



Josephine

Deliverable 1.4: Data Management Plan



**European
Commission**

Horizon Europe
European Union funding
for Research & Innovation

Document ID:	D 1.7	Main author:	C. Zederman, CNRS
Work Package:	WP1: Coordination		
Document version:	1	Dissemination level:	Public
Project:	High-Tc Josephson Neurons and synapses: towards ultrafast and energy efficient superconducting neuromorphic computing		
Short name:	JOSEPHINE	Grant number:	101130224
Coordinator:	CNRS	Project coordinator:	Dr. Javier E. Villegas
Call identifier:	HORIZON-EIC-2023-PATHFINDEROPEN-01		

Disclaimer:

The Josephine project has received financial support from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101130224. However, this document reflects only the Josephine Consortium's view and the Commission cannot be held responsible for any use that may be made of the information it contains.



Table of content

Introduction	4
1 Data Summary	4
1.1 Purpose of the data collection / generation	5
1.2 Data types and formats	5
1.3 Origin of the data and possible re-use of existing data	6
1.4 Expected size of the data	7
1.5 Data utility	7
1.6 Data storage	7
2 FAIR Data	8
2.1 Making Data findable, including provisions for metadata	8
2.2 Making data openly accessible	9
2.3 Making data interoperable	9
2.4 Increase data re-use (through clarifying licences)	9
3 Allocation of resources	10
4 Data security	10
5 Ethical aspects	11
6 Other issues	11
7 Annex 1: Josephine metadata template form	12



Introduction

The Josephine project has been thought with a focus on open science practices. Our consortium will be fully committed to make the data from its project results findable, accessible, interoperable and reusable, following the FAIR principles adopted by the European Commission.

This deliverable is a first version of the Data Management Plan (DMP) which will be updated during the course of the project (deliverables 1.5 & 1.6). It provides guidelines for the data management policy to be applied by the Partners to datasets generated within the Project. In particular, the DMP identifies the main data to be generated within Josephine, outlining the handling of research data during the project as well as how and what parts of the datasets will be openly shared. This document is intended mainly for consortium internal use, aiming to provide guidance to Project Partners on data management. The DMP is the central tool to agree on data processing within the consortium, facilitating the creation of a common understanding and common practices.

This deliverable is submitted to the European Commission (EC) at Month 6 of the project and has been produced following the EC guidelines. The DMP reflects the provisions established by the project Grant Agreement (GA) and will be an integral part of the exploitation and dissemination strategies.

1 Data Summary

The Josephine project will generate data – including associated metadata – in a wide range of R&D activities, including those needed to validate the results of the project that will be presented in scientific publications and those associated to reports and other documents.

The format of the data and associated metadata will be mainly electronic and may include laboratory measurements and records, schemes, technical protocols, datasheets and performances of the technological developments of the project, meeting reports and presentations, demonstrator videos, pictures, device design and validation data.

In this paragraph we list the types of dataset that will be produced by each Josephine partner and the structure of their content.



1.1 Purpose of the data collection / generation

All partners may generate data related, but not limited to the characterization and optimization of hybrid interfaces for neuromorphic applications, but also the different structures needed for the neuromorphic demonstrator and the modelling of such structures.

Partners will provide:

- data background for reports
- data to share between the partners
- data background for scientific publications

1.2 Data types and formats

In particular, data will be related to:

- Morphology of films and interfaces- AFM, STM, SEM and optical images (.flt, .gwy, tif, .tiff, .gif, .jpeg, jpg, .pdf, .mvd, .psd, .bmp)
- Magnetic characterization of AFM layers, MOKE magnetization data (.dat,.opj, .opju,)
- Crystallographic characterization - XRD (.asc, .ras., .opj, .opju,)
- Morphology and crystallinity of YBCO films and interfaces- AFM, SEM images and XRD maps (.gwy, tif, .tiff, .gif, .jpeg, jpg, .pdf, .mvd, .psd, .bmp)
- Current voltage characteristic (.opj, .opju,.dat)
- Superconducting properties - critical temperature, critical current density (.dat, .opj, .opju,)
- Magnetotransport of superconducting devices (.dat, .opj, .opju,)
- Raman Spectra (.wdf, .txt, .opj, .opju,)
- STS spectra (.txt, .dat, .opj, .opju,)
- Spectroscopic data: x-ray photoemission and absorption spectroscopy measurements (.txt, .csv, .opj, .opju,)
- Surface structure: LEED (.jpg)
- Spectroscopic imaging ellipsometry (SIE) spectra, SIE maps and ellipsometric contrast images (.dat, .csv, .xml, .png, .pdf, .txt, .log, .config, .opj, .opju,)
- Raman/PL Spectra and Raman/PL Maps (.wdf.txt, .csv, .dat, .json, .svg, .tif, .png, .pdf, .log, .tvb, .opj, .opju,)
- Photocurrent maps (.txt, .dat, .csv, .svg, .pdf, .log, .opj, .opju,)
- Transport measurements (.dat, .txt, .log, .csv, .opj, .opju,)
- Design of devices (.dwg)
- Text (.docx, .pdf, .rtf, .tex)
- Tables (.xls, .xlsx, .opj, .opju, .ASCI, .opj, .opju,l)
- presentations (.pptx, .pdflatex, .dvips, .LyX)
- Code, in particular code generating analytical expressions describing spectral and transport properties in terms of materials parameters (.py, .osg, .mlx, .dat)

All data is expected to be stored in digital form, either in the format in which it was originally generated or in converted files for analysis.



The exact experimental conditions for each data set will be documented in a physical or electronic laboratory notebook and we will use a clear and specific identifier for each sample/reaction.

1.3 Origin of the data and possible re-use of existing data

Origin of the data:

The origin of existing and/or new data are our own-developed in-house characterization methods and setups, in-house computations and theories develops. The majority of data will be originated from software used for experimental activities/simulations. We will also use numerical calculations from the codes we plan to write.

In particular, for experimental activities:

The analytical data are collected by the instruments that generated them; later they are processed by either native programs associated with the instruments or analysed in other softwares.

Depending on the equipment, data can be firstly saved on each working machine for each measuring device (workstation, laptop). In this case, the data from each measuring machine will be regularly backed up on institutional file drive and it is possible to safely share data if needed via file permissions.

Reproducibility of the experimental findings is assured by running sample preparation and analysis several times. All generated data is expected to be digital, with samples and experimental procedures persistently identified.

Re-use of existing data:

The project is also expected to re-use existing data from literature or from previous lab activities (measurements, characterization, modelling, simulation, partners' codes...) for comparison to the newly developed ones and for further analysis.



1.4 Expected size of the data

The expected size of generated data will be reasonable according to the normal practices of the partners' researchs, which are well served by existent infrastructure for data management and storage

We expect to generate data up to 4 GB per month per partner. The fraction relevant for publications, however, will be significantly lower.

1.5 Data utility

The data are of fundamental importance for partners of the JOSEPHINE consortium and can be useful for possible scientific production (publications, patents, etc.).

1.6 Data storage

(Meta)data related to the project will be stored on partners' local servers. Each partner is responsible for data created within its lab. The project data will be stored in a specific folder. The storage convention for the project datasets is as follows:

Josephine_Beneficiary_Date_Title_Version

Title: short description of the content

Partners agree to preserve and cure data for at least 5 years after the end of the project (i.e until **April 2033**).

The CNRS is providing a **Virtual Research Environment** as planned in the Grant Agreement (D1.11). The consortium agreed that it takes the form of a shared cloud (**sDrive**: <https://sdrive.cnrs.fr/>) for the course of the project. A secured access was granted to the partners. The platform will be used to share all kinds of data. The use of the platform will also allow the project data to be more secured by storing it in several locations.

sDrive servers are hosted in France and are owned by the CNRS. The project cloud will also remain open 5 years after the end of the project. The storage convention will be the same as for local storage.

Whenever a set of data, a protocol or a project result is generated, an associated metadata/report will be made containing a set of information useful to make the associated data FAIR prior to data deposition and storage. A template for metadata management will be provided to all the partners (see Annex 1).

The template for metadata meets the minimum requirements established by the Dublin Core initiative. Metadata reports will be attached to datasets and stored in a folder of the project cloud.



All the datasets will be complemented with information regarding methods, tools and instruments used to generate or collect the data, as stipulated in the Annex 5 of the GA.

As specified in the GA (Annex 5), datasets related to scientific publications will be posted on a repository at the same time as the publications. The chosen repository is the **Zenodo** platform. Zenodo is fully compliant with the European Open Science Cloud (EOSC) requirements and the FAIR principles. In section 2, we will explain more in details the use of the repository and how we will make our data Findable, Accessible, Interoperable and Reusable.

2 FAIR Data

2.1 Making Data findable, including provisions for metadata

Datasets related to the results presented in scientific publications will be easily findable. **Zenodo** provides a Digital Object Identifier (DOI). This DOI is unique and persistent, enabling users to possibly update the shared material.

Datasets will be supported by rich metadata. The metadata fields mentioned in the template (Annex 1) will be reported on Zenodo. Metadata will follow the most used standards and it will be possible to index it to other resources. The description of the dataset will mention the DOI of the scientific publication or at least a link to the repository.

It will be possible to find datasets thanks to the information provided in the metadata, which will mention, among other things, the DOI of the dataset. Zenodo allows the addition of keywords that make it easier to find publications. This possibility will be used to add keywords related to the project and the subject of the publication.

Zenodo can be used as a searchable resource and our datasets will be easy to find thanks to the metadata associated. Web search engine indexing is also facilitated with Zenodo.

To make datasets even more findable, a community has been created on Zenodo to index datasets on the community page and to group them together:

<https://zenodo.org/communities/josephine>

Zenodo also allows direct indexing of datasets on Openaire thanks to the European grant reference (Grant No. 101130224). Results will be available and findable on Openaire and on the Openaire page of the project (Openaire-josephine).

Finally, the project website, which will be used for communication and dissemination activities, includes a section for publications. This section will contain publications, open access to peer-reviewed versions and all datasets available. (<https://www.josephine-project.eu/>)



2.2 Making data openly accessible

Datasets related to the **project results shared in scientific publications will be available online**. It will be possible to access them very easily thanks to the DOI provided by Zenodo. Access will not be restricted and will be fully open access.

At this stage, there are no particular restrictions to be mentioned in the Data Management Plan. However, in accordance with the principle "as open as possible, as closed as necessary", if there is any doubt about the possibility of exploiting the results, the WP leaders should anticipate and discuss it with the governing board before taking a decision on whether to publish it or not.

Although the embargoed or closed access option provided by Zenodo could be a valid option, the consortium agrees that research data linked to possibly exploitable results will not be deposited to avoid compromising their protection or commercialization prospects.

Project data to be shared between consortium members will not be published on Zenodo but put on the Virtual Research Environment.

2.3 Making data interoperable

Most of the data generated by each partner will adhere to standard/open formats. In case of the incompatibility of the generated data with the available (open) software, the data will be converted to standard formats.

The data formats will be as far as possible non-proprietary formats.

2.4 Increase data re-use (through clarifying licences)

Metadata will be provided, as it is necessary to re-use the data. Metadata will notably include the origin of the data, the version of the dataset thanks to the persistent DOI, methods, tools and instruments used to generate the data.

Zenodo follows cross-domain standards.

Materials physics do not have any specific standards but we will follow simple and clear rules to organize and describe datasets in order to facilitate its re-use.



3 Allocation of resources

Financial Resources:

The implementation of the DMP should not imply any additional costs.

Indeed, all-internal data management and storage systems are already established in each institution. The project's cloud (sDrive) does not involve any additional cost either as it is provided free of charge by the CNRS. Finally, the Zenodo platform and Creative Commons licencing are also free of charge.

Technical and Human Resources:

Each partner will be technically and legally responsible for keeping, disseminating and preserving data created within their labs, even if the work is carried out by researchers from different partners.

Each partner will be responsible to make accessible the data related to its publications. The quality of the data will be monitored by the coordinator and the WPs leaders. They will make sure that the data can be opened and plotted in good order and that all datasets contain the promised information.

The CNRS as project coordinator, will control the overall monitoring of the Data Management Plan. In particular, the CNRS team will verify that the datasets related to the publications and project results are available with the corresponding metadata and that they have all been shared on Zenodo, in the absence of Intellectual Property concern.

4 Data security

As previously stated, each partner is in charge of backing up its data. All partners have strongly secured IT systems protecting their local storage servers.

The shared cloud provided by the CNRS is fully compliant with security requirements. The Cloud servers are hosted in France in private data centres owned by the CNRS. This platform will be the only means of data exchange used by the consortium for internal purpose.

Regarding the repository, Zenodo is a trustworthy platform compliant with security requirements. Zenodo's servers are located on CERN's premises in Switzerland and managed via OpenStack and Puppet configuration management system which ensures that its servers always have the latest security patches applied.

(Meta)data will be stored at least on both local and shared cloud storage in order to avoid possible losses and better secure the data.



5 Ethical aspects

The Josephine project does not raise any ethical issues related to the management of its data.

6 Other issues

This Data Management Plan is compliant with the Grant Agreement of the project and is relying on well-established procedures for data management within all its partners.



7 Annex 1: Josephine metadata template form

DATA SET TITLE	
Creator:	
Contributors:	
Date:	
DOI (if assigned):	
DOI (or link or title of related publication)	
Dataset type:	
Files format(s):	
Description (including relation to project objectives):	
Instructional Methods (tools, instruments, ...):	
Rights:	

