MANNGA Deliverable

D1.2 Data Management Plan (DMP)

Version 1.0 – updated M20 version

Grant Agreement number	101070347
Action Acronym	MANNGA
Action Title	MAGNONIC ARTIFICIAL NEURAL NETWORKS AND GATE ARRAYS
Call	HORIZON-CL4-2021-DIGITAL-EMERGING-01
Version date of the Annex I against which the assessment will be made	31.5.2022
Start date of the project	1.9.2022
Due date of the deliverable	28.02.2023 – updated M20 version 30.4.2024
Actual date of submission	28.02.2023 – updated M20 version 29.4.2024
Lead BEN / AP for the deliverable	AALTO
Dissemination level of the deliverable	Public

Coordinator, PI
Sebastiaan van Dijken
AALTO KORKEAKOULUSÄÄTIÖ SR, Aalto University School of Science

Scientific coordinator
Volodymyr Kruglyak
THE UNIVERSITY OF EXETER

Co-funded by the European Union

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

Authors in alphabetical order		
Name	Beneficiary	e-mail
Sebastiaan van Dijken	Aalto University	sebastiaan.van.dijken@aalto.fi
Volodymyr Kruglyak	The University of Exeter (UNEXE)	V.V.Kruglyak@exeter.ac.uk

Document reviewers		
Name	Beneficiary	e-mail
Paweł Gruszecki	AMU	gruszecki@amu.edu.pl

Abstract

This Data Management Plan formalises MANNGA's commitment to the FAIR principles and the obligations set out in the Grant Agreement. This includes description of the data to be generated in the project and the ways how the data will be made findable, accessible, and interoperable.

Contents

1. Introduction 5
2. Data Summary 5
3. FAIR Data 6
   3.1 Making data findable, including provisions for metadata 6
   3.2 Making data accessible 7
   3.3 Making data interoperable 7
   3.4 Increase data re-use 8
4. Other research outputs 8
5. Allocation of resources 8
6. Data security 9
7. Ethics 9
8. Other issues 9

1. Introduction

In MANNGA, we are already committed to open science practices, which form important part of our national research ecosystems. The purpose of this deliverable is to formalise this commitment at the project level, in form of a Data Management Plan ("the Plan"). The present version of this living document includes updates made in M20 due to the periodic reporting phase as well as a new partner (AMU) having joined the MANNGA project via a hop-on facility call. The plan will be further detailed and fine-tuned via updates as MANNGA progresses, e.g. when significant changes occur. The coordinator (AALTO) will be responsible, with contributions from the other beneficiaries, for the revisions and updates of the Plan, as well as data management and quality assurance in general. This Plan complies with principles of the FAIR data management concept, ensuring that the data be findable, accessible, interoperable, and re-usable, and created using a FAIR template to develop, in compliance with INSPIRE and GDPR.

2. Data Summary

Data generated and used in MANNGA will be inherently technical and will not lead to emergence of any ethical issues. Table 1 summarises the types, formats, and origins of data that MANNGA will generate, while we do not foresee re-using any existing data (since such data does not exist). MANNGA's data will primarily have form of experimental and numerical data sets. Primarily, the datasets will be generated for the purpose of demonstration and / or validation of scientific and technological concepts and ideas, as well as to benchmark performance of MANNGA's devices. The concepts, ideas, and devices constitute and will adhere to the objectives of MANNGA.

Experimental datasets have relatively small size, generated at speeds of 1 GB per day at most. The datasets as CSV files will be stored and shared across MANNGA in full, followed by sharing to the wider research community in parts sifted to exclude any redundancies. Numerical datasets are considerably larger, generated at speeds reaching a few TB per day, but are reproducible provided the script and software used to generate them are available. Hence, the scripts will be archived and shared, together with (i) some data samples in processed form, and (ii) full trainable datasets (still smaller than the raw datasets). The latter datasets will represent, together with the experimental data, the main data types that will expect to find use outside of MANNGA, with the potential users to include academic and commercial entities and individuals interested in MANNGA's research.

Table 1. A summary of the data to be generated and reused in MANNGA.

Type of data	Origin of data	Instrument or software used	File formats	Total disk space required across MANNGA	Shared / reused across MANNGA	Shared externally
VNA[1] data sets (raw or processed)	VNA[1] measurements and data post-processing	UNEXE: ZNA26 from Rohde & Schwarz	CSV	1 TB	Yes	Yes

[1] "VNA" stands for "Vector Network Analyser".

TRSKM[2] data sets (raw or processed)	TRSKM measurements and data post-processing	Custom setups at AALTO, MLU, and UNEXE	CSV	1 TB	Yes	Yes
Raw numerical data sets	Numerical micromagnetic simulations	MuMax3, OOMMF, Comsol	OVF, ZARR, TXT, mx3	1000 TB	Data samples only	Data samples only
Processed numerical datasets	Post-processing of numerically simulated results	Python, MATLAB, C++	OVF CSV, ZARR, TXT, NPZ	1000 TB	Yes	Trainable datasets

3. FAIR Data

3.1 Making data findable, including provisions for metadata

As a rule, the full associated datasets will accompany all research articles published by MANNGA. In each case, the article and accompanying datasets (with necessary metadata) will be assigned persistent identifiers. Publisher's Digital Object Identifiers (DOIs) and established preprint servers (ArXiv, OpenAir) will be used to ensure findability of our publications, which will contain permanent links ("handles") to the shared datasets. As an exception, some articles may be published before the full dataset is ready for publication. In such cases, the article will be linked to the dataset later when the dataset is published. Usually, the latter datasets will be located in the institutional repositories of the involved beneficiaries. Prior to the article's publication, the datasets will be stored in and shared via a searchable database, located at MLU. The datasets will be accompanied by rich metadata to allow their discovery. Tables 2 and 3 present provisional lists of searchable numerical and string parameters. These will be revised once the database is up and running.

Table 2. A provisional list of searchable numerical parameters for MANNGA's datasets.

Parameter name	Examples of possible values (optional)
Magnonic waveguide or medium thickness (nm)	
Magnonic waveguide width (µm): in-plane dimension orthogonal to the spin-wave propagation direction	
Resonator to medium spacing (nm)	
Resonator thickness (nm)	
Resonator width (nm): in-plane dimension parallel to the (typical) spin-wave propagation direction	
Resonator length (µm): in-plane dimension orthogonal to the (typical) spin-wave propagation direction	
Number of resonators	0, 1, 2, …
Number of inputs	
Number of outputs	

Table 3. A provisional list of searchable string parameters for MANNGA's datasets.

[2] "TRSKM" stands for "Time-Resolved Scanning Kerr Microscopy".

Parameter name	Examples of possible values (optional)
Data origin	'experiment', 'micromagnetic simulations', 'theory', 'machine learning'
Entry type	'dataset', 'graph', 'image', 'movie', 'summary', 'code', 'script'
Magnonic medium type	'film', 'waveguide'
Magnonic medium material	'YIG', 'NiFe', 'CoFeB'
Resonator shape	'stripe', 'rectangle', 'square', 'ellipse'
Resonator role	'input', 'control', 'output', 'network'
Resonator material	'YIG', 'NiFe', 'CoFeB'
Resonator composition	'single layer', 'bilayer'
Device type	'CMR', 'MFPR', 'CMLG', 'mFPGA', 'RC', 'RNN'
Free form notes	

3.2 Making data accessible

The accessibility of MANNGA's datasets is ensured via depositing them in trusted repositories (selected after exploring appropriate arrangements). This includes the beneficiaries' institutional repositories for datasets accompanying published articles, the project-wide repository being set up at MLU for data storage and sharing prior to publication, and GitHub for codes and scripts. The repositories enable creation of DOIs for the data and code, either directly or using third-party tools, such as the data archiving tool Zenodo on GitHub.

As a rule, MANNGA will make all useful data openly available, free of charge, via downloading from the repository. Here, "useful data" means data that has not been identified by beneficiaries as erroneous or violating accepted scientific standards of quality. MANNGA will not attempt to identify the person or tool accessing the data shared. Furthermore, we do not foresee a need for a data access committee. The metadata containing information to enable the user to access the data will also be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement. The data and metadata will remain available and findable for as long as the repositories will remain functional. Although not foreseen at present as a need, any specialised software to access or read the data of non-standard types will either be included with the datasets or links will be provided to the open-source code and documentation, if required.

3.3 Making data interoperable

The interoperability of the data will be ensured by committing to storing it in the common formats: CSV, OVF, and NPZ, while ZARR will also be used for sharing results of micromagnetic simulations within MANNGA. The formats are standard for and already endorsed by the communities that could make use of our data. So, there will be no need for additional vocabularies, standards, formats or methodologies to make our data interoperable, to allow data exchange or re-use within and across disciplines. In case it becomes unavoidable that we use uncommon data formats or generate project specific ontologies or vocabularies, we will provide mappings to more commonly used ontologies. The generated ontologies or vocabularies will be openly published to allow their reuse, refinement, and

extension. MANNGA's data have not needed to include qualified references to other data so far, but this will be reconsidered if a need arises.

3.4 Increase data re-use

In MANNGA, we would like to see our results published in scientific literature and used as widely as possible. The research articles published, with properly documented methodology sections on how the data was generated, will serve as natural methodological and provenance references for the associated data. The same concerns data included within Master and PhD Theses. The majority of orphaned datasets, which have not been included in a journal publication or a Thesis, will only require a short accompanying readme file with a reference to the publication(-s) where the methodology is systematically described and with the provenance of the data thoroughly documented using the appropriate standards. If necessary, however, more detailed stand-alone readme files describing the relevant methodology and the provenance of the data will be made available with the data. These readme notes will be made freely available on the same terms as the data, licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement. This will ensure that the data produced in MANNGA will be reusable by third parties, both during and after the end of the project. All datasets, whether included in a publication or orphaned, will undergo standard rigorous data quality assurance processes, adopted in the community. These processes will include self-consistency checks, comparisons with results published elsewhere, while we will also every effort to compare results produced different experimental and theoretical techniques available within the MANNGA consortium.

4. Other research outputs

In addition to data per se, MANNGA is expected to generate outputs in form of new codes and algorithms. The outputs will be handled, made accessible and interoperable, and published using exactly the same principles as those described above, in sections 3.1-3.4, in line with both the FAIR principles and the obligations set out in the Grant Agreement. This includes the use of Open Source licenses for openly distributed software source codes, which are easily accessible to anyone, anytime and anywhere, and encourage a culture of collaboration, are very low cost, and scale very well. All software donated by MANNGA will be provided as an open source and will be made available as a clone of the original source code, which can be freely modified at no cost.

5. Allocation of resources

As a rule, MANNGA will make use of the open-access infrastructure, either existing within the participating institutions or publicly available (e.g. GitHub), while each beneficiary will also have their own measures for data storage and backup. The associated costs are expected to be relatively modest and will be met by the beneficiaries from their allocated budgets. The only exception is the pre-publication storage and sharing of data within MANNGA via a searchable database located at MLU (section 3.1). The latter costs will be met by MLU. The same arrangements will apply after the end of MANNGA, to ensure long term preservation of the datasets.

The coordinator (AALTO) will bear the overall responsibility for data management in MANNGA.

6. Data security

The data security in MANNGA will be ensured via (i) backing up data prior to sharing it across the consortium, (ii) password protected access and backup of the data stored and shared using the searchable database at MLU, and (iii) using trusted repositories for long term preservation and curation post-publication (or after making public the orphaned datasets). The datasets progress from stage (i) to (iii), their previous copies will be retained for as long as practicable, both to ensure easy access for the data re-use within MANNGA and for added security.

7. Ethics

Data generated and used in MANNGA will be inherently technical, and they are not associated with technology which could be used for military purposes or cause harm for the environment, and so, no ethical issues are expected to emerge.

8. Other issues

The data management in MANNGA will be governed by this Plan, in strict accordance with the FAIR principles and the obligations set out in the Grant Agreement.