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	**Sustainability Plan**
Deliverable No.	**D3.7**
Delivery Date	**24 December 2019**
Authors	**Valerio Grossi (CNR), Beatrice Rapisarda (CNR)**

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	48 months
Funding	H2020-INFRAIA-2014-2015
Grant Agreement No.	654024
DOCUMENT	
Deliverable No.	D3.7
Deliverable Title	Sustainability Plan
Contractual Delivery Date	30 Nov 2019
Actual Delivery Date	24 Dec 2019
Author(s)	Valerio Grossi (CNR), Beatrice Rapisarda (CNR)
Editor(s)	Beatrice Rapisarda (CNR)
Reviewer(s)	Marco Braghieri (KCL), Pasquale Pagano (CNR), Paolo Ferragina (UNIPI), Kalina Bontcheva (USFD)
Contributor(s)	Pasquale Pagano (CNR), Massimiliano Assante (CNR), Kalina Bontcheva (USFD), Paolo Ferragina (UNIPI), Antulov-Fantulin Nino (ETHZ), Avishek Anand (LUH), Jurek leonhardt (LUH), Marlon Dumans (UT), Rajesh Sharma (UT)
Work Package No.	WP3
Work Package Title	NA2_Dissemination, Impact, and Sustainability
Work Package Leader	USFD
Work Package Participants	CNR, USFD, UNIPI, FRH, UT, IMT, LUH, KCL, SNS, AALTO, ETHZ, TUDelft
Dissemination	Public
Nature	Report
Version / Revision	V1.1
Draft / Final	Final
Total No. Pages (including cover)	20
Keywords	Business Model, VRE, Support, Computational Resource

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
RI	Resource Infrastructure
VRE	Virtual Research Environment

TABLE OF CONTENT

DELIVERABLE SUMMARY

This deliverable reports on some sustainability actions related to the future management of SoBigData Research Infrastructure. Sustainability is an important factor in order to explain how a research infrastructure can become economically self-sustainable without funding from the European Union. This document focuses on sustainability proposing solutions at different levels: first hand showing how the entire RI can sustain itself as a whole; secondly outlining how some platforms which are already integrated into the Research Infrastructure could develop independent business models, contributing to the overall sustainability of the SoBigData Research Infrastructure.

This deliverable contains two main sections:

- Section 2 overviews strategies for governance and sustainability of SoBigData RI;

- Section 3 outlines some business-models that have been developed for specific platforms which are already integrated into the SoBigData Research Infrastructure and D4Science (which is the development software engine at the heart of the SoBigData E-Infrastructure).

EXECUTIVE SUMMARY

This report provides a detailed sustainability plan for the pan-European SoBigData Research Infrastructure. The purpose of this document is to describe sustainability activities, actions and events that have been undertaken in order to provide some guidelines about SoBigData's governance and long-term sustainability. Our description involves a detailed overview of the main steps in order to evaluate the impact of sustainability-related actions which have been undertaken in the SoBigData project.

1   RELEVANCE TO SOBIGDATA

One of the aims of the SoBigData project is to serve a large community of data scientists, researchers, and industrial stakeholders which base their needs on the storage, processing, analysis and searching of big data and social mining. In order to achieve this goal, we have set up and run a set of training activities, events and actions which have involved students, researchers and academics affiliated with the SoBigData partners and beyond. Moreover, we have benefited from the contribution of professionals, entrepreneurs and innovation leaders as teachers, coaches and mentors. Finally, in relation to Task "T3.3 Sustainability Planning", we complete the picture showing the main actions related to the sustainability of the research infrastructure.

1.1   PURPOSE OF THIS DOCUMENT

This document reports on activities related to "WP3 - NA2_Dissemination, Impact, and Sustainability" with special reference to "T3.3 Sustainability Planning".

T3.3 describes a sustainability plan and strategy for the future development of the project results. This document details on how the pan-European SoBigData Research Infrastructure and services will be sustained after the end of the project; how the collaboration has been enhanced and deepened; how project results are crucially embedded in all partner institutions. Furthermore, the task monitored that the development of services has been created on an open source basis and published on relevant platforms such as GitHub. Currently, it is possible to integrate a new method into the SoBigData E- Infrastructure, importing the code directly from GitHub (see deliverable "D7.3 VA e-Infrastructure Service Provision and Operation Report 3" which will report several details on the integration of new methods into the SoBigData E-Infrastructure).

The purpose of this document is to describe sustainability activities, actions and events that have been undertaken in order to provide some guidelines on the governance and sustainability of SoBigData Research Infrastructure (RI). Our description involves a detailed overview of the main steps undertaken in order to assess the impact of sustainability actions on SoBigData project and overviews some of the specific targets for the project. As far as the goals of SoBigData project are concerned, Task 3.3 has contributed to the following:

- studying different strategies for sustainability; including ESFRI and EUDAT29 governance and sustainability models;
- studying sustainability models already available inside the members of the consortium, and considering the integrated platform;
- supporting sustainability models related to the new-accepted H2020 project called SoBigData++ (grant agreement n. 871042);
- monitoring the accesses and the integration tools from relevant open source platforms.

1.2   RELEVANCE TO PROJECT OBJECTIVES

The SoBigData Research Infrastructure and tools allow significant innovation opportunities aimed at diverse stakeholders. Therefore, a careful planning of the sustainability actions is of the highest importance in order to provide a clear vision of the future of the SoBigData RI.

1.3   SOBIGDATA PROJECT DESCRIPTION

SoBigData RI serves the wide cross-disciplinary community of data scientists, researchers, and industrial stakeholders concerned with all aspects of societal complexity, both from a data-driven and a model-driven perspective.

The SoBigData RI is based on three pillars:

- an ever-growing, distributed data ecosystem for procurement, access and curation of big social data, to underpin social data mining research within an ethic-sensitive context;
- an ever-growing, distributed platform of interoperable, social data mining methods and associated skills: tools, methodologies and services for mining, analysing, and visualising complex and massive datasets, harnessing the techno-legal barriers to the ethically safe deployment of big data for social mining;
- a community of scientific, industrial, and other stakeholders (e.g. policy makers), supported by transnational and virtual access activities, and brought together by extensive dissemination, networking and innovation actions (in particular workshops, summer schools, datathons, training resources in social data mining, knowledge transfer, industrial partnerships).

1.4   RELATION TO OTHER WORKPACKAGES

This document belongs to the "WP3 - NA2_Dissemination, Impact, and Sustainability" work package related to dissemination activities, with special reference to task "T3.3 Sustainability_Planning". Therefore, there is a clear interaction with all other work packages related to:

- promoting events, supporting research and community, i.e. "WP4 – NA3_Training" and "WP11 - NA5_Evaluation" with special reference to tasks "T11.3 Sustainability of the framework and engagement of commercial providers" and "T11.4 Cross-disciplinary social mining exploratories";
- accelerating innovation through SoBigData services for Industrial stakeholders, i.e. "WP5 - NA4_Accelerating Innovation";
- promoting research infrastructure accesses, i.e. "WP6 - TA1_Transnational Access" and "WP7-VA1_Virtual Access";
- developing and maintaining the social mining ecosystem and the E-infra platform, i.e. "WP8 - JRA1_Big Data Ecosystem", "WP9 - JRA2_Integrating Big Data Analytics Methods and Techniques" and "WP10 - JRA3_SoBigData e-Infrastructure".

1.5   STRUCTURE OF THE DOCUMENT

This report contains two main sections:

- Section 2 provides an overview of the governance and sustainability of SoBigData RI;
- Section 3 outlines some business models related to other existing research infrastructure with special reference to D4Science that is the engine of SoBigData E-infra.

2   SOBIGDATA SUSTAINABILITY

The SoBigData RI is a community of data scientists originating from Mathematics and ICT experts with the goal of creating a "distributed, multidisciplinary European infrastructure on Big Data and social data mining". In regards to sustainability models, the majority of analysed platforms are based on similar business models, such as community based (member fees); pay-per-use (service charge) or public funding and most of them struggle to generate revenues as they are conceived by end-users as a "public good".

The governance and sustainability of SoBigData RI is strictly correlated to the forthcoming SoBigData++ project. SoBigData++ as an advanced community will strengthen services it offers, providing a wide range of new tools, widen its user-base toward new key areas (such as social science, data journalism, digital humanities, political science, demography, data-driven medicine), address scalability through European Open Science Cloud (EOSC) and High-performance computing (HPC) and sustainability. SoBigData++ will start on 01 January 2020 and its services will be based on the ones designed and deployed for SoBigData RI.

In this context, we can state that SoBigData has the three functional layers:

- *on site: training and hosting visitors*. This aspect is not only related to TransNational Access, but also to training events or specific community events such as datathons;
- *on line: e-infrastructure services*. This aspect is related to all the services required for providing Virtual Access to the RI resources, but also the online tools for community support;
- *research community facilities and consulting*. In this context, the RI supports and stimulates innovation; promoting its tools among in other communities and/or into companies as testified by the Tuscan Data challenge.

2.1   DESIGNING A GOVERNANCE FOR SUSTAINABILITY

In order to understand the sustainability plan for SoBigData, we can state that SoBigData++ will create a non-profit organisation named **SoBigData Association** and will study the viability and possible governance of a future **SoBigData Foundation**. More specifically, during the first year of the SoBigData++, the project will establish the SoBigData Association, which will include consortium members and additional experts and will be tasked with organising large community events. The association will act as the focal point of the SoBigData community, with members from interested organisations in Europe and beyond, especially targeting new entities which will not be part of the SoBigData++ consortium. In this context, the main objective is to create a *SoBigData Association* including all major actors in Social Mining research in Europe.

In the second part of the SoBigData++ project, the association will study how to transform into a different legal entity, such as a foundation or other legal entity forms that might be deemed fit for the purpose. This new entity will seek partners in the research community, private companies, public sector bodies, Non-Government Organisations, existing foundations and will include the core project partners acting as founders. The new entity will be tasked with ensuring the sustainability of the SoBigData RI and further propel its mission: to build, maintain and enhance an open, sustainable platform and ecosystem for ethical social data mining. In any case, CNR – D4Science will guarantee the platform's operability until two years after the end of the SoBigData project. In brief, we can state that a further step toward sustainability will be the creation of the *SoBigData Foundation* where leading institutions from the consortium will join forces to maintain and grow the platform in the future. This does not preclude the option of individual users joining

the SoBigData community and contributing on an *ad hoc* basis. This policy of openness and inclusiveness towards new contributors will thus enable the new RI to grow organically, well beyond the bounds of its initial user community.

2.2 SUSTAINABILITY THROUGH E-INFRA SERVICES

Another important way enhance SoBigData RI sustainability will rely on using the services related to E-Infrastructure. In this context, cloud applications can be viewed a distributed Software-as-a-Service, by automating several expensive, mission critical and time-consuming tasks. SoBigData E- Infrastructure will also exploit business models inhered from to the ones already integrated into RI and exposed in Section 3, where we will briefly outline the business model related to D4Science that represents the computational core of the E-Infrastructure of SoBigData.

Furthermore, the *SoBigData Foundation* capacity to attract private investments from commercial organisations will represent another possible form of long-term sustainability. In this context, the research infrastructure needs to support a federation of resources, providing an uniform standards-based transnational cloud access as a new capability to attract new research communities; support the routine operational needs of the infrastructure operations and maintenance of domain specific software environments. Finally, several incomes for maintaining the RI covering personnel and computational cost can derive from patent development which can also be done in collaboration with industrial stakeholders.

2.3 OTHER MODELS: ESFRI & EUDAT

Other RI-related governance and sustainability models include EUDAT (European Data Infrastructure) [4] and ESFRI roadmap (European Strategy Forum on Research Infrastructures) [5]. The main challenge is the development of a suitable funding model for an EU organisation, which is at the same time sustainable, equitable and realistic, and securing the necessary funding from Member States. Furthermore, one of the requirements is the design of an appropriate governance model, which provides sufficient level of independence to different entities involved and yet still ensures an integrated and effective management of the infrastructure as a whole.

General recommendations and minimal key requirements are available by ESFRI, based on the experience of 59 project and landmarks. With the contribution of University of Sheffield, we have extracted those that may be useful for the elaboration of a sustainability model for SoBigData RI. Some of the general recommendations and minimal key requirements have been deployed since the project's beginning, such as:

- *involvement of all stakeholders*, in particular funding organisations, from an early stage;
- *transnational access* is generally a good instrument to ensure the best use of available resources dependent on the specific situation of the provider and its services;
- a *centrally managed budget* for "supported access to services" added to the current Transnational Access model;
- *development of general model agreements*, statutes etc. in a standard form, with clauses on accountability, open access, governances, etc. that can be adapted to specific situations would greatly facilitate the process;
- establishment of a high-level *International Advisory Board*;

- introduction of *ethical* and *legal expert* boards to reduce legal and other barriers and provide cross-border access to the research infrastructure.
- development and preparation of an *integrated investment* and *business plan*, showing the required investments over the next 5 to 10 years and the way this will be financed (a clear investment plan can reduce the time needed to convince funders).

3   BUSINESS MODELS FROM THE CONSORTIUM

This section reports some sustainability and business models already available from the partners of the consortium. Furthermore, the section recalls and updates some information related to the services provided by the different institutions and fully reported into deliverable "D11.4 SoBigData Evaluation Framework: Engagement and Sustainability Report"[1], with the aim of showing how the different services can be used for both *i)* maintain the service itself and *ii)* contribute to the sustainability of SoBigData Research Infrastructure.

3.1   GATE PLATFORM BY USFD

Currently, GATE Platform, which is part of the SoBigData Research Infrastructure, is funded partly by research grants and partly by commercial sponsorship. GATE proposes a specific model for business. If a company wants to build products that need to analyse language, the University of Sheffield and its commercial partners offer a complete set of software and services for all budgets; from zero costs community-supported open-source solutions to enterprise packages with 24/7 professional support, on-site training and development. The business model is based on zero fixed costs, as a company does not buy software licences or server hardware, just pays for the compute time that it uses.

As reported in the GATE Platform main site different kind of services are defined [2]. Three types of solutions are available accordingly to the tasks the customer needs:

- semi-automatic workflow based using GATE Teamware;
- fully-automatic GATE-based extraction (accessed via parallel distributed web services or tightly coupled with your application as appropriate);
- a wide range of other bespoke language processing systems.

These three solutions involve different costs and support. For example, a fully-automatic GATE-base extraction requires three stages:

1. task analysis and prototype specification (which typically requires less than a person-week of time and involves discussing the business case with the client, assessing if client requirements are feasible using GATE, and how much resource a full-blown solution would require);
2. development of a working prototype to serve as proof-of-concept (typically several weeks of developer time done over the course of a month or 6 weeks, plus several days of training time for the client's organisation);
3. deployment and maintenance of the working system (typically in-house at the client's organisation, with ongoing support as required).

3.2   CNR, UNIPI, IMT & SNS SERVICES

The Italian node of SoBigData provides different services and application, here we report the main ones with specific update for sustainability:

- QuickRank is an efficient Learning-to-Rank toolkit providing several C++ implementations of LtR algorithms. QuickRank is designed and developed with efficiency in mind provided by HPC Lab at ISTI-CNR;

- Twitter Monitor is an interactive Web application designed to access the Twitter stream by exploiting the public TwitterStreaming APIs provided by WAFI at IIT-CNR;
- Human Mobility tools & Community Discovery algorithms provided by KDD Lab at ISTI-CNR;
- TagMe, WAT & SWAT for end-to-end entity linking by annotating a natural language text developed by A3 Lab at UNIPI;
- IMT & SNS offer methods concerning network science and time series analysis with tailored to interested stakeholders in economics, finance and policy makers.

All these services are fully integrated into SoBigData E-Infrastructure. In particular, TagMe services received support in terms of awards, grants from companies, and industrial papers that mention the use of the TagMe suite. The business model for these applications can be defined as Freemium[1]. Each basic service is provided at no cost for the users, but it can also offer more advanced services (or additional features) at a premium with specific costs. It is important to highlight that this business model deployment manages to cover all the personnel costs to update, operate and maintain the service.

3.3   ETHZ SERVICES

The services offered from ETHZ and integrated to Research Infrastructure are:

- NervousNet: an open, privacy-preserving and participatory platform designed to be collectively built by citizens and for citizens.
- Innovation accelerator: an open and editable database for metadata publications, that enables monitoring and analyzing social processes around science innovation by visualizing social relations and communities in the network of (scientific) publications.

The sustainability of ETHZ services in the next years is driven by the financial support of the Computational Social Science Group, led by Professor Dirk Helbing. The Computational Social Science Group budget is financed through ETHZ Institution, Swiss National Foundation and EU grants.

3.4   L3S RESEARCH CENTER SERVICES AT LUH

L3S Research center at LUH has updated and integrated the following services:

- BoilerNet service in the list; a deep learning based tool for web content extraction. The source code in now on GitHub now (available here: https://github.com/mrjleo/boilernet). The repository includes usage instructions as well as a proof-of-concept Chrome browser extension for demonstration purposes (at the present stage this extension is still being developed);
- ArchiveSpark: Processing archive collections using Apache Spark framework for easy data processing, extraction as well as derivation for archival collections. Originally developed to be used with Web archives, it has now been extended to support any archival dataset through Data Specifications.

L3S provides these services as free to use in terms of trained models (for boilerNet). Currently, the L3S services serve as common resource across multiple projects and are hence sustainable. However, if the code was to used for commercial purpose then there is potential to monetize it under the using Apache licences.

---

[1] a combination of the words free and premium. The freemium is business model that involves offering customers both complementary and extra-cost services.

3.5   UNIVERSITY OF TARTU SERVICES

The University of Tartu offers the following main services:

- Training on business data analytics. University of Tartu has developed a course on business data analytics, which has been delivered both in a 32-hour and in a 64-hours format, as well as in a 16-hours format, suitable for corporate training.
- Consultancy services on use of social media datasets in business data analytics. Tartu University has built the social media analysis datasets, particularly those available in the SoBigData infrastructure. By combining these sources with existing expertise in business data analytics and social network analysis, Tartu University is in a position to offer consultancy services to companies seeking to extend their internal business data analytics pipelines with external datasets.

The training packages on business data analytics developed as part of SoBigData are going to become a core part of University of Tartu's Masters of Data Science starting from September 2020. This MA program has a strong professional orientation. It was created with a two-years grant from the Estonian Ministry of Economy. The latter grant will fund the transfer of the SoBigData training package into the format required by the Data Science MA. Subsequently, the MA programme will sustain itself via tuition fees (particularly for students coming from companies) and base teaching funding. Additionally, the training material will remain available for corporate training to companies. These arrangements will ensure the sustainability of this output for the next 10 to 15 years.

The consultancy services will continue to be provided in cooperation with the Estonian Competence Centre in Data Science (STACC). Tartu University already supporting STACC in the provisioning of data science services, particularly in the healthcare and government sectors, using expertise accumulated during the SoBigData project. This collaboration will continue for several years to come.

3.6   D4SCIENCE SUSTAINABILITY

D4Science proposes different terms of use and costs based on the needs of the final users. More than a single user, we can state that the model is based on the different types of user groups. In this context, it is relevant to stress that the services related to D4Science are based on the concept of Virtual Research Environment. Each VRE can have different access policies that can be selected at any time during the operation of the VRE:

- *private access*: the VRE is private and a user can access it only by invitation issued by the VRE Manager. Upon acceptance of the user of the invitation, the user becomes member of the VRE with a user role;
- *restricted access*: a user registered to the gateway and with a valid identity can request access to the VRE. The VRE Manager can approve or reject any user request. In case of approval, the user becomes member of the VRE with a user role;
- *public access*: a user registered to the gateway and with a valid identity can become member of the VRE by simply accessing it. The VRE Manager is just notified.

Furthermore, four different roles are supported by default and additional ones can be defined as per request of the client:

-   *VRE Manager*: this kind of user manages user registration/deregistration to the VRE and assign/remove roles by accessing the Administration dashboard;
-   *user*: any user with this role can access the VRE data and share private data with other selected members of the VRE;
-   *processor* (Optional): any user with this role can register its own process and execute a process by accessing the Method Engine service;
-   *editor*: the users with this role can publish data to all members of the VRE by exploiting the Catalogue service.

The following three subsection outlines the three main classes of services D4Science proposes.

3.6.1   AS-A-USER

This model is conceived for a single user requiring immediate access to data, tailored applications, wanting to use a service available in the E-Infrastructure. She/he can follow the activities performed by the other users of a large community and join one of the existing applications offered by the existing VREs. Furthermore, each single user has a workspace where she/he can store, access, and optionally share files and datasets.

This service is provided for free.

3.6.2   AS-A-GROUP

This class of services is conceived for a small set of users focusing on a specific scientific topic requiring a common infrastructure to store, maintain, and process data. In this context, the group needs to design a specific new VRE in order to create a dedicated environment with specific applications and to manage the users by authorizing them to join.

This service is for free **without installing** any custom and private service.

3.6.3   AS-A-COMMUNITY

This class of services is designed in the presence of medium/large set of users focusing on many scientific topics ranging from the analysis of statistical and biodiversity data to the management of geo-referenced data. In this context, the community needs a common infrastructure to store, maintain, and process data through the creation of focused applications, each of which is focusing on a specific scientific goal. The community will be set-up in order to offer members an integrated storage and computational platform for the execution of daily tasks.

A user may exploit any of the available capabilities and may add specific applications and data. Data and applications are kept private to your community and confidentiality and security are guaranteed by encryption.

The cost of this agreement has to be negotiated, and varies according to the number of users partaking in the community and the desired service quality. The expected actual costs are based on:

-   *computational resources*: a set of computational resources will be reserved for this each community supported. In particular, a minimum equivalent Amazon m3.xlarge – 15 GiB of

memory, 13 EC2 Compute Units EBS storage only, 64-bit platform - is assigned to the community. This minimum configuration can grow up to 2 m3.2xlarge – 30 GiB of memory, 26 EC2 Compute Units. In any case, additional resources can be assigned on-demand to scale up distributed computations and up to the negotiated quota as established in the negotiated phase between the D4Science and the customer;

- *deployment bundles*: different deployment and configuration schemas are defined based on the bundle of services requested by the customers. Currently, three main bundles are available. Each bundle has different costs based on the *number of users* to be supported and *quality of services* and include the following basic services: VRE Management, Messaging, notification and Social Networking; Workspace quota; Gateway support for min number of users, logos, template and so on; Catalogue; the method engine; the support to geospatial data.

It is important to highlight that the cost is connected to the operation and maintainment of the VRE and to provide support upon request. It has been calculated as a reimbursement of the personnel cost needed to maintain the service operational, estimated in 5 working days per month of a skilled senior engineer, while the operation and maintenance and operation costs are typically classified into routine and periodic costs. Routine maintenance costs involve small tasks that are undertaken frequently, while periodic maintenance costs are related to works undertaken at intervals of either several weeks or months.

Finally, the business model includes also costs for:

- *training personnel for using D4Science services*. Training modules include i.e. *VRE Management and Operation*, a half day training module covering the following services: VRE Management Service, Messaging Service, Social Networking Service, Notification Service, Gateway Service; *Catalogue Service*, a half day training module covering the Catalogue Service, and/or *Data Analytics Platform and SDI that requires* a two full-day training module covering both the Data Analytics Platform and the Spatial Data Infrastructure. Each module has different costs based on topic and duration and, the costumer can host training modules upon compensation of the travel cost;
- *support*. This service is provided via D4Science.org Service. Malfunctioning, are handled according to a subscribed "Quality of Support" agreement that is estimated according to the impact of the outage or service quality degradation. The actual supporting service is typically included into deploying scheme.

4   CONCLUSIONS

This deliverable reported all the different actions taken into account in order to guarantee the sustainability of SoBigData RI.

Section 2 has described the sustainability actions related to the whole infrastructure considering several general aspects and not only the ones related to E-Infrastructure services. The Section 2 reported also consideration on accessing and to be part of grant roadmaps available in the context of European Community. Section 3 has shown several examples of applications integrated into SoBigData RI that can be used for supporting sustainability, furthermore it outlined the sustainability model related to D4Science Infrastructure that is the engine of SoBigData E-infrastructure.

Under this view, the report proposed different perspective of sustainability related both to short and long term. First the sustainability on short term is guaranteed by the D4Science terms of use, that provides computational support for free for two years and provides full supports to scientists, research communities in according to D4Science policies.

Secondly, the sustainability on long term is related to the creation of SoBigData Association during the first year of the forthcoming SoBigData++ Project. Moreover, SoBigData++ will evaluate the creation of a foundation in order to manage all the aspect inside the research infrastructure. Finally, also the applications integrated inside the SoBigData E-infrastructure can contribute to support also RI services. Part of the costs get from a customer/user can be reserved to RI sustainability for both short and long terms.

REFERENCES

[1] N. Antulov-Fantulin and T. Guo, "D11.4 - SoBigData Evaluation Framework: Engagement and Sustainability Report", https://data.d4science.net/jhmW, 2019.

[2] University of Sheffield - Gate Cloud, https://gate.ac.uk/business.html, 2019.

[3] D4Science Infrastructure - https://www.d4science.org/exploitation-models, 2019.

[4] EUDAT - Research Data Services, Expertise & Technology, https://eudat.eu/, 2019.

[5] ESFRI, the European Strategy Forum on Research Infrastructures, https://www.esfri.eu/, 2019.