D1.3

Data Management Plan RP1 update
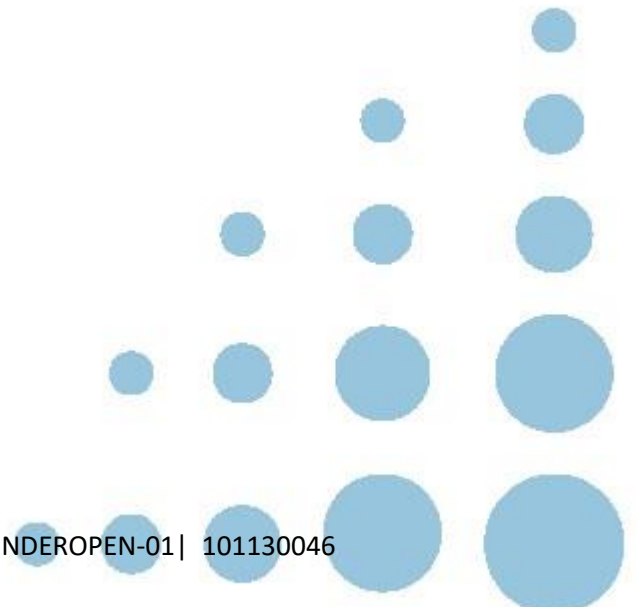
Deliverable Details

Lead beneficiary: BIU

Work package: WP1

Dissemination level: *PU (Public): Fully open (automatically posted online)*

Nature: *DMP: Data Management Plan*

Due date: 31/01/2025

Submission date: 13/02/2025

Resubmission date:

Authors: Anna Madar & Lior Klein, BIU

Contributors: All partners

Reviewers: All partners

Version History:

Date	Version No.	Author	Notes
13/02/2025	1.0	Anna Madar	Final version for submission

MULTISPIN.AI

Project Details

Acronym: **MultiSpin.AI**

Title: **n-ary spintronics-based edge computing co-processor for artificial intelligence**

Coordinator: **BAR ILAN UNIVERSITY** (ISRAEL)

Reference: 101130046

Type: HORIZON EIC Grants

Program: HORIZON EUROPE

Topic-ID: HORIZON-EIC-2023-PATHFINDEROPEN-01-01

Start: 01.02.2024 – 31.01.2027

Duration: 36 months

Consortium:

Id	Participant Name	Short name	Country
1	BAR ILAN UNIVERSITY	BIU	ISRAEL
2	INESC MICROSISTEMAS E NANOTECNOLOGIAS - INSTITUTO DE ENGENHARIA DE SISTEMAS E COMPUTADORES PARA OS MICROSISTEMAS E AS NANOTECNOLOGIAS	INESC MN	PORTUGAL
3	UNIVERSITE CATHOLIQUE DE LOUVAIN	UCLouvain	BELGIUM
4	SPINEDGE LTD	SPINEEDGE	ISRAEL
5	INTERACTIVE FULLY ELECTRICAL VEHICLES SRL	IFEVS	ITALY
6	~~VRIJE UNIVERSITEIT BRUSSEL~~	~~VUB~~	~~BELGIUM~~
7	AMIRES SRO	AMI	CZECHIA

MULTISPIN.AI

Table of Contents

Executive Summary

The vision of the MultiSpin.AI project is to improve neuromorphic computing by increasing the energy efficiency and processing speed by at least three orders of magnitude over digital computing and >10x compared to the most advanced neuromorphic devices to reach an unparalleled 2,000 Tera operations per second per watt (TOPS/W). To achieve this, MultiSpin.AI will develop an AI co-processor based on a crossbar of multi-level magnetic tunnel junctions (M2TJ) cells/ n-ary state cells. The use of M2TJs reduces the number of cells, simplifies circuity, and reduces the number of digital-to-analog conversions (DAC) at the input of the crossbar, and analog-to-digital conversions at the crossbar output. The combined effect is realising much higher energy efficiency and faster AI inference at the Edge. This breakthrough will help provide a significant impact by enabling transformative applications like autonomous vehicles, robots, and medical devices, will help strengthen strategic autonomy for the EU chips industry, and will reduce CO_2 emissions from AI inference.

This Data Management Plan (DMP) for the MultiSpin.AI project provides a framework for how data material will be handled during and after the end of the project. This document provides an overview of the datasets that will be collected, used and re-used within Multispin.AI. It is not a fixed document; it is a living document that will evolve during the lifespan of the project to include the research data that will be generated.

This version of the DMP, D1.3, is the first update of the plan, submitted at the end of the first project year. The changes compared to the original DMP, D1.1, are:

- the division of the dataset mentioned for WP2 into two separate datasets to highlight differences in their management, and the partners involved;
- the further development of the Zenodo community for the project.

1. Data Summary - Datasets to be collected, used and delivered

Dataset description

Collected in WP	WP2: Multispin.AI post-von Neumann architecture - design & fabrication
Data set Persistent identifier (PID) and name	Characterization and simulation of memory elements
Data type & format	Data in text and image format: ascii, *.txt; *.jpg, *.bmp. Csv files from the instruments used in the measurements
Data origin (if applicable)	We plan to generate our own data, and do not expecting to re-use data.
Partners / Stakeholder(s) / user(s) involved	BIU, UCLouvain, SpinEdge
Data set description	The data consists of the output of experimental measurements and computer simulations.
Expected data volume & frequency	From few kB (text) up to tens of MB (e.g. electrical characterization)
Standards and metadata relevant for the data (if relevant)	Not relevant
Data sharing target(s) outside the project	Experts in the field of thin film magnetic memory devices, guiding the experimental and theoretical activities.

Collected in WP	WP2: Multispin.AI post-von Neumann architecture - design & fabrication
Data set Persistent identifier (PID) and name	Nanofabriction of memory elements
Data type & format	Data in text and image format: ascii, *.txt; *.jpg, *.bmp. Csv files from the instruments used in the metrology measurements Dxf files for the masks used in the two lithography tools.
Data origin (if applicable)	We plan to generate our own data, and do not expecting to re-use data.
Partners / Stakeholder(s) / user(s) involved	BIU, INESC
Data set description	The data consists of the description of nanofabrication process, output of experimental measurements and quality control during nanofabrication.
Expected data volume & frequency	From few kB (text) up to tens of MB (e.g. data through the process run sheet microscope inspection, materials quality control and metrology for plasma-based processes)
Standards and metadata relevant for the data (if relevant)	Not relevant
Data sharing target(s) outside the project	Experts in the field of thin film magnetic devices, guiding the materials optimization and nanofabrication activities.

Collected in WP	WP3: SLMMS and M2TJs characterization & testing
Data set Persistent identifier (PID) and name	PIDs will be created and managed throughout the project. Name(s) to be defined later on.

MULTISPIN.AI

Data type & format	Data will be provided in text and image format: ascii, *.txt; *.jpg, *.bmp.
Data origin (if applicable)	We are planning to generate our own data, so in general, are not expecting to re-use data.
Partners / Stakeholder(s) / user(s) involved	BIU, UCLouvain, INESC MN, SpinEdge
Data set description	The data consists of magnetic simulations using COMSOL Multiphysics and Mumax3 and experimental magneto-transport measurements. Both simulations and measurements will explore properties of SLMMS and M2TJs. The measurements will be DC, AC and fast (pulse).
Expected data volume & frequency	From few kB (text) up to tens of MB
Standards and metadata relevant for the data (if relevant)	N/A
Data sharing target(s) outside the project	Experts in the field of spintronics

Collected in WP	WP4: AI algorithm design, validation, and specs
Data set Persistent identifier (PID) and name	PIDs will be created and managed throughout the project. Name(s) to be defined later.
Data type & format	Depending on data accessibility requirements, data will be stored in text or binary format. The fundamental nature of the research conducted by UCLouvain is such that specific data types and formats will be defined on a case-by-case basis. The description and universal means of reading the data will be provided in the form of open-source code to guarantee its long-term reuse.
Data origin (if applicable)	Some data will be generated specifically by this project, in particular for the simplified recognition task linked to the proof of concept. For the more complex recognition tasks, standard datasets, such as the MNIST database, will be re-used for comparison with other solutions.
Partners / Stakeholder(s) / user(s) involved	UCLouvain Also SpinEdge might be involved in some parts of the dataset.
Data set description	The dataset will be composed of pattern databases (simplified version and MNIST) which will serve as a reference for the proof of concept (WP5) as well as for the scaling-up of our solution.
Expected data volume & frequency	The expected order of magnitude is 5-10 GB of data. The dataset will be updated after each corresponding deliverable.
Standards and metadata relevant for the data (if relevant)	Not applicable yet.
Data sharing target(s) outside the project	The standard databases (MNIST, etc.) will be processed and will create new data that will potentially serve as a basis for other research projects. The specific database of simplified patterns created for the PoC will also be of interest to those who compare it to our innovative solution.

MULTISPIN.AI

Collected in WP	WP4: AI algorithm design, validation and specs
	WP5: PoC demonstration
Data set Persistent identifier (PID) and name	Making data easily findable is useful for future researchers and the public as they are more likely to be able to understand the project and use its outputs if they can easily access the different documents. Additionally, the project has an obligation under the Grant Agreement to promote the action by disseminating its results "as soon as feasible, in a publicly available format, subject to any restrictions due to the protection of intellectual property, security rules or legitimate interests."
	Digital Object Identifiers (DOI) as standard identification mechanisms would be implemented according to the needs of the MultiSpin.ai progress. The data will be hosted in a stable and trusted repository to ensure that the research data can be found by others. MultiSpin.ai will identify a suitable repository for open data sets, for example the Zenodo Open European Research Database.
Data type & format	Data formatted to be elaborated by the n-ray cell cross bar that will be sued in the PoC.
	- Text: plain text (TXT), HTML, XML, PDF/A.
	- Databases: XML, CSV.
	- Image: JPEG, JPG-2000, PNG, TIFF.
	- Audio: AIFF, WAVE.
	- Containers: TAR, GZIP, ZIP.
	- Other: Video.
Data origin (if applicable)	Data will be generated by the on board IoT devices of the 2-3-4 wheel road vehicles selected for the PoC.
	There are two types of origin of data identified in the datasets processed by MultiSpin.AI: newly collected data and re-used data.
Partners / Stakeholder(s) / user(s) involved	IFEVS, SpinEdge
Data set description	End user perspective. Data will be collected from the on-the-vehicle IoT network aiming at monitoring the vehicle functionalities and providing situational awareness (safety of driver-rider and vulnerable road users). Data and algorithms formatted for the Proof of concept using a limited version of the n-ray cell cross bar.
	Benchmarking data: tables and spreadsheets
	Electrical design of the PCB board: Bill of Materials, Gerber Files, Centroid Data, Mechanical Files
	PCB driver – source code, text manual
Expected data volume & frequency	The PoC will be performed using two types of data volumes and frequency (bandwidth): 1) three multicamera system for situational awareness 3 x 25 frames second in each of the typical 5-million-pixel camera high bandwidth in a typical ethernet protocol. Algorithms requiring low latency (<30ms) and low energy consumption 2) Control of the on-board functionalities requiring

MULTISPIN.AI

	limited bandwidth in a typical CAN line protocol. Latency of the order of 100ms and low energy consumption. Gigabyte
Standards and metadata relevant for the data (if relevant)	Data formatted to be elaborated by the n-ray cell cross bar. The datasets processed by MultiSpin.AI will have appropriated metadata compiled, which will provide a summary of the characteristics of a dataset. A good metadata record enables the user of a dataset or other information resource to understand the content of what they are reviewing, its potential value and its limitations.
Data sharing target(s) outside the project	The data could be shared within the community of micromobility manufacturers. The algorithms developed could be shared in a wider context than only road mobility.

Collected in WP	WP6 – Communication, dissemination and exploitation
Data set Persistent identifier (PID) and name	/
Data type & format	Mainly two types of data: - Performance data: results of communication productions (click-through rates, opening of campaigns, number of visitors to the various communication media) - People data: on target audiences (age, gender, geographical location)
Data origin (if applicable)	The data are collected through Google Analytics, Twitter Analytics and WordPress (project website)
Partners / Stakeholder(s) / user(s) involved	AMI, (VUB until departure)
Data set description	Collecting and analysing this data enables communication strategies to be adjusted, improving the relevance and effectiveness of messages and maximising the overall impact of the research project.
Expected data volume & frequency	The data will be collected every month and stored in an Excel document.
Standards and metadata relevant for the data (if relevant)	/
Data sharing target(s) outside the project	Nobody

2. Data collection/generation - FAIR data

Methodologies for data collection / generation

Each datatype at each involved partner site will be collected via in-house systems. At the current DMP this information is not yet available. Details of the collection site and system of each dataset will be updated during each revision (new version) of this DMP when more information is available. Besides the data, associated metadata (depending on each datatype) will be also collected at the data collection sites and stored in a central meta-data repository. Files and folders will be named according to local custom in clearly marked MultiSpin.AI folders. Datasets will be named with date and version.

Data quality and standards

Data standards for each data type collected in this project will be evaluated during the annual meetings of the project by means of structured interviews with key researchers in the project.

MultiSpin.AI will manage its data outputs in line with the FAIR principles: findable, accessible, interoperable and reusable. This includes alignment with all the country-level and EU-level guidelines for data privacy and data protection. The following principles will be observed:

Making data findable & accessible

Repository

MultiSpin.AI is oriented towards the Zenodo solution. Zenodo is built and developed by researchers, for Open Science. The OpenAIRE project, for open access and open data movements in Europe was commissioned by the EC to support their nascent Open Data policy by providing a catch-all repository for EC funded research.

For documents sharing within the consortium the project has established a TEAMS group to which all involved team members have access. On this platform all internal project materials and data are stored, from the proposal stage and until today, accessible at any time for the consortium members but protected from non-authorised access. The TEAMS group is managed by the project coordination team at BIU.

For **findability,** we will implement early and open sharing by preregistering the project and its preprints. Zenodo provides a Digital Object Identifier (DOI), which is issued to every published record. Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms, with a few additional enrichments. The DOI is a top-level and a mandatory field in the metadata of each record. Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing. Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

As a rule for **accessibility**, MultiSpin.AI implements the principle "as open as possible, as closed as necessary". The main restrictions will refer to: i) the security-sensible IP of participants, ii) pending patents, and iii) intermediate test results requiring further validation and verification. Zenodo provides metadata for individual records as well as record collections are harvestable using the OAI- PMH protocol by the record identifier and the collection name. Metadata is also retrievable through the public REST API. OAI-PMH and REST are open, free and universal protocols for information retrieval on the web. Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it. Data and metadata will be retained for the lifetime of the repository.

The coordination team opened a community of the project in Zenodo with the name "MultiSpin.AI project", and all partner organisations have been added as members of the community. Whenever a partner wants to upload

an article or data the material will be uploaded through the community. All material will be approved by the Coordination team as the community manager prior to publication.
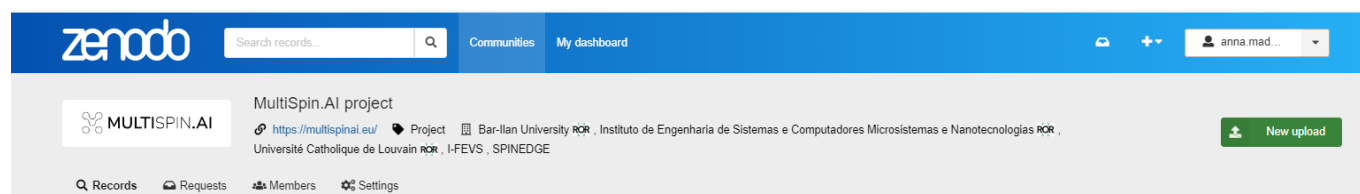
Currently available in the MultiSpin.AI Community in Zenodo are the public deliverables submitted up to date - D6.1, D1.1 and D6.2. As soon as this very deliverable is final it will also be uploaded to the repository. We have also started populating the community with conference material such as posters and presentations, as well as scientific publications.

While the activities carried out so far have resulted in a large number of results from the measurements carried out, it has been decided not to upload that data yet, since it is not of value on its own to stakeholders outside the consortium without the context. The moment there will be a publication describing these activities and results, extensive additional data, not included in the publication, will be made available through Zenodo.

Making data interoperable

MultiSpin.AI will ensure **interoperability** by using commonly used data formats, and ontologies, and will ensure a sufficient metadata framework. Zenodo provides a formal, accessible, shared, and broadly applicable meta(data) language. Zenodo uses JSON Schema as internal representation of metadata and offers export to other popular formats such as Dublin Core or MARCXML. Each referenced external piece of metadata is qualified by a resolvable URL.

Increase data re-use

To enable **reusability** the project's metadata and data will be well-described so that they can be reused, replicated, and/or combined by third parties in different settings. In Zenodo each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments. License is one of the mandatory terms in Zenodo's metadata and is referring to an Open Definition license. Data downloaded by the users is subject to the license specified in the metadata by the uploader. All data and metadata uploaded is traceable to a registered Zenodo user. Metadata can optionally describe the original authors of the published work. Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meets one of the broadest cross-domain standards available.

3. Other research outputs

INESC-MN plans to create a good protocol for memory device nanofabrication in the Clean Room, based on pre-existing know-how. During the project, we will generate several testing devices to be measured by the partners involved in this WP2. The results will be shared during project meetings, and data obtained during characterization shared and made accessible for the partners.

As lead of the WP6 "Communication & Dissemination" of the research project, the AMI is in charge of the management and promotion of the research results, ensuring that they are well accessible and reusable. This involves clear communication of results to external stakeholders via various digital channels, collaboration with

other project teams to ensure adequate, accurate and complete dissemination of research results, and facilitating the understanding and reuse of results obtained by researchers.

By focusing on findability, accessibility, interoperability and reuse, the AMI contributes to maximizing the visibility, impact and usefulness of MultiSpin. AI research results, whether digital or physical.

4. Data security

For data security, we are relying on provisions placed by Zenodo, which has appropriate security measures in place that include recovery and archiving. Zenodo will also be the means to transfer the data between the partners, so we don't need to install additional advanced data sharing tools. If storage in BIU's servers will be needed, we will use password protection to access and retrieve the data.

The data generated by UCLouvain in the framework of WP4 will be fundamental and open data.

5. Ethics

The data collected and generated by UCLouvain is not expected to be subject to ethical issues, and UCLouvain will ensure that this is the case throughout the project.

In WP6 "Communication and Dissemination", it is important to consider the ethical and legal aspects of Multispin.ai project data sharing. Communication activities may involve the dissemination of sensitive data (research results, publications, personal data of registrants, ...) in accordance with the applicable EU, international and national law on data protection (in particular, Regulation 2016/679 - e.g. requiring informed consent from participants). In addition, communication data may be subject to copyright, requiring special attention to ensure legal compliance. It is also crucial to consider ethical issues related to the presentation and interpretation of research results.

WP6 will also ensure clear specifications regarding which data will be preserved and for how long in questionnaires on this topic.

In summary, WP6 will pay attention to the ethical and legal implications of data sharing and take the necessary steps to comply with applicable EU, international and national standards and regulations.

6. Conclusion

The current version of the DMP will be updated as needed along the MultiSpin.AI project, as data formats or storage requirements may change during the acquisition of data.

All project related material will be made available via the "MultiSpin.AI project" community in Zenodo, following approval by the coordination team.