ARIADNE - Structure and dynamics in active glass-forming liquids

Data Management Plan

Project Name	Structure and dynamics in active glass-forming liquids
Project Acronym (number)	ARIADNE (840195)
Funding	European Commission - MSCA-IF-EF-ST (2018)
Coordinator	Sándalo Roldán-Vargas
Beneficiary	University of Granada, Spain
Deliverable Number	1
Deliverable Title	Data Management Plan
Due Date	01/12/2019
Issue Date	29/11/2019
Status	Submitted
Dissemination level	Public
Contact	sandalo@ugr.es

Version	Modifications	Date	Author
1.0	Initial Document Creation	29/11/2019	Sándalo Roldán-Vargas

Contents

1. Data Summary

1.1. Introduction and purpose

This report describes the initial **Data Management Plan** (DMP) for the project **ARIADNE (840195)**, which is funded by the European Commission through the Marie Skłodowska-Curie Actions (MSCA) as part of the Individual European Standard Fellowship Programme (IF-EF-ST). The coordinator of this DMP, who is also the experienced researcher associated to the funded fellowship, is Dr. Sándalo Roldán-Vargas from the department of applied physics 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, archiving, 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 **ARIADNE_01_DMP_V1.0_WP4.pdf** (see section 2.1 for naming conventions).

1.2. Relevance to project targets

ARIADNE will generate several datasets of different types both quantitative and qualitative. 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:

1- Research objectives. The datasets associated to this category will allow any potential user to replicate the main scientific results of the project. This includes data from experiments and computer simulations as well as codes to produce and analyze data.

2- 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 project. This includes preprints, technical reports (*e.g.* protocols), and conference presentations.

3- Communication activities for non-experts 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 distinct non-expert audiences and teaching material for undergraduate and graduated students.

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

ARIADNE aims to investigate the dynamic and structural collective patterns arising in systems of active particles (in particular bacterial colonies) at high densities. To this purpose, ARIADNE proposes two complementary approaches. On one hand, during the lifetime of the project ARIADNE will generate data coming from three different experimental techniques: light scattering (static and dynamic), rheology, and microscopy (optical and transmission electron microscopy, TEM). On the other hand, ARIADNE will generate data coming from Molecular Dynamics (MD) simulations whose codes will be created from scratch and will be open to any potential user. Apart from the research objectives, ARIADNE will generate data to disseminate its results between different expert audiences and data for educational purposes. A brief description of the main objectives (both scientific and pedagogical) can be found in the work package table presented in Annex A.

ARIADNE 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.

Type of Data	Description	Format
Compressed data	Apart from saving data storage, compression will be used for packaging files with similar and/or complementary content	**TAR.GZ** (other formats inside)
Crude numerical data	Experimental and computational	**dat**
Images	Scientific and educational purposes	**JPEG, PNG, TIFF, PDF**
Codes (programing languages)	Used for both running simulations and analyzing data	**FORTRAN** and **Python** (codes), **dat** (parameter files and readme's)
Reports	Scientific, technical, and educational purposes	**PDF**
Paper preprints	Green open access according to the H2020 guidelines	**PDF**
Graphical representations	Used for graphical analysis and Molecular Visualization	**XMGRACE, VMD**
Presentations	Scientific and educational purposes	**PDF** and **Keynote** (talks and posters), **mp4** (video)

ARIADNE will create data from **five distinct origins**: **experimental**, **computational**, **reports** (technical and pedagogical), **paper preprints**, and **presentations** (technical and pedagogical). In the following four tables we summarize in a tentative scheme the different types of datasets that will be generated by ARIADNE linked to the four project work packages presented in Annex A. The total estimated amount of data is around 7GB, with datasets ranging from 10MB to 1000MB.

Work Package 1. Location of the Glass Transition (Research Objectives O1, O2, and O3)			
Type of dataset	**Origin**	**Format**	**Quantity**
Dynamic Light Scattering (DLS) measurements	Experimental	TAR.GZ (dat)	< 200 MB
Static Light Scattering (SLS) measurements	Experimental	TAR.GZ (dat)	< 200 MB
Rheology measurements	Experimental	TAR.GZ (dat)	< 100 MB
Optical microscopy images	Experimental	JPEG, PNG, TIFF, PDF	< 300 MB
Optical microscopy data	Experimental	TAR.GZ (dat)	< 500 MB
TEM images	Experimental	JPEG, PNG, TIFF, PDF	< 200 MB
Data from MD simulations (Isotropic systems)	Computational	TAR.GZ (dat)	< 1000 MB
Data from MD simulations (Non-isotropic systems)	Computational	TAR.GZ (dat)	< 500 MB
Data from MD simulations (Polar systems)	Computational	TAR.GZ (dat)	< 500 MB
MD codes (Isotropic systems)	Computational	TAR.GZ (FORTRAN, dat)	< 50 MB
MD codes (Non-isotropic systems)	Computational	TAR.GZ (FORTRAN, dat)	< 50 MB
MD codes (Polar systems)	Computational	TAR.GZ (FORTRAN, dat)	< 50 MB
Codes for analyzing experimental Data (see previous data with an experimental origin)	Computational	TAR.GZ (FORTRAN, Python, dat)	< 50 MB
Codes for analyzing computational Data (see previous data with a computational origin)	Computational	TAR.GZ (FORTRAN, Python, dat)	< 50 MB
Files containing graphical representations coming from experimental and computational data	Experimental, Computational	TAR.GZ (XMGRACE, VMD)	< 100 MB

Work Package 2. Study of Collective Properties (Research Objectives O4 and O5)			
Type of dataset	**Origin**	**Format**	**Quantity**
DLS measurements	Experimental	TAR.GZ (dat)	< 200 MB
SLS measurements	Experimental	TAR.GZ (dat)	< 200 MB
Optical microscopy images	Experimental	JPEG, PNG, TIFF, PDF	< 200 MB
Optical microscopy data	Experimental	TAR.GZ (dat)	< 500 MB
Codes for analyzing dynamic collective properties in experimental Data	Computational	TAR.GZ (FORTRAN, Python, dat)	< 50 MB
Codes for analyzing dynamic collective properties in computational Data	Computational	TAR.GZ (FORTRAN, Python, dat)	< 50 MB
Codes for analyzing Static collective properties in experimental data	Computational	TAR.GZ (FORTRAN, Python, dat)	< 50 MB
Codes for analyzing static collective properties in computational data	Computational	TAR.GZ (FORTRAN, Python, dat)	< 50 MB
Files containing graphical representations coming from experimental and computational data	Experimental, Computational	TAR.GZ (XMGRACE, VMD)	< 100 MB

Work Package 3. Training and Management			
Type of dataset	**Origin**	**Format**	**Quantity**
Research training on protocols for sample preparation (Bacterial colonies)	Report	PDF	< 20 MB
Research training on protocols for DLS measurements	Report	PDF	< 20 MB
Research training on protocols for SLS measurements	Report	PDF	< 20 MB
Research training on protocols for rheology measurements	Report	PDF	< 20 MB
Research training on protocols for optical microscopy measurements	Report	PDF	< 20 MB
Career Development Plan (with a description of the hands-on-training and management activities)	Report	PDF	< 20 MB

Work Package 4. Dissemination and Communication			
Type of dataset	**Origin**	**Format**	**Quantity**
DMP (including updates)	Report	PDF	< 10 MB
Final preprint versions of papers for expert audiences (Estimation: 3 preprints)	Paper preprint	PDF	< 20 MB
Presentations in international conferences for expert audiences (including talk presentations, videos, and posters)	Presentation	TAR.GZ (PDF, Keynote, mp4)	< 500 MB
Teaching material (undergraduates/graduates, including lessons on Biological Physics and short courses)	Report	TAR.GZ (PDF, Keynote, mp4	< 500 MB
Presentations in internal seminars	Presentation	PDF, Keynote, mp4)	< 300 MB
Presentations in events for non-expert audiences (including talk presentations, videos, and posters)	Presentation	TAR.GZ (PDF, Keynote, mp4)	< 500 MB

1.4. Data value

ARIADNE aims to reveal how active systems organize and cooperate at different time and length scales at high packing fractions. This general problem presents its culmination in the physical study of archetypal biological systems: bacterial colonies. The data supporting the project scientific results and their dissemination will reach a broad range of expert audiences. This includes: physicists, biophysicists, biologists, chemists, engineers, computer scientists, and professionals from biotechnological companies (*e.g.* nano-medicine and design of soft materials).

ARIADNE will also generate data for pedagogical purposes that will reach different non-experts audiences. This includes: children and professors of elementary education, students and professors of secondary education, undergraduate and graduate students (particularly in the broad fields of physics and biology), and general public.

ARIADNE will also create new standardization activities that will be supported by the project data management. This includes the creation of technical reports describing new experimental protocols (based on an interplay between biological and physical techniques) as well as a new data context management to maintain and extend the produced computational resources.

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

Each dataset generated by ARIADNE will be recorded by a dataset identifier in the general file **ARIADNE_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	
Dataset Identifier	The ID will result from the naming convention provided in the next table
Tile of the Dataset	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 (*e.g.* WP2, see work package table in Annex A)
Dataset Description	A brief description of the dataset (*e.g.* DLS measurement, specifying the experiment conditions)
Dataset Benefit	What are the benefits of the dataset (*e.g.* the data will serve to reproduce some particular results and/or perform further analysis)
Dataset Dissemination	Where will the dataset be disseminated (*e.g.* peer reviewed journal)
Type Format	See table with formats in section 1.3 (*e.g.* **dat**, **PDF**, **TIFF**, ...)
Expected Size	Dataset size (see size in the work package tables in section 1.3)
Source	How was the dataset generated (*e.g.* experimental data)
Repository	DIGIBUG (for final preprints we will also have an arXiv version, section 3)
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 on the repository
Keywords	Keywords associated with the dataset (*e.g.* light scattering, *E coli* bacteria)
Version Number	Version number to keep track of changes (*e.g.* V1.0)

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	ARIADNE (always)
Two digits chronological number (corresponding to the order of appearance in ARIADNE_DATASET_LIST.pdf)	12
Title of the dataset	MD-code-polar-systems
Version of the dataset (DIGIBUG will allow to keep several versions)	V1.0
Work Package associated to the dataset	WP1
Format of the dataset	TAR.GZ
Example dataset file identifier: **ARIADNE_12_MD-code-polar-systems_V1.0_WP1.TAR.GZ** Example metadata file identifier: **ARIADNE_12_MD-code-polar-systems_V1.0_WP1_META.dat**	

2.2. Making data openly accessible

The data hosted in the institutional repository DIGIBUG will be accessible to the complete research community. In this respect, the data generated by ARIADNE 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 ARIADNE 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 (Ubuntu).

Format	Examples of software (operating system) to open/use them
TAR.GZ	**TAR.GZ** package (Ubuntu), **tar** utility (MacOS), **7-Zip** (Windows)
dat	**GNU emacs** (Ubuntu, MacOS), **Notepad** (Windows)
PDF	**Adobe Reader** (Ubuntu, Windows), **Preview** (MacOS)
JPEG, PNG, TIFF	**gThumb** (Ubuntu), **Preview** (MacOS), **Adobe Illustrator** (Windows)
FORTRAN	**gfortran** (Ubuntu, MacOS, Windows)
Python	**Python sources releases** (Ubuntu, MacOS, Windows) in www.python.org
XMGRACE	**XMGRACE** (Ubuntu), **XQuartz** (MacOS), **QtGrace** (Windows)
VMD	**Visual Molecular Dynamics (VMD)** (Ubuntu, MacOS, Windows) in www.ks.uiuc.edu/Research/vmd/ (Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign)
Keynote	**Keynote** (MacOS) (another pdf version will uploaded for each Keynote presentation)
mp4	**VLC** (Ubuntu, Windows), **MPlayerX** (MacOS)

2.3. Making data interoperable

ARIADNE 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. Allocation of Resources

All the datasets generated by ARIADNE 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 the arXiv (a private not-for-profit educational repository owned and operated by Cornell University). For the arXiv version 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. Sándalo Roldán-Vargas, department of applied physics at the UGR (Spain), will be responsible for the data management within the ARIADNE project, in particular for the creation of the DMP and its subsequent updates, and for recording and updating the datasets generated by ARIADNE.

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 ARIADNE 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.

5. Ethical Aspects

As stated in the Ethics Issues of the MSCA-IF-EF-ST grant agreement (840195 - 2019), there is no requirement for ethical review since ARIADNE 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

[1] Guidelines on FAIR Data Management in Horizon 2020, version 3.0, 26 July 2016

[2] Data Management Plan Template, UGR Library, version 1.0, 1 July 2017

[3] J. Horst and C. Lynch, AMECRYS Data Management Plan, version 3.0, 29 March 2017

7. List of Acronyms

Acronym	Meaning
DLS	Dynamic Light Scattering
DMP	Data Management Plan
GT	Glass Transition
IP	Intellectual Property
MD	Molecular Dynamics
MSCA-IF-EF-ST	Marie Skłodowska-Curie Actions - Individual Fellowship - European Fellowship - Standard
SLS	Static Light Scattering
TEM	Transmission Electron Microscopy
UGR	University of Granada

8. Annex A. Work Packages

The table below summarizes the main research objectives, tasks, training, management, and documents associated to each work package of the ARIADNE project (see tables in section 1.3).

Work Package 1. Objectives O1, O2, O3	**Title: Location of the Glass Transition (GT)**	**Duration: month 1 to 14**
Research Objectives. • **Objective O1**: Characterization of the GT in isotropic models (both experimental and computational). • **Objective O2**: Characterization of the GT in non-isotropic models (computational). • **Objective O3**: Characterization of the GT in polar models (both experimental and computational).		
Main Tasks. Computer programming and testing. Computer data collection and analysis. Experiment data collection and analysis.		
Work Package 2. Objectives O4 and O5	**Title: Collective Properties**	**Duration: month 14 to 23**
Research Objectives. • **Objective O4**: Characterization of collective dynamic properties in experiments and simulations. • **Objective O5:** Characterization of collective static properties in experiments and simulations.		
Main Tasks. Computer programming and testing. Computer data collection and analysis. Experiment data collection and analysis.		
Work Package 3. Management	**Title: Training and Management**	**Duration: month 1 to 24**
Documents to be produced. Four Career Development Plan reports: one for the initial document and three updates. One report on protocols for sample preparation (bacterial colonies). One report on protocols for DLS measurements. One report on protocols for SLS and DLS measurements. One report on protocols for rheology measurements. One report on protocols for optical microscopy measurements. One Final Report summarizing the results of the project.		
Main Tasks. Hands-on training activities. Report writing. Research training on isotropic sample preparation for rheological measurements and light scattering measurements. Research training on polar sample preparation for rheological and light scattering measurements. Research training on sample preparation for optical microscopy measurements.		
Work Package 4. Dissemination	**Title: Dissemination and Communication**	**Duration: month 6 to 24**
Documents to be produced. One DMP (with 2 repository updates). Three final preprint papers. Presentations for: five international conferences, three internal seminars, seven events for non-experts, two specific lab courses, and material for biological physics lessons for undergraduate students.		
Main Tasks. Report writing. Repository updates. Paper writing. Preparation of material for conferences, labs, lessons, seminars, and events for non-experts.		