BioMeld

Project:

BioMeld

Grant Agreement (GA) No. 101070328

"A MODULAR FRAMEWORK FOR DESIGNING AND PRODUCING BIOHYBRID MACHINES"

Call: HORIZON-CL4-2021-DIGITAL-EMERGING-01
Type of action: Research and Innovation action (RIA)
Start date of project: 01/10/2022 Duration: 36 months

D5.2: DATA MANAGEMENT PLAN

DELIVERABLE FACTSHEET

Project title | Acronym | Number		A Modular Framework for Designing and Producing Biohybrid Machines | BioMeld | 101070328	
Due Date:	31/03/2023	**Date of submission:**	30/03/2023
Month of Project	M01	**Month of submission:**	M06
Title of deliverable:	D5.2 – Data Management Plan	**Work Package:**	WP5 – Management and Impact
Dissemination level:	PU	**Version/Status**	1.0/FINAL
Deliverable leader (Name|Organisation)	Igor Balaz|UNSPF	**Editor(s)**	Entire consortium
Contribution of partners	UNSPF compiled report, all other partners gave comments and suggestions for improvements.		
Final review and approval	All consortium partners		
Keywords	Data management plan, metadata, repository		
Abstract	The Data Management Plan for the BioMeld project outlines principles for managing publicly-funded research data. These principles include making data openly available with few		

	restrictions, adhering to relevant standards and best practices, preserving data with long-term value, recording sufficient metadata for discoverability and re-use, providing information on accessing supporting data, considering data release throughout the research process, allowing limited privileged use of collected data for publishing results, acknowledging the sources of research data, using public funds to support data management, and using cost-effective mechanisms to maximize research benefits.

Document change history

Date	Authors	Description
06/03/2023	Igor Balaz	Initial version of the report created
30/03/2023	Igor Balaz	Final version of the report after consultations and comments from all consortium members.

CONSORTIUM

	Name	Short Name	Country
1.	UNIVERZITET U NOVOM SADU, POLJOPRIVREDNI FAKULTET NOVI SAD	UNSPF	Serbia
2.	SCUOLA SUPERIORE DI STUDI UNIVERSITARI E DI PERFEZIONAMENTO S ANNA	SSSA	Italy
3.	FUNDACIO INSTITUT DE BIOENGINYERIA DE CATALUNYA	IBEC-CERCA	Spain
4.	SMART SENSING S.R.L.	SMART SENSING	Italy
5.	UNIVERSITA DEGLI STUDI DI CAGLIARI	UNICA	Italy
6.	LEVERETTE LANCE	Lance Leverette	Belgium
7.	The University of the West of England	UWE Bristol	United Kingdom

EXECUTIVE SUMMARY

The Data Management Plan of the BioMeld project adheres to the following principles: (i) The publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner. (ii) Institutional and project-specific data management policies and plans should be in accordance with relevant standards and community best practice. (iii) Data with acknowledged long-term value should be preserved and remain accessible and usable for future research. (iv) To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. (v) Published results should always include information on how to access the supporting data. (vi) To ensure that the research process is not damaged by the inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process. (vii) To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake funded

work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils. (viii) In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed. (ix) It is appropriate to use public funds to support the management and sharing of publicly-funded research data. (x) To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

TABLE OF CONTENTS

1 DATA SUMMARY

Long term vision of BioMeld project is to create a framework for designing and producing biohybrid machines (BHMs). Within BioMeld we defined four specific objectives:

Objective 1: To develop a modelling and simulation framework for the digital design of BHMs.

Objective 2: To fabricate a set of BHM-based modules for a reconfigurable modular catheter.

Objective 3: To validate the framework.

Objective 4: To construct the demonstration BIMC (bio-intelligent manufacturing cell).

Reaching any of these objectives will require generation and/or collection of specific data:

- Source code data (Objectives 1, 3, and 4) ;
- Simulation output files (Objectives 1 and 3) ;

- Analysis of simulation output files (Objectives 1 and 3) ;
- Characterization and collection of available data on physico-chemical properties of materials and living cells that could be used for the construction of BHMs;
- Collection of experimental results of the behaviour and control of constructed BHMs (Objectives 2 and 3); and
- Collection and analysis of BIMC functioning data (Objective 4).

In summary, data collection/generation follows different procedures for each of the proposed objectives and within BioMeld we will create at least 6 separate dataset types. Since reusability of datasets between different research groups is essential for the success of the project, the development of DMP is equally important.

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

Yes, we will use some of the existing data. Primarily from previous experiments, and some segments of existing code.

What types and formats of data will the project generate or re-use?

- Data from computer models, and simulations, represented as text, binary, or graphics files and videos.
- Measurements of BHM behaviour and control, represented as text of graphic files and videos.
- Parameter datasets, represented as text files.

Data and metadata will be requested, stored, and transferred in a comma-separated values (CSV) format. To facilitate the data exchange, MS Excel-compatible files including comma-separated and .xls(x) format will be also accepted. For statistical purposes, other formats include .sas7bdat (SAS), .RData (R), .SAV (SPSS), .mat (matlab). Where applicable data formats may be migrated when new technologies become available and are proven robust enough to ensure digital continuity and continued availability of data.
Specifically:

- SSSA: Microscope images in .png/jpg, Videos in .avi or .mp4, graphs in .png/.svg, raw data in .txt, or in Excel as .xlsx and in GraphPad as .pzfx;
- UNICA: Origin files, Excel files, images (.png, .tiff), videos (.mp4, .avi), PCB files (.grb), 3D schematics (.svg, .stl);
- IBEC: Microscope images in .png/.jpg. Videos in .avi, graphs in .png/.svg, raw data in excel and .svg. 3D designs (.svg, .stl), microscopy images (.tiff), data sets from simulations (.csv or .txt);
- UWE & UNSPF: Program codes. Videos and photos of simulated environment. Numerical values and graphs of optimization process.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

Generated and re-used data will be used

- as data storage for material characterization;
- to lead the fabrication of flexible electronic platform and readout electronics;
- for reproducing the samples/prototypes;
- To make design of future models efficient.

What is the expected size of the data that you intend to generate or re-use?

- Numerical data related to optimisation c. 50 Gb per project's lifetime;
- Full history of the 3D models: c. 10Gb;
- Videos of the computer models and animation of results: c. 100Gb;
- Videos of the BHM experiments: c. 100Gb;
-

The above are estimates. To be evaluated during the course of the project. The expected size depends on the extend and the nature of the data that are made available.

What is the origin/provenance of the data, either generated or re-used?

- Written source code;
- Published scientific articles;
- Outputs of *in silico* experiments;
- Outputs of tools for mathematical analysis;
- BHM experimental test;
- BHM characterization results.

To whom might your data be useful ('data utility'), outside your project?

- BioMeld consortium;
- European Commission services and European Agencies;
- EU National Bodies;
- The general public including the broader scientific community
- Manufacturers of BHMs

2 FAIR DATA

2.1 MAKING DATA FINDABLE INCLUDING PROVISIONS FOR METADATA

Will data be identified by a persistent identifier?

Yes, each dataset wih have assigned an unique DOI.

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Yes. All EVO-NANO generated data are stored at Zenodo repository (https://zenodo.org/). Zenodo closely follows FAIR principles:

- Each uploaded dataset gets DOI;
- Metadata for individual records are retrievable by their identifier using a standardized communications protocol;
- Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it;
- Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation;
- Metadata from microscope images and videos will bein .txt format;
- Metadata from images and from data sets in .xml format.

All data will be discoverable via medata provision. All data will be identifiable and referable via standard identification mechanism (DOI). A unique online naming conventions are adopted. The data will be searchable by keywords. Clear versioning will be in place.

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

Yes. The dataset information reported into the metadata fiche will be published in BioMeld, where specific filters, based on the metadata elements, will allow to refine the search across datasets.

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

Data will be stored on Zenodo that allows harvesting entire repository via the Open Archives Initiative Protocol for Metadata Harvesting (OAI- PMH). OAI-PMH is a widely used protocol for harvesting metadata and most popular repository software provide support for this protocol.

All metadata is licensed under Creative Commons Zero, while the data files may be either open access and subject to a license described in the metadata or closed access and not available for download.

2.2 MAKING DATA ACCESSIBLE

Repository:

Will the data be deposited in a trusted repository?

All generated datasets within BioMeld will be uploaded to Zenodo repository (https://zenodo.org/).The data sharing should occur in a timely fashion.This means that the data resulted from the research conducted in the project should become available close to the project results themselves. Furthermore, it is reasonable to expect that the data will be released in waves as they become available or as main findings from waves of the data are published.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Yes.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Yes.

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Restriction is foreseen only for data whose open availability can interfere with planned IP protection measures. They will be released in two cases: (i) if the consortium decides that they will not pursue IP protection for results connected to the dataset in question, or (ii) if the IP protection will be activated, then the release of a dataset will be done only to the extent that do not jeopeardize IP protection.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Creation of each dataset will be accompanied by the Consortoim meeting, at most within two weeks, to discuss whether its publishing is relevant to IP protection.

Will the data be accessible through a free and standardized access protocol?

Since Zenodo stores data as publicly accessible, the only requirement is internet access. With regards to open software, all the data needed to create and maintain the marketplace

is being made openly accessible through the GitHub repository, along with the corresponding technical documentation.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

Zenodo protocol ensures there are no restrictions on use.

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

The data will be accessible upon open access principles. The initial access to the data will be without any login or user account

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

All validated data will be publicly available, therefore a data access committee is not necessary.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

Yes, all metadata in Zenodo will be freely used under the CC0 waiver.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

At Zenodo, metadata are accessible, even when the data are no longer available. Data and metadata will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least. Metadata are stored in high-availability database servers at CERN, which are separate to the data itself.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

Software will be shared via GitHub, which is directly linked with Zenodo.

2.3 MAKING DATA INTEROPERABLE

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

We will aim to develop an ontology to enable users to reason if their outputs are suitable for analysis or visualisation using a particular technique, and to use standard provenance reasoning techniques to gain some understanding of the processes by which the output was generated.

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 reusing, refining or extending them?

Yes, if needed.

Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

Yes, if data from previous research published or not published may be relevant to the project scope, and, if applicable data from GitHub and Zenodo will be referenced.

2.4 INCREASE DATA RE-USE

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

A README text file will be included with the data files and will contain all necessary information about the data files to allow others to understand the data. The README.txt file for each database will be created and will contain:

- title of the dataset
- dataset overview (abstract)
- file structure and relationships between files

- methods of data collection
- software and versions used
- standards
- specific information about data (units of measurement, explanations of abbreviations and codes, etc.)
- possibilities and limitations of data reuse
- contact information for the uploader of the dataset

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Yes.

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Yes.

Will the provenance of the data be thoroughly documented using the appropriate standards?

Yes, as explained above in the format of a README.txt file.

Describe all relevant data quality assurance processes.

The data quality is ensured by different measures. Quality assurance concerning accuracy and completeness of data and metadata will be performed by the BioMeld uploader. The data will be checked automatically during the upload process and manually by BioMeld uploader.

3 OTHER RESEARCH OUTPUTS

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

4 ALLOCATION OF RESOURCES

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

Exact costs estimated will be known and adjusted dynamically during the project's lifetime.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

The costs for depositing the dataset with the project, and subsequent resources required to make the dataset publicly available have been included within specific WPs within the project.

Who will be responsible for data management in your project?

The project coordinator has the ultimate responsibility for the data management in the project and so, for the platform management.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Zenodo ensures data storage for the next 20 years. Due to the data being shared via public repositories, the preservation beyond lifetime of the project does not involve any costs.

5 DATA SECURITY

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Due to the data volume, Zenodo also hold a copy of their own processed data, effectively acting as a second distributed database and additional backup. Locally, within each partner, all data will be stored at backup external hard-disks.

Will the data be safely stored in trusted repositories for long term preservation and curation?

The digital signature of the whole dataset, or the storage of the dataset in a git repository could provide support for the correct duplication and preservation. In addition zenodo operates with 12-hourly backup cycle with one backup sent to tape storage once a week.

6 ETHICS

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

All currently identified ethical parameters to be obeyed will be dealth with, in the WP1 “Ethics Requirements” This work package sets out the 'ethics requirements' that the project must comply with, which will be reported in D1.1. This Deliverable have not yet been finalized and additional details will be reported, as needed, in future versions of the DMP.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

Yes. Full details on when informed consent is needed and how to get them, e.g. addressing informed consent procedure for communication with stakeholder will be provided D1.1 on “Ethics requirements“. This deliverable is currently work in progress.

7 OTHER ISSUES

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

No