

D8.2

Initial Data Management Plan



Unitelma Sapienza
università degli studi di Roma



Funded by
the European Union

www.star4bbs.eu
info@star4bbs.eu

@STAR4BBS





D8.2

Initial Data Management Plan

DELIVERABLE TYPE

DMP- Data Management Plan

MONTH AND DATE OF DELIVERY

February, 28/02/2023

WORK PACKAGE

WP8

LEADER

TUB

DISSEMINATION LEVEL

Public

AUTHORS

Marco Cardinaletti (UnitelmaSapienza)

Programme

HORIZON
EUROPE

Grant agreement

101060588

Start

Sept.2022

Duration

36 Months



Contributors

NAME	ORGANISATION
Marco Cardinaletti	UnitelmaSapienza
Enrica Imbert	UnitelmaSapienza
Giulia Borghesi	UnitelmaSapienza

Peer Reviews

NAME	ORGANISATION
Luana Ladu	TUB
Nikola Matovic	TUB

Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
1	17/02/2023	UnitelmaSapienza Team	UnitelmaSapienza sent the first version to Coordinator
2	22/02/2023	Luana Ladu (TUB)	Minor inputs and comments
3	24/02/2023	UnitelmaSapienza Team	Addressing comments raised by TUB
4	27/02/2023	Luana Ladu(TUB)	Final minor comments
5	27/02/2024	UnitelmaSapienza Team	Unitelma final version

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf.



Index of Contents

Executive Summary.....	7
1 Introduction.....	8
2. Data Summary.....	9
2.1. Purpose of data collection and relation to the project objectives.....	9
2.2. Data Management Plan (DMO) guiding Principles.....	9
2.3. STAR4BBS Data Management Strategy.....	10
2.4. Types and formats of the data.....	11
2.5. Potential users of published research data.....	12
3. FAIR (FINDABLE, ACCESSIBLE, INTEROPERABLE AND REUSABLE).....	13
3.1. Making data findable, including provisions for metadata.....	13
3.1.1. Discoverability of data (metadata provision).....	13
3.1.2. Identifiability of data.....	13
3.1.3. File naming convention and versioning.....	14
3.1.4. Approach towards keywords.....	15
3.2. Making data openly accessible.....	15
3.2.1. Data to be made publicly available and rationale for keeping some data closed	15
3.2.2. Approach to providing public access to datasets.....	17
3.3. Making data interoperable.....	17
3.4. Increase data re-use (though clarify licenses).....	18
3.4.1. Data licensing.....	18
3.4.2. Reusability during and the end of the project.....	19
3.4.3. Re-usability duration.....	19
4. Allocation of resources.....	20
4.1. Costs for making STAR4BBS data FAIR.....	20
4.2. Responsibilities for data management in the project.....	20
5. Data security.....	20
6. Ethical aspects to data management.....	21



Index of Tables

Table 1	16
---------------	----

Index of Figures

Figure 1.	10
----------------	----



Partners short names

TUB	(Technische Universität Berlin)
UNITELMA	(Università degli studi Unitelma di Roma)
UNI	(Ente Italiano di Normazione)
AUA	(Geoniko Panepistimion Athinon)
USC	(Universidad de Santiago de Compostela)
APRE	(Agenzia per la promozione della Ricerca Europea)
NOVA	(Nova - Institut für politische und Ökologische Innovation GMBH)
BAM	(Bundesanstalt für Materialforschung und Prüfung)
RSB	(Roundtable on Sustainable Biomaterials Association)
ISEAL	(ISEAL Alliance)
BB	(Better Biomass)

Abbreviations

DMP	Data Management Plan
EC	European Commission
EU	European Commission
PC	Project Coordinator
SCS	Sustainability Certification Schemes



Executive Summary

This document defines the overall strategy that will guide the management of data of STAR4BBS Consortium throughout the duration of the project. According to the European Commission (EC), data management plans are fundamental elements for good data management. In fact, they describe the life cycle of data management, showing how data is collected, generated and processed. According to the EC, a data management plan should include the following information:

- the handling of research data during & after the end of the project;
- what data will be collected, processed and/or generated;
- which methodology & standards will be applied;
- whether data will be shared/made open access;
- how data will be curated & preserved (including after the end of the project).

STAR4BBS DMP has been developed by UnitelmaSapienza during the first months of the project and it is part of WP8– Project Management & Internal Project Communication, led by TUB. UnitelmaSapienza will also support TUB in the data management process throughout the project. This DMP was prepared using the template and the indications provided by the EC. The Initial DMP is a living document and will be updated throughout the project. It will result in the D8.4 Final Data Management Plan due in M36.

The present document outlines:

- purpose of data collection;
- DMP guiding principles;
- data management strategy;
- how to make data FAIR (Findable, Accessible, Interoperable, Reusable);
- allocation of resources;
- data security;
- ethical aspects related to data management.



1 Introduction

STAR4BBS is a three-year multidisciplinary and multi-actor collaborative project, involving six partners, four associated partners and one associated third party. The overall aim of STAR4BBS is to maximize the potential of Sustainability Certification Schemes (SCS) and labels to support a successful transition to sustainable bio-based economy.

At the core of the STAR4BBS project is the development of indicators and a new monitoring system for assessing the effectiveness and robustness of existing international and EU SCS, B2B labels, and related trace ability systems applicable to biological feedstock and bio-based materials and products. This information will create the foundations to support achieving the much-needed harmonization between schemes and transparency in global and EU trade flows.

The project will involve important stakeholders (including scheme owners, policy makers, and industry) in the design of research and the monitoring system and in the development of practical recommendations emerging from the research and analysis in order to ensure that the project achieves its ultimate goal.

This STAR4BBS Data Management Plan (DMP) is part of the WP8 - Project Management & Internal Project Communication. Particularly it is related to the tasks 8.2 Ethical requirement and Data Management. This task deals with the management of data to be collected, processed and generated by STAR4BBS, with the aim of ensuring compliance with ethical and EC requirements.

UNIT, in collaboration with all partners, prepared this Data Management Plan (DMP). The plan, developed following the recommendation provided by the EC and the Horizon Europe template for the DMP, includes information on: (i) what data will be collected, processed and/or generated; (ii) which methodology and standards will be applied; (iii) whether data will be shared/made open access; and (iv) how data will be handled and preserved (including after the end of the project). UNIT will ensure that collected data will strictly comply with the Regulation (EU) 2016/6791 - the European Union's new General Data Protection Regulation (GDPR). Transparency, legal and social implications of information and knowledge will be considered.



2. Data Summary

2.1. Purpose of data collection and relation to the project objectives

The STAR4BBS Data Management Plan (DMP) aims to provide a strategy for managing key data generated and collected during the project and to optimize access to and re-use of research data. The DMP is intended to be a ‘living’ document that will outline how the STAR4BBS research data will be handled during and after the project, and so, if needed, it will be reviewed and updated at regular intervals.

The main purpose of the DMP is to ensure the accessibility and intelligibility of the data generated during the project. Each data set created during the project will be assessed and categorized as open, embargo or restricted by the owners of the content of the data set.

All the data sets, regardless of their categorization, will be stored in each of the participant entities databases and in the TUB Cloud created as internal database of the partners. In addition, those categorized as open or embargo will be publicly shared (in the case of embargo, after the embargo period is over) through the public section of the project website and ZENODO (<https://zenodo.org/>).

2.2. Data Management Plan (DMO) guiding Principles

The Data Management Plan of STAR4BBS is implemented within the Work Package 8 Project Management & Internal Project Communication. The STAR4BBS project data management plan follows the principle of Open Access guideline summarized in the diagram here below.

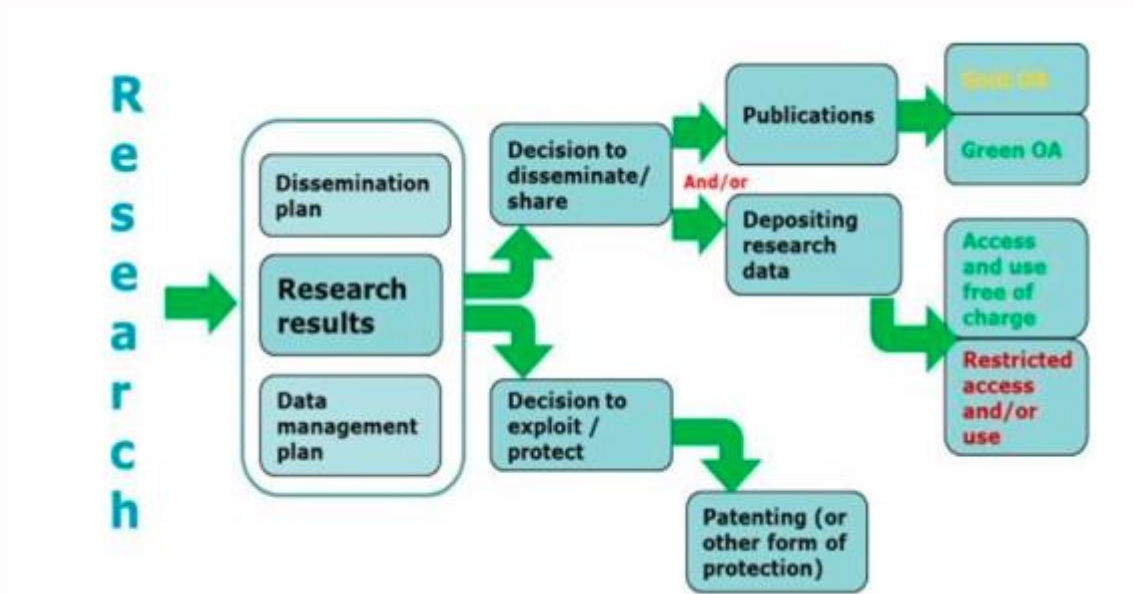


Figure 1- Open access to research data and publication decision diagram (from Guidelines to the Rules on Open Access to Scientific publications and Open Access to Research Data in Horizon 2020).

In particular, the STAR4BBS DMP:

- Will be drafted taking into account the template of the “Guidelines on Data Management in Horizon 2020”
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf and its updates
- will support the consortium to be compliant with the requirements of Regulation (EU) 2016/679 and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Guidance on how these regulations interact with open-access data policy can be found here: <https://www.openaire.eu/ordp/>
- will assure that all type of data, storage, confidentiality, ownership, management of intellectual property and access: procedures are in line with EU standards as described in the Grant Agreement and the Consortium Agreement.

2.3. STAR4BBS Data Management Strategy

The Data Management Strategy of the STAR4BBS project is focused on the observation of FAIR (Findable, Accessible, Interoperable and Reusable) Data Management



Protocols. This document addresses for each data set collected, processed and/or generated in the project the following elements:

- Dataset reference and name: Internal project Identifier for the data set to be produced. This will follow the format:
 - o PNumber_TaskNumber__PartnerName_DataSubset_Dataset-Name_Version__DateOfStorage, where the project name is STAR4BBS, the Partner Name represents the name of the data custodian (WP Lead/Task Leader).
- Dataset description: description of the data generated or collected, including its origin (in cases where data is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the potential for integration and reuse.
- Standards and metadata: reference to existing suitable standards. If these do not exist, an outline on how and what metadata will be created.
- Data sharing: description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling reuse, and definition of whether access will be open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating the type of repository (institutional, standard repository for the discipline, etc.). In cases where the dataset cannot be shared, the reasons for this will be stated (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).
- Archiving and preservation (including storage and backup): description of the procedures to be put in place for long-term preservation of the data, including an indication of how long the data should be preserved, the approximate end volume, associated costs, and how these are planned to be covered.

2.4. Types and formats of the data

The data that STAR4BBS will generate will be primarily sourced from existing relevant knowledge from (scientific and non-scientific) publications and experts. In most cases, the collected data will be organized in a spreadsheet, which besides some numerical



data typically contained short or long text entries. These spreadsheets have the .xlsx (Microsoft Excel) format.

Other data formats will be text: documents (.pdf and .docx).

Some complementing formats for graphical and audio-visual data could also be generated. Some examples include but are not limited to:

- .jpg (images);
- .ppt (Microsoft PowerPoint);
- .avi (video).

Lastly, in case specific software or models will be used, datasets may be stored in formats specific software or models. A detailed information will be duly updated throughout the course of the project, when more information on the datasets will be available for STAR4BBS.

2.5. Potential users of published research data

Datasets generated by STAR4BBS can be interesting for:

- Members of the scientific community;
- Professionals with a link to bio-based product sustainability;
- Those active in the specific research field to which the dataset contributes;
- Other research projects that can continue to build on or combine their work with STAR4BBS datasets.

A distinction can be made between research datasets supporting peer reviewed academic publications and datasets forming the basis on which the project deliverables are elaborated. The research data underlying project deliverables requires a big investment in terms of human resources, therefore the main aim of making this data public is for others to use and build on this data, preventing duplication of work and improving efficiency, quality and speed of research. This also holds true for scientific publications, but for these validation and reproducibility are key consideration as well.



3. FAIR (FINDABLE, ACCESSIBLE, INTEROPERABLE AND REUSABLE)

3.1. Making data findable, including provisions for metadata

Metadata is data on the research data themselves. It enables other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers, can better determine whether the dataset is relevant and useful for their own research. Metadata (type of data, location, etc.) will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data.

3.1.1. Discoverability of data (metadata provision)

A multi-disciplinary project such STAR4BBS requires a broad, commonly used standard that is not restricted to a specific research area. The default option in STAR4BBS will be to apply the DataCite Metadata Schema. This standard compatible with the OpenAIRE platform and the project's use of Zenodo as a repository.

However, some specific STAR4BBS datasets may be of interest to a scientific discipline for which a dedicated metadata protocol exists. Usage of a dedicated metadata standard will be preferred over the broad protocol. In case a dataset consists of multiple files a readme.txt is included explaining the intention of each included file and the relevant metadata, unless a more appropriate means of including such information is available. The selected metadata standard is recorded per dataset in the key characteristics tables by following the scheme/format proposed in the 2.2.1 Section.

Similarly, data findability can in some cases be enhanced by depositing the dataset in a repository, which is specific to the discipline. Therefore, for each dataset the Registry of Research data Repositories (www.re3data.org) should be consulted, in order to determine most appropriate repository. At the end of the project no such dedicated metadata standard was found for any of the project's datasets.

3.1.2. Identifiability of data

The STAR4BBS datasets (see section 2.2) that will be published as part of the data management approach form the supporting data for scientific manuscripts and project deliverables. In order to establish a clear link between the supporting data and the documents based on them, a broadly used identifier is needed. Each published STAR4BBS dataset will have a unique Digital Object Identifier (DOI) attributed to it.



Unless a different platform is considered more appropriate (such as in the case of scientific publications that have their own DOI), the DOI will be generated using the STAR4BBS community on the Zenodo repository.

3.1.3. File naming convention and versioning

STAR4BBS_D[x.y.]_v[Version]_[ShortTitle]_[Type]_[Date]_[Status]_[Free].[extension]

It consists of:

STAR4BBS	Project name, fixed
D[x.y.]	Deliverable identifier, if relevant
V[Version]	Version number in x.y format, should match a version number with a short description inside the document, such as the Document History table in this document
[Short Title]	Short descriptor for easy identification, maximum 40 characters
[Type]	Describes the type of data (e.g. publication, inventory, etc.)
[Date]	Date in format YYYYMMDD
[Status]	Draft, Final, Public, Restricted, Confidential
[Free]	Free text field for internal communication purposes (e.g. initials of reviewer). This field should not be included in the name of published files.
[extension]	File extension

For the completed individual datasets generated by the project, a specific file naming convention was used:

STAR4BBS_[Short Title]_v[Version]_[Type]

It consists of:

STAR4BBS	Project name, fixed
[Short Title]	Short descriptor for easy identification, maximum 40 characters



V[Version]	Version number in x.y format, should match a version number with a short description inside the document, such as the Document History table in this document
[Type]	Describes the type of data (e.g. publication, inventory, etc.)
[extension]	File extension

These minimal metadata schemas can be extended by arbitrary subjects from a taxonomy or controlled vocabulary as described in the Zenodo API documentation.

3.1.4. Approach towards keywords

For each dataset, the responsible beneficiary should indicate a set of selected keywords aiming to maximise findability. The default STAR4BBS repository is Zenodo, which follows the DataCite Metadata Schema. The keywords are also included in the key characteristics' tables per dataset, in the Tab1 section 2.2.1.

3.2. Making data openly accessible

3.2.1. Data to be made publicly available and rationale for keeping some data closed

In order to maximise the impact of STAR4BBS research data, the results are shared within and beyond the consortium. Selected data and results will be shared with the scientific community and other stakeholders through publications in scientific journals and presentations at conferences, as well as through open access data repositories.

The STAR4BBS project datasets are first stored and organized in a database by the data owners (personal computer, or on the institutional secure server) and on the project database (project repository – TUB CLOUD). All data are made available for verification and re-use, unless the task leader can justify why data cannot be made openly accessible. To protect the copyright of the project knowledge, Creative Commons license will be used in some cases.

Not publishing, or under embargo or other restrictions is allowed, but only if there are sufficient grounds to do so. An example of such a case may be confidential data received for case studies or data gathered through interviews. Via the STAR4BBS letter of consent (see annex 1-2-3 on Deliverable 8.3), the project guarantees anonymity to



respondents. If anonymized data from an interview was to be made public via open access, each respondent would need to sign a second informed consent regarding this step.

Key information was collected for each dataset, by the lead beneficiary responsible for the dataset. This key information is recorded for each dataset using the template provided in the Table 1.

Table 1 - Template for the key information log for each dataset

Dataset identifier	STAR4BBS_[Short Title]_v[Version] _[Type]
Dataset name	
Dataset description	
Dataset DOI	
Dataset version history	
Key contact [Beneficiary]	
Dataset file format and size	
Associated WPs/Tasks/Deliverables/Milestones Other contributing beneficiaries	
Dataset (to be made) public?	[Yes/No]
If No, justification: Published version(s)	
Repository(-ies)	
Keywords	
Licence	[Default: Attribution-ShareAlike (CC BY-SA)]
Useful for whom?	
Key data sources	



3.2.2. Approach to providing public access to datasets

The STAR4BBS dataset are both public (data access policy unrestricted) and they will be accessible by:

- Project web site
- Partners database
- ZENODO (<https://zenodo.org>)
- Open access journals
- Other platforms if needed

All data deposited on ZENODO are accessible without restriction for public. For other data, potential users must contact the Project Coordinator (PC) or the data owner in order to gain access. If necessary, appropriate procedure (such as non-disclosure agreement) will be used.

Basically, until a dataset will be fully finalised and ready for publication, the private area of the STAR4BBS will be the default platform (TUB CLOUD) for data exchange.

3.3. Making data interoperable

It is essential that published datasets are unequivocally interpretable by third persons without any link to the project. Therefore, each dataset needs to be accompanied with a description of the methodology, sources, definitions and scope of the data contained in it.

As discussed in section 2.1, whenever possible, datasets should be structured in such a way that it can, in full or in part, be combined with another dataset, from the project or any other data source. For some fields of research specific definitions, metadata and/or vocabulary exists to enable this (see [RDA's Metadata Standards](#)). For each dataset, the responsible beneficiary needs to determine if this exists for the relevant field(s) of research and comply with the relevant standards.

In order to ensure good interoperability of datasets, it is imperative that standards and methods commonly used in the same field are used. [Fairsharing.org](#) is a valuable resource where researches can identify relevant standards, as well as databases and repositories.



Moreover, the OpenAIRE guidelines for online interoperability, including OpenAIRE Guidelines for Literature Repositories, OpenAIRE Guidelines for Data Archives, OpenAIRE Guidelines for CRIS Managers based on CERIF-XML can be a valuable resource to be consulted. These guidelines can be found at: <https://guidelines.openaire.eu/en/latest/>.

3.4. Increase data re-use (though clarify licenses)

3.4.1. Data licensing

[Creative Common Licensing](#) will be used to protect the ownership of the datasets. Both Share-Alike and NonCommercial-ShareAlike licenses will be considered for the parts of datasets for which the decision of making that part public has been made by the Consortium.

However, an embargo period may be applied if the data (or parts of data) are used in published articles in “Green” open access journals. The recommended maximum embargo period length by European Commission is 6 months.

For datasets deposited on a public data repository (ZENODO) the access is unlimited. Restrictions on re-use policy are applied for all protected data (see Figure 1 – Section 1.2), whose re-use will be limited within the project partners.

Other restrictions could include:

- the “embargo” period imposed by journals publication policy (Green Open access);
- some or all of the following restrictions may be applied with Creative Commons licensing of the dataset:
 - Attribution: requires users of the dataset to give appropriate credit, provide a link to the license, and indicate if changes were made.
 - Non-Commercial: prohibits the use of the dataset for commercial purposes by others.
 - Share-Alike: requires the others to use the same license as the original on all derivative works based on the original data.

Internal process of quality evaluation is activated throughout the entire project duration to assess both project data /products and project process (See the task 8.1 Project



Management including quality assurance and timely delivery of project results). An internal peer review is performed for the main project deliverables to guarantee the deliverable is developed with a high level of quality.

3.4.2. Reusability during and the end of the project

Final datasets will be uploaded in the Zenodo repository (<https://zenodo.org>) which ensures long-term archiving of the final research data. Dataset published on Zenodo will be retained for the lifetime of the repository, which is currently indicated as indefinitely. Uploaded data files and metadata are backed up on a 12-hourly basis, as well as replicated in multiple copies in the online system. Where relevant, datasets will be made available through the [STAR4BBS website](#), which will remain online for at least three years after the end of the project.

During the project, the lead beneficiary in charge of a dataset aimed to make the dataset public as early as possible, while adhering to the restrictions of consent of all contributing beneficiaries and sufficient time for pre-publication quality control. Publishing to Zenodo means a DOI will be obtained, which could be used in the scientific publications and project deliverables. Hence if possible, datasets were published ahead of the documents for which the dataset provide the supporting data.

Beneficiaries provided repository address and basic access instructions as part of any dissemination related to the datasets.

The same quality assurance performed on the STAR4BBS Deliverables specified in the grant agreement shall be performed on its datasets, including extensive review by contributing partners.

3.4.3. Re-usability duration

Given the fast-changing nature of both bio-based product development and sustainability certification, the useful lifetime of the majority of STAR4BBS datasets is likely to be limited. This is especially true if the data sources used are subject to change, such as the rules guiding existing sustainability schemes. Inventories using scientific publications are more stable over time; future research can use published STAR4BBS data and build on it by updating and completing such data.



The default STAR4BBS approach is for its datasets, once published, to remain re-usable indefinitely.

4. Allocation of resources

4.1. Costs for making STAR4BBS data FAIR

Cost for making data FAIR are estimated to be zero. Public project deliverables and datasets will be published on the Zenodo repository; therefore, no additional costs are foreseen. However, repositories of specific institutions will not be free. Consortium partners will use their own budgets to archive personal data in their own repositories.

4.2. Responsibilities for data management in the project

Each beneficiary leading Work Packages is responsible for preparing the datasets to make FAIR the data collected within its own activities, following the instructions provided in this Data Management Plan. TUB as Project Coordinator and UNIT as responsible for the preparation of the different version of Data Management Plan will be responsible for general coordination and supervision.

Furthermore, consortium partners have the responsibility to make sure their activities are in line with all applicable local, government and international laws, regulations and guidelines.

5. Data security

The following guidelines will be followed in order to ensure the security of the data:

- Store data in at least two separate locations to avoid loss of data;
- Encrypt data if it is deemed necessary by the participating researchers;
- Limit the use of USB flash drives.
- Label files in a systematically structured way in order to ensure the coherence of the final dataset.

All project deliverables and data will be stored and shared in the TUB Cloud restricted to the project consortium. Only the Consortium Partners will have access to the cloud storage where dataset and metadata are filed.

Following, scientific publications and articles, the dataset deliverables and the final demonstrator research results will be shared through ZENODO and other database to promote the data making FAIR.



6. Ethical aspects to data management

STAR4BBS will collect expert knowledge and opinions by means of interviews, questionnaires and Delphi exercises. Moreover, a very wide stakeholder engagement process throughout laboratories, workshops and seminars is envisaged during the project life, where information, sensitive data, pictures, images will be capture for the scope of dissemination and results exploitation.

Sensitive issues will be carefully avoided and, if any personal information will be given, researchers going to be ensuring that the opinions, personal information cannot be directly associated to individuals.

In general, individual names and organisations will be not identified in the research. In the specific case of interviews, the minutes of each of them will not mention names but will be labelled with an identifying code to allow easier consultation of documents. If individuals (speakers, interviewed, general people, etc.) agreed to be quoted, researchers first verified the accuracy of quotes with the respondent before they were used.

Finally, for managing all personal and sensitive data in accordance with the General Data Protection Regulation 2016/679 dedicated formats will be used and submitted to the persons of interest. Such formats, included in the Ethical issues report [see D.8.3], are the following:

- ANNEX 1 - Informed consent form
- ANNEX 2 – Information sheet form
- ANNEX 3 - Informed Consent Form for Processing Personal Data, Photographs, videos and recorded speeches/interviews collected in the course of STAR4BBS Project events/meetings/workshops
- ANNEX 4 - STAR4BBS Database subscription form
- ANNEX 5 – STAR4BBS Privacy Policy

“ Sustainable bio-based systems via effective certification & labelling ”

Consortium:



UnitelmaSapienza
università degli Studi di roma



Funded by
the European Union

www.star4bbs.eu
info@star4bbs.eu

@STAR4BBS

