D5.2 | Data management plan

Author(s): Tero Heikkilä and Heli Lehtivuori
Delivery date: 27.2.2019
Version: 1.0

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800923.

Project Acronym:	SUPERTED
Project Full Title:	Thermoelectric detector based on superconductor-ferromagnet heterostructures
Call:	H2020-FETOPEN-2016-2017
Topic:	FETOPEN-01-2016-2017
Type of Action:	RIA
Grant Number:	800923
Project URL:	https://superted-project.eu/

Editor:	Tero Heikkilä, JYU
Deliverable nature:	Data Management plan (DEC)
Dissemination level:	Public (PU)
Contractual Delivery Date:	28.02.2019
Actual Delivery Date:	27.2.2019
Number of pages:	8
Keywords:	Data collection, Open data, Raw data, Documentation, Metadata
Author(s):	Tero Heikkilä, JYU Heli Lehtivuori, JYU
Contributor(s):	Sebastian Bergeret, CSIC
External contributor(s):	Juuso Marttila, JYU

Abstract

This deliverable provides the SUPERTED project data management plan. The Deliverable outlines how the research data collected or generated will be handled during and after the SUPERTED project, describes which standards and methodology for data collection and generation will be followed, and whether and how the data will be shared.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800923.

1. Motivation and definitions

H2020 principles

Data Management Plan (DMP) is a **key element** of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project[1]. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a **DMP should include information** on:

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

A DMP is required for all projects participating in the extended Open Research Data (ORD) pilot, unless they opt out of the ORD pilot. However, projects that opt out are still encouraged to submit a DMP on a voluntary basis.

Open science principles at the University of Jyväskylä

The University of Jyväskylä (JYU) promotes open science[2]. The main objective is to advance open access to research results, research data as well as scientific methods. JYU follows the national and international principles and requirements of openness in science and research, thus, JYU requires that all research publications be self-archived in the university repository JYX. JYU encourages publishing in open publication series in those fields of science where esteemed publication series exist. Freeing individual articles from the payment (hybrid model) is not recommended, unless specifically required by the research funder. JYU requires that research data be deposited and, whenever possible, published on an open access basis so that they will be freely searchable and usable. Researchers are encouraged to explain the methods used in the research. Similar basic services are offered to all research projects financed by supplementary funding for managing their materials, for the work of the research groups, and for publishing the research results on an open access basis. Researchers are encouraged to actively inform the academic community about the research results and make the general public aware of them.

Data Management Plan in SUPERTED project

This DMP explains how the SUPERTED project manages the data obtained in the course of research done within the project. By the data we primarily mean the following:

1. **Experimental data:** Results of the different types of characterization measurements carried out on the samples either prepared within the project or obtained from the collaborators of the project participants, for the purpose of fulfilling the project's goals. These results are typically stored first in data files indicating the measured quantities (for example currents or voltages) with respect to the controlled quantities (currents, voltages, radiation powers, magnetic fields, etc.). Often the numerical values stored in those data files as such do not mean much to people outside the individuals or groups carrying out the measurements. Therefore, the more relevant data is obtained when such data is plotted in figures with axis labels indicating the proper units and normalizations of the measured and controlled quantities, along with short text explaining the conditions under which the data was obtained. This work is routinely carried out in lab books in each partner's site, containing the data and the explanations.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800923.

2. **Theory data:** Theoretical predictions about different quantities of interest, typically obtained from a derivation underlying a sequence of assumptions and approximations, and a subsequent numerical (or in some cases analytical) solution of the resulting equations. Here the data can mean (i) the numerical code used to obtain those results, or (ii) the curves produced in the numerical simulations.

The purposes of the data management within the SUPERTED project are:

a. Takes care of the integrity of the data, its proper storage, and documentation.
b. Efficient communication of intermediate data among the project members.
c. Identifying three categories of data: publishable data (PU), supporting data (SU), and data that may require IPR protection (IPR).
d. In the spirit of open data initiatives, attempting to publish as much data as possible along with the research publications based on that data along with metadata helping to search that data.

2. Data collection

Data collection takes place in work packages 1-4 as part of the research targeting the SUPERTED goals. We identify slightly different ways of data collection for the experiments and theory work.

1. **Experimental data:** All experimental partners of the project maintain electronic lab books on their experimental activity. These form the basis for collecting experimental data. Based on these lab books, we gather regularly a set of relevant information regarding each project on the SUPERTED wiki site. This site is open only to the SUPERTED project partners. To ensure useful representation of data, gathering the information into the wiki site requires some selection by the researchers. This is because the way experimental data is gathered in our field of science exhibits some level of uncertainty: a large part of data turns out to be not relevant for the general outcome because of some trivial or uninteresting reason - for example, that the sample under study was not of the intended type, or in case of equipment failure. The selected set of information is presented in the wiki site under the following generic form, structured within the different work packages/project deliverables in the wiki site.

Acronyms: PU=publishable as such after relevant paper has been submitted, PE=publishable after submitting the relevant paper, but needs editing, CO=confidential, do not publish, IPR=possibly relevant for patenting

PA - part of the published manuscript, SU - supplementary information, EX - published separately

Overall aim:

1. First aim
2. Second aim

Specific aims:

1. First specific aim
2. Second specific aim

Process description - for example, different sample batches:

(For each item, add the month/year, a reference to your lab book, and your name and institute)

- Month/year

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800923.

- Some overall description of this set of data
- Link to a summary of the data with pictures/descriptions - lab book reference - (name, institute)

Conclusions (update during the process):

- Conclusion for the process, or overall conclusions from the research

Open questions:

- Possible new open questions whose solutions would help carry out the process

---

The main purpose of this website is to serve the efficient communication within the project partners. However, at the same time it is possible to directly identify parts of data that can be used either as part of publication, part of supplementary information material, or part of an external set of data to be published on an archival site. In addition, we will actively aim at identifying data that needs to stay confidential for future research projects or because it is linked to protecting intellectual property rights. The different types of data will be identified via the acronyms and colour codes listed above.

2. **Theory data:** Theoretical physics work does not produce lab books. However, it may support presenting the experimental data by producing predictions of expected behaviour of the experimentally studied observables. This is often not as such new publishable theory, but rather using existing theory within the parameter regime of the experiment. Such predictions often arise from some generic numerical codes, and can be presented in a graphical form as curves. SUPERTED project will include such curves and codes from the theory work within the wiki site as parts of the experimental project. This set of data then includes identification related to their publishing capability similarly as with the experimental data.

   In addition, SUPERTED includes separate theory collaboration projects among two of the partners. For such projects, we will create separate wiki site pages, with the following format:

Acronyms: PU=publishable as such after relevant paper has been submitted, PE=publishable after submitting the relevant paper, but needs editing, CO=confidential, do not publish, IPR=possibly relevant for patenting

PA - part of the published manuscript, SU - supplementary information, EX - published separately

Overall aim:

1. First aim
2. Second aim

Specific aims:

1. First specific aim
2. Second specific aim

Intermediate results:

(For each item, add a title of the subproject, month/year, and your name and institute)

- Title of the subproject (month/year)
- Some overall description
- Link to a summary of the data with pictures/descriptions - (name, institute)
- Link to codes (matlab/python/C++/Mathematica) producing this behavior

Conclusions (update during the process):

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800923.

- Conclusion for the process, or overall conclusions from the research

Open questions:

- Possible new open questions whose solutions would help carry out the process

3. Data storage, preservation and data sharing

The SUPERTED wiki site will act as an intermediate-stage repository for the data gathered during the project, in addition to the lab books of the participating experimental groups. Access of the participants to this wiki site will be maintained until at least 3 years after the end of the project. The wiki site is provided by the University of Jyväskylä for this project.

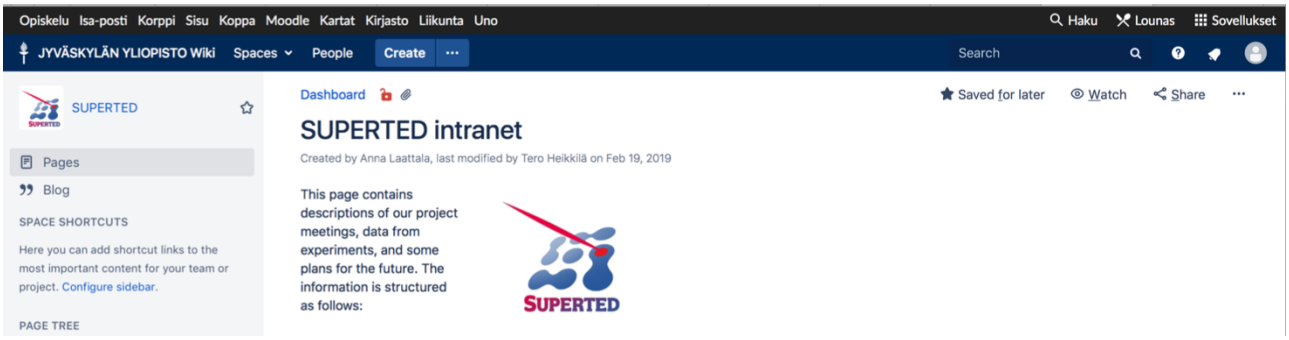

Figure 1: Main page of SUPERTED project intranet or wiki site. Confluence wiki is used in this project.

In addition, access to the lab books will be maintained until at least 3 years after the project. The data format within the wiki site is in terms of description (ascii text) and pictures (png/jpg/pdf and other often-readable formats). For the data chosen to be part of published scientific papers (identified with acronyms PU/PE and/or PA/SU/EX), we will use one or more of the following way to ensure data sharing and open access:

PA: part of the published manuscript or SU: supplementary information

If the data is part of the published manuscript or the supplementary information published together with the manuscript, it typically does not need further efforts in terms of preservation as we will publish our results in well-known journals that have long-term storage of the information. In addition, if not forbidden by the journal, we will submit the manuscripts to the arXiv repository (https://arxiv.org)[3].

Self-archived at JYU. Research conducted at the JYU is self-archived (parallel published) in the JYX publication archive. In JYU, researcher submits a research article (both the final PDF and the final draft version) using the TUTKA form in connection of registering the publication data for the University Library. (Final draft = Author's Accepted Manuscript (AAM) = the authors' version = post-print = post-review = the version after peer-review changes but before copy editing and formatting by the publisher.) The University Library verifies the publication permission of the article, checks the archived version and possible embargo and saves the article in JYX.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800923.

EX: published separately

Besides supplementary information material to the published works, in certain cases we will opt to move part of the background content of the project wiki page directly into a public data repository (for example, University of Jyväskylä JYX publication archive or the zenodo.org service provided by the OpenAIRE project) for long-time storage and with open access. This data is then linked to and from the published scientific articles.

4. Documentation and metadata

The following table indicates an example structure of the metadata included in the beginning of the self-archived datasets. It is compatible with the DataCite 4.0 scheme[3]

Project and GA number	**SUPERTED 800923**
Identifier	Superted.xxx (xxx is a running number)
Creator	Name and institute (+ contact details)
Title	Title of the data
Publisher	Name and institute (+ contact details)
Publication Year	Year
Resource type	Experimental/simulation/code
Description	Short description of the presented data
Data source	E.g. experiment performed in Pisa on date
Version	Possible version information (number)
Rights	Licence (if any) or use constraints

The SUPERTED project maintains up-to-date documentation and metadata to ensure that the research data produced in project will be 'FAIR', that is findable, accessible, interoperable and re-usable[5].

5. Intellectual property rights

The Intellectual Property Rights (IPR) Management Plan (*Separate Deliverable 4.1*) of SUPERTED project has been elaborated as a set of rules and protocols. In SUPERTED project, we follow three main principles:

1. Both the recognition of knowledge brought in (background) and generated within the project (foreground) by each partner has been assigned by default, based in the information contained in the Grant and Consortium Agreements. This will allow ensuring respect of partners' rights without causing administrative burden.
2. There is a dedicated Task (*Deliverable 4.1* carried on by partners JYU and BIHUR) who will actively monitor the IPR generation and propose paths to exploit and disseminate the results avoiding conflicts. This will allow to systematically dealing with all knowledge generation while allowing partners to focus onto the technical activity.
3. Management of IPR (including conflicts) will be in the first instance carried out by the teams involved in the project. This is to take the most of the technical knowledge of the teams for discussing the issue and finding solutions.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800923.

6. Responsibilities and resources

All SUPERTED partners are responsible for complying with the DMP and its procedures for data collection, handling and preservation. JYU is responsible for overseeing this implementation, along with ensuring that the plan is reviewed and revised during the project. This DMP is a plan, and its actual implementation may reveal better ways to operate. We hence leave the possibility of improving the policies during the project. Any changes in DMP will however require approval by the SUPERTED board, either in a common Skype meeting, or in an annual project meeting. The contact person for communication: Tero Heikkilä, Tero.T.Heikkila@jyu.fi

Bibliography

1. European Commission, Data management in Horizon 2020 Online Manual Available from: http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template
2. University of Jyväskylä, Open Science and Research, Available from: https://openscience.jyu.fi/en/open-science-jyu
3. arXiv ®, Cornell University, Available from: https://arxiv.org
4. DataCite, https://schema.datacite.org/meta/kernel-4.0/doc/DataCite-MetadataKernel_v4.0.pdf
5. European Commission. Guidelines on FAIR Data Management in Horizon 2020, July 2016. Available from: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800923.