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	***VRE specification and software 1***
Deliverable No.	***D10.12***
Delivery Date	***31 August 2016***
Authors	***Massimiliano Assante (CNR), Leonardo Candela (CNR), Luca Frosini (CNR), Lucio Lelii (CNR), Francesco Mangiacrapa (CNR), Pasquale Pagano (CNR)***

SoBigData receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654024

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.	D10.12
Deliverable Title	VRE specification and software 1
Contractual Delivery Date	31 August 2016
Actual Delivery Date	08 September 2016
Author(s)	Massimiliano Assante (CNR), Leonardo Candela (CNR), Luca Frosini (CNR), Lucio Lelii (CNR), Francesco Mangiacrapa (CNR), Pasquale Pagano (CNR)
Editor(s)	Leonardo Candela (CNR)
Reviewer(s)	Valerio Grossi (CNR), Paolo Manghi (CNR), Pasquale Pagano (CNR)
Contributor(s)	Valerio Grossi (CNR)
Work Package No.	WP10
Work Package Title	JRA3_SoBigData e-Infrastructure
Work Package Leader	CNR
Work Package Participants	USFD, UNIPI, FRH, UT, IMT, LUH, KCL, SNS, AALTO, ETHZ
Dissemination	Public
Nature	Other
Version / Revision	1.0
Draft / Final	Final
Total No. Pages (including cover)	21
Keywords	Virtual Research Environment, Resource Catalogue

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
Research Infrastructure	Facilities, resources and services that are used by a research community to conduct research and foster innovation in their fields. Include: major scientific equipment (or sets of instruments), knowledge-based resources such as collections, archives and scientific data, e-infrastructures, such as data and computing systems and communication networks and any other tools that are essential to achieve excellence in research and innovation. They may be 'single-sited', 'virtual' and 'distributed'.
RI	Research Infrastructure
VA	Virtual Access
Virtual Access	**Open and free access** through communication networks to resources needed for research, without selecting the researchers to whom access is provided.
Virtual Research Environment	Innovative, web-based, community-oriented, comprehensive, flexible, and secure working environments conceived to serve the needs of modern science.
VRE	Virtual Research Environment

TABLE OF CONTENT

DELIVERABLE SUMMARY

This deliverable complements “D10.5 SoBigData e-Infrastructure software release 1” by describing how such a software has been deployed to serve the current needs of the SoBigData community. In particular, it describes how such a software has been exploited to make available the components envisaged in “D10.2 SoBigData e-Infrastructure release plan 1”, i.e. the SoBigData portal (and the underlying Virtual Organisation), the SoBigData Catalogue, and the SoBigData Virtual Research Environments.

EXECUTIVE SUMMARY

This deliverable describes how the software documented by "D10.5 SoBigData e-Infrastructure software release 1" has been exploited to make available the components envisaged in "D10.2 SoBigData e-Infrastructure release plan 1", i.e. the SoBigData portal (and the underlying Virtual Organisation), the SoBigData Catalogue, and the SoBigData Virtual Research Environments. It is just a placeholder of the actual deliverable that being of type "other" has been implemented via a set of web sites and pages: (i) the SoBigData portal https://sobigdata.d4science.org/ representing the gateway for accessing and exploiting the SoBigData infrastructure and its services, (ii) the SoBigData.eu Catalogue available at https://ckan-sobigdata.d4science.org/ for searching and browsing the SoBigData.eu published products, and (iii) the Virtual Research Environments, namely (a) SoBigData.eu VRE – conceived to provide the SoBigData.eu project members with a VRE-based working environment, (b) SoBigData.it VRE – conceived to provide the SoBigData.it initiative members with a VRE-based working environment, (c) Resource Catalogue VRE – conceived to enable SoBigData.eu members to populate the catalogue by registering products (datasets and methods) worth sharing in the community, and (d) TagMe VRE – conceived to provide its users with an environment for exploiting the TagMe facility.

1 INTRODUCTION

This document accompanies the actual D10.12 "VRE specification and software 1" that being of type "other" is implemented by a series of web sites and pages. In practice the deliverable consists in documenting how the software released and described in D10.5 "SoBigData e-Infrastructure software release 1" [1] has been exploited to provide the SoBigData community with a set of facilities and Virtual Research Environments.

The enabling software is described in D10.5 [1], it is part of the gCube 4.0 release[1] and includes the following major subsystems:

- *Resource Catalogue*: it is a set of portlets and software libraries to discover, index and search all the datasets accessible through the D4Science infrastructure. It is based on CKAN technology (http://ckan.org);
- *Portal*: it is a set of software libraries to realise a Liferay based web-portal customized to interface with the D4Science infrastructure and equipped with gCube portlets. Once installed and configured it will act as the "one stop shop" for the entire SoBigData e-Infrastructure. Through it users will have access to the resources and Virtual Research Environments created to serve the needs of the SoBigData.eu community and scenarios;
- *Accounting Framework*: a set of services and portlets supporting the collection and consumption of resources usage metrics, i.e. it is a distributed system comprising an array of services automatically collecting per-resource usage metrics, integrated into the gCube SmartGears container that enables automatic accounting of user calls, plus a user interface for their visualisation and analysis;
- *VO Management Framework*: it is a set of basic services and portlets created and operated in the context of D4Science to serve the needs of SoBigData. This Virtual Organisation (VO) realizes the actual operational context for implementing and operating the SoBigData e-Infrastructure and its resources in autonomy with respect to the other communities and initiatives supported by D4Science;
- *VRE Management Framework*: it is a set of services and portlets supporting the creation and operation of VREs.

This software has been deployed and configured leading to three major services:

- the SoBigData.eu portal (cf. Sec. 2);
- the SoBigData.eu Catalogue (cf. Sec. 3),
- the initial set of SoBigData.eu Virtual Research Environments (cf. Sec. 4).

[1] https://github.com/gcube-system/gcube-releases/tree/v4.0.0

2 SOBIGDATA.EU PORTAL

This is expected to be the end-user access point to the SoBigData.eu services and Virtual Research Environments. It is available at https://sobigdata.d4science.org/. A screenshot of the home page is in Figure 1.

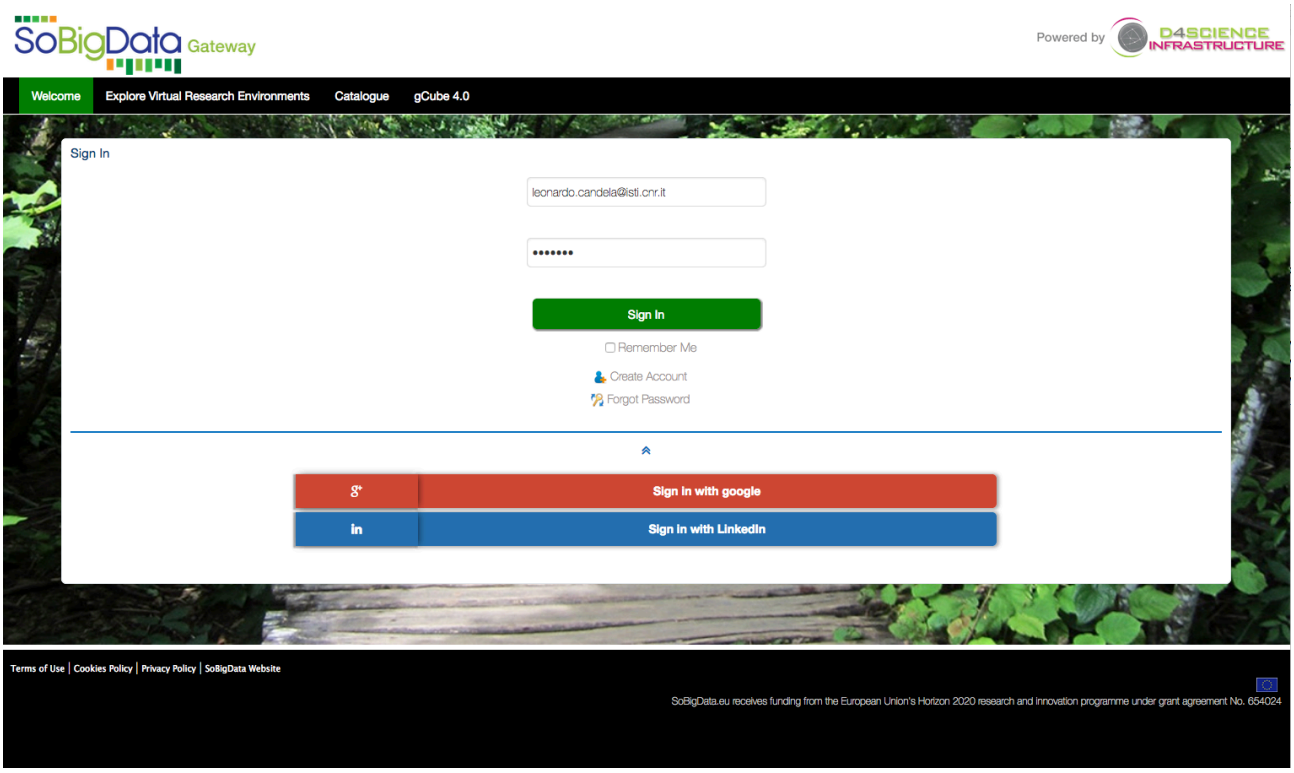

Figure 1. SoBigData Portal Home Page

Users are provided with log-in facilities and once logged in they are provided with a dashboard (Figure 2) realising a user friendly working environment.

Figure 2. SoBigData.eu Portal user dashboard

The dashboard provides users with:

- Access to all the VREs a user is member of;
- A news feed reporting all the posts and discussions occurring in the VREs a user is member of;

- A shared workspace for managing the objects shared in the VREs a user is member of as well as for managing his/her own objects organising them in folders;
- A communication area for exchanging private messages with other users and co-workers.

3 SOBIGDATA CATALOGUE

This is a portal for publishing, sharing and finding SoBigData.eu products, namely datasets and methods worth sharing in the SoBigData.eu community. It is It is available at https://sobigdata.d4science.org/catalogue as well as at https://ckan-sobigdata.d4science.org/. A screenshot of the home page is in Figure 3.

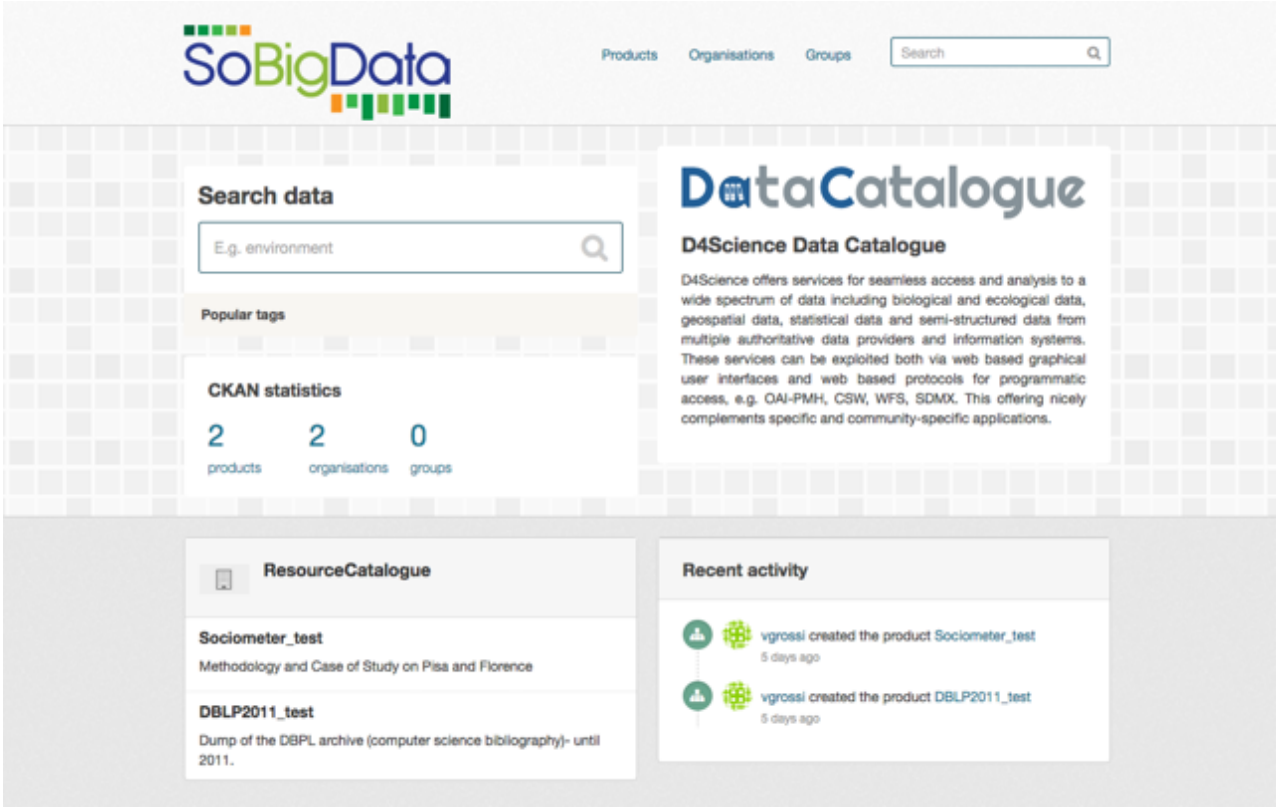

Figure 3. SoBigData Catalogue Home Page

Such catalogue will be populated by relying on the SoBigData.eu Resource Catalogue (cf. Sec. 4.3), i.e. a Virtual Research Environment specifically conceived to provide its users with facilities for creating, publishing and modifying product records, i.e. datasets and methods.

The catalogue is conceived to be nicely integrated in the overall SoBigData.eu Portal (cf. Sec. 2). Furthermore, it will be public and on-line searchable from the main site www.sobigdata.eu. With this integration in the website we are actually realising an enhanced version of what was documented by "D8.1 Data Management Report" and implemented by a wiki. The publication and modification of each product in the catalogue is allowed by the dedicated VRE only (cf. Sec. 4.3).

4 SOBIGDATA VIRTUAL RESEARCH ENVIRONMENTS

At the time of writing this document four Virtual Research Environments have been created and are operational, i.e. SoBigData.eu VRE (cf. Sec. 4.1), SoBigData.it VRE (cf. Sec. 4.2), Resource Catalogue VRE (cf. Sec. 4.3), and TagMe VRE (cf. Sec. 4.4).

4.1 SOBIGDATA.EU VIRTUAL RESEARCH ENVIRONMENT

This Virtual Research Environment is conceived to provide the SoBigData.eu project members with a VRE-based working environment.

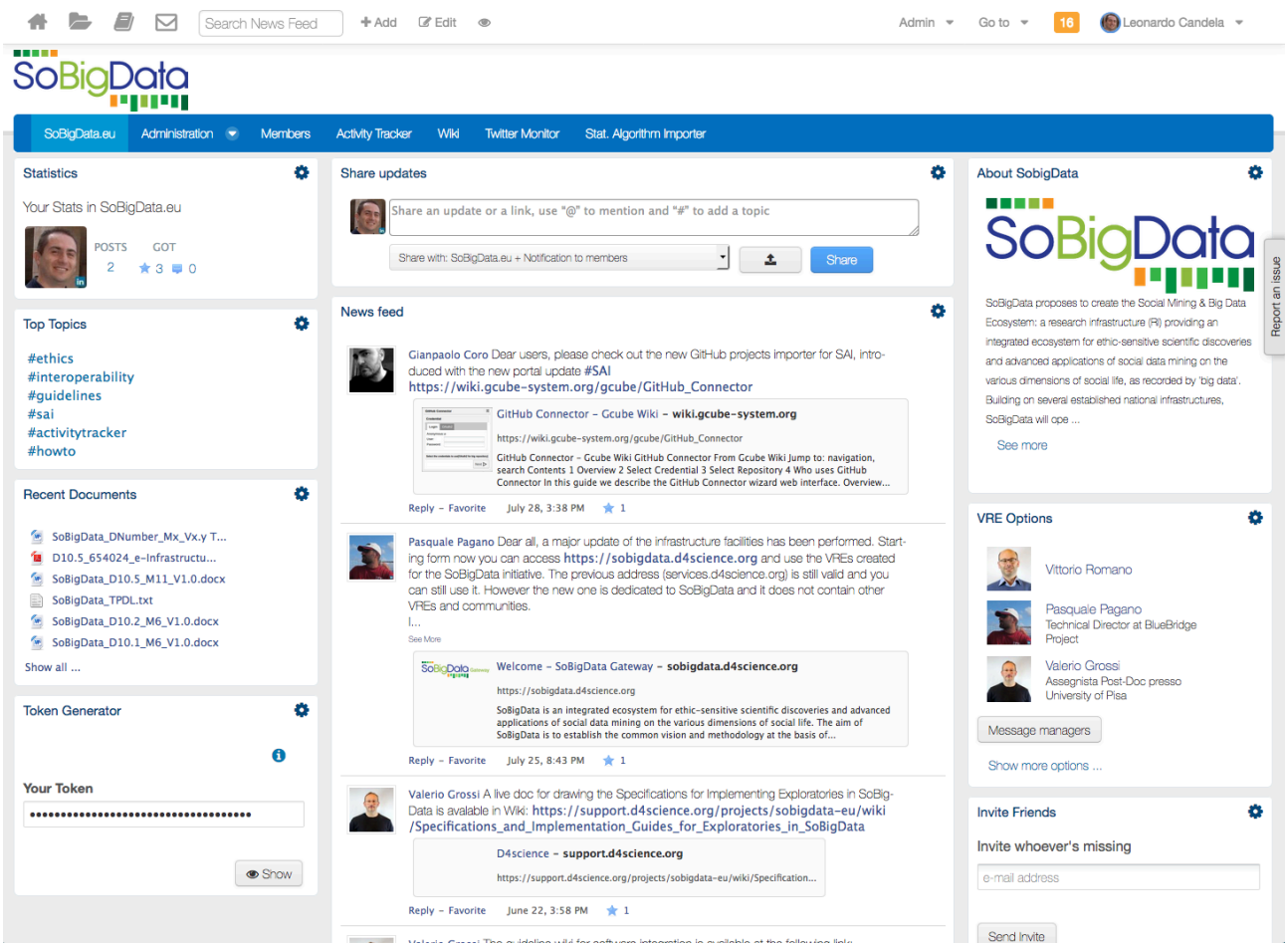

Figure 4. SoBigData.eu VRE Home Page

The main functionality offered by the VRE include:

- A dashboard (cf. Figure 4) for (i) collaborating with other VRE members by posting messages or being informed and reacting to (e.g. commenting, favouring) co-workers' posts, (ii) being informed on the most recent objects added to the VRE workspace, (iii) invite colleagues to participate, (iv) have information on the VRE including a description and the list of managers, (v) having some indicators on the user activity in the VRE, and (vi) acquire an authorization token for exploiting the services offered by the VRE in a programmatic way;

- A Wiki for collaboratively documenting project related activities;
- An issue tracking system (or ticketing system), to support the planning and monitoring of project related activities including work packages, tasks, deliverables and milestones as well as technology development tasks;
- A shared area in the workspace, for making available objects of interests, e.g. project deliverables, presentations, working notes;
- A Members area, for enabling each VRE member to be informed on the rest of VRE members and acquire details for contacting them;
- Twitter Streaming Monitor, a community tool integrated in the SoBigData.eu infrastructure. It is a tool for gathering data from Twitter in a focused way;
- Algorithm Importer, a tool enabling the integration of existing algorithms (e.g. R scripts) in the SoBigData.eu infrastructure;
- A user management area, to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers can (i) authorise users in accessing the VRE and its services, (ii) assign or withdraw roles to users, (iii) remove users, and (iv) send a communication to the current users;
- An accounting area, to enable authorised users (e.g. VRE Managers) to analyse usage records pertaining VRE services, e.g. most used service, most active user.

4.2 SOBIGDATA.IT VIRTUAL RESEARCH ENVIRONMENT

This Virtual Research Environment is conceived to provide the SoBigData.it initiative members with a VRE-based working environment.

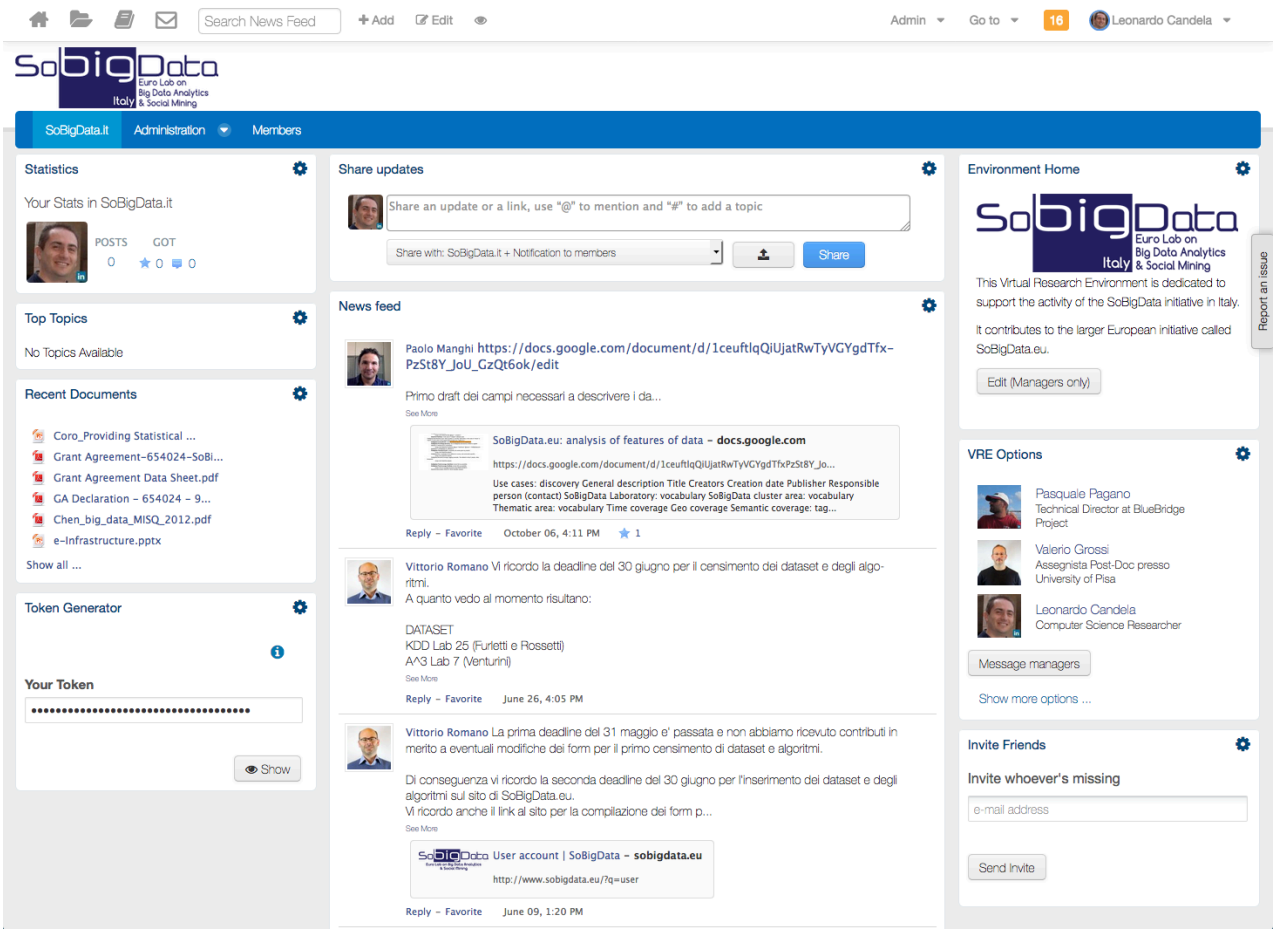

Figure 5. SoBigData.it VRE Home Page

The main functionality offered by the VRE include:

- A dashboard (cf. Figure 5) for (i) collaborating with other VRE members by posting messages or being informed and reacting to (e.g. commenting, favouring) co-workers' posts, (ii) being informed on the most recent objects added to the VRE workspace, (iii) invite colleagues to participate, (iv) have information on the VRE including a description and the list of managers, (v) having some indicators on the user activity in the VRE, and (vi) acquire an authorization token for exploiting the services offered by the VRE in a programmatic way;
- A shared area in the workspace, for making available objects of interests, e.g. documents, working notes, datasets, scripts;
- A Members area, for enabling each VRE member to be informed on the rest of VRE members and acquire details for contacting them;
- A user management area, to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers can (i) authorise users in accessing the VRE and its

services, (ii) assign or withdraw roles to users, (iii) remove users, and (iv) send a communication to the current users;

- An accounting area, to enable authorised users (e.g. VRE Managers) to analyse usage records pertaining VRE services, e.g. most used service, most active user.

4.3 SOBIGDATA RESOURCE CATALOGUE

This Virtual Research Environment is conceived to provide the SoBigData.eu members with a VRE-based working environment for populating the Resource Catalogue (cf. Sec. 3).

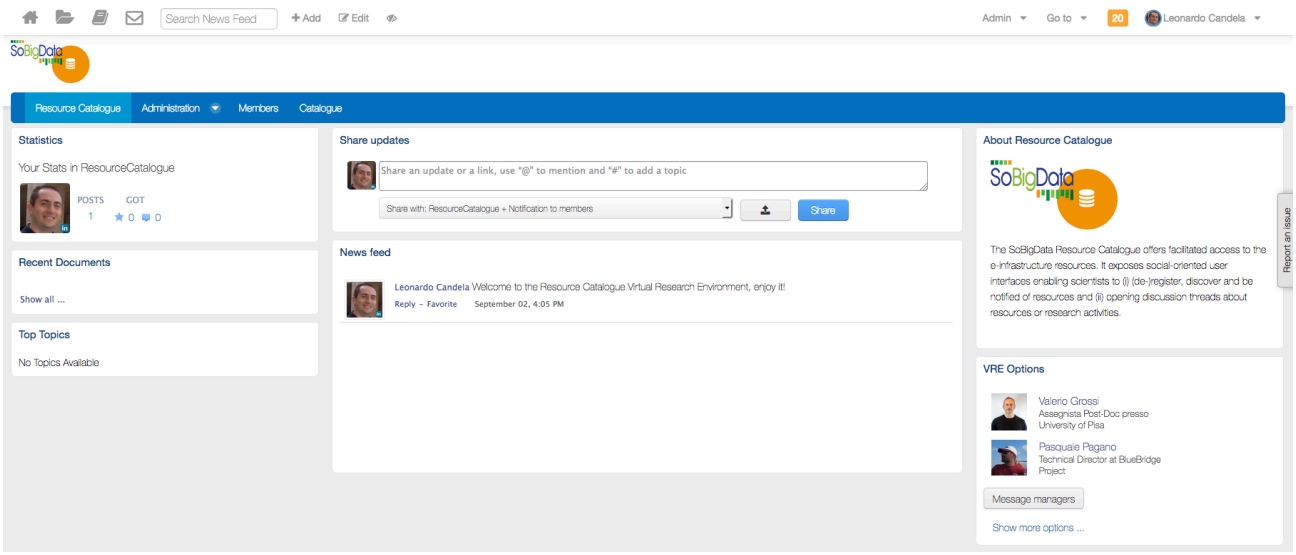

Figure 6. Resource Catalogue VRE Home Page

The main functionality offered by the VRE include:

- A dashboard (cf. Figure 6) for (i) collaborating with other VRE members by posting messages or being informed and reacting to (e.g. commenting, favouring) co-workers' posts, (ii) being informed on the most recent objects added to the VRE workspace, (iii) invite colleagues to participate, (iv) have information on the VRE including a description and the list of managers, (v) having some indicators on the user activity in the VRE, and (vi) acquire an authorization token for exploiting the services offered by the VRE in a programmatic way;
- A Catalogue area (cf. Figure 7), where users can mainly publish and/or modify products (datasets and methods) by providing reach characterisations of them aiming at maximising their exploitation and reuse in accordance to Open Science practices. For each product typology, a community specific set of attributes characterising them have been identified and documented by dedicated Wiki pages[2]. Dataset attributes include: general description attributes, e.g. title and creator(s); accessibility properties, e.g. VA or TNA; technical properties, e.g. size and format; legal and ethical

[2] https://wiki.gcube-system.org/gcube/GCube_Data_Catalogue#SoBigData.eu:_Dataset_Metadata and https://wiki.gcube-system.org/gcube/GCube_Data_Catalogue#SoBigData.eu:_Method_Metadata.

attributes, e.g. whether containing personal data; intellectual properties, e.g. licences. Method attributes include: general description attributes, e.g. title and creator(s); accessibility properties, e.g. VA or TNA; technical properties, e.g. programming language; intellectual properties, e.g. licences. Such attributes are compiled by using a specific portlet (cf. Figure 8);

- A shared area in the workspace, for making available objects of interests, e.g. guidelines, working notes;
- A Members area, for enabling each VRE member to be informed on the rest of VRE members and acquire details for contacting them;
- A user management area, to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers can (i) authorise users in accessing the VRE and its services, (ii) assign or withdraw roles to users, (iii) remove users, and (iv) send a communication to the current users;
- An accounting area, to enable authorised users (e.g. VRE Managers) to analyse usage records pertaining VRE services, e.g. most used service, most active user.

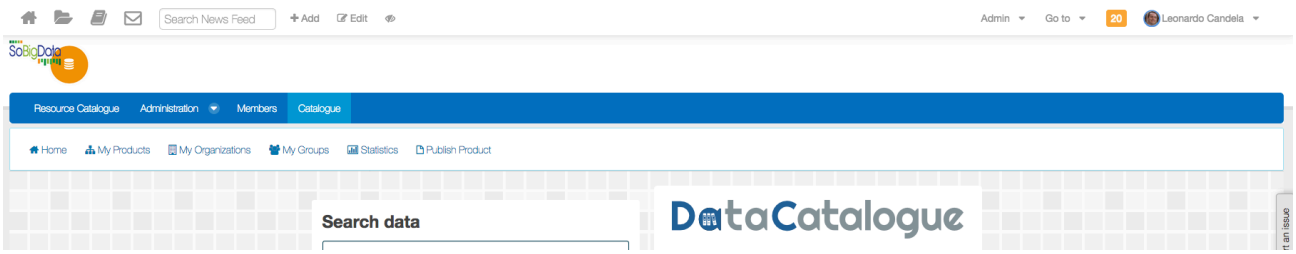

Figure 7. Catalogue Management Menu

Figure 8. Product Publishing Form

4.4 TAGME

This Virtual Research Environment is mainly conceived to provide its members with a VRE-based working environment for exploiting the TagMe facility [5] hosted by the SoBigData.eu Infrastructure (cf. Sec. 3). In the reality, it is expected that the majority of users with exploit TagMe in a programmatic way thus the VRE is mainly exploited for managing membership (i.e. authorise users), enable users to acquire a proper token for using the service, and collect statistics. However, it is equipped with the standard set of collaboration oriented facilities that can be exploited by its members.

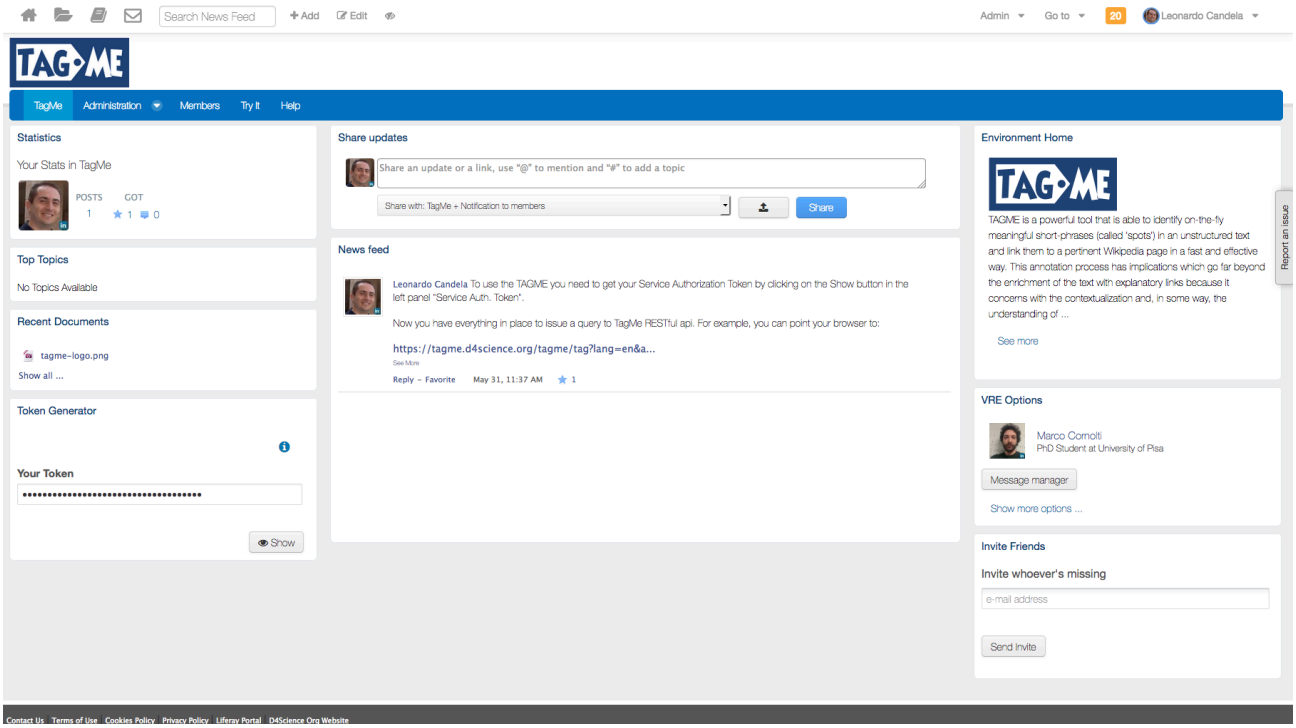

Figure 9. TagMe VRE Home Page

The main functionality offered by the VRE include:

- A dashboard (cf. Figure 9) for (i) collaborating with other VRE members by posting messages or being informed and reacting to (e.g. commenting, favouring) co-workers' posts, (ii) being informed on the most recent objects added to the VRE workspace, (iii) invite colleagues to participate, (iv) have information on the VRE including a description and the list of managers, (v) having some indicators on the user activity in the VRE, and (vi) acquire an authorization token for exploiting the services offered by the VRE in a programmatic way;
- A shared area in the workspace, for making available objects of interests, e.g. project deliverables, presentations, working notes;
- A Members area, for enabling each VRE member to be informed on the rest of VRE members and acquire details for contacting them;
- A user management area, to enable authorised users (i.e. VRE Managers) to manage other users using or willing to access the VRE. VRE Managers can (i) authorise users in accessing the VRE and its

services, (ii) assign or withdraw roles to users, (iii) remove users, and (iv) send a communication to the current users;

- An accounting area, to enable authorised users (e.g. VRE Managers) to analyse usage records pertaining VRE services, e.g. most used service, most active user.

5 CONCLUSION

This document is a placeholder of the actual deliverable that is implemented via a set of web sites and pages, namely:

- The SoBigData Gateway https://sobigdata.d4science.org/
- The SoBigData Catalogue https://sobigdata.d4science.org/catalogue
- The SoBigData.eu VRE https://sobigdata.d4science.org/group/sobigdata.eu
- The SoBigData.it VRE https://sobigdata.d4science.org/group/sobigdata.it
- The Resource Catalogue VRE https://sobigdata.d4science.org/group/resourcecatalogue
- The TagMe VRE https://sobigdata.d4science.org/group/tagme

REFERENCES

[1] Assante, M., Candela, L., Frosini, L., Lelii, L., Mangiacrapa, F., Pagano, P. (2016) SoBigData e-Infrastructure software release 1. SoBigData Project Deliverable D10.5, July 2016

[2] Candela, L., Castelli, D., Manzi, A., Pagano, P. (2014) Realising Virtual Research Environments by Hybrid Data Infrastructures: The D4Science Experience. International Symposium on Grids and Clouds (ISGC) 2014, Proceedings of Science PoS(ISGC2014)022

[3] Candela, L., Castelli, D., Pagano, P. (2013) Virtual Research Environments: An Overview and a Research Agenda. Data Science Journal, Vol. 12, pp. GRDI75-GRDI81, DOI 10.2481/dsj.GRDI-013

[4] Candela, L., Manghi, P., Pagano, P. (2016) SoBigData e-Infrastructure Release plan 1. SoBigData Project Deliverable D10.2, March 2016

[5] Cornolti, M., Ferragina, P., Ciaramita, M. (2013) A framework for benchmarking entity- annotation systems. In *Proceedings of the 22nd international conference on World Wide Web* (WWW '13). ACM, New York, NY, USA, 249-260. DOI 10.1145/2488388.2488411