Ref. Ares(2021)5841164 - 24/09/2021

Social Mining & Big Data Analytics

SoBigData

RESEARCH INFRASTRUCTURE ++

Deliverable D9.6

Supercomputing Network management and access reports

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

DOCUMENT INFORMATION

PROJECT	
PROJECT ACRONYM	SoBigData-PlusPlus
PROJECT TITLE	SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics
STARTING DATE	01/01/2020 (48 months)
ENDING DATE	31/12/2023
PROJECT WEBSITE	http://www.sobigdata.eu
TOPIC	INFRAIA-01-2018-2019 Integrating Activities for Advanced Communities
GRANT AGREEMENT N.	871042

DELIVERABLE INFORMATION	
WORK PACKAGE	WP9 JRA2 - E-Infrastructure and Supercomputing Network
WORK PACKAGE LEADER	CNR
WORK PACKAGE PARTICIPANTS	USFD, UNIPI, FRH, UT, LUH, AALTO, ETH Zürich, TUDelft, EGI, OpenAIRE, BSC, Nubisware
DELIVERABLE NUMBER	D9.6
DELIVERABLE TITLE	Supercomputing Network management and access reports
AUTHOR(S)	Nadia Tonello (BSC)
CONTRIBUTOR(S)	Ingrid Garcia Solera (BSC), Simon Carroll (BSC)
EDITOR(S)	Beatrice Rapisarda (CNR), Valerio Grossi (CNR)
REVIEWER(S)	Valerio Grossi (CNR)
CONTRACTUAL DELIVERY DATE	31/12/2020
ACTUAL DELIVERY DATE	15/01/2021
VERSION	1.4
TYPE	Websites, patents filling, etc
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	15
KEYWORDS	Supercomputing, web portal, mediators

EXECUTIVE SUMMARY

This deliverable provides the current design of the Portal accessible through the SoBigData++ platform for users to find the computing resources more suitable to their research objective, their availability and the administrative contacts to access them.

The work has been realized in the context of task “T9.5: Supercomputing Network management and access”, with the collaboration of the institutions providing computing resources to the project. This document is describing the design of the web tool to collect and publish the policies governing the access of the Supercomputing centres that constitute a network connected to the SoBigData++ infrastructure. Given the nature of the information and the heterogeneous procedures of the supercomputing centres involved, the role of institution mediator has been introduced, as a key part for the agile approach of the user to the Supercomputing centre selected.

The design presented in this deliverable is going to be implemented as a full operational tool in the next months, with a first prototype phase to be realized at short term, and a more complete version to be delivered according to the feedback received by the users on the prototype.

DISCLAIMER

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

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

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

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

Copyright © The SoBigData++ Consortium 2020. See http://www.sobigdata.eu/ for details on the copyright holders.

For more information on the project, its partners and contributors please see http://project.sobigdata.eu/. You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: "Copyright © The SoBigData++ Consortium 2020."

The information contained in this document represents the views of the SoBigData++ Consortium as of the date they are published. The SoBigData++ Consortium does not guarantee that any information contained herein is error-free, or up to date. THE SoBigData++ CONSORTIUM MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

GLOSSARY

EU	European Union
EC	European Commission
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
CMS	Content Management Systems
GUI	Graphical User Interface

TABLE OF CONTENTS

1 Relevance to SoBigData++ 7
1.1 Purpose of this document 7
1.2 Relevance to project objectives 7
1.3 Relation to other work packages 7
1.4 Structure of the document 7
2 Objectives of the portal 8
2.1 Functionalities for the users 8
2.2 Role of the mediators 8
3 Portal design 9
3.1 Proposed Structure of the wiki 9
4 Conclusions 13
Appendix A. Institution Mediators 14
Appendix B. Template of the institution´s information 15

1 Relevance to SoBigData++

This document and the portal outcome are part of work package (WP) 9.

1.1 Purpose of this document

The purpose of this document is to provide the design of the Portal for the computing network of the project.

1.2 Relevance to project objectives

One of the main objectives of the project is the transnational access and the access to the project resources by users. The design and implementation of a portal will make the users' selection and the institution approach simple and effective, minimizing the reaction time from the user's selection of the institution to their effective access.

1.3 Relation to other work packages

Transnational access (WP6) modalities and information will be coordinated with T6.1

1.4 Structure of the document

The content of this deliverable has been organized in 4 sessions, that describe the objective of the portal, the role of the institution's mediators, the web portal design and the procedure to collect information and keep it up to date.

2 Objectives of the portal

The objectives of the portal are:

- Give support to the user in finding the appropriate computational resources
- Facilitate the matching between resource availability and the users' needs
- Facilitate the procedures to access the resources

The portal will collect the technical information the users will need to select the proper computing facility for running their jobs, and the administrative information with the contacts that will provide access conditions and facilitate the access process. While the selection of the institution to host the user's jobs will be a simple self-service, the access procedure will be followed by a person (Institution mediator) who will follow and guide the user in the approach of the administrative process.

2.1 Functionalities for the users

The design of the portal is based on the objectives of the portal itself. The information provided by each institution has to be displayed, as well as comparison tables, to find:

- Facility technical characteristics; information about the infrastructure of each center, type of computing machines, their processor type, and amount of memory.
- Updated availability of the resources; size, description and subject for each resource available. The list of resources is updated by each institution mediator.
- Administrative access mechanisms; information on both in person and virtual access terms.
- Technical information to use the computing resources; information on the services, systems and code supported by each center in order to access the resources.
- Direct link to the mediator of each institution; institutional personal email for each of the institution´s mediators.

2.2 Role of the mediators

Each Institution that provides the possibility to use computing resources through the project has been asked to identify an Institution Mediator collaborating to T9.5. The list of mediators collected so fat is reported in Annex A.

The Institution Mediator provides information and administrative support for the access procedure to the users. Mediators are invited to participate actively, providing content, collaborating with the user's support, and facilitating the user's bureaucratic process to access the provided facilities. The responsibility of the mediator is also to keep the resources available for its institution up to date.

3 Portal design

In the inception, the portal was thought to be designed as a web application or a website and then be deployed in the already existing infrastructure and link it in the SoBigData++ gateway. At this point, the web page creators would get the rights to edit and modify the pages as needed. However, once the portal is deployed, every center mediator needs to have editing rights and access to the page in order to keep their own information updated. This would mean that the page´s contents would need to be within a CMS that provides a simple document editing GUI.

With the editing limitations presented above, the best option to accomplish the task's goals would be to create a wiki within the SoBigData++ gateway. The wiki would need to be publicly available without a specific login. In the future, all center mediators will need to have editing access since every center/mediator needs to be able to update his/her own information and the info related to their institution in the future.

A wiki within the SoBigData++ gateway is the solution that will more adequately accomplish the goals of the T9.5 task. A wiki can be a very efficient way to gather and distribute information. Anyone with the necessary editing rights can update their own information. The flexible structure of a wiki will allow us to adapt as the information itself changes, and will also allow the information within to grow into whatever possible future needs.

The template with the information that will be integrated for each institution can be found in Appendix B of this document.

3.1 Proposed Structure of the wiki

The main page will have a contents menu with five expandable sections. The first section will be named “Institutions,” where the name of each institution that forms part of SoBigData++ will be displayed. When you click an institution name, the information contained in the template shown in Appendix B will open. All the institutions information will follow the same template.

In order to make facts comparisons easier to view for the user, the main menu will also have tabs for “Computing”, “Storage”, “Datasets”, and “Mediators”. The comparison tables will have information of specific elements of all the centers. Figures 3.1.1 - 3.1.5 report the views of the expected tabs.

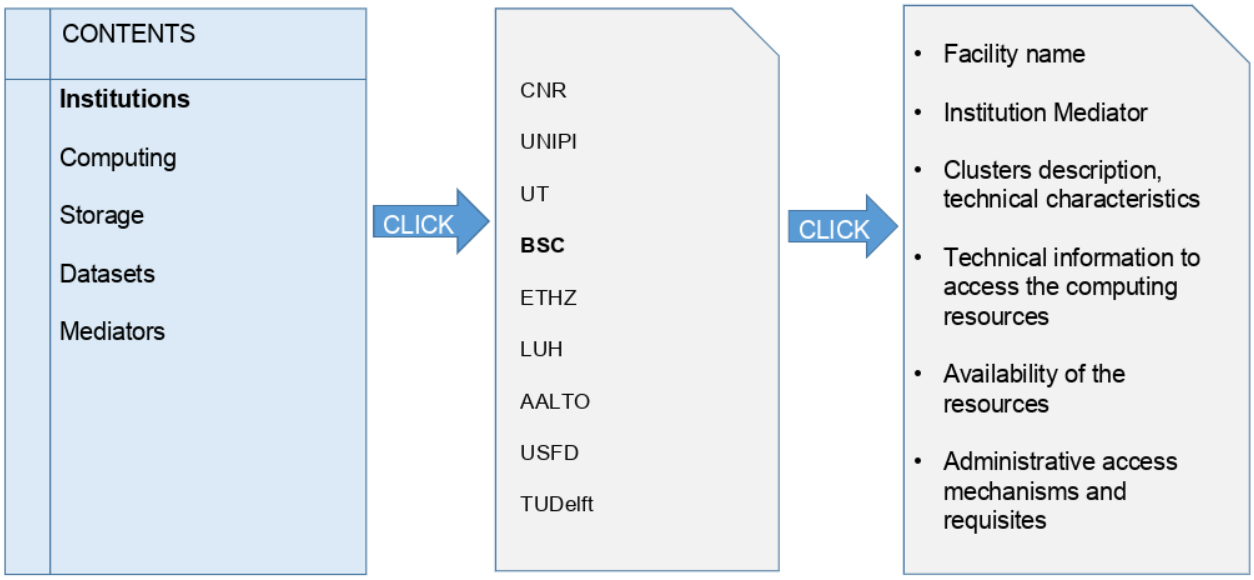


Figure 3.1.1 - Diagram 1: institutions tab view

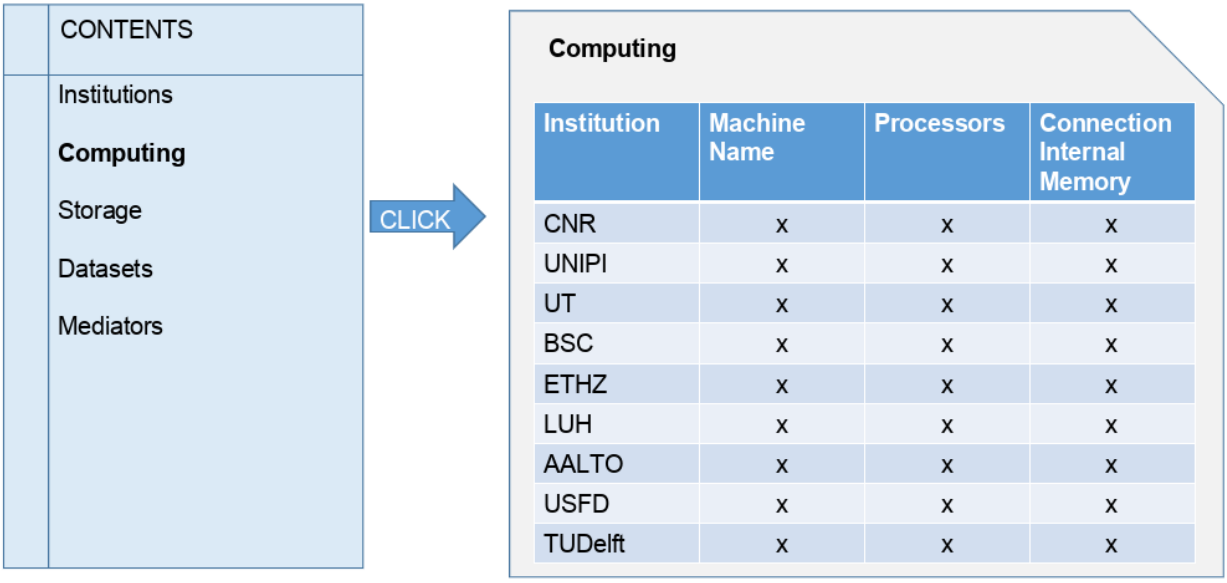
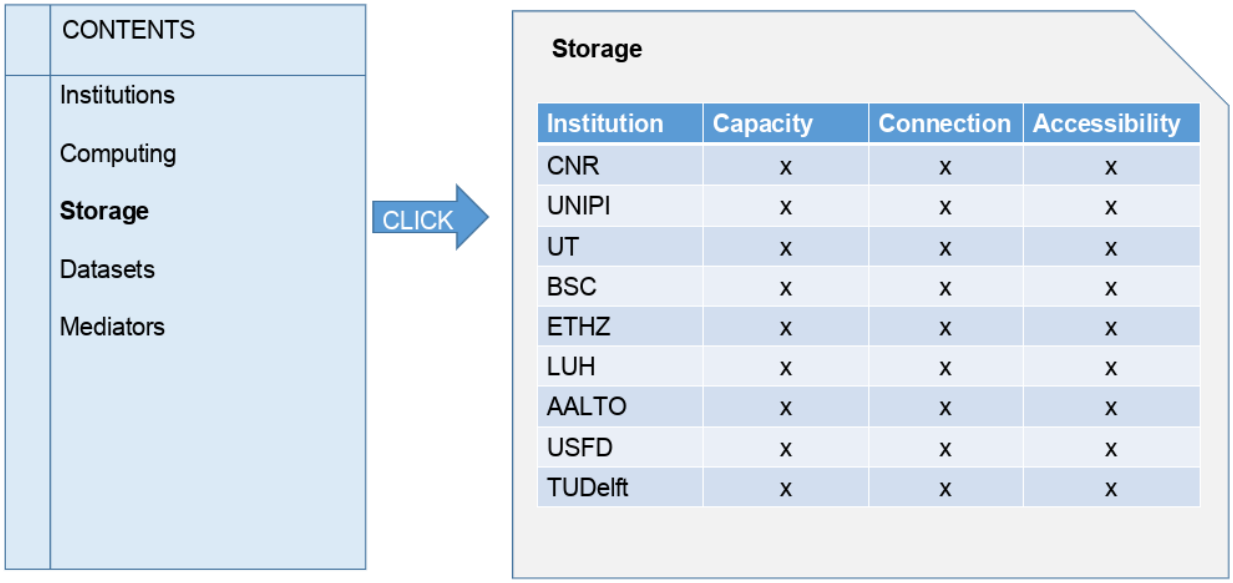
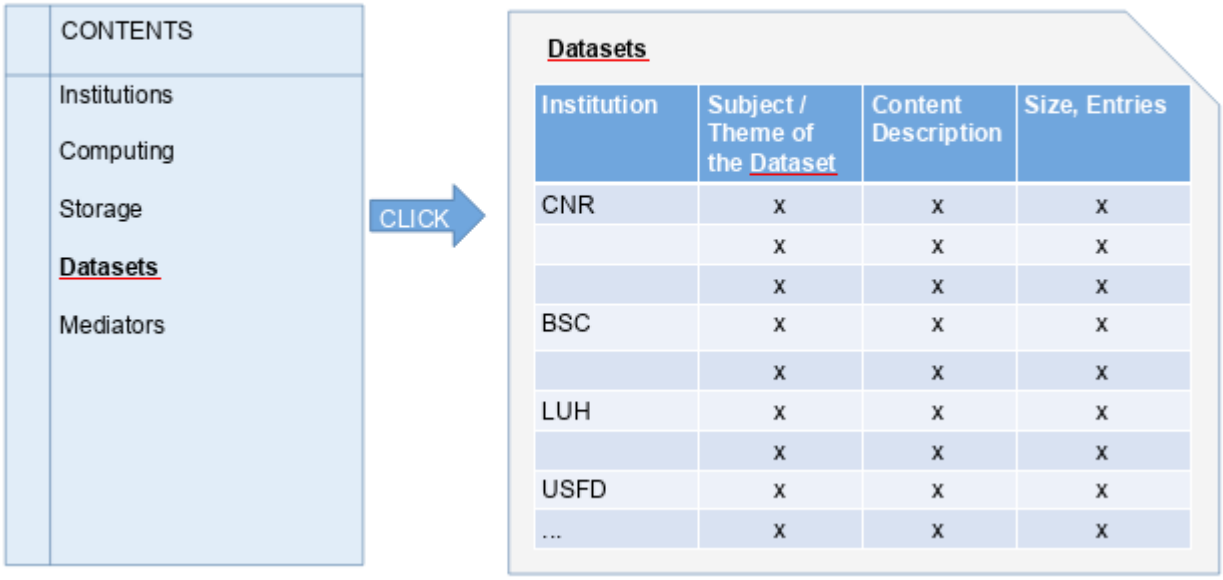


Figure 3.1.2 - Diagram 2: computing tab view



Institution	Capacity	Connection	Accessibility
CNR	x	x	x
UNIPI	x	x	x
UT	x	x	x
BSC	x	x	x
ETHZ	x	x	x
LUH	x	x	x
AALTO	x	x	x
USFD	x	x	x
TUDelft	x	x	x

Figure 3.1.3 - Diagram 3: storage tab view



Institution	Subject / Theme of the Dataset	Content Description	Size, Entries
CNR	x	x	x
	x	x	x
	x	x	x
BSC	x	x	x
	x	x	x
LUH	x	x	x
	x	x	x
USFD	x	x	x
...	x	x	x

Figure 3.1.4 - Diagram 4: datasets tab view

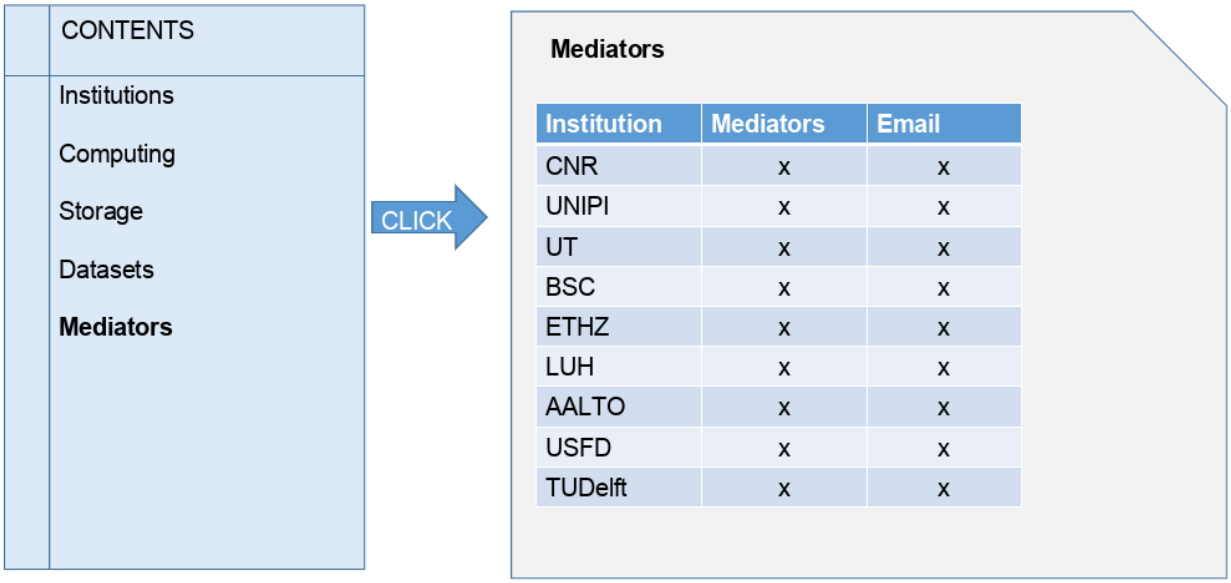


Figure 3.1.5 - Diagram 5: mediators tab view

4 Conclusions

The portal's objective is to support new users in finding the appropriate computational resources, facilitate matching resource availability and the user's needs, and facilitate the procedures to access the resources.

The objectives will be accomplished by creating a wiki in the already existing SoBigData++ gateway where the facilities' characteristics, availability of resources, administrative access mechanisms, technical information to access the computing resources, and the contact links to the institution mediators will be displayed.

The information contained within the wiki is of public interest and therefore needs to be publicly available without registration or login.

Once the wiki is deployed, the centre mediators will take charge of their own information update and, therefore, will need access to the wiki as editors. With such a system, we assure the content will be easily kept updated, without need of any technical intermediate.

The format and platform supporting the portal can be adapted and updated if the needs and objectives are not met, or the users will request changes or a different functionality in the future.

Appendix A. Institution Mediators

Institution	Mediator	email
CNR	Vittorio Romano Pasquale Pagano	vittorio.romano@isti.cnr.it pasquale.pagano@isti.cnr.it
UNIPI	Roberto Pellugrini	roberto.pellugrini@unipi.it
UT	Rajesh Sharma Ivar Koppel	rajesh.sharma@ut.ee ivar.koppel@ut.ee
BSC	Nadia Tonello	nadia.tonello@bsc.es
ETHZ	VaivaVasiliauskaite	vaiva.vasiliauskaite@gess.ethz.ch
LUH	Jurek Leonhardt	leonhardt@l3s.de
AALTO	Kimmo Kaski	kimmo.kaski@aalto.fi
USFD	Kalina Bontcheva	k.bontcheva@sheffiled.ac.uk
TUDelft	Juan M. Duran	j.m.duran@tudelft.nl

Appendix B. Template of the institution´s information

Facility:	Barcelona Supercomputing Center - Centro Nacional de Supercomputación (link)
Institution Mediator:	Nadia Tonello
Clusters description, technical characteristics	Computing: - MareNostrum 4 (3456 nodes, 165888 processor cores and 390 Terabytes of main memory); - NordIII (756 compute nodes, 12,096 Intel SandyBridge-EP E5–2670 cores at 2.6 GHz, with at least 24.2 TB of main memory.) - CTE-Power 9 (52 compute nodes, each of them: 2 x IBM Power9 8335-GTH @ 2.4GHz, 512GB of main memory, 2 x SSD 1.9TB as local storage, 2 x 3.2TB NVME, 4 x GPU NVIDIA V100 (Volta) - HTC- Cloud (Virtual Machines): NordIII Data: - Storage: HSM (disk, tape) quota? - Available Datasets: [Dataset description|entries, size]
Technical info to access the computing resources	- MN4: MPI, Singularity, git, Slurm - NordIII: OpenStack system, plain Virtual Machines, CUDA
Availability of the resources Last update: 09/12/2020	- Visitors places: 2 (transnational access only) - Computing hours: ... - Storage quota: ... - Relevant links (if applicable): N/A