Deliverable D1.7

Data management plan 2

Project number	101070290
Project name	Nonlinear Magnons for Reservoir Computing in Reciprocal Space
Project acronym	NIMFEIA
Work package	WP1 Management, dissemination and exploitation
Type	Report
Dissemination level	Public
Lead beneficiary	HZDR
Due date of delivery	Month 18 – March 2024

Authors: Katrin Schultheiss (HZDR), Gerhard Jakob (JGU), Johan Mentink (RU)

1. Introduction

This document is the second version of the Data Management Plan (DMP) of the project *Non-linear Magnons for Reservoir Computing in Reciprocal Space* (NIMFEIA), funded by the European Union under the call *HORIZON-CL4-2021-DIGITAL-EMERGING-01-14 – Advanced spintronics: Unleashing spin in the next generation ICs (RIA)* with grant agreement number 101070290.

2. Data Management Plan – Version 2.0

2.1. Project Coordinator

Katrin Schultheiss
Helmholtz-Zentrum Dresden-Rossendorf (HZDR)
Bautzner Landstrasse 400, 01328 Dresden, Germany
k.schultheiss@hzdr.de

2.2. Data summary

2.2.1. Data type

What kind of dataset is it?

Raw data: Several activities in the NIMFEIA project involve the collaborative work of several partners in different disciplines. Therefore, a large variety of data will be generated using different, often unique experimental and numerical techniques:

- Spin-wave spectra are measured optically by Brillouin light scattering spectrometry or electrically using a spectrum analyser.
- Images and videos are taken by Kerr microscopy.
- Ferromagnetic resonance spectra are measured by a vector network analyser.
- Transport measurements are taken by magnetoresistive setups.
- Magnetic material parameters will be determined by vibration sample magnetometry.
- Deposition protocols are generated by a deposition tool computer.
- Images of microstructures on silicon chips are taken by optical microscopes and/or scanning electron microscopy.
- Data related to documentation, such as pictures, calibration data, and mechanical drawings not intended for publication or reuse.
- Data is collected from computer simulations, together with tailored computer code and scripts used to generate this data.

All of this data stem from monitoring magnetic physical systems. For their generation, different means and experimental setups are used which are often standard in the line of magnetisation studies but are typically in-house-built in the partners' labs.

Publication data: This dataset includes relevant raw data, processed and analysed experimental and numerical data, and simulation codes and scripts as used for publications.

Use case data: Anonymous open-source data and anonymous proprietary radar data from industrial collaborators (Infineon) which will be used to test NIMFEIA's approaches for reservoir computing.

Which file formats are used?

Raw data / Publication data: Data will be stored in:

- Generic .dat format such that it can be analysed and stored cross-platform using open-source and commercial software (NumPy, Origin, Matlab, Excel, Mathematica).
- Standard image and video formats .jpg, .tiff, .png, and/or .avi.
- Hierarchical data format .hdf5 (https://www.hdfgroup.org), a self-describing data format that can be read cross-platform by a wide range of open-source and commercial tools like HDF View and Origin or via the programming interfaces in all major languages including Python, Matlab, and LabVIEW.

Is the dataset being created or re-used?

Raw data / Publication data: This data is created using different means and experimental set-ups, which are often standard in the line of physics but in-house-built in the partners' labs.

Use case data: This data is re-used. NARMA datasets were presented first in 1994 (https://ieee-explore.ieee.org/document/279188) and are typically generated based on a differential equation (e.g. https://github.com/Devrim-Celik/simple_echo_state_network). Micro-Doppler data were discussed in 2018 (https://ieeexplore.ieee.org/document/8333730) with details given in GitHub (https://github.com/baptist/idrad) and the anonymous dataset available at https://www.imec-int.com/en/IDRad. The anonymous proprietary data from industrial partners (Infineon) is currently not available for general use.

If created, are already existing, similar research data available and why is their subsequent use not possible or useful here?

Raw data / Publication data: No similar data exists, and no data will be reused. Pushing the current state-of-the-art towards novel fundamental insights into the underlying physical principals is at the heart of the NIMFEIA project. Hence, it is necessary to perform experiments and simulations that have not been done before.

2.2.2. Purpose

How / for what purpose will this dataset be used during the project?

Raw data: Data will be generated to optimise materials and check the functionality of devices (WP4), advance the operational capabilities of the magnon reservoir (WP2), and realise its electrical readout (WP5). Furthermore, simulation data will be generated to model the magnon reservoir in the GHz range (WP3) and in the THz range (WP7).

Publication data: This data will include relevant raw data, processed and analysed data, as well as relevant simulation codes and scripts to make it accessible to the scientific community through publications.

Use case data: This data will be used to test the operational capabilities of the magnon reservoir on the basis of real-world use cases (WP6).

2.2.3. Data volume

What is the actual or expected size of the dataset?

Raw data: less than 1 TB

Publication data: a few GB

Use case data: a few GB

2.2.4. Provenance

What is the provenance of the data?

Raw data / Publication data: The physical objects studied in the NIMFEIA project will be silicon chips containing continuous thin films as well as magnetic microstructures and metallic mi-

crowave antennas. These chips will be fabricated collaboratively by the partners HZDR, JGU, UPS, and GF. The magnetic materials, layer thicknesses, and the design of the microstructures will vary throughout the project.

2.2.5. Data utility

Which individuals, groups, or institutions could be interested in re-using this dataset?

Raw data: Most of the generated raw data is only of interest to researchers working on the NIMFEIA project. Material and sample optimisation involves many different test measurements and simulations that are not of interest to a broader community. It is currently not foreseen to make this data available outside the collaboration.

Publication data: Portions of the raw data as well as processed and analysed raw data is the basis for scientific publications and, thus, may be of interest to the broader magnetism community.

2.3. FAIR data

2.3.1. Findability

Will persistent identifiers (PIDs) be used for this data set?

Raw data: No

Publication data: Yes, digital object identifiers (DOI) will be used.

Which metadata are collected automatically?

Raw data / Publication data: Metadata logging the day, time, and experimental measurement parameter (e.g. microwave frequency and power, magnetic field amplitudes, applied direct currents) will be created automatically by the data acquisition system (homemade LabVIEW or Python software or commercially available software thaTEC:OS: https://thatec-innovation.-com) and are contained in the .hdf5 files.

Which metadata are collected semi-automatically?

Raw data / Publication data: Metadata logging the author name, project number, sample name, device name, and control settings of the experimental setup which have to be set manually and cannot be logged remotely via PC (e.g. laser power, interferometer settings, physical mounting or bonding) are entered manually before starting the data acquisition and are contained in the .hdf5 files.

Which metadata are collected manually?

Raw data: In addition to metadata stored by the data acquisition system, a lab book entry will be generated. The output from the deposition system logging needs to be converted to .dat or .xls manually.

Publication data: When uploading the data to a repository, standard metadata such as author names, affiliations, ORCIDs, and keywords will be collected.

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

Raw data: No search keywords will be provided.

Publication data: Search keywords will be provided via the upload in the repository.

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

Raw data / Publication data: No.

2.3.2. Accessibility

2.3.2.1. Accessibility – Repository

Where will the data (including metadata, documentation, and, if applicable, relevant code) be stored or archived after the end of the project?

Raw data: At the partner institution where the data is acquired.

Publication data: At the partner institution where the data is acquired and in the data repositories.

Will the data be deposited in a trusted repository?

Raw data: No.

Publication data: Yes, using Rodare https://rodare.hzdr.de or Zenodo https://zenodo.org/

- The repositories meet the basic expectations of a research data repository: easy and free access while observing legal and ethical barriers, detailed metadata, and protection against unauthorised access.
- The repositories offer expert curation for long-term archiving.
- The repositories provide persistent identifiers, e.g. DOI.
- The repositories are internationally acknowledged for the specific discipline/domain.
- The repositories are managed by trustworthy institutions, HZDR and CERN
- The repository's policies are available online: https://rodare.hzdr.de/about/policies/ and https://about.zenodo.org/policies/

Will the repository resolve the identifier to a digital object?

Publication data: Yes.

Have you explored appropriate arrangements with the identified repository?

Publication data: No, the data will be uploaded via standard procedures and require no special arrangements.

2.3.2.2. Accessibility – Data

Will this dataset be published or shared?

Raw data: Yes, the data will be shared internally with everyone in the NIMFEIA project, as long as they do not pass on the data.

Publication data: Yes, in the repositories Rodare or Zenodo.

Shall there be an embargo period before the data are made available?

Raw data: The data will only be made available internally.

Publication data: Yes, until:

- publication of the results in a peer-reviewed journal
- achievement of intellectual property rights such as patents
- finishing a bachelor/master/doctor thesis

If yes, under which terms of use or license will the dataset be published or shared?

Publication data: Creative Commons Attribution-NonCommercial (CC-BY-NC)

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

Publication data:

- Users are required to register to use the repository.
- If necessary, an authentication system or a data-on-demand function will be provided.

Is a data access committee needed to handle access requests to the published data of the project?

Publication data: No

2.3.2.3. Accessibility - Metadata

Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why.

Publication data: Metadata is integrated into the .hdf5 files and published with the data.

Will metadata contain information to enable the user to access the data?

Publication data: A readme file will be made available together with the data.

How long is it intended that the data remains re-usable?

Raw data / Publication data: at least 10 years

Is documentation about relevant software needed to use the data?

Raw data / Publication data: self-generated software or script: LabView, Python scripts

2.3.3. Interoperability

Which standards, ontologies, classifications, etc. are used to describe the data and context information?

Raw data / Publication data: No fixed system for the description is used.

Is this dataset interoperable, i.e. allowing data exchange and reuse between researchers, institutions, organisations, countries, etc.?

Raw data / Publication data:

- The dataset adheres to standardised formats: Generic .dat files or .hdf5 files are used.
- The dataset is usable with available open-source software applications (Python, HDF View) or commercial software applications that are established and widely used in the respective community (LabView, Origin, Matlab, Excel, Mathematica)

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?

Raw data / Publication data: No mappings are necessary, as the datasets are described using standard terminologies. There is no need for mapping since the used terminology is chosen to be compatible with the existing literature.

2.3.4. Re-usability

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

Raw data / Publication data: Lab books and readme files will be generated.

Which components of the data documentation are available together with the dataset?

Raw data / Publication data:

- Information on the methodology used for data generation (creation or re-use) and transformation
- Description of variables used and respective units of measurement

Where is the data documentation provided?

Raw data / Publication data: Some data documentation, e.g., the settings of the experimental setup, will be provided in the .hdf5 files. Additional documentation will be provided in lab books and readme files.

Which measures of quality assurance are taken for this dataset?

Raw data / Publication data::
- Repeated or comparative measurements
- Adherence to standard procedures for data recording

2.4. Other research outputs

What other research outputs are generated or re-used throughout your project?
- Software
- Workflows
- Samples

2.5. Allocation of resources

2.5.1. Costs

What are the personnel costs associated with long-term preservation for the project?

It is assumed that the NIMFEIA partner institutions will provide the personnel for the long-term preservation of the data, so no funds from the NIMFEIA project are needed.

What is the amount of non-personnel costs regarding long-term preservation for the project?

It is assumed that the NIMFEIA partner institutions will provide the funds for the long-term preservation of the data, so no funds from the NIMFEIA project are needed.

How will the data management costs of the project be covered?

Since data FAIRness is in the interest of all partner institutions, the institutions will support the NIMFEIA project by contributing the general resources to make data FAIR (archiving infrastructure, repositories, person months for IT personnel).

2.5.2. Responsibility

Who is responsible for the backups?

Raw data:
- The IT departments of the partner institutions

Publication data:
- The repositories
- The institutional libraries
- The IT departments of the partner institutions

Who is responsible for data documentation and for checking if metadata and data documentation are correct and complete?

Raw data: All researchers who work on the NIMFEIA project

Publication data: The researcher who uploads data to the repository

Who is responsible for curating the data once the project has ended?

Raw data:
- The IT departments of the partner institutions

Publication data:
- The repository
- The institutional library
- The IT department of the partner institutions

2.5.3. Preservation

How long will the data be stored?

Raw data / Publication data: at least 10 years

2.5.4. Selection

What are the criteria/rules for the selection of the data to be archived (after the end of the project)?

Raw data / Publication data: All project data will be archived.

2.6. Data security

2.6.1. Security measures

Which measures or provisions are in place to ensure data security (e.g. protection against unauthorized access, data recovery, transfer of sensitive data)?

Raw data / Publication data:

- Protection against unauthorized access
- Data recovery: backups

2.7. Ethics

2.7.1. Personal data

Does this dataset contain personal data?

Raw data / Publication data: No.

Does the dataset contain "data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership" or genetic or "biometric data to uniquely identify a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation" (Art. 9 GDPR)?

Raw data / Publication data: No

2.7.2. Other sensitive data

Does this dataset contain sensitive data other than personal data? If yes, please describe the non-personal sensitive data used in the project.

Raw data / Publication data: No

2.7.3. Official approval

Has the project been approved by a research ethics committee?

The nature of the project means that no approval is required from an independent research ethics committee. The project is however working closely with our external ethics advisor.

Is a statutory approval/permit needed for the research?

No

2.7.4. Intellectual property rights

Does the project use and/or produce data that is protected by intellectual or industrial property rights?

No

Does copyright law apply to this dataset?

No

Do other intellectual property rights apply to this dataset?

No