

DEEP-IBERIAN project

Data Management Plan

Version 1.0

Authors:

José Manuel Fuertes García (jmf@ujaen.es@ujaen.es),
Antonio Jesús Rueda Ruiz (ajrueda@ujaen.es)

Grant No.: <PENDING>

Call: Proyectos de Generación de Conocimiento 2024 -
Ministerio de Ciencia, Innovación y Universidades de España

Project Start Date: <PENDING>

Change control

Version	Date	Author	Organization	Description/Comments
1.0	2025/01/24	José Manuel Fuertes García	University of Jaén	First draft

Contents

1. Identification and Description of Data
2. Data Storage and Security
3. Data Organization and Management
4. Data Publication and Accessibility
5. Responsibilities and Resources
6. Data lifecycle

Summary

This is the preliminary version of the Data Management Plan (DMP) for the project DEEP-IBERIAN. It contains a description of the data that will be produced during the execution of the project, as well as how this data will be used, stored and published.

Since this project is publicly funded, and according to the Open Science principles¹, all the data and scientific publications that are not protected by copyright laws will be publicly accessible, applying the guidelines described in this document.

This document is not final, as it might be modified during the research lifecycle.

¹ <https://www.unesco.org/en/open-science/about>

1. Identification and Description of Data

1.1. Types of Data:

The project will generate various types of data, including:

- 3D Models of Iberian Pottery: Polygonal meshes, point clouds, and voxelized models of complete artifacts and fragments.
- 2D Images: Photographs of pottery decorations.
- Data Associated with 3D Models and Images: Metadata, contextual, and typological information.
- Neural Network Outputs: Parameters from trained networks and results from classification and reconstruction tasks.
- Software Usage Data: Information on tool usage by experts, validation results, and output quality.
- Data Formats: Standard formats will be used to ensure interoperability and accessibility:
 - 3D Models: Formats such as OBJ, STL, and PLY.
 - Images: Formats such as JPG, PNG, and TIFF.
 - Metadata: Formats such as XML and JSON.
 - Neural Network Data and Results: Formats such as CSV, TXT, and HDF5.

1.2. Data Volume:

The expected data volume will be significant, given the number of 3D models, images, and training datasets for neural networks. Appropriate tools for large-scale data storage and management will be required.

2. Data Storage and Security

2.1. Storage Infrastructure:

The project will utilize the storage infrastructure of the Computational Archaeology Laboratory at the University Institute for Iberian Archaeological Research, with additional cloud-based solutions if local capacity is insufficient.

2.2. Backups:

Periodic backups will be conducted following a defined protocol to prevent data loss. Copies will be stored in multiple locations, including at least one offsite backup.

2.3. Access Control:

Access will be restricted to authorized project personnel. Roles and permissions will be assigned to ensure secure data handling.

2.4. Security Measures:

Data will be protected from unauthorized access, loss, or corruption through firewalls, antivirus software, and other security protocols.

3. Data Organization and Management

3.1. Metadata:

Metadata will be generated for each dataset using standards such as Dublin Core or ISO 19115, including descriptions, creation dates, authors, typologies, sites, formats, and licensing information.

3.2. Folder Structure:

A logically structured and consistent folder hierarchy will be implemented to facilitate data retrieval and accessibility.

3.3. Document Management System:

A document management system, including the CATA database, will be used to streamline data access and management. This system will support data migration to large cultural heritage digital repositories like EUROPEANA.

3.4. Version Control:

Version control systems will be employed to track changes in 3D models and software code, enabling rollbacks to previous versions if needed.

4. Data Publication and Accessibility

4.1. Open Access:

All generated data will be published in an open-access repository, adhering to FAIR principles (Findable, Accessible, Interoperable, Reusable), to ensure usability by the scientific community.

4.2. Licensing:

Open-access licenses such as Creative Commons (e.g., CC BY) will define usage terms.

4.3. Repository:

The data will be hosted on the repository of the University Institute for Iberian Archaeological Research or the institutional repository of the University of Jaén.

4.4. Dissemination:

Access to the data will be promoted via the Computational Archaeology Laboratory website, conferences, and publications.

5. Responsibilities and Resources

The data collected and/or created during the execution of the project will be under custody of the research team while they are used for experimental purposes. Every person in the team could be assigned the custody of some data, as long as it is related to the project-related tasks they have been assigned.

The University of Jaén is responsible for RUJA, in terms of keeping the site working and preserving the uploaded data. This repository assigns a permanent URI to each uploaded work and complies with the Open Archives Initiative (OAI) standards², in order to make the data comply with the FAIR principles.

5.1. Data Manager:

A dedicated data manager will coordinate plan implementation and ensure compliance.

² <https://www.openarchives.org>

5.2. Technical Staff:

Experienced technical personnel will support infrastructure setup and data migration to publication platforms.

5.3. Training:

Project staff will receive training on data management principles, the use of the document management system, and best practices to comply with this plan.

6. Data lifecycle

This section explains how the data described above will be dealt with, not only during the project execution, but also once it has finished.

In application of the FAIR principles³, a key element for the management of the data will be the infrastructure provided by the University of Jaén; to be precise, the RUJA⁴ repository provided by this institution to publish research results with Open Access (OA). All the information published in this resource, unless stated otherwise, is available under a Creative Commons (CC by-nc-nd) license⁵.

6.1. During the project execution

The information elements collected in the repository in their various formats will be structured according to the document management system defined during the project's development. Proper metadata will also be added, in application of the FAIR principles.

These datasets will be kept private during the development of our experiments, but once they are finished and the research results are submitted for publication, they will be made available through RUJA, so that reviewers can verify our achievements.

Regarding research papers, once they have been accepted for publication, they will be made publicly available, either through OA journals or the institutional repository RUJA. In any case, copyright infringements will be avoided by checking the OA policies of the journals using resources like Sherpa Romeo⁶. Together with the paper, corresponding metadata will be added to the corresponding repository item, in order to make it easily findable and accessible, in application of the FAIR principles.

³ <https://www.go-fair.org/fair-principles>

⁴ <https://ruja.ujaen.es>

⁵ <https://creativecommons.org/licenses/by-nc-nd/4.0>

⁶ <https://v2.sherpa.ac.uk/romeo>

6.2. After the end of the project

Once the project execution has reached its deadline, the resources uploaded to RUJA will remain publicly available indefinitely, as this repository is intended to be permanent in time.

Those datasets and papers that at the end of the project have not yet been published will be treated in the same way as described above, i.e. the efforts for publication will continue until success, and once they have been published, both the datasets and the published papers will be uploaded to RUJA and made publicly available.

This research data management plan aims to ensure that data generated by the project are accessible, secure, and reusable by the scientific community. The plan prioritizes generating future-reusable data and promoting transparency in research. Data will be published in open formats to ensure accessibility for the scientific community and the general public.