“MIETMAN”

Modeling of Ionic and Electronic Transport in 2D Materials Toward Memristive Applications in Neuromorphic Computing

Grant Agreement: 101032701

Project Acronym: MIETMAN

Horizon 2020

H2020-MSCA-IF-2020

Marie Skłodowska-Curie Individual Fellowships

MSCA EF-ST

Standard European Fellowships

WP5 – “MANAGEMENT & COORDINATION”

Deliverable D2.1

“DMP - Version 2.0”

Table of Content

1. Data Summary

1.1. Introduction and Purpose

This report describes the initial **Data Management Plan (DMP)** for the project **MIETMAN**, which is funded by the European Commission through the Marie Sklodowska-Curie Action (MSCA) as part of the Individual European Standard Fellowship Programme (IF-EF-ST).

MIETMAN aims to demonstrate the ability to and the importance of combining different levels of abstraction to understand and propose optimized memristors, i.e. solid state nano-devices specifically designed to emulate the synaptic behavior of neurons in the human brain. The project comprises a comprehensive computational study from first principles atomistic simulations, to numerical calculations at the device level and circuit simulations aiming at final neuromorphic applications.

During the progress of the project, MIETMAN is expected to generate considerable amounts of data from the experimental and theoretical research activities. In this sense, committed to the FAIR policies MIETMAN will make its research data findable, accessible, interoperable and reusable. To this aim, this DMP provides a detailed description of the procedures and protocols followed to manage the datasets generated during the lifetime of the project, as well as the main data management principles in terms of data and metadata standards, sharing, archiving, preservation and security.

The responsible of the DMP is Dr. Mohit D. Ganeriwala, recipient of the MSCA fellowship and member of the Electronics and Computer Technology department at the University of Granada (UGR), Spain, beneficiary host institution.

This is an active document which will be updated in a regular basis during the lifetime of the project and allocated in the institutional repository of the UGR, DIGIBUG, under the name of **MIETMAN_01_DMP_V#.#-Wp5.pdf** (see section 2.1 for naming conventions). The current document is the first version of the DMP and it includes an overview of the datasets to be produced by the project, and the specific conditions related with it. More details about the procedures implemented within the MIETMAN project will be included in future versions.

1.2. Relevance to Project Targets

MIETMAN will generate a considerable number of datasets of different types both quantitative and qualitative. Three main data categories have been identified in order to classify the proposal results:

1. Research objectives: These datasets will enable the replication of the main scientific outcomes of the project, including data from experiments and computer simulations as well as the codes to produce and analyze data. They are mainly directed to the scientific community with expertise in the field of research of the proposal.
2. Dissemination activities for expert audiences: These datasets will provide access to those documents summarizing the main scientific results of the project. This includes preprints, technical reports, and conference presentations. They are aimed at a specialized audience, i.e. a technical background will be required to achieve a fill comprehension of the content of these documents.
3. Communication activities and dissemination for non-experts audiences: These datasets will give access to any potential user to those documents dedicated to educational purposes. This includes presentations used in events for non-expert audiences and teaching material for undergraduate and graduate students related with the project.

1.3. Type, Origin, Format and Quantity of Data

MIETMAN seeks the development of a multi-scale modeling and simulation framework, combined with the experimental characterization of working prototypes of memristors based on two-dimensional materials (2DMs), eventually aiming to implement neuromorphic functions potentially able to emulate brain inspired computation processes.

Consequently, **MIETMAN** will produce scientific data in the form of: i) simulation codes and simulation results, and ii) different experimental characterization techniques. In addition to the scientific objectives, **MIETMAN** will also produce dissemination and communication documents aimed at both: i) expert audiences as well as ii) the general public with educational purposes. A brief description of the main objectives (both scientific and pedagogical) can be found in the work package table included in Appendix A.

In more detail, MIETMAN will generate data with different formats that will be accessible using free software (see table in section 2.2). The different formats are presented in the table below.

Table 1.3.1.

Type of Data	Description	Format[1]
Compressed Data	Compression will be used for sorting and packaging files with similar and/or complementary content	.tar (typical compression format managed by numerous free programs)
Raw Data	Experimental measurements and computational results	.txt, .dat, .csv, .xlsx. xyz
Images	Illustrations, schemes, and figures scientific and educational purposes	.jpeg, .png, .pdf
Codes	Scripts, functions and software packages ssed for both, running simulations and analyzing data in different programming languages	.py, .cmd, .v (codes) and .dat, .txt (parameter files and readme's)
Reports	Scientific, technical, and educational documents	.pdf
Paper preprints	Pre-edited versions of journal and conference manuscritps compliant with green open access according to the H2020 guidelines	.pdf
Software specific files	Input files to replicate computational results in specific software	.ini, .hdf5, .cif, .cmd, .sp

[1] Formats shown in this column are considered as the most appropriate for each type of data. However, a set of data will be saved only in one format so as to avoid duplications.

Presentations	Scientific and educational talks and slides	.pdf, .key, .pptx (talks and posters), .mp4

MIETMAN is structured around 5 work packages that deal with the numerical and experimental research activities as well as with the communication and dissemination actions. The next 5 tables summarizes, the different types of datasets that will be generated by **MIETMAN** in each of the 5 work packages (that are presented in more detail in Appendix A).

Table 1.3.2.

Work Package 1: Ab initio Study of 2D Material Properties			
Type of Dataset	**Origin**	**Format**	**Size**
Crystal structure files	Computational	.tar (.cif, .hdf5)	≈ 20 G
Results (data)	Computational	.tar (.txt)	≈ 700 MB
Results (images)	Computational	.tar (.jpeg)	≈ 50 MB

Table 1.3.3.

Work Package 2: Numerical Modeling of Memristive Systems Based on 2D Materials			
Type of Dataset	**Origin**	**Format**	**Size**
Python Codes (for visualization and analysis of data)	Computational	.tar (.py)	≈ 10 MB
Results (data)	Computational	.tar (.txt)	≈ 10 MB
Results (images)	Computational	.tar (.png)	< 10 MB

Table 1.3.4.

Work Package 3: Device Fabrication and Experimental Characterization (secondment)			
Type of Dataset	**Origin**	**Format**	**Size**
AFM Images	Experimental	.tar (.jpeg)	≈ 300 KB
Electrical characterization data	Experimental	.tar (.xlsx)	≈ 50 MB
Python Codes (for visualization and analysis of data)	Computational	.tar (.py)	≈ 10 KB

Table 1.3.5.

Work Package 4: Development of Compact Model for Circuit Level Simulations			
Type of Dataset	**Origin**	**Format**	**Size**
Codes (Verilog A)	Computational	.tar (.v)	≈ 5 KB
SPICE and Model extraction files	Computational	.tar (.mdl, .dds)	≈ 200 KB
Results (data)	Computational	.tar (.xlsx)	≈ 20 KB
Results (images)	Computational	.tar (.jpeg)	≈ 200 KB

Table 1.3.6.

Work Package 5: Management, Coordination and Dissemination			
Type of Dataset	**Origin**	**Format**	**Size**
DMP	Report	.pdf	≈ 200 KB
Final or preprint versions of papers for expert audiences	Paper Preprint	.pdf	≈ 20 MB
Presentations in international conferences and collaborators for expert audiences (talks and posters)	Presentations and Posters	.pdf	≈ 20 MB
Presentations in events for non-expert audiences (including talks, videos, and posters)	Presentation	.tar (.pdf, .key, .pptx) .mp4	< 1000 MB

1.4. Data Value and Utility

The successful achievement of the scientific main goals of **MIETMAN** will require of comprehensive computational studies as well as comprehensive sets of characterization data extracted from fabrication and characterization of the prototype devices. The data supporting the main scientific results of this project and their dissemination will be of interest to expert audiences across multiple disciplines including engineers, physicists, biologists (and neuroscientist), chemists, computer scientists, and professionals from semiconductor industry.

In particular, a repository of the project scientific data will be fundamental for the training of new researchers in the field and specifically to those joining the research group at the host institution, so to reduce the learning curve and to allow a faster integration of new researchers within the project activity.

The data obtained from experimental measurements and simulations will also be useful for theoretical groups as input for theoretical modelling or as a reference for future comparison and for the reproducibility of published results. These scientific data will be specially important for the private sector and for other

academic research groups working in nanomaterials, neuromorphic computing, simulation of 2D materials and devices, etc.

MIETMAN will also generate data useful for pedagogical purposes targeting to non-expert audiences, including students and teachers of elementary and secondary education, undergraduate and graduate students (particularly in the broad fields of physics, computer science and electrical engineering), and general public. The tools employed to disseminate the generated knowledge will make use of different media, such as “Canal UGR” belonging to the press release office of the UGR, events focused on attracting STEM students to UGR from the surrounding high schools, and social networks such as LinkedIn, Instagram, and/or Twitter.

2. FAIR Data

2.1. Making Data Findable

To ensure data visibility, the metadata system used for the description of the materials hosted in the UGR repository, DIGIBUG, is Dublin Core Qualified. This is a metadata initiative adopted by the European repository OpenAIRE. DIGIBUG (https://digibug.ugr.es/?locale-attribute=en) assigns a unique identifier (handle) to each document and/or dataset, which allows the identification and citation of electronic documents. The guidelines provided by DIGIBUG will be used by MIETMAN to ensure the right format of data uploaded to comply with FAIR principles.

Each dataset generated by **MIETMAN** will be recorded by a dataset identifier in the general file **MIETMAN_DATASET_LIST.pdf**, which will be hosted (and regularly updated) in DIGIBUG. The information of each individual dataset will be included in a linked metadata file, which will be updated (when necessary) and will contain the information appearing in the table below.

Table 2.1.1.

Contents of a Generic Metadata File Associated to a Given Dataset	
Dataset Identifier	**ID in accordance to the naming convention in the next table**
Title of the Dataset	The title of the dataset, which could be used to locate the data
Responsible Partner	Partner institution responsible for the creation of the dataset (UGR or AMO)
Work Package	Project work package with which the dataset is associated (e.g. WP1)
Dataset Description	A brief description of the dataset
Dataset Dissemination	Where will the dataset be disseminated (e.g. peer reviewed journal)
Format	Format/type of dataset (e.g. .pdf, .jpeg, etc)
Source	Origin of the dataset (e.g. experimental, computational)
Version	Version of the current dataset

The dataset identifier will be created according to the convention presented in the table below. The corresponding metadata file will be suffixed with META. All the metadata files will have extension .pdf.

Table 2.1.2.

Convention for Creating the Dataset Identifier	
Components	**Example**
Project Name	MIETMAN (fixed prefix)
Serial number (corresponding to the order of appearance in MIETMAN_DATASET_LIST.pdf)	10
Title of the Dataset	PtSe2_Memristor_IV_data
Version	V1.0
Work Package	WP3
Format of the Dataset	.dat
Example dataset filename : MIETMAN_10_PtSe2_Memristor_IV_data_V1.0_WP3.dat **Example metadata filename : MIETMAN_10_PtSe2_Memristor IV_data_V1.0_WP3_META.pdf**	

2.2. Making Data Openly Accessible

The data hosted in the institutional repository DIGIBUG will be accessible to the whole research community. In this respect, the data generated by **MIETMAN** does not entail Intellectual Property (IP) rights such as patents, trademarks, or copyrights. The agreement, supervised by UGR's Research Results Transfer Office, between the beneficiary institution (UGR) and the DMP coordinator complies with the IP-MSCA rules for access rights (royalty-free basis) and results ownership. The data generated will be accessible through the DIGIBUG website and will be open to any user without restrictions.
In the table below we provide examples of software packages to open and/or use the data generated by **MIETMAN** for different operating systems (Linux, MacOS, and Windows) and for all the formats. These examples are not unique and the potential user might employ other standard packages and/or platforms. At least one of the options for each format is open source for an open source Linux based OS (e.g. Ubuntu).

Table 2.2.1.

Format	Example of Software (OS) to access
tar.gz	uzip (Linux), tar (MaxOS), winrar (Windows)
.dat, .txt, .cmd, .v, .sp	gedit, emacs (Linus, MacOS), Notepad (Windows)
.pdf	Adobe Reader (Linux, MacOS, Windows)
.jpeg, .png	Image viewer (Linux), Photo Viewer (Windows), Preview (MacOS)
.py	gedit, emacs (Linus, MacOS), Notepad

	(Windows), jupyter (Linux, MacOS, Windows)
.xlsx, .pptx	WPS office, MS-Office (Linux, MacOS, Windows)
.key	Keynote (MacOS)
.mp4	VLC (Linux, MacOS, Windows)
.cif	Xcrysden (Linux, MacOS)
.ini, .hdf5, .cif	Quantum ATK (Linux, MacOS, Windows)
.cmd	Synopsys TCAD (Linux, MacOS, Windows)
.sp, .v	HSPICE (Linux, MacOS, Windows)

All the datasets generated by MIETMAN will be allocated in DIGIBUG without costs (neither in the short nor in the long term), time limitation, or access restrictions. In particular, for what concerns the published articles, and specifically those published in journals which do not offer green open access, we will upload in public repositories the final version of the pre-edited manuscript, where we will also attach the raw data included in the generation of the published figures. An example of these public repositories would be arxiv.org, a private not-for-profit educational repository owned and operated by Cornell University. In case arxiv is finally used, we would upload the same preprint title as the one used for the version allocated in DIGIBUG, ensuring and enhancing green open access according to the H2020 guidelines.

2.3. Making Data Interoperable

MIETMAN aims to collect and document all the data generated in a standardized way (Dublin Core Qualified) to ensure that all datasets, which will be accompanied by the corresponding metadata file, can be interpreted and shared.

In this respect, a metadata file will be created and linked to each dataset. These metadata files will include all the information detailed in the first Table of section 2.1 (contents of a generic Metadata file associated with a given Dataset).

2.4. Increase Data Re-use

The datasets will be made available for their re-use and stored in DIGIBUG without any cost. If datasets are updated, the coordinator of this DMP will be responsible for managing the different data versions, making sure that the latest version is available. The policies adopted by DIGIBUG concerning licenses, availability periods, and quality are detailed as follows:

A. License

All materials published in the DIGIBUG repository incorporate different licenses of the nonprofit organization Creative Commons, "**Attribution-Noncommercial-No Derivative Works 3.0**". This license allows the reuse of data at the end of the project and their use by third parties.

B. Availability

The data will remain reusable after the end of the project with no time limitations. There are no restrictions on the use of the published data nor access restrictions, unless embargo or access restrictions are eventually indicated.

C. Quality

The quality of the datasets is guaranteed by the DIGIBUG operating software, which performs routine backups and checking of the material hosted.

3. Allocation of Resources

All the datasets generated by **MIETMAN** will be allocated in DIGIBUG without costs (neither in the short nor in the long term), time limitation, or access restrictions. Another version of the produced paper preprints will be uploaded on the arXiv to ensure and enhance green open access according to the H2020 guidelines. The associated costs for dataset preparation for archiving will be covered by the project itself. Resources for long term preservation, associated costs and potential value, as well as how data will be kept beyond the project and how long, will be provided by DIGIBUG as it is totally maintained and updated by the UGR.

Mohit D. Ganeriwala, member of the Electronics and Computer Technology department at the UGR (Spain), will be responsible for the data management within the **MIETMAN** project; in particular for the creation of the DMP and its subsequent updates, and for recording and updating the datasets generated by MIETMAN.

4. Data Security

The Scientific Documentation Service of the UGR, located in the Library of the Hospital Real (Granada, Spain), coordinates the electronic management of the DIGIBUG repository. In particular, DIGIBUG incorporates a program for backup and preservation. In this respect, the **MIETMAN** datasets hosted in DIGIBUG will receive the same security treatment as the rest of the documents in this repository. All the responsibilities concerning data recovery and secure storage will go to the Scientific Documentation Service of the UGR, which is in charge of the storing of the datasets hosted in DIGIBUG.

In addition to this safety arrangements, the experienced researcher will perform a second backup in an independent server for a period well beyond the lifespan of the project.

5. Ethical Aspects

As stated in the Ethics Issues of the MSCA-IF-EF-ST grant agreement (101032701), there is no requirement for ethical review since **MIETMAN** does not involve the use of human participants, human cells or tissues, personal data collection and/or processing, animals, potential for misuse of research results, or elements that may cause harm to the environment, animals or plants.

6. References

Guidelines on FAIR Data Management in Horizon 2020, version 3.0, 26 July 2016
Data Management Plan Template, UGR Library, version 1.0, 1 July 2017
J. Horst and C. Lynch, AMECRYS Data Management Plan, version 3.0, 29 March 2017