



**European Research Council**  
Executive Agency

Established by the European Commission



**European Research Council (ERC)**

**ERC Data Management Plan**

**Template**



# ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

European Research Council

Established by the European Commission

Project Acronym	Project Number
STONE-MASTERS	101040152

***Template for the ERC Open Research Data Management Plan (DMP). The following sections should describe how you plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR). Each of the following five issues should be addressed with a level of detail appropriate to the project.***

**SUMMARY** (*dataset<sup>1</sup> reference and name; origin and expected size of the data generated/collected; data types and formats*)

Purpose: The project aims at building a Digital Atlas of Workshops in making the inscriptions between the third and fifth century CE, and to explore the impact of artisans on the vertical cultural transfer in the Later Roman Empire. In order to complete these goals, the project will draw on published data regarding the collections of Greek and Latin inscriptions from the Roman period, and mentions of artisans in the literary sources which are also published. There is no restraint in accessing any of these publications.

The dataset generated by the project will include raw and compiled data regarding different aspects of the activity of workshops making the inscriptions on stone and mosaic between third and fifth century CE. The data collected will include prosopographical information on particular individuals, ancient stonemasons and mosaicists, their signatures and stonemason marks, styles of the support and lettering of the inscriptions, the geographical outreach of specific workshops, and diachronic changes in the practices pursued by these workshops.

Provenance of data: print publications, digital databases of inscriptions (mentioned below), digital corpus of Greek texts (Thesaurus Linguae Graecae), digital corpus of Latin texts (Library of Latin Texts).

Re-use of data: The project will re-use existing data on Greek and Latin inscriptions from the Roman period, published in open access online databases, such as Packard Humanities Institute database (PHI), Epigraphic Database Heidelberg (EDH), Epigraphic Database Roma (EDR), Hispania Epigraphica (HE), Epigraphic Database Bari (EDB), Epigraphic Database Clauss-Slaby (EDCS), The Inscriptions of Sicily (I.Sicily), Inscriptiones Christianae Graecae (ICG – Topoi Cluster project), Inscriptions of Israel/Palestine (IIP – Brown University), and Epigraphische Datenbank zum antiken Kleinasien (EDAK, Universität Hamburg). Existing data of the same type from print-only editorial series of inscriptions will also be used. In addition, a corpus of literary sources from the Roman period will be searched for data on ancient stonemasons, mosaicists, and their workshops. The data will be analyzed in order to distinguish specific markers of style of the visual appearance of inscriptions, and of the textual formulae, allowing for the recognition of workshops making the inscriptions in the Mediterranean between the third and fifth century C.E.

Formats and types of data: The project will use data recorded as texts in different digital formats such as .txt, .doc, .docx, .pdf, .xml, .xlsx, .csv, .sql. The project will also generate textual data in these same formats. Images recorded as the following files will also be used: in open formats such as .svg, .dwg, .dxf, .shp, .jpg, .png, .tiff, and in proprietary formats (where needed) such as .cdr. The project will generate visual data in these same file formats. The project necessitates the use of proprietary formats such as .txt, .doc, .docx, .xlsx, and .sql since the data produced and exchanged by the epigraphers are normally recorded in these formats. Adoption of only open formats would result in their reduced re-usability and their gradual abandonment by the community's researchers. Some information would also be misrepresented and data quality reduced during the process of saving to open formats. However, whenever possible open and lossless formats will be used, especially for the backup service.

Size of data: A typical text .xml or .txt file recording one inscription is usually up to 100kB. A typical photograph of an inscription greatly varies, from 1 MB to 300 MB. This depends on the quality of the image, the date of its creation, and format chosen by the photographer.

<sup>1</sup> Several datasets may be included into a single DMP.

# ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

The project will analyze approximately 500.000 inscriptions of which a smaller sample will be chosen for a closer study.

## 1. MAKING DATA FINDABLE (*dataset description: metadata, persistent and unique identifiers e.g., DOI*)

Metadata: Metadata will be provided within text and image files, describing their origin and contents. The Digital Atlas of Workshops will also provide metadata on the provenance of the objects listed, and specific record metadata, listing, for example, the history of the processing of data, the dates of the creation of records, authors of records, etc. The standard used will be OGC for the spatial metadata and IIIF for photographs, and Dublin Core for other types of metadata.

Metadata keywords: Keywords will be provided in the metadata in the Digital Atlas of Workshops in accordance with the EAGLE/Europeana controlled vocabularies for epigraphy <https://www.eagle-network.eu/resources/>

Data identifier: The data generated on the project will be identified by Uniform Resource Locators, and by internal database record numbers. Publications produced on the project will be usually identified by the Digital Object Identifier System. Authors will be identified with the ORCID identifiers.

## 2. MAKING DATA OPENLY ACCESSIBLE (*which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?)*)

### Repository:

During the project, the data will be deposited in a GitLab repository established on one of the servers of the University of Warsaw only for the use by the team. After the project, the data will be deposited in two open repositories, GitHub (for open source code) and Zenodo (for open data), in accordance with the 2022 Report from the Scoping Survey of the FAIR Epigraphy Project recommendations. [https://fair-epigraphy.github.io/scoping\\_survey\\_report/scripts/01\\_FAIR\\_epi\\_report.html](https://fair-epigraphy.github.io/scoping_survey_report/scripts/01_FAIR_epi_report.html) Some data, especially publications, will be deposited in the CeON Repository (Centre of Open Science, Poland).

These repositories are usually used by projects in the humanities departments at the University of Warsaw and the terms of their use are known to the PI and the university staff.

The mentioned repositories will ensure Digital Object Identifiers.

### Data:

During the project, software generated on the project will be stored at a GitLab repository started by the University of Warsaw. After the project, the GitHub and Zenodo repositories will be used to publish the data. As for new materials: the project may also produce hand drawings, and squeezes of some inscriptions. These will be digitized and uploaded to the Zenodo repository, together with a catalogue of the original copies and information on their signatures and the storage place. This will be most probably the Library of Papyrology, Roman Law, and the Ancient Laws at the University of Warsaw.

Documentation about the new software developed on the project will be made and deposited in the university's GitLab repository during the project. After the project, the documentation and the source code will be open and published alongside the documentation in two open repositories, Zenodo and GitHub.

The data will be made openly available through the mentioned repositories and the project's Digital Atlas of Workshops, and in open access publications.

The data will be accessible through a free and standardized access protocol, most probably REST, though GraphQL will also be considered.

There will be restrictions of use during the duration of the project but data will be provided upon request to any third parties. These

# ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

restrictions are necessary due to the ongoing process of data curation in order to provide its high quality. After the project, there will be no restrictions. It will be possible to access the data via the mentioned repositories, the Digital Atlas of Workshops, and project publications.

The identity of the persons accessing the data will not be collected.

There will be no need for any data access committee since the project does not operate on sensitive data for research purposes.

Metadata:

The metadata will be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement, and will contain the information to enable the users the access to data.

The data will be available and findable in the mentioned repositories for a period of at least five years, very probably much longer. The longer access depends on the future agreements at the Warsaw University upon maintenance and supporting data from archival projects. The metadata will remain accessible even after the erasure of data.

Metadata harvesting and indexing: Metadata will be exportable via an API interface.

## **3. MAKING DATA INTEROPERABLE** (*which standard or field-specific data and metadata vocabularies and methods will be used*)

The epigraphical data will follow the vocabularies of the Leiden convention, and of the preliminary Epigraphic Ontology suggested by the EAGLE/Europeana controlled vocabularies for epigraphy <https://www.eagle-network.eu/resources/> This will allow interoperability, data exchange, and re-use according to the best community-endorsed practices, described in the digital epigraphy manifesto "Digital Epigraphy in 2022. A Report from the Scoping Survey of the FAIR Epigraphy Project" by Petra Heřmánková, Marietta Horster, and Jonathan Prag. See: [https://fair-epigraphy.github.io/scoping\\_survey\\_report/scripts/01\\_FAIR\\_epi\\_report.html](https://fair-epigraphy.github.io/scoping_survey_report/scripts/01_FAIR_epi_report.html)

The project will also use university procedures for data management recommended by the University Research Office (Biuro Obsługi Badań UW). The document was published on 25 November 2019.

<http://bob.uw.edu.pl/plan-zarzadzania-danymi-ogolne-informacje-o-pojeciach-oraz-przykladowy-plan/>

It includes a description of general types and ontology of research data, Open Research Data policy, metadata and project documentation, recommended treatment of sensitive data, copyright, and licensing issues, and recommended practices for the archiving of data.

Any project specific domain and upper-level ontologies or vocabularies which may be needed for the construction of the Digital Atlas of Workshops will be published, reusable, and compatible with existing vocabularies.

The project's data will include qualified references to other data where possible and applicable, i.e. in a form of identifiers nested as foreign keys, which is consistent with the policy of Linked Open Data (LOD).

## **4. INCREASE DATA RE-USE** (*what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures*)

Data utility: The data will be made freely available in the public domain to permit the widest re-use possible, and will be licensed using standard reuse licences, in line with the obligations set out in the Grant Agreement. The data generated on the project will be useful to epigraphers attempting to establish the provenance or plausible date of Greek and Latin inscriptions from the Later Roman Empire, based on their style, lettering, and formulae. The published outcomes of the project will also be useful to researchers of vertical cultural transfer between the elites and the middle and lower social groups, of commemorative practices, and of cultural memory.

Validating data analysis and data re-use: The types of data gathered and sources of data will be carefully described on a dedicated subwebsite of the Digital Atlas of Workshops. Another subwebsite will describe the basic methods used to clean and analyze the data, and visualize them with the use of layers in the user interface. The book published by the PI will contain a methodological chapter on data analysis. The project will also produce an edited volume tackling the methods of the identification of workshops with contributions by speakers at a conference organized by the project. The documentation of the software developed for the project will be published with an open source code in a digital repository (mentioned above).

The data produced on the project will be kept in the mentioned repositories during at least five years after the completion of the project

# ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

(very probably much longer).

No embargo periods are envisaged for project publications. Papers and books publishing the results of the project will be immediately granted Open Access status (Gold Open Access). The texts (published papers and books) will also be deposited in their AAM or VoR forms in the open repository Zenodo, using a CC BY 4.0 or an equivalent license. The authors' copyrights will be respected. As for the images, the relevant licenses will apply, in accordance with the policy of the institutions owning copyrights.

The data produced in the project will be easily usable by third parties, in particular after the end of the project. It will be facilitated through a system of exporting data either as .csv, .xlsx, or .xml files. The user interface will also allow for the use of unexported, visualized and structured data in the Digital Atlas of Workshops.

The provenance of data will be documented in metadata in accordance with the standard citation systems (e.g., Chicago Manual of Style or others). Where necessary, the provenance and former edition of inscriptions will be documented with proper genetic lemma. The provenance and authors of images will be documented appropriately in the digital files.

The quality of data will be assured through team work on data collecting and data cleaning which will involve the PI, two postdocs, and two predocs. First, the PI will carefully design the guidelines for data collecting, establishing the main strands of surveys which will bring representative, non-anecdotal sets of data, and assure their completeness. Then the predocs and postdocs will use a digital instrument, the backend interface of the Digital Atlas of Workshops, to collect the data according to the said guidelines, will use the visualization component, and will look for distorting data requiring further cleaning. The PI will use the workshops organized on the project, especially the methodological conference, to consult the methods of data cleaning with the world leading experts.

## **5. ALLOCATION OF RESOURCES and DATA SECURITY** (*estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation*)

Direct and indirect costs related to storage, archiving, re-use, and security of data are covered by the University of Warsaw independently of the project's direct costs. The necessary infrastructure is provided by the University of Warsaw Centre of Digital Competences and the Faculty of History's digital lab. Some procedures regarding the procedures of data sharing and starting a GitLab repository dedicated for the project will be conducted by the software engineer employed on the project. The envisaged cost of software engineer's employment amounts to EUR 122,732.00. These employment costs are covered by the project's direct costs. The cost of keeping the data will be covered by the institutions running the mentioned depositories in accordance with their usual policy of projects' support (University of Warsaw's GitLab repository, GitHub repository, CeON repository, Zenodo will also be considered in accordance with the 2022 Report from the Scoping Survey of the FAIR Epigraphy Project recommendations).

The persons responsible for data management will be the PI (Pawel Nowakowski) and software engineer (Maciej Krawczyk).

The data collected on this project will be stored in a custom database underlying the Digital Atlas of Workshops, protected by password. The database will be run from a secure university server, with a backup tool allowing for the recovery of data from the last five days. All the correspondence regarding the project will be done with the use of university e-mail which is based on a secure commercial version of Gmail for education institutions. The Google drive storage provided within this same service agreement will be used by the PI and project members for temporary storage of data currently processed. All the names of project participants, including the non-members of the team (invited speakers, participants of two intensive workshops and conferences), will be kept on university computers with access secured by passwords. The data generated on the project will be kept after its completion in a dedicated GitLab repository started by the Faculty of History, and in the trusted, internationally recognized CeON repository which ensures necessary safety protocols. Zenodo will also be considered in accordance with the 2022 Report from the Scoping Survey of the FAIR Epigraphy Project recommendations.

Industry standards for security from DDoS and other kinds of attacks will be implemented.

There are no ethics or legal issues that can have an impact on data sharing, as specified in the Description of the Action.

Personal data will be collected only for the recruitment procedure, and the organisation of project events. The recruitment procedure applications include an informed consent form for limited sharing of data with university admin offices. Admin office employees are trained in secure handling of any possibly sensitive data. All the personal data gathered, for example, for the recruitment procedure, the reimbursement of travel costs of invited participants, and others, will be securely stored in a licensed cloud service. Access to computers processing these data will be protected by passwords. The regulations of the General Data Protection Regulation will be used.

## ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

**DISCLAIMER.** Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project.