BIO-SUSHY

Sustainable surface protection by glass-like hybrid and biomaterials coatings

Deliverable D.6.3
Data Management Plan

Deliverable Information	
Responsible partner:	7P9-SI
Work package No and Title:	WP6 - Data Management Plan
Contributing partner(s):	7P9-DE
Dissemination level[1]:	PU
Type:	R
Due date:	30/06/2023
Submission date:	30/06/2023
Version:	V2

[1] PU = PUBLIC fully open ((warning) automatically posted online on the Project Results platforms)
SEN = Sensitive — limited under the conditions of the Grant Agreement
EUCI = EU classified under Decision 2015/444

Funded by the European Union

Funded by the European Union under the Grant Agreement 101091464. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

Project Profile

Programme	Horizon Europe
Call	HORIZON-CL4-2022-RESILIENCE-01
Topic	HORIZON-CL4-2022-RESILIENCE-01-23: Safe and sustainable by design chemicals and materials (RIA)
Number	101091464
Acronym	BIO-SUSHY
Name	Sustainable surface protection by glass-like hybrid and biomaterials coatings
Start Date	1 January 2023
Duration	48 months
Type of action	HORIZON Research and Innovation Actions
Granting authority	European Health and Digital Executive Agency
Project Coordinator	MATERIA NOVA

Document History

Version	Date	Entity	Remarks
V1	8 June 2023	7P9-SI, 7P9-DE	First draft
V2	26 June 2023	MANO, SIK, ZSI, AXIA, RESCOLL, CNR, PQSAR, WOODK+	Partner input

Publishable Summary

This deliverable report describes the creation of the BIO-SUSHY Research Output Management Plan (ROMP) as an extension of a Data Management Plan (DMP) and contains the first version as Annex I. With the success of the FAIR principles to foster making more and more data available for re-use, it became obvious that the high-quality knowledge management approaches should also be applied to other research outputs to increase their re-usability. To spearhead such a comprehensive approach, BIO-SUSHY has started to go beyond the DMP, mandatory for all projects of the Horizon Europe Framework Programme, by covering all research outcomes including but not limited to data, protocols, models, software, surveys and responses, reports and publications in the ROMP and specifying common management recommendations and guidelines across all research outputs as well as specific approaches and tooling for individual types. To be able to cover the requirements of all output types, BIO-SUSHY, in collaboration with the MACRAMÉ project (EU Horizon Europe research and innovation programme, grant agreement No. 101092686), performed an evaluation of existing DMP templates proposed in the DMPonline, DS Wizard and ARGOS tools. None of them is providing all functionality and flexibility needed by BIO-SUSHY, especially with respect to covering computational methods and research software. Thus, the BIO-SUSHY ROMP was newly designed by starting from the structure of the ARGOS tools, transferring sections of the other tools to strengthen specific aspects and adding additional sections based on the FAIR for Research Software (FAIR4RS), TRUST and CARE data guidance principles.

Table of Contents

Publishable Summary 3

List of Figures 5

Table of Abbreviations 6

1. Objectives 8
2. Creation of the first version of the Research Output Management Plan (ROMP) 8
3. BIO-SUSHY data management infrastructure 10
4. Conclusions 11
5. References 12
6. Annexes 13

Annex I: Initial Version of the BIO-SUSHY Research Output Management Plan (ROMP) 13

List of Figures

Figure 1: Data management concept in which all data from experiment, computational approaches and public resources is integrated to build the data lake used as input for data-driven modelling to guide the SSbD decisions for the development of the new coatings as well as to prepare upload to public databases) 11

Table of Abbreviations

Abbreviation	Definition
EU	European Union
EC	European Commission
H2020	Horizon 2020 research and innovation programme
HE	Horizon Europe research and innovation programme
WP	Work Package
PFAS	Per- and polyfluorinated alkyl substance
DMP	Data Management Plan
ROMP	Research Output Management Plan
RO	Research Output
TBD	To Be Defined
FAIR	Findable, Accessible, Interoperable and Re-usable (data guidance principles)
FAIR4RS	FAIR for Research Software
FIP	FAIR Implementation Profile
FER	FAIR Enabling Resource
FSR	FAIR Supporting Resource
TRUST	Transparency, Responsibility, User focus, Sustainability, and Technology (data guidance principles)
CARE	Collective benefit, Authority to control, Responsibility, and Ethics (data guidance principles)
GDPR	EU's General Data Protection Regulation
DOI	Digital Object Identifier
ORCID	Open Researcher and Contributor ID
ROR	Research Organisation Registry (identifier)
ERM	European Registry of Materials (identifier)
CAS	Chemical Abstract Service Registry Number
IUPAC	International Union of Pure and Applied Chemistry
InChI	IUPAC International Chemical Identifier
ELIXIR	European life sciences infrastructure

TeSS	ELIXIR Training eSupport System
SOP	Standard Operating Procedure
SSbD	Safe-and-Sustainable-by-Design
MODA	MOdelling DAta reporting format
CHADA	CHaracterisation DAta reporting format
QSAR	Quantitative Structure Activity Relationship
QMRF	QSAR Model Reporting Format
QPRF	QSAR Prediction Reporting Format
API	Application Programming Interface
EOSC	European Open Science Cloud
OECD	Organisation for Economic Co-operation and Development
VAMAS	Versailles Project on Advanced Materials and Standards
ISO	International Organization for Standardization
CEN	European Committee for Standardization
ECHA	European Chemicals Agency

1. Objectives

As a part of its open science philosophy, BIO-SUSHY is implementing high-quality knowledge and data management with the goals to facilitate, harmonise and accelerate information sharing within the project and, at the same time, prepare all BIO-SUSHY research output for open and FAIR (findable, accessible, interoperable, and re-usable) release. One important tool to organise data collection, documentation, preservation and preparation for sharing is the Data Management Plan (DMP). It is the central guidance document for the consortium specifying, first, the types of data to be generated within the project and what data management concepts, approaches and tools are recommended and, in subsequent versions, is providing more and more information on the data produced (data formats, size), applied (meta)data standards, annotations and solutions for long-term storage and indexing. Due to this central role, the DMP is now mandatory for projects funded under Horizon Europe. However, BIO-SUSHY is going even one step further. Already during the proposal writing, the consortium agreed that it will apply the high-quality management standards not only to data but to all research outputs and will document these in the Research Output Management Plan (ROMP), which integrates and extends the DMP. This was seen as necessary to be able to clearly document all steps in the development, production and safety and sustainability evaluation of the new coatings. Since many novel and non-standard methods are used in BIO-SUSHY, detailed and complete documentation of the design process, procedures (method descriptions, protocols, standard operating procedures), processing, analysis and extracted evidence and conclusions have to be provided as (meta)data to understand and evaluate all potential influences on the results. Additionally, the strong integration of physics-based and data-driven modelling and simulations extends the types of research outputs to be covered to computational models, workflows and software.

To develop the best way to guide the consortium in the task of documenting all these research outputs and selecting appropriate approaches, BIO-SUSHY was collaborating with the MACRAMÉ project to create the structure of the ROMP. Besides the complete adoption of the open science and FAIR guidance principles[1], addressing requirements of specific other research outputs and integrating aspects such as quality assurance, responsibility, trustworthiness and ethics were major objectives of the ROMP development. To achieve these, data management was completely aligned with BIO-SUSHY's overall quality and risk management (see Deliverable D6.2) and the FAIR for Research Software (FAIR4RS)[2], TRUST[3] and CARE[4] principles, besides the newest recommendations from the WorldFAIR project and the GO FAIR AvancedNano Implementation Network, were considered.

2. Creation of the first version of the Research Output Management Plan (ROMP)

As introduced above, details of and guidance on the adopted management principles, FAIRification approaches, data security and ethics considerations as agreed on by the consortium are documented in the Research Output Management Plan (ROMP) as an extended version of a Data Management

Plan (DMP) covering not only data but also all other outputs including but not limited to protocols, models, software, surveys and responses, reports and publications. It is designed as a living document being updated constantly during the runtime of the project and is provided in its first version as Annex I to this report. Besides the initial version presented here, we anticipate that at least two additional stable versions of the plan (intermediate and final) will be compiled and publicly shared to accommodate reporting of the data generation at the midterm and end of the project but also to be able to adapt to the fast-changing field of Open and FAIR data in Europe.

According to the EC recommendation to make data findable, accessible, interoperable and re-usable (FAIR), a Data Management Plan and, thus, also its extension, the ROMP, includes information on the following details:

1. Handling of research outputs (RO) during & after the end of the Project;
2. What RO are collected, processed and/or generated;
3. Which methodology and standards are applied;
4. Whether RO are shared/made Open Access;
5. How RO are curated and preserved (including after project end).

The initial version of the BIO-SUSHY ROMP mainly documents decisions made by the consortium and guidelines to be implemented by the data providers and the data managers in collaboration with and under the supervision of the BIO-SUSHY data shepherd. This will subsequently be complemented throughout the project with specific information on the research outputs generated and resources re-used in the project as well as updates and refinements of the used FAIR Enabling and Supporting Resources (FERs and FSRs) and improvements in harmonisation and interoperability. Thus, the ROMP has, as already stated above, to be understood as a living document, in which changes are clearly tracked to show additions as well as decisions, which had to be changed, revised or replaced as a reaction to emerging new standards, tools and guidelines. In cases where the content of the document needs more extensive adaptations for a specific type of research output, a specific section for this type will be added or even a specific ROMP for this type will be created and referred to in this general ROMP covering the project globally.

For creating the ROMP, three different online DMP tools were evaluated according to criteria that evaluate if they covered all necessary aspects and provided the appropriate structure needed to define the management approach and the generalisation to all research output in close collaboration with the MACRAMÉ project. The tools evaluated are:

1. DMPonline from the Digital Curation Centre,
2. DS Wizard and
3. ARGOS.

Results of this evaluation have already been described in the public deliverable D3.1 of the MACRAMÉ project, which is currently still under review by the European Commission. Therefore, we will reproduce them here again and want to stress the importance covering computational models, workflows, and software as specific research outputs in the ROMP. This was a specific requirement of BIO-SUSHY because of the much more central role of physics-based and data-driven modelling and simulation approaches compared to MACRAMÉ.

DMPonline provides a quite general structure with questions associated with specific categories like FAIR data, Data security and Ethics. However, most of the questions are related to FAIR and are specific to data. Other research outputs (e.g. protocols, SOPs, software) can only be addressed by answering two very generic questions. Additionally, there is no way to describe the reasons for data collection, re-use and the overall concept, of how links between different research outputs are established.

DS Wizard uses the concept of following the data life cycle from creation to processing to preservation and sharing. The advantage is that the user gets more guidance through the detailed structure of questions, follow-up questions based on previous answers and suggestions for answers for specific details like file formats and ontologies. However, some of the questions are very specific and should, in our opinion, be answered outside of a DMP or ROMP and other aspects cannot be provided at all. One example is the questions about quality control addressing important points but are not able to represent the complex quality control needed for Safe-and-Sustainable-by-Design (SSbD) projects, which is better covered outside of the DMP/ROMP as part of the protocols and SOPs.

Finally, ARGOS seems to be the closest match for the BIO-SUSHY and MACRAMÉ's requirements. Unfortunately, the tool expects that a DMP is created for each research output individually. This makes clear that it is designed for documentation of the approaches taken during or after generation and FAIRification of the research outputs and not for conceptualisation and giving guidance to the project partners at the beginning of a project. For the latter, the description should be able to cover groups of research outputs and should be more in the form of a guidance document. Therefore, it was decided not to use one of the online tools but to generate the ROMP using the structure from ARGOS as the starting point. This was then amended and complemented by adding individual questions or topics from the other two systems and considering the FAIR Principles for Research Software (FAIR4RS) besides the original FAIR principle for data. Sections added to the ARGOS structure or additions and clarifications to ARGOS sections introduced by BIO-SUSHY and MACRAMÉ are marked in green in Annex 1.

3. BIO-SUSHY data management infrastructure

Implementation of the guidelines provided by the ROMP is supported by the solution provided by the BIO-SUSHY knowledge and data exchange infrastructure used for first sharing all information within the project and then preparing the research outputs for public and FAIR sharing and long-term storage. The concept of a data lake, covering all experimental and computational data and providing it to further modelling applications and safety and sustainability evaluation as outlined in the description of action and visualised in Figure 1, was established in its first preliminary version and made accessible to all BIO-SUSHY partners.

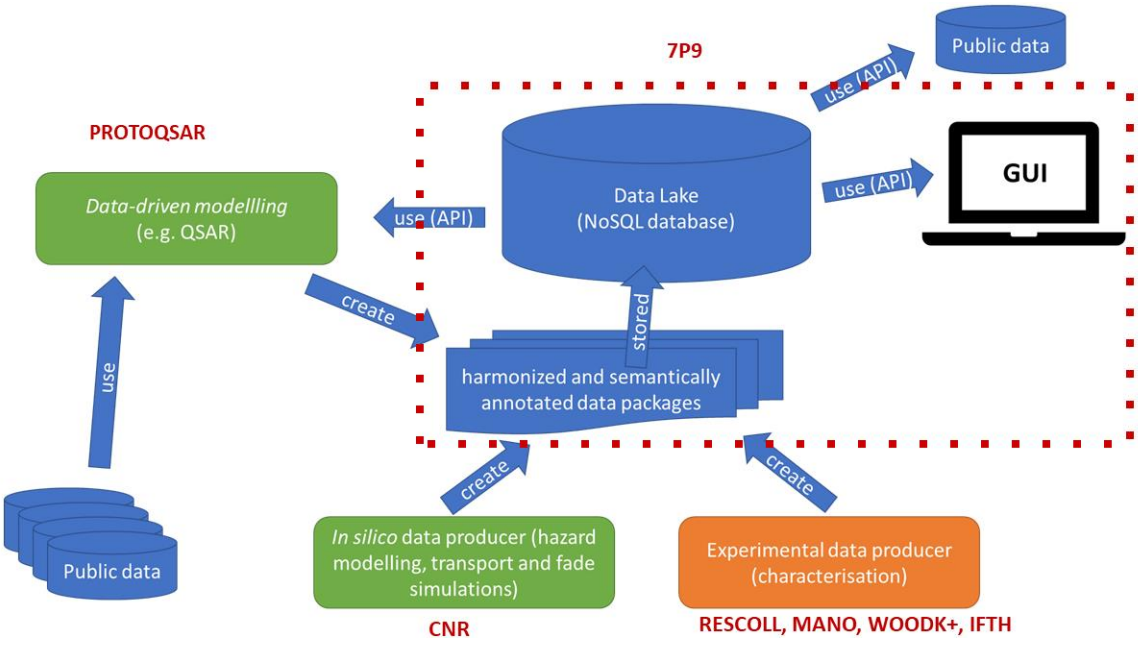

Figure 1: Data management concept in which all data from experiment, computational approaches and public resources is integrated to build the data lake used as input for data-driven modelling to guide the SSbD decisions for the development of the new coatings as well as to prepare upload to public databases)

It is composed of the central BIO-SUSHY Registry, in which all research outputs are indexed at the time work on them is started (use case and study design stage), the instance map tool, which visualises study designs and links relevant research outputs, and the BIO-SUSHY Collaboration Platform and Google Shared Folder for storage and sharing of intermediate and final research outputs. This will now be complemented with more specialised solutions for specific research outputs (e.g. electronic lab notebooks) and populated by the outputs and corresponding metadata from the BIO-SUSHY partners. Further information on the interplay between these solutions, efforts to improve the FAIRness of the research outputs and examples of research outputs of different types will be described in more detail in deliverable D3.1: "Knowledge and data exchange infrastructure specifications" due in Month 12 (December 2023) of the project.

4. Conclusions

BIO-SUSHY has created the first version of the Research Output Management Plan (ROMP) as an extension of a Data Management Plan (DMP) attached to this report as Annex I. It contains the agreed recommendations and guidelines to guarantee high-quality information and knowledge management covering all research outcomes including but not limited to data, protocols, models, software, surveys and responses, reports and publications and, thus, constitutes an important part of BIO-SUSHY's overall quality and risk management (see also Deliverable D6.2). To be able to cover all these different types of output, the ROMP was developed based on existing DMP questionnaires and then extended by considering the FAIR4RS, TRUST and CARE principles. The first version intended to define the requirements for the data management infrastructure and facilitate the knowledge collection and sharing within the project will now be continuously revised to document more and more details on the outputs generated within BIO-SUSHY and finally, function as the reference to all

research outputs available at the end of the project. These future updates of the ROMP will be discussed in relation to the knowledge exchange infrastructure in deliverable D3.1: Knowledge and data exchange infrastructure specifications, the periodic reports and D3.2: FAIR data sharing implementation and final Research Output Management Plan.

5. References

1. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
2. Barker, M. *et al.* Introducing the FAIR Principles for research software. *Sci. Data* **9**, 622 (2022).
3. Lin, D. *et al.* The TRUST Principles for digital repositories. *Sci. Data* **7**, 144 (2020).
4. Carroll, S. R. *et al.* The CARE Principles for Indigenous Data Governance. **19**, 43 (2020).

6. Annexes

Annex I: Initial Version of the BIO-SUSHY Research Output Management Plan (ROMP)

0 Main information

0.1 Project

BIO-SUSHY Sustainable surface protection by glass-like hybrid and biomaterials coatings

0.1.1 Grant information

Project Start: 01.01.2023

Project End: 31.12.2026

Funder: European Commission – Horizon Europe

Grant URL: https://doi.org/10.3030/101091464

0.1.2 Contributors

Name	ORCID	Affiliation	Role
Thomas Exner	0000-0002-1849-5246	7P9	Data Manager Principal Investigator
Joh Dokler	0000-0002-8053-8198	7P9	Data Manager
Maja Brajnik	0000-0002-7420-1388	7P9	Data Manager

0.2 Description

0.2.1 Objective of the project

Per- and polyfluorinated alkyl substances (PFASs) are a large group of manufactured chemicals that do not break down easily and can accumulate in the environment, food and human bodies. They are used in many consumer and industrial products including coatings. The EU-funded BIO-SUSHY project will develop innovative repellent organic and hybrid coatings, harnessing bio-based additives for advanced functionalisation as an alternative to PFAS-based coatings. A safe-and-sustainable-by-design strategy will be powered by physics-based and data-driven modelling tools for predicting the repellent properties of coating surfaces and the leaching mechanisms of the materials. A life-cycle assessment will determine the economic, social, and environmental impacts along the value chains of the case studies considered for the validation of the BIO-SUSHY coating materials.

0.2.2 Purpose of data generation and re-use

The BIO-SUSHY project will bring scientific advances by engaging in the development of new coatings for controlled wettability properties without PFAS at laboratory and pre-industrial scales. The modulable character of the coating materials will make it possible to easily adapt them to an extended area of substrates and end-uses. Moreover BIO-SUSHY will provide new material concepts that will improve recyclability of the coating materials thanks to the replacement of fluorine-based components. Thus, the development of the new coatings will be based and guided by Safe-and-Sustainable-by-Design methodology, through the life cycle of the new products. To this purpose, 3 R&I processes will be conducted to provide competitive coating solutions to replace PFAS with lower toxicities and environmental footprints in 3 use cases.

Data will be generated as part of a full physicochemical and functional characterisation and to be able to perform health and environmental risk assessments of the produced materials and additives, resulting in the establishment of criteria for safe- and sustainable-by-design coating materials, additives and products. To best support the design of the new coatings, a large variety of computational methods is provided all together building the BIO-SUSHY set of computational tools. These can be grouped into two categories, physics-based simulations and data-driven approaches like QSAR and read-across. Both, experimental (wet lab) and computational data together with associated resources including but not limited to study designs, protocols, computational models and software build the corpus of (meta)data discussed in this Research Output Management Plan.

0.3 Tags

To be defined.

0.4 Template

This ROMP is based on the Horizon Europe template taken from the ARGOS DMP platform (https://argos.openaire.eu/) with extensions (marked in green) by BIO-SUSHY needed to be applicable to all research outputs.

0.5 FAIR Implementation Profile (FIP) associated with the project

Selection of the FAIR Enabling and Supporting Resources (FERs and FSRs) is based on the general FIP for nanomaterial/nanosafety research developed by the WorldFAIR project. A FIP more specific to the BIO-SUSHY project is under development.

0.6 National / funder / sectorial / departmental policies and procedures for data management relevant to research output management

With this ROMP, BIO-SUSHY is fully implementing the transition from the Open Data initiative of H2020 to the Open Science initiative of HE especially by extending the management concepts from data to all research outputs including but not limited to study designs, method descriptions, protocols/SOPs, computation models and workflows, and software.

0.7 Support provided

Creation and implementation of the ROMP is supported by the (Meta)Data Shepherding service established as part of the BIO-SUSHY project. Additional support is provided by the WorldFAIR project as well as the GO FAIR AdvancedNano Implementation Network.

1 Summary

1.1 Brief description of the described research output

Areas of data collection, generation and curation around the three BIO-SUSHY use cases (textiles, food trays packaging, and cosmetic glass packaging) are:

1. Physicochemical characterisation of novel coatings
2. Functionality evaluation in comparison to existing PFAS- and non-PFAS-based coating
3. Human health and environmental safety assessment (*in vitro*, *in silico*) of raw materials, additives, catalysts
4. Interactions of coatings with substrates and biological systems on an atomistic level (*in silico*)
5. Safety and sustainability evaluations of coatings and final products including end-of-life.

1.1.1 What kind of research output are you describing?

This ROMP is meant to give a general description of all research output produced and collected by BIO-SUSHY. This includes, besides raw, processed, and robust summary data both experimental and computational, outputs like sampling plans, study designs, method specifications, protocols, SOPs, computational models, workflows and software as well as guidelines, reports, training materials and publications. Management and FAIRification are harmonised across all these kinds as much as possible. However, each of these outputs requires some specific treatment including e.g. selection of appropriate (meta)data structures, exchange file formats and public repositories for long-term storage. These are described either as separate sections for individual output kinds in this ROMP or in more specific future ROMPs if necessary to enhance readability.

1.1.2 Is it physical or digital?

This ROMP covers the digital objects generated or re-used in BIO-SUSHY. If required, physical objects produced by the project and managed using specific centralised and standardised resources (e.g. reference material repositories, biobanks) will be addressed in a specific ROMP on physical objects aligned to the ROMP on digital objects presented here as much as possible.

1.1.3 Are you generating or re-using it?

Digital research objects are generated as part of the BIO-SUSHY coating developments and SSbD evaluation. These are complemented by data and accompanying metadata mined and curated from public databases and scientific literature. Additionally, existing computational models, workflows and software will be re-used for simulating molecular interactions and predicting physicochemical characteristics, functionality, and human health and environmental safety.

1.1.4 What is the type of the described research output?

As described above, this ROMP covers all research output from BIO-SUSHY. It can be grouped into the following general categories (more specific categories will be added during the continuous updating of this living document), which is used (including numbering) for the further structuring of the answers in this ROMP:

1. Study designs
2. Method specifications
    a. *experimental*
    b. *computational*
3. Protocols / SOPs
4. Computational models
5. Software and computational workflows
6. Data
7. Surveys and survey responses
8. Guidelines
9. Reports and knowledge collections
10. Training materials
    a. Manuals / tutorials
    b. Videos
    c. Handbooks / reference material
11. Publications
    a. Dissemination material
    b. White papers
    c. Peer-reviewed papers

1.1.5 What is its format?

The variety of research output types listed above clearly shows that it is not possible to cover all of them in one exchange format. Additionally, different BIO-SUSHY partners entered the project with

very different knowledge management approaches applied in their internal processes and workflows, which have to be considered when creating the recommendations for project-wide data management and public sharing. Therefore, this initial ROMP lists all data formats currently in use and uses the (meta)data reported in these formats to develop recommendations for improvement of the data management practices at specific partners, select appropriate FAIRification tools for each setting, and develop (semi-)automated approaches to translate internal formats into harmonised and interoperable data documentation formats aligned with community and regulatory standards.

1. *Study designs*: instance maps
2. *Method specifications*: text documents, data templates (MODA, CHADA)
3. *Protocols / SOPs*: text documents, electronic lab notebooks
4. *Computational models:* data templates (QMRF, QPRF, MODA)
5. *Software and workflows*: data templates (QMRF, QPRF, MODA), notebooks (jupyter, colab), API definitions.
6. *Data*: customised spreadsheets, data templates (nanoFASE, eNanoMapper, MODA, CHADA), proprietary formats, data serialisation formats (json, yaml)
7. *Surveys and survey responses:* text documents, spreadsheets, Google forms
8. *Guidelines*: text documents
9. *Reports*: text documents
10. *Training materials*: text documents, slides, videos
11. *Publications*: text documents

1.1.6 What is its expected size?

Limited information on expected size is available at the current state. Similar projects had relatively low requirements for data storage capacities. However, the distributed data storage approach adopted by BIO-SUSHY will also be able to cover much higher demands if necessary. This can be achieved by e.g. storing large raw data files only locally at the data providers or using data storing solutions specialised for a specific data type (e.g. omics) or for general storage of big data as e.g. provided by the European Open Science Cloud (EOSC).

1.1.7 Why are you collecting/generating or re-using it?

Data is primarily generated and collected for re-use to satisfy the data needs of the BIO-SUSHY coating development in the three use cases including functionality and SSbD evaluation. The experimental and computational datasets will then be provided to guide defining SSbD criteria for coatings and, subsequently, as input for the standardisation and regulatory validation project initiated at the relevant bodies (e.g. OECD, VAMAS, ISO, CEN).

1.1.8 What is its origin / provenance?

Each research output, both generated by BIO-SUSHY and collected from third parties, will specifically document its origin and provenance as part of the set of standardised metadata.

1.1.9 To whom might it be useful ("data utility")?

Primary users of the data will be the BIO-SUSHY consortium partners performing the development and optimisation as well as SSbD evaluation on the use case coatings. Secondary users will then be developers of SSbD criteria for bio-based and hybrid coating materials as well as labs (academia and industry) interested in performing Safe-and-Sustainable-by-Design (SSbD) material development. The data will then also be made available for starting the standardisation and regulatory validation projects for integrating specific SSbD criteria for coatings supporting registration of such new materials.

2 Links Between Outputs

2.1 Publications

2.1.1 Does the described output support any scientific publication?

Due to the recent project start, BIO-SUSHY has not yet produced scientific publications supported by the newly generated research outputs. For data collected and mined from databases and literature, the corresponding references are stored as metadata providing clear provenance trails for the extracted information.

2.1.2 Is there a data availability statement provided along with the publication?

Data availability statements will be provided in all future publications resulting from the BIO-SUSHY project. These will include unique, persistent identifiers (DOIs or data-specific identifiers), licensing information, clear provenance trails, and references to the data models used if applicable.

2.2 Qualified references to other objects

2.2.1 Does the described output support any other research output?

BIO-SUSHY has adopted the approach of providing the different research output types as individual resources. Instead of creating complex (meta)data templates storing information on the methods, protocols and the generated raw and processed data, as e.g. implemented in the NIKC-NanoFASE templates or the MODA/CHADA system, all these components are managed using individual tools customised and optimised for each type (e.g. electronic lab notebooks, method-specific data formats and databases, version control systems). However, this results in the fact that many research outputs are needed to provide all information a specific piece of evidence or conclusion is based on. Managing all research output using the harmonised approach described in this ROMP has the advantage that all different types can be handled individually and, at the same time, outputs supporting each other can be linked together giving full access to all information relevant for e.g. a specific method or a BIO-SUSHY use case. For example, data include references to the methods and protocols used to generate it as metadata and methods can list all dataset used for testing and validation.

2.2.2 Is the research output integrated into a system of qualified references cross-linking outputs?

The concept of individual resources has the advantage of access to optimal data management tools for each type. However, such a distributed system is putting more demands on keeping track of all the resources supporting each other and consistency of the resource references since losing links between the resources would destroy data completeness, understandability, interpretability, and, thus, trust and ultimately re-usability. BIO-SUSHY has established an advanced system for research output cross-linking composed of two services, the BIO-SUSHY Registry and the instance map tool. The first assigns a unique, even if only internal identifier to each research output. Additionally, basic metadata on data provenance and accessibility, references to use cases and project partners are provided. The second service is then providing ways to build further cross-links between the research outputs based on the unique identifiers, e.g. linking protocols to the generated data, and to visualise these as instance maps representing life-cycle stages of the materials and the experimental workflow of production and assessment. For public sharing and long-term storage, all relevant research, e.g. supporting a scientific publication, can be extracted from the internal management solution by:

- Replacing the internal identifiers automatically with global and persistent identifiers from authorities like DataCite/Zendo (DOIs), European Registry for Materials (ERM), ORCID, Research Organisation Registry (ROR) or public data/protocol repositories;
- Packing all resources specified in one instance map into a data package following the frictionless data or RO-Crate specifications either as (meta)data files or as links to other public resources; and
- Publicly sharing of the data packages in FAIR data storing solutions.

2.3 Pre-existing data

2.3.1 Are you using any pre-existing research output?

Pre-existing research outputs will be re-used in the form of existing method descriptions, protocols/SOPs, computational models, workflows, and software as well as data mined and curated from public databases and scientific literature.

2.3.2 Is the pre-existing research output handled according to this ROMP including additional FAIRification if necessary? If not, provide links to DMP describing the treatment of these pre-existing research outputs.

The pre-existing data will be handled in the same way as newly generated data as described in this ROMP. Additional data management steps will include integration into the BIO-SUSHY Registry, provision of data provenance trails, and harmonisation of data models and transfer formats. Additionally, FAIRification steps will be described here if they become necessary for specific pre-existing resources.

3 Quality control, FAIR Practices and openness

3.1 Making research outputs findable, including provisions for metadata

3.1.1. What type(s) of persistent identifier(s) are used for the described research output?

BIO-SUSHY is using an internal set of identifiers, which are assigned by the BIO-SUSHY Registry or, in specific cases, the project's Google Shared Folder, are unique within the project and can be translated into globally unique and persistent identifiers at a later stage. In this way, the different (future) research outputs can be clearly identified starting from the planning phase on and global uniqueness and also indexing in the relevant identifier services is achieved for the final versions of the outputs meant for public sharing and long-term storage. Globally unique, persistent identifiers are currently available and integrated into community standards for:

1. *Sampling plans*: DOI (potentially)
2. *Study designs*: DOI (potentially)
3. *Method specifications*: DOI (Zenodo)
4. *Protocols / SOPs*: DOI (Zenodo, DataCite)
5. *Computational models, software and workflows*: GitHub
6. *Data*: DOI (Zenodo)
    a. *Materials:* ERM, NInChI
    b. *Chemicals*: CAS, InChI, (CAS)
    c. *Providers*: ORCID
    d. *Institutions*: ROR
    e. *Projects*: DOI (EU)
7. *Surveys and survey responses:* DOI (potentially)
8. *Guidelines, reports, training materials*: DOI, TeSS
9. *Publications*: DOI (from publisher)

3.1.1.1 Are components of the research output representing levels of granularity (software modules, experimental steps, materials) assigned distinct identifiers?

Study designs, protocols, materials/chemicals, models, workflows, software as well as surveys and their results, reports and publications are referred to by individual identifiers in the internal system. These can thus also be mapped to individual distinct global identifiers later. The system also encourages splitting the experimental procedure into multiple protocols, e.g. for sample preparation, measurement, and processing, all with their own identifiers. Splitting into even smaller parts (individual protocol steps) is in principle also possible but currently not envisioned. This decision will be periodically reviewed.

3.1.1.2 Are different versions of the method descriptions, protocols, software assigned distinct identifiers?

Development of all research outputs, which might exist in different versions, are managed in solutions with automatic version control (GitHub, Google documents). Stable versions used to generate results for specific studies will be specifically marked (named versions) and provided with distinct internal identifiers first and then global identifiers if they support specific public research outputs.

3.1.2 Will you provide metadata for the described research output? What metadata will be created?

All research outputs will be accompanied by metadata, which will be standardised and made richer over the runtime of the project. At the current stage, the following high-level metadata fields are mandatory for all research output referred to as resources in the BIO-SUSHY Registry:

1. Unique internal identifier
2. Resource name
3. Type of resource (e.g. Material, Test method, protocol)
4. Status (e.g. Scheduled, in preparation, under internal review)
5. Resource link (access path)
6. Short description
7. Contributors and their roles
8. Licence

3.1.2.1 What disciplinary or general standards will be followed?

The final metadata schema of BIO-SUSHY will be composed of archetypes, describing specific aspects like contributors, publications, (meta)data schema, and method-specific metadata. These archetypes will be constructed following existing standards. For example, contributors, institutions and publications will use the DataCite and Dublin Core specifications. For disciplinary components, standards are currently under development or revision (MODA/CHADA, eNanoMapper-based templates) and will be integrated when available.

3.1.2.2 In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

The new production processes and test methods developed in BIO-SUSHY will require method-specific metadata to fulfil minimum reporting requirements. These are currently under development based on the expertise from earlier projects and the first results provided by the BIO-SUSHY partners. They will be based on existing minimal reporting guidelines and regulatory requirements but will provide the flexibility to report metadata specific to the process/method. To guarantee interoperability to the highest possible extent, the metadata models/schemas implemented in the reporting guidelines will be provided as high-level metadata to the research output (data but also structured information from e.g. protocols, study designs).

3.1.3 Will search keywords be provided in the metadata to optimise the possibility for discovery and then potential re-use?

Provision of structured (meta)data with well-defined (meta)data schemas can be used for data discovery within BIO-SUSHY and potential re-use using advanced searching/browsing features. These will heavily rely on the unique identifiers for materials, methods and protocols as well as semantic annotated metadata. Additional full text search will provide additional means to find relevant research outputs. If this does not satisfy all needed for information discovery, additional search keywords will be provided.

3.1.4 Will metadata be offered in such a way that it can be harvested and indexed?

Metadata is available in a structured format in the BIO-SUSHY Registry and by providing the (meta)data schemas as high-level metadata. Options on how to perform harvesting and indexing will be provided as part of the metadata accessible via the application programming interfaces (APIs) of the Registry (project internal) and potentially of the long-term storage solutions (external).

3.2 Making research outputs accessible

3.2.1 Repository

3.2.1.1 In which repository will the dataset / output be deposited?

The research output is currently managed and stored in the internal BIO-SUSHY Registry and different data storage solutions (mainly Google Shared Drive). Options for long-term storage are currently evaluated and selected based on their fitness for the specific output and their FAIRness considering general solutions like Zenodo, open institutional repositories, and domain-specific data warehouses.

3.2.1.2 Is the selected repository a trusted source?

Long-term storage solutions will be selected based on community recommendations at the point of time. Since the research output is prepared according to the high BIO-SUSHY FAIR standards already for internal data sharing, it will be ready to be publicly released after semi-automatic transformation into the format requested by the storage solution. This allows a flexible selection of solutions, which will only consider trusted sources, according to the needs of the individual research output.

3.2.1.4 Are appropriate arrangements made with the repository(ies) where the described dataset will be deposited

Arrangements will be made when the selection of the long-term storage solution is finalised.

3.2.1.5 Does the repository(ies) assign research outputs with persistent identifiers?

Only solutions providing persistent identifiers will be considered in the selection process.

3.2.1.7 Does the repository support versioning?

Only solutions providing versioning (when required by a specific research output like software) will be considered.

3.2.2 Data

This section will be complemented whenever new research outputs become available since the answers need to be specific to these outputs. Currently, only general aspects of the internal data management system (BIO-SUSHY Registry and instance maps) are given.

3.2.2.1 What is the described research output title?

"BIO-SUSHY (meta)data" (Additional and more specific titles will be added during the integration of specific research outputs.)

3.2.2.2 How is the research output shared? Specify reasons for the type of sharing selected (fully open, restricted, confidential) and embargo period, if applicable, separating legal and contractual reasons from intentional restrictions.

Currently, all research outputs are only internally shared via the BIO-SUSHY Registry. This is the case since none of the outputs have already matured to their final version. When this status is reached for a specific output, sharing decisions are made on a case-by-case basis with preference for fully open sharing and licences allowing re-use in most situations as outlined in the consortium agreement.

3.2.2.4 Will the research output be accessible through a free and standardised access protocol?

Final, long-term access will be provided using existing web-based FAIR data solutions with free and standardised access protocols (http, ftp and specific API endpoints).

3.2.2.5 Are there any methods or tools required to access the research output?

All outputs will be provided without requiring any specific methods or tools other than standard applications (e.g. pdf). However, if relevant, raw data will be provided in proprietary formats since they offer re-use in advanced, data-type-specific analysis software.

3.2.2.8 Is the described research output supported by a data/research output access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

The BIO-SUSHY General Assembly will function as the research output access committee guaranteeing that no confidential data is shared unauthorised. Since no sensitive personal data is meant to be collected in BIO-SUSHY, an additional external committee is deemed non-essential.

3.2.2.9 Please specify how the research output will be accessed during and after the project ends especially if restrictions are in use.

TBD

3.2.2.10 Please specify how long after the project has ended the research output will be made accessible for.

TBD on a case-by-case basis.

3.2.2.11 How will the identity of the person accessing the data be ascertained?

The BIO-SUSHY Registry, instance map tool and the Google shared drive used for temporal data storage are all protected by state-of-the-art authentication and authorisation management and accessible only after logging in using a personal account. In this way, ascertaining and protocolling the identity of persons accessing research outputs is assured.

3.2.3 Metadata

3.2.3.1. Will you provide metadata even if the described research output cannot be openly shared?

Metadata for all research outputs are collected independent of their final usage and planned sharing options. This will also include information on how to get access to the output and the person responsible to handle all requests (see Section 3.1.2 above).

3.2.3.2. Under which licence will metadata be provided? Specify reasons for selecting the licence separating legal and contractual reasons from intentional restrictions.

Metadata will be shared under the Creative Commons Attribution 4.0 License (CC-BY-4.0) if not prohibited for legal or contractual reasons. If such cases become relevant in the future, the reasons will be specified here.

3.2.3.3. Do metadata provide information about how to access the described research output?

Information on the access routes and the person responsible for handling all requests will be provided as metadata to every research output. Currently, this is managed by the BIO-SUSHY registry providing links to the resources and responsible partners. This will be replaced with information on long-term storage solutions once the research output is publicly shared.

3.2.3.4. Do metadata provide a full description of the data model used for the research output (including input and output formats) or a link to such a description?

It is anticipated that all research outputs provide the underlying (meta)data model as part of their high-level data documentation. This will be established as part of the additional FAIRification

following the currently ongoing evaluation of the reporting and management approaches at the individual BIO-SUSHY partners.

3.2.3.5. Will metadata remain available after the dataset/output is no longer available?

Indexing of the metadata in standard FAIR Supporting Resources like Zenodo is planned guaranteeing availability even after the output is no longer available.

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

Controlled vocabularies will be used whenever available. On one hand, the DataModel Ontology and/or the Information Artefact Ontology will be used for the high-level data documentation including the description and semantic annotation of the (meta)data model. On the other hand, for the low-level annotation of method-specific metadata, different ontologies like the eNM ontology, the EMMO as well as multiple chemical and biological ontologies are available. However, it is expected that these will not cover all relevant aspects and BIO-SUSHY will collaborate with other projects to increase the ontological coverage.

3.3.2 If you created the vocabulary, where can it be found?

Terminology resulting from BIO-SUSHY's ontology work will be integrated into existing ontologies and will therefore be available from the services providing these ontologies (e.g. BioPortal).

3.3.3 Have you applied a standard schema for your (meta)data?

As described above, harmonisation has been started with standardisation of high-level metadata re-using (parts of) existing standards (DataCite, Dublin Core). This will be complemented by schemas for the method-specific (meta)data documented as part of the low-level metadata. These are based on existing standards and minimum reporting guidelines and the enhanced versions created by BIO-SUSHY (in intensive collaboration and alignment with other projects) will be proposed for standardisation.

3.3.5 What is the methodology followed?

The following concepts, partly already described in previous sections, define the BIO-SUSHY methodology:

- Individual management of research outputs according to their types allows optimal selection of tools for curation, documentation and sharing (e.g. electronic lab notebooks for protocols).
- Separation of high- (biographical metadata, access options, licences, documentation of data model) and low-level (method-specific metadata) data documentation provides interoperability and computer-actionability for different applications (data discovery vs. data integration into computational workflows)

- Metadata is structured into sections described by archetypes based on existing standards (e.g. DataCite for biographical metadata)
- FAIRification will be continuously improved over the runtime of the project by aligning with and integrating emerging concepts and standards and proposing new or improved standards when necessary.

3.3.6 What community-endorsed interoperability best practices are followed?

BIO-SUSHY partners have been / are engaged in multiple activities defining interoperability best practices, which form the basis for this ROMP. This has started with the OpenRiskNet and NanoCommons projects defining FAIRification and FAIRness assessment approaches as well as the roles responsible for specific tasks and is now continued in the WorldFAIR project (cross-domain interoperability) and GO FAIR AdvancedNano Implementation Network. Additionally, collaboration with the European Materials Modelling Council and the OntoCommons project have been started.

3.3.6.1 What domain-relevant community standards are used for reading, writing, and exchanging data?

QMRF/QPRF and MODA/CHADA formats, eNanoMapper-based and NIKC/NanoFASE data curation templates, ISA-TAB-nano format as well as the OECD harmonised templates have been proposed as community standards for data exchange. However, all these formats are currently under evaluation by many projects to increase their FAIRness and computer-actionability. Therefore, BIO-SUSHY decided to postpone the decision on using a specific standard and to use first project-internal, highly structured reporting formats with clearly defined (meta)data models provided as high- and low-level metadata. Mapping of these internal (meta)data models to the models implemented in the improved standard file formats will allow automatic translation into these new formats as soon as they have been endorsed by the community.

3.3.6.2 In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow re-using, refining, or extending them?

Ontology and vocabulary development in BIO-SUSHY will only be performed, if at all, in intensive collaboration with other projects (e.g. OntoCommons, ELIXIR), which have established solutions for ontology publishing, re-using and mapping.

3.4 Increasing data and other outputs re-use

3.4.1 What internationally recognised licence will you use for your research output?

A licence is assigned to each individual BIO-SUSHY research output by the producer. They can choose between different Creative Commons licences. CC-BY-4.0 will be given preference if no reasons dictate otherwise.

3.4.1.1 How is the ownership of the collected data arranged?

Ownership of the generated research outputs is fully in the control of the producer. This includes licensing as well as selection of the sharing and long-term-storing options. Guidance will be given by the BIO-SUSHY (meta)data Shepherd.

3.4.1.2 Is there a clear procedure for the decision process selecting the licence, sharing option and embargo periods?

The BIO-SUSHY procedure for licensing and sharing is applied to each individual research output. Corresponding contractual obligations for the implementation of the action are specified in the BIO-SUSHY Consortium Agreement and, in the cases of Key Exploitable Results (KERs), in the Dissemination and Exploitation Plan.

3.4.2 What re-usability and / or reproducibility methods are followed?

Clear licences, rich metadata and especially the linking of all resources (study design, method description, protocols for sample preparation, measurement and processing, models, workflows and software, and raw and processed data) supporting each other via persistent identifiers or even combining all these resources into one self-contained data packages is dramatically increasing repeatability, re-usability and reproducibility.

3.4.3 Will you provide the described research output in the public domain?

BIO-SUSHY is following the policy of "as open as possible, as closed as necessary". All research outputs generated developing the coatings and for the internal validation of the new BIO-SUSHY methods will be publicly shared.

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

The BIO-SUSHY use cases including all generated and collected data/knowledge regarding the production of the coatings and their Safe-and-Sustainable-by-Design (SSbD) evaluate are meant to be provided as case studies for the SSbD development of 2D materials according to well-defined criteria. This will guarantee that the knowledge available from the BIO-SUSHY research outputs (data but even more important descriptions of formulation optimisation procedures, computational models and SSbD criteria for raw materials, formulations and final coating products) will be re-used in the future.

3.4.5 Is provenance well documented?

As already described above, provenance trails are important metadata associated with each research output, both generated in BIO-SUSHY and re-used in BIO-SUSHY coming from model libraries, database and text mining activities.

3.4.6 What documented procedures for quality assurance do you have in place?

All BIO-SUSHY protocols and SOPs will provide detailed descriptions of the quality-control and assurance procedures that were applied and are recommended when re-using them. This follows

the high standards of Good Practices (GxP) and requirements from bodies like OECD and ECHA. The same level of quality-assurance measures will also be applied to data documentation and management facilitated and guaranteed by the (meta)data shepherding service.

3.4.7 How will you provide documentation needed to validate data analysis and facilitate re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, training sets, unit tests, etc.)?

Study designs, protocols/SOPs, methodology and model descriptions including training and test datasets, analysis and modelling workflows, and documentation of quality assurance are central components of the BIO-SUSHY research outputs and give detailed descriptions of all steps including but not limited to sample preparation, measurement, and processing / data cleaning, and model development and validation all with corresponding descriptions of quality control and assurance measures. Rich, structured metadata will be provided as part of these research outputs providing important settings and parameters. Software used and customised as part of the BIO-SUSHY project will be documented, shared as open source if possible, and accompanied by training and test sets. BIO-SUSHY will provide functionality to pack all this information into self-containing data packages, which will provide all necessary information to facilitate validation and re-use.

3.4.8 Will you monitor data integrity once it has been collected?

During the generation of the self-contained data packages, integrity of the information is thoroughly checked. These will be used to provide the data to the long-term storage facility and indexing services and can be revisited if integrity issues become evident (e.g. if a long-term-storage service is going out of business). They are also the primary place if further data curation is needed to improve the completeness of the data or correct errors identified during re-use of the data. All services to which the data is uploaded or in which it is indexed can then reinsert the data clearly marking the changes or providing it as a new version.

4 Allocation of Resources

4.1 Collaborative platform

4.1.1 Are you using a shared working space to work with your data?

The BIO-SUSHY Collaboration Platform, BIO-SUSHY Registry, instance map tool and Google Shared Drive form the basis for the shared working environment. This can be extended with additional shared services like electronic lab notebooks, version control systems (e.g. GitHub), and database systems if required.

4.1.2. Are you using a centralised or distributed data storage system? How are the resources connected in the distributed system?

BIO-SUSHY is using a mixture of centralised and distributed data storage. The BIO-SUSHY Registry is the central place providing information on existing research outputs (both final and under

development) and how they can be accessed. The actual resources are then stored in a distributed system allowing integration of optimal solutions for the different research output types and customisation reflecting the different data management approaches at the different partners.

4.1.3. Are you using centralised and harmonised approaches for developing or modifying the workflow for data processing and analysis?

No centralised approaches are used at the moment. This is necessary to not slow down the method development and sharing of preliminary data for the use cases. The provided data and other research outputs are currently being analysed to find similarities in the used approaches and start harmonisation of the data reporting including processing and analysis. This analysis will then be used to generate centralised services whenever possible.

4.1.4 How are you monitoring progress?

The combination of the BIO-SUSHY Registry and the instance maps are used to monitor progress. Instance maps are used to plan the needed experiments and other data collection activities. The resources (research outputs) documenting these activities are then accessible in the BIO-SUSHY Registry by direct links from the instance maps and are clearly showing the status of completion.

4.2 Data management and FAIRification costs

4.2.1 What is the cost of making the described output FAIR?

BIO-SUSHY approach is to base data management and FAIRification on the internal processes established by the partners and, in this way, minimise additional effort. However, to also show the importance of these tasks and especially making the research outputs interoperable, parts of the budget reserved for experimental work has been assigned to research output documentation.

4.2.2 How is this cost covered?

Budget for the experimental/computational work in BIO-SUSHY including the required data management activities is split between WP2, WP3, and WP4 with parts of the budget of data producers reserved for data management. Additionally, budget for data management infrastructure provision and the data shepherding service are allocated in WP3.

4.3 Identify the people who are responsible and their role(s) in the management of the described output

- Data reporting, curation, and FAIRification: all partners of WP2, WP3 and WP4
- Data quality control: all partners of WP2, WP3, WP4
- Data completeness evaluation: all partners of WP3
- Data infrastructure provision, data transformation: 7P9-SI, 7P9-DE
- Data shepherding service: 7P9-DE, 7P9-SI

5 Security

5.1 What security measures are followed?

Strong data security measures are applied in BIO-SUSHY. All collaborative services and the data stored in them are secured by state-of-the-art authentication and authorisation mechanisms. Hosting on European servers is also guaranteed. Data transfer is using secure and encrypted communication (https). Additionally, the distributed knowledge management approach allows local storage of highly sensitive information. Finally, backup solutions are implemented for disaster recovery.

5.2 What conditions do the security measures meet?

- Authentication and authorisation mechanisms
- Monitoring of access
- Privacy protection and clear conditions of use
- Encrypted data transfer
- Backup solutions
- Constant application of security updates

5.3 How will you preserve the described research output in the long term?

Long-term storage will be using existing and emerging standard solutions (e.g. NanoCommons Knowledge Base, Zenodo, ELIXIR, and EOSC). Negotiations with the corresponding services providers are ongoing.

6 Ethical Aspects

6.1 Are there any ethical or legal issues that can have an impact on sharing the described research output?

Full ethical approval has and will be obtained for all experimental work performed within BIO-SUSHY. Sharing research outputs is not expected to add any ethical or legal issues.

6.2 Does the described research output contain sensitive information?

Confidential data might be included in research outputs supporting the use cases. If and when this data can be shared and with whom will be decided by the use case partners on a case-by-case basis.

6.3 Does the described research output contain personal data?

Personal data on the providers of the research outputs is collected as part of the data provenance trails. This includes names, addresses, and email. These will be handled according to EU's General Data Protection Regulation (GDPR).

7 Other Issues

7.1. Do you make use of other procedures for data management?

Currently, no other data management procedures are envisaged. If this changes, these will be described in detail here.