

## DATA MANAGEMENT PLAN

PROJECT	
Project number:	101064402
Project acronym:	HICODY
Project name:	Mathematical Challenges of Higher-Order Interactions in Collective Dynamics, and Applications

DATA MANAGEMENT PLAN	
Date:	09/01/2023
Version:	1.0

### 1. Data Summary

#### 1.1. Introduction and purpose

This report describes the initial **Data Management Plan** (DMP) for the project **HICODY (101064402)**, which is funded by the European Commission through the Marie Skłodowska-Curie Actions (MSCA) as part of its Postdoctoral Fellowship Programme (MSCA-PF). The coordinator of this DMP, who is also the recognized researcher associated to the funded fellowship, is Dr. Jesús David Poyato Sánchez, from the department of Applied Mathematics at the University of Granada (UGR), Spain (the beneficiary institution).

The purpose of this DMP is to provide a detailed description of the procedures and protocols for the management of the datasets generated during the lifetime of the project. This DMP will describe the main data management principles in terms of data standards and metadata, sharing, achieving, preservation, and security.

This is an alive document that will be updated at regular intervals during the lifetime of the project and be allocated in the institutional repository of the UGR, DIGIBUG, under the name of **HICODY\_01\_DMP\_V1.0\_WP1.pdf**.

#### 1.2. Relevance to project targets

HICODY will generate several datasets of different types. The data management will serve to support the project scientific objectives and spread the project results. This includes the management of three main data categories:

- A. **Research objectives.** The datasets associated to this category will allow any potential user to replicate the main scientific results of the project. These include codes and numerical simulations.
- B. **Dissemination activities for expert audiences.** The datasets associated to this category will give access to any potential user to those documents summarizing the main scientific results of the projects. This includes scientific papers, preprints, and conference presentations.
- C. **Communication activities for non-expert audiences.** The datasets associated to this category 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.

### 1.3. Type, origin, format, and quantity of data

HICODY aims to investigate patterns of emergent behaviour and collective dynamics arising in large multi-agent systems (flock of birds, swarms of bacteria, brain neurons, etc) which are governed by higher-order interactions, in contrast with most classical approaches based on binary interactions. The project will study various scales of description of the problem, from microscopic, to mesoscopic and macroscopic (which are more adapted to populations with high density), using scaling limit methods like mean-field and hydrodynamic limits. In doing so, the project HICODY will generate data to disseminate results between different expert audience, data for educational purposes, and data arising from the numerical simulations, which may become necessary in some stages of the project to find intuition to guide the mathematical outcomes. A brief description of the main objectives (both scientific and pedagogical) can be found in the work package table presented in **Annex A**.

HICODY will generate data with different formats that will be accessible using free software (see **Section 2.2**). The different formats are presented in the table below.

Type of Data	Description	Format
<b>Compressed data</b>	Apart from saving data storage, compression will be used for packaging files with similar and/or complementary content	<b>TAR.GZ</b> (other formats inside)
<b>Images</b>	Scientific and educational purposes	<b>JPEG, PNG, TIFF, PDF</b>
<b>Codes</b> (programming languages)	Used for both running simulations and analyzing data	<b>PYTHON</b> (codes)
<b>Reports</b>	Scientific, technical, and educational purposes	<b>PDF, TEX</b>
<b>Paper preprints</b>	Green open access according to the H2020 guidelines	<b>PDF, TEX</b>
<b>Presentations</b>	Scientific and educational purposes	<b>PDF</b> (talks), <b>mp4</b> (video)
<b>Websites</b>	Scientific, technical and educational purposes	<b>HTML</b>

HICODY will create data from **four distinct origins: computational** (codes), **reports** (technical and pedagogical), **paper preprints**, and **presentations** (technical and pedagogical). In the following six tables we summarize in a tentative scheme the different types of datasets that will be generated by HICODY linked to the six project work packages presented in **Annex A**. The total estimated amount of data is around 5GB, with datasets ranging from 10MB to 200MB.

<b>WP1. Management of the action</b>			
Type of dataset	Origin	Format	Quantity
Data Management Plan	Report	PDF	< 10 MB
Midterm report to EU	Report	PDF	< 10 MB
Financial report to EU	Report	PDF	< 10 MB
Final report to EU	Report	PDF	< 10 MB

<b>WP2. Scaling limits and emergent phenomena in CDM with HOI</b>			
Type of dataset	Origin	Format	Quantity
Mean-field limit of CDM with HOI	Report	PDF	< 10 MB
Hydrodynamic limit of CDM with HOI.	Report	PDF	< 10 MB
Long-time behavior of CDM with HOI.	Report	PDF	< 10 MB

<b>WP3. Computational study of activity patterns in large neural networks with HOI</b>			
Type of dataset	Origin	Format	Quantity
Codes for the generating numerical simulations of large neural networks with HOI.	Computational	PYTHON	< 50 MB
Files containing graphical representations coming from computational data.	Computational	JPEG	< 1 GB
Search for activity patterns: phase transitions, abrupt transitions to synchronization and cluster states in large networks with HOI.	Report	PDF	< 10 MB

<b>WP4. Macroscopic PDE for filopodia-mediated morphogenesis</b>			
Type of dataset	Origin	Format	Quantity
Mean-field limit of the filament-based CDM for orientation.	Report	PDF	< 10 MB
Derivation of macroscopic PDE for filopodia-mediated morphogenesis.	Report	PDF	< 10 MB

<b>WP5. Training and Career Development</b>			
Type of dataset	Origin	Format	Quantity
Initial teaching training course	Report	PDF	< 10 MB
Training of the UGR Research Promotion Plan (guidelines for project development, digital identity and reputation, promotion of interdisciplinary research, etc)	Report	PDF	< 10 MB
Career Development Plan (with a description of the hands-on-training and management activities)	Report	PDF	< 10 MB
Research training on Python (scientific computing)	Report	PDF	< 40 MB
Research training on complex networks and neuroscience (scientific computing and modeling)	Report	PDF	< 1 MB

<b>WP6. Dissemination, exploitation, communication and outreach activities</b>			
Type of dataset	Origin	Format	Quantity

Final preprint versions of papers for expert audiences	Paper preprint	PDF, TEX, JPEG	< 20 MB
Presentations in international conferences for expert audiences (including talk presentations, videos, and posters)	Presentation	TAR.GZ (PDF, TEX, JPEG, mp4)	< 20 MB
Outreach material (primary/secondary schools and the general public)	Presentation	TAR.GZ (PDF, TEX, JPEG, mp4)	< 20 MB
Teaching materials for undergraduate and graduate courses (MSc in Physics and Mathematics, BSc in Administration and Direction of Enterprises, BSc in Civil Engineering, summer schools, etc)	Report	TAR.GZ (PDF, TEX, JPEG, mp4, Python)	< 200 GB
Book of abstracts of organized workshop, special sessions at conferences and summer schools	Report	PDF	< 10 MB
Entries (titles and abstracts) on the website of the weakly “Seminar of Differential Equations” (UGR)	Website	HTML	<10 MB

#### 1.4. Data value

HICODY aims to investigate scaling limits along different scales and patterns of collective dynamics arising in large multi-agent systems which are governed by higher-order interactions. This problem is common to the classical approach in kinetic theory and statistical mechanics, but however non-exchangeability of agents and the addition of higher-order interactions in this class of systems makes difficult to understand how multi-agent systems self-organize and cooperate at different time and length scales. We shall conduct the study in the case of flocking, swarming and synchronization dynamics (e.g., flock of birds, swarms of bacteria, brain neurons, etc).

The data supporting of the project scientific outcomes and their dissemination will reach a broad range of expert audiences. This includes: physicists, biophysicists, biologists, engineers and computer scientists among others.

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 for a faster integration of new researchers within the project activity.

The project will also generate data for pedagogical purposes targeting non-expert audiences. This includes: students and professors of elementary education, secondary education, undergraduate and graduate students (particularly in mathematics and physics), and general public.

## 2. FAIR data

### 2.1. Making data findable, including provisions for metadata

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 assigns a unique identifier (handle) to each document and/or dataset, which allows the identification and citation of electronic documents.

Each dataset generated by HICODY will be recorded by a dataset identifier in the general file HICODY\_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 (if necessary) and will contain the information appearing in the table below.

Contents of a generic metadata file associated to a given dataset	
Data identifier	The ID will result from the naming convention provided in the next table
Dataset title	The title of the dataset, which will be easily searchable and findable
Responsible partner	Partner institution responsible for the creation of the dataset (always UGR)
Work package	Project work package (see work package table in <b>Annex A</b> )
Dataset description	A brief description of the dataset
Dataset benefit	What are the benefits of the dataset
Dataset dissemination	Where will the dataset be disseminated
Format type	Dataset format (see table with formats in <b>Section 1.3</b> ).
Expected size	Dataset size (see table with formats in <b>Section 1.3</b> )
Source	How was the dataset generated
Repository	DIGIBUG (arXiv and/or HAL also for final preprints, see <b>Section 4</b> )
DOI (if known)	The DOI will be entered once the dataset has been deposited
Date of Submission	The date of submission will be added once the dataset has been uploaded to the repository
Keywords	Keywords associated with the dataset
Version Number	Version number to keep track of changes

The dataset identifier (first field in the previous table) will be created according to the convention presented in the table below. The corresponding metadata file will be named by adding **META** at the end of the identifier of its linked dataset file. All the metadata files will have extension **.dat**.

Convention for creating the Dataset Identifier	
Components	Example
Project name	HICODY (always)
Two digits chronological number (corresponding to the order of appearance in HICODY_DATASET_LIST.pdf)	01
Title of the dataset	python_code_order_parameter
Version of the dataset (DIGIBUG will allow to keep several versions)	V1.0
Work Package associated to the dataset	WP3
Format of the dataset	TAR.GZ

Example dataset file identifier: **HICODY\_01\_python\_order\_parameter\_V1.0\_WP3.TAR.GZ**

Example metadata file identifier: **HICODY\_01\_python\_order\_parameter\_V1.0\_WP3\_META.dat**

## 2.2. Making data accessible

The data hosted in the institutional repository DIGIBUG will be accessible to the complete research community. In this respect, the data generated by HICODY 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 MSCA-PF 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.

The table below provides examples of software packages to open and/or use the data generated by HICODY for different operating systems (Ubuntu, MacOS, and Windows) and for all the formats. These examples are not unique, and the potential user might use other standard packages and/or platforms. At least one of the options for each format is free software based (e.g., Ubuntu).

Format	Example of software (OS) to access
TAR.GZ	TAR.GZ package (Ubuntu), tar utility (MacOS), WinRAR (Windows)
DAT, TXT	GNU emacs (Ubuntu, MacOS), Notepad (Windows)
PDF	Adobe Reader (Ubuntu, Windows, MacOS)
XLSX, PPTX	WPS office, MS-Office (Ubuntu, MacOS, Windows)
JPEG, PNG, TIFF	gThumb (Ubuntu), Preview (MacOS), Adobe Illustrator (Windows)
PYTHON	Python sources releases (Ubuntu, MacOS, Windows) in <a href="http://www.python.org">www.python.org</a>
MP4	VLC (Ubuntu, MacOS, Windows)

All the datasets generated by HICODY 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 open access 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, or hal.science, a public not-for-profit educational repository owned and operated by the French National Centre for Scientific Research (CNRS). In the case that either arXiv or HAL are 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

HICODY 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 to a given dataset).

## 2.4. Increase data re-use

The datasets will be made available for their re-use and be 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:

1. **License:** All materials published in the DIGIBUG repository incorporate different licenses of the nonprofit organization **Creative Commons**, in particular, the **Creative Commons 4.0 version of the Non-Commercial-ShareAlike CC BY-NC-SA license**, which is the one recommended by **OpenAIRE**. This license allows the reuse of data at the end of the project and the use by third parties.
2. **Availability:** The data will remain reusable after the end of the project with no time limitations nor access restrictions, unless embargo or access restrictions are eventually indicated.
3. **Quality:** The quality of the datasets is guaranteed by the DIGIBUG operating software, which performs routine backups and checking of the material hosted.

## 3. Other research outputs

HICODY does not involve further research outputs different from the ones described in Section 2, neither digital (e.g., software, workflows, protocols, models, etc) nor physical (e.g., new materials, antibodies, reagents, samples, etc).

## 4. Allocation of resources

All the datasets generated by HICODY will be allocated in **DIGIBUG** without costs (neither in the short nor in the long term), time limitation, or access restrictions. In particular, another version of the produced preprints will be uploaded on **arXiv** (owned and operated by Cornell University) and **HAL** (owned and operated by CNRS), see **Section 2.2**. For the arXiv and HAL versions we will use the same preprint title as the one used for the version allocated in DIGIBUG. With this we will ensure and enhance **green open access according to the H2020 guidelines**.

Dr. **David Poyato**, **Department of Applied Mathematics** at the **UGR** (Spain), will be responsible for the data management within the HICODY project, in particular for the creation of the DMP and its subsequent updates, and for recording and updating the datasets generated by HICODY.

## 5. 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 HICODY 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.

## 6. Ethics

As stated in the Ethics Issues of the **HORIZON-MSCA-2021-PF-01** grant agreement **101064402** there is no requirement for ethical review since HICODY 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.

## 7. Other issues

HICODY will not make use of other national, funder, sectorial or departmental procedures for data management.

## 8. Annex A: Work packages

WP #	1	Start-End Months	M1-M24	Total Months	24	Person Month (PM)	1.2
Title	<b>Management of the action</b>						
Tasks	Kick-off meeting; Group meetings and contingency planning; Administrative meetings with UGR. Preparation of reports for REA.						
WP #	2	Start-End Months	M1-M12	Total Months	12	Person Month (PM)	8.4
Title	<b>Scaling limits and emergent phenomena in CDM with HOI</b>						
Tasks	Mean-field limit of CDM with HOI; Hydrodynamic limit of CDM with HOI; Long time behaviour of CDM with HOI.						
WP #	3	Start-End Months	M13-M18	Total Months	6	Person Month (PM)	4.2
Title	<b>Computational study of activity patterns in large neural networks with HOI</b>						
Tasks	Numerical simulation of the mean-field equation; Search for activity patterns: phase transitions, abrupt transitions to synchronization and cluster states.						
WP #	4	Start-End Months	M19-M24	Total Months	6	Person Month (PM)	4.2
Title	<b>Macroscopic PDE for filopodia-mediated morphogenesis</b>						
Tasks	Mean-field limit of the filament-based CDM for orientation; Derivation of macroscopic PDE for filopodia-mediated morphogenesis.						
WP #	5	Start-End Months	M1-M24	Total Months	24	Person Month (PM)	3.6
Title	<b>Training and Career Development</b>						
Tasks	Meeting with supervisor; Training in complex networks; Teach at undergraduate courses; Teach at graduate courses; Attend Python course; Co-organize summer schools, workshops, congresses and the <i>Seminar of Differential Equations</i> at UGR (weekly); Enrol in <i>Research Promotion Plan</i> of UGR; Supervise MSc & PhD students.						
WP #	6	Start-End Months	M1-M24	Total Months	24	Person Month (PM)	2.4
Title	<b>Dissemination, exploitation, communication and outreach activities</b>						
Tasks	Prepare articles and lectures; Write entries in website; Give lectures at seminars; Give communications at international conferences; Engage in outreach activities.						

## 9. List of Acronyms

Acronym	Meaning
CDP	Career Development Plan
CDM	Collective Dynamics Models
DMP	Data Management Plan
HOI	Higher-order interactions
IP	Intellectual Property
UGR	University of Granada

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0	09/01/2023	Initial Document Creation