and visualization for Smart health and Smart mobility in Smart Cities. I expect that it will result in further scientific achievements through publication of joint scientific papers, development of prototype/demo solutions and submission of joint project proposal".

"The stay resulted in a seminar held at the KDDLab at the University of Pisa, which resulted in invaluable feedback on my current work. Furthermore, research meetings with Professor Franco Turini, Anna Monreale, and Mattia Setzu has paved the way for further collaborations, and a trip to Aarhus University by Mattia Setzu has been suggested".

"Gave my first ever talk as a PhD student. Gained further insight to the data that was collected, prompting further questions for analysis".

"We have started three different projects on the proposed theme. The preliminary results are very promising and we hope to be able to turn some of them in a preprint soon. The visit has been useful to start a long-term collaboration between the two research groups and future visits are envisaged".

"We successfully applied the EVA algorithm to the Facebook data set, including using edge attributes (in this case the topics of the posts) to give the results more meaning. I also explored using different visualisation methods to better communicate the results to a wider audience. This involved working with other researchers in Pisa and using d3 for the first time".

"TNA helped me to use resources available at L3S Research centre for benchmarking and to collaborate with experts there. Gained the exposure of working in different setting\environment and to learn and Collaboration with L3S hosts was pretty fruitful in terms of brainstorming and strengthening the idea".

2.6.4 LEGAL AND ETHICAL ISSUES

Prior to the visits commencing, each successful project is reviewed by the host institution from a data protection, ethical, and legal perspective. Where ethical approval is required we have developed the use of an Ethical Board that screens all applications and gives necessary feedback for applicants to address. This is a change from the first call where visitors applied for approval through their home institution and ensures we can be sure the ethical procedures in place are uniform and robust.

Upon completion of this process, the visitor is requested to sign the formal agreement (see Appendix D), which sets out their rights and obligations formally.

Each host institution assists their visitors with finding accommodation, obtaining computer registration, and access to the data and algorithms. Each visitor is assigned at least one supervisor for the duration with their visit, with whom they meet regularly. Training materials and scientific papers are provided to the visitors in advance of commencing their STSM visit.

3   CONCLUSIONS

The project has been highly successful with positive feedback from researchers who undertook the visits. The collaboration and broadening of working relationships within the EU and further afield has been exceptional and will deepen as SoBigData++ continues from early next year for the following 4 years.

We are hoping that there will be even more female participation as the next project progresses and we will be looking to expand the participation of the hosts by increasing the number of visits to the hosts who didn't attract as many applications during this project.

Although some expenses remain outstanding, the costs incurred have fallen below expectations and are not expected to increase significantly from the provisional figures included in this report.

APPENDIX A. FINAL CALL FOR SOBIGDATA-FUNDED TRANSNATIONAL ACCESS

Final Call for

SoBigData-funded Transnational Access

The SoBigData project invites researchers and professionals to apply to participate in Short-Term Scientific Missions (STSMs) to carry forward their own big data projects. These opportunities are offered as part of SoBigData's Transnational Access (TNA) activities and applications can be submitted at any time.

You should submit your project application by the end of each calendar month to ta-admin@sobigdata.eu.

We welcome applications from individuals with a scientific interest, professionals, startups and innovators that may benefit from training in data science and social media analytics. In order to apply you have to fill the Project Application Form.

Applications will be reviewed within 2 months of submission. Please plan your visit start to be at least 3 months from the end of the month when you submitted your application.

Under no circumstances should you start your visit without an email confirming your application has been successful, that an ethical approval has been granted and a contract has been signed.

Funding for a short-term scientific mission (2 weeks to 2 months) is available up to 4500 euros per participant (to cover the cost of daily subsistence, accommodation, and economy flights). STSM bursaries are awarded on a competitive basis, according to the procedure described in the application pack and eligibility criteria below, and based upon the quality of the applicant, the scientific merit of the proposed project, and their personal statement.

Success rates from previous calls have been very high (65-75%). Applications from female scientists are particularly encouraged.

Visitors are welcome between 1 November 2018 and 31 July 2019 extended to 30 November 2019.

Pre-requisites for projects to carry out hosted research:

Good understanding of social data and, ideally, track record of prior social data analysis projects

Experience with using at least one of machine learning, natural language processing, and/or complex networks algorithms

Pre-requisites for projects to integrate new tools/datasets/services:

An already existing open-source tool for social media mining to be integrated

  OR

An already created openly licensed dataset of relevance to SoBigData, that can be integrated within the infrastructure

Please, note that the user group leader and the majority of the users must work in a country other than the country(ies) where the installation is located. This rule does not apply:

if the applicant is from an International organisation, the Joint Research Centre (JRC), an ERIC or similar legal entities;

in case of remote access to a set of installations located in different countries offering the same type of service

The goal is to provide researchers and professionals with access to big data computing platforms, big social data resources, and cutting-edge computational methods. STSM visitors will be able to:

- Interact with the local experts
- Discuss research questions
- Run experiments on non-public big social datasets and algorithms
- Present results at workshops/seminars

The STSM visits will enable multi-disciplinary social mining experiments with the SoBigData Research Infrastructure assets: big data sets, analytical tools, services and skills.

DESCRIPTION OF THE TNA OFFERED

The SoBigData RI manages vertical, thematic environments, called exploratories, on top of the SoBigData infrastructures, for performing cross-disciplinary social mining research. The Transnational Activities offered in this call will be for Short-Term Scientific Missions (STSM), between 2 weeks and 2 months.

Under this call, there will be two kinds of proposals funded: STSM research proposals and STSM tool/data integration proposals. Each kind is described in more details next.

Each STSM research proposal needs to be aligned with **one of the following exploratories** and also specify which of the centres listed above they wish to access:

CITY OF CITIZENS:

Participating Partners: **SoBigData.it, Aalto University**

Brief description: The exploratory is concerned with investigating city mobility. Our data scientists already study the traffic in the Italian cities of Pisa and Florence by analyzing Big Data sources such as mobile phone traces, veicular gps and social media data as proxy of human behaviour. New STSM proposals on complementary topics are welcome.

WELL-BEING AND ECONOMY:

Participating Partners: **UT, SoBigData.it, ETHZ**

Brief description: This exploratory uses data of purchases in supermarkets and investigates the changes in people's behavior after the economic crisis. This study allows to work out an early indicator of diseases. We also study the measurement of the real cost of life by studying the price variation. Furthermore we try to correlate people's well being with their social and mobility data. We also focus on systemic risk and credit risk measures and modeling using a complex network approach. Furthermore, we work on the topic of network reconstruction from partial information, focusing on social, economic and financial systems. New STSM proposals on complementary topics and/or methods are welcome.

SOCIETAL DEBATES, ONLINE MISINFORMATION, AND RUMOURS:

Participating Partners: **Gate, LUH, Fraunhofer IGD, Aalto University, SoBigData.it, ETHZ**

Brief description: By analysing discussions on social media and newspaper articles, this exploratory studies public debates to understand which are the most discussed topics, key opinions and stances expressed, and the characteristics of the various debate participants. By automatically analyzing and understanding text documents, we identify themes, follow the discussions around them, and track them through time and space.

MIGRATION STUDIES:

Participating Partners: **SoBigData.it**

Brief description: This exploratory will try to answer key questions around migration within Europe and worldwide, with particular focus on economic models of migration. Social media is being used as a key information source.

SPORTS ANALYTICS:

Participating Partners: **SoBigData.it, FRH**

Brief description: This exploratory aims at investigating new ways of measuring sports performance from Big Data sources, allowing for monitoring and possibly predict the activity of professional athletes. The exploratory also focuses on investigating the relation between sports performance and success, intended as both success of players, tactics and strategies in sports competitions and success of players and teams in terms of popularity and revenues.

Each STSM integration proposal can focus on the integration of already existing open source tools for social media mining or social datasets within the SoBigData infrastructure as a whole or via integration within one of the national facilities listed above. Applications focused on interoperability and integration with other European Research Infrastructures are also strongly encouraged. Applicants for integration proposals are strongly encouraged to contact informally their target host institution, to discuss and ensure technical feasibility of the proposal prior to submitting the application.

APPLICATIONS ARE INVITED AT THE FOLLOWING CENTRES (INFRASTRUCTURES):

GATE (TEXT AND SOCIAL MEDIA MINING), UNIVERSITY OF SHEFFIELD

The Natural Language Processing (NLP) group at the University of Sheffield is one of the largest and most successful research groups in text and social media mining in the EU. We develop and maintain the world-leading open-source GATE text and social media mining infrastructure (http://gate.ac.uk), its GATE Cloud deployment, and its vibrant user community.

Contact: Kalina Bontcheva k.bontcheva@sheffield.ac.uk

Location: SHEFFIELD, UNITED KINGDOM

SOBIGDATA.IT, PISA, ITALY

The European laboratory on Big Data Analytics and Social Mining (www.SoBigData.it) is aimed at pursuing interdisciplinary research initiatives connected to the impetus that "big data" and the ICT's are having on science, and the socio-economic sciences. Participating groups in the call are: Knowledge Discovery and Data Mining Lab (KDD – ISTI), 'Networked Multimedia Information System' Lab. (NeMIS - ISTI), High Performance Computing Lab (HPC– ISTI), the Web Applications for the Future Internet (WAFI-IIT), Ubiquitous Internet groups (UI-IIT), and the Acube Lab (Acube-UNIPI), Quantitative Finance group at the Scuola Normale Superiore (SNS), IMT School for Advanced Studies Lucca (NETWORKS Unit)

Contact: Roberto Trasarti roberto.trasarti@isti.cnr.it

Location: PISA, ITALY

FRAUNHOFER IGD, DARMSTADT, GERMANY

The Competence Center for Information Visualization and Visual Analytics (IVA) at Fraunhofer IGD is a world-leading research centre for the interactive visualization of big data. We offer access to visual analytics and information visualization technologies and methods for multidimensional data, visual text analysis, and we offer advice in visualization, and interaction design.

Contact: Thorsten May thorsten.may@igd.fraunhofer.de

Location: DARMSTADT, GERMANY

UT, UNIVERSITY OF TARTU, ESTONIA

UT brings in a curated and inter-linked dataset of Estonian e-government and e-health service descriptions, detailed statistics of usage of said services, and data related to societal and economic development in Estonia over the past decade. The dataset is complemented with relevant automated analysis methods.

Contact: Marlon Dumas marlon.dumas@ut.ee

Location: TALLIN, ESTONIA

L3S RESEARCH CENTER / LEIBNIZ UNIVERSITY HANNOVER

L3S provides access to innovative and cutting-edge datasets, methods and technologies for Web Science. The Alexandria infrastructure is based on a number of unique datasets like the German and UK Web Archives

which span around 20 years of Web history, ArchiveIT collections, etc.. Researches focus around Web Search (especially temporal and Entity-centric search), Web Information Management (including semantic technologies) and the Web of People (including Personalization and Social Web).

Contact: Avishek Anand anand@l3s.de

Location: HANNOVER, GERMANY

AALTO UNIVERSITY

The Data Mining group in ICS focuses on developing novel methods to extract knowledge from data, designing algorithms to summarize large volumes of data efficiently and effectively, and exploring new ways of using the extracted information. The research conducted in the Sociophysics Laboratory in BECS focuses on: (i) living and other complex systems, their measurement, analysis, modeling, understanding and control, (ii) detection of communities and social dynamics, with a focus on the dynamics of scientific interactions and human behavior in social and information systems.

Contact: Aristides Gionis aristides.gionis@aalto.fi

Location: AALTO, FINLAND

ETHZ

The Computational Social Science group at ETH Zurich aims to integrate social research by bringing modeling and computer simulation of social processes and phenomena together with related empirical, experimental, and data-driven work, while combining the perspectives of different scientific disciplines (e.g. computer science, socio-physics, social and complexity science). Big Data analytics, data-driven socio-systems, social mining, real-time data mining, the creation of self-organizing systems, innovation and the analysis of how science works, are core subjects of interest.

Contact: Nino Antulov-Fantulin nino.antulov@gess.ethz.ch

Location: ZURICH, SWITZERLAND

The goal is to provide researchers and professionals with access to big data computing platforms, big social data resources, and cutting-edge computational methods. STSM visitors will be able to:

Interact with the local experts

Discuss research questions

Run experiments on non-public big social datasets and algorithms

Present results at workshops/seminars

The STSM visits will enable multi-disciplinary social mining experiments with the SoBigData Research Infrastructure assets: big data sets, analytical tools, services and skills.

APPENDIX B. APPLICATION FORM

SoBigData Transnational Access Application form

Application form to be submitted to ta-admin@sobigdata.eu , accompanied by a 2 page CV as an appendix. Please make it all 1 PDF file.

Family Name		Nationality	
First Name		Birth year	
Gender (*tick the appropriated item*): ❑ female ❑ male			
Home Institution:			
Legal Status of Home Institution Code[3]		Home Institution Country Code[4]:	
Function / Job / Title:		Position Code[5]:	
Mailing Address:			
Phone (office):		Phone (mobile):	
Fax:		E-mail:	

Short Biography*(max 500 words)*

Please provide details of your academic qualifications and experience in your research area

---

[3]**UNI**=University, **RES**=Public Research Organisation, **SME**=Small or Medium Enterprise, **PRV**=Other and/or profit or not profit Private Organisation, **OTH**= Other Organisation

[4]**AL**=Albania, **AT**=Austria, **BE**=Belgium, **BG**=Bulgaria, **CH**=Switzerland, **CY**=Cyprus, **CZ**=Czech Republic, **DK**=Denmark, **EE**=Estonia, **FI**=Finland, **FR**=France, **DE**=Germany, **GR**=Greece, **HR**=Croatia, **HU**=Hungary, **IS**=Iceland, **IE**=Ireland, **IL**=Israel, **IT**=Italy, **LV**=Latvia, **LI**=Liechtenstein, **LT**=Lithuania, **LU**=Luxembourg, **MT**=Malta, **MK**=Macedonia, **ME**=Montenegro **NL**=Netherlands, **NO**=Norway, **PL**=Poland, **PT**=Portugal, **RO**=Romania, **SK**=Slovakia, **SI**=Slovenia, **ES**=Spain, **SE**=Sweden, **SR**=Serbia, **TR**=Turkey, **GB**=United Kingdom

[5]**UND**=Undergraduate, **PGR**=Post graduate (student with a first University degree or equivalent), **PDOC**=Post-doc researcher, **TEC**=Technician, **EXP**=Experienced researcher (professional researcher).

Note: Please supplementthis with a 2 page CV, as an appendix to your application.

Recent Publications (please list 5)

Web page and open code:

Please provide pointers to your web page and any open source code you have on github or elsewhere.

SoBigData Infrastructure access being applied for:
Name of the host institution:
Start date: Location:
Specific requests for preparation or training *(optional)*

NOTE: PROJECT TITLE, PROJECT OBJECTIVES and DESCRIPTION OF THE PLANNED WORK, *AND* REFERENCES are intended to provide information on the project that the applicants are proposing to work on during the trans-national access at the above named host institution.

Project title:
Project objectives (max 500 words)
Description of the planned work (max 2 pages, figures and tables included)
Personal statement: expected achievements and impact (max 500 words) *Outline the reasons why you are interested in the TNA, what you wish to gain from it, and how you expect it to contribute to your own research in the future.*
Dissemination plans (max 500 words) *Describe how you plan to bring lessons learned in the TNA to the attention of members of your project team, academic department or organization, or the research community at large. This could be in the form an informal talk, a brief report, a blog post, or a publication.*
Ethical self-reflection *Please provide some reflection on the ethical integrity and implications of your project. Describe the dataset(s) you are planning to use in your project, with particular emphasis on the ethical implication. Does it contain personal data? Is it public? Do you think that real identities of data subjects can be revealed by the access of this dataset or sensitive information can be inferred by the publication of your results?*
References

Main scientific field of the project

Give the particular field of research in your project, e.g., "Social network analysis", "Natural Language Processing", or "Machine Learning".

1. **Request for SoBigData sponsorship?**   ❑ Yes   ❑ No

2. **Any other EU project related to this proposal?**   ❑ Yes   ❑ No

In case of positive response, specify:

3. **How did you hear about SoBigData?**

        ❑ personal contact   ❑ web page   ❑ conference   ❑ twitter

        ❑ article ❑ other (*specify*)

I confirm that, I will send a Summary Report no later than 2 months after the end of the access visit, and that the results of the proposed project will be published specifying that the project leading to the publication has received funding from the European Union's Horizon 2020 programme: "*The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654024 (SoBigData)"*

Date:

_____

Signature:

_____

APPENDIX C. EVALUATION FORM

SoBigData Transnational Access: Proposal Evaluation form

Completed evaluation forms to be sent to ta-admin@sobigdata.eu AND K.Bontcheva@sheffield.ac.uk

Reviewer Name		Nation ality	
Gender (*tick the appropriated item*): say	q female q male q prefer not to		
Affiliation:			
Home Institution Country Code[1]		Job Title:	
E-mail:			

Project title:
Applicant Name:
Applicant's Institution:
Quality of the applicant: Please provide a brief summary on the quality of the applicant, based on their CV.

Applicant score (delete as appropriate): 1 – unsatisfactory; 2 – adequate; 3 – good; 4 – very good; 5 – excellent
Quality of the proposed research project: *Please comment.*   **Quality score (delete as appropriate): 1 – unsatisfactory; 2 – adequate; 3 – good; 4 – very good; 5 – excellent**
Originality of the proposed research project: *Please comment.*   **Originality score (delete as appropriate): 1 – unsatisfactory; 2 – adequate; 3 – good; 4 – very good; 5 – excellent**
Feasibility and Impact on SoBigData: *Discuss whether the proposal is feasible given the short duration of the visit. Also comment on how the proposed project would help enhance SoBigData's impact through joint publications, outreach to new countries/research groups, open source software, or any other planned contributions that can **realistically** be expected to arise.*   **Feasibility Score (delete as appropriate): 1 – unsatisfactory; 2 – adequate; 3 – good; 4 – very good; 5 – excellent**

Overall Score	**(delete as appropriate) 1 – unsatisfactory; 2 – adequate; 3 – good; 4 – very good; 5 – excellent**

In the event of equal scoring, weighting will be given to users who have not previously used the SoBigData infrastructure, to young researchers and to those working in countries where no research infrastructure is available. The selection panel shall also pay regard to issues of gender equality.

[1]**AL**=Albania, **AT**=Austria, **BE**=Belgium, **BG**=Bulgaria, **CH**=Switzerland, **CY**=Cyprus, **CZ**=Czech Republic, **DK**=Denmark, **EE**=Estonia, **FI**=Finland, **FR**=France, **DE**=Germany, **GR**=Greece, **HR**=Croatia, **HU**=Hungary, **IS**=Iceland, **IE**=Ireland, **IL**=Israel, **IT**=Italy, **LV**=Latvia, **LI**=Liechtenstein, **LT**=Lithuania, **LU**=Luxembourg, **MT**=Malta, **MK**=Macedonia, **ME**=Montenegro **NL**=Netherlands, **NO**=Norway, **PL**=Poland, **PT**=Portugal, **RO**=Romania, **SK**=Slovakia, **SI**=Slovenia, **ES**=Spain, **SE**=Sweden, **SR**=Serbia, **TR**=Turkey, **GB**=United Kingdom

APPENDIX D. AGREEMENT TEMPLATE

<table>
<tr><td>

The University Of Sheffield.

</td><td>

Department
Of
Computer
Science.

</td></tr>
</table>

29 gennaio 2020

Head of Department
Professor Guy Brown

Lucy Moffatt
Regent Court
211 Portobello
Sheffield
S1 4DP

Tel: +44 (0) 114 222 1901
Fax: +44 (0) 114 222 1810
Email: l.moffatt@sheffield.ac.uk

Dear USER NAME

INSERT PROJECT TITLE

The SoBigData Consortium are pleased to inform you that your project INSERT PROJECT TITLE (annex 1) has been accepted by the selection panel.

The duration of your visit to INSERT HOST is for INSERT DURATION, starting INSERT START DATE and ending INSERT END DATE.

In line with University financial regulations please refer to annex 2 for information with regards to expenditure and allowable costs. Please ensure that these guidelines are adhered to so we can ensure that your funding claim is eligible in line with EC and University legislation.

You may be required to receive Ethical Approval from the Host Institution. If so full details will be provided to you before your arrival.

In addition you will be required to sign a non-disclosure agreement which will be provided and signed by you on the day of arrival for your visit.

I should be grateful if you would confirm that you are happy to accept this offer on the above terms by signing and returning a copy of this letter to me at the above address.

Accepted on behalf of   Accepted by

SoBigData Consortium

Signature:

                                                   Signature:

Name: **Lucy Moffatt**   Name:

Date: Date:

APPENDIX E. FINANCIAL GUIDELINES

SoBigData – Trans-National Access

Travel Expense Guide for User Groups

Trans-National Access (TNA) is granted by seven national infrastructures:

Gate:	Kalina Bontcheva (k.bontcheva@sheffield.ac.uk)
SoBigData.it:	Roberto Trasarti (roberto.trasarti@isti.cnr.it)
Fraunhofer:	Thorsten May (thorsten.may@igd.fraunhofer.de)
UT:	Marlon Dumas (marlon.dumas@ut.ee)
L3S:	Avishek Anand (anand@l3s.de)
AALTO UNIVERSITY:	Aristides Gionis (aristides.gionis@aalto.fi)
Nervousnet:	Antulov-Fantulin Nino (nino.antulov@gess.ethz.ch)

Researchers should be advised that travel costs incurred while undertaking a Trans-National Activity must be made by the most economical method available.

- **Air Fares**
  - o   Should be Economy Class and be by the cheapest route where possible (whilst considering reasonable travel durations).

- **Rail Fares**
  - o   Standard Class Rail Fares or discounted rail should be used where possible and bookings should be made at least 7 days in advance of travel. First class travel will not be reimbursed.

- **Public transport**
  - o   Should be used where possible. Car mileage will normally be reimbursed to the equivalent value of a standard class rail fare.

- **Subsistence**
  - o   Will only be reimbursed for participants who require overnight accommodation (i.e. not for the host organisation). In such cases, subsistence we would expect claims at the following maximum rates: €7.50 for one meal, €12 for two meals and €22.00 for all meals.

- **Accommodation**
  - o   Should be a reasonable price for the location of the meeting and estimations for standard hotel prices (per person per night) are listed below.

Partner	Standard accommodation cost per night
GATE – Sheffield	€ 95

SoBigData.it – Italy	€105
Fraunhofer – Germany	€115
UT – Estonia	€ 75
L3S – Germany	€ 80
AALTO UNIVERSITY	€ 90
Nervousnet	€140

If you have any queries regarding costs of travel then please contact com-resadmin@sheffield.ac.uk

The maximum you can be refunded for your whole stay is €3,500. If you incur more than this then the funding provided by the TNA SoBigData Consortium will be capped.