Funded by the European Union

Ant — Com From Antiquity to Community

DATA MANAGEMENT PLAN

PROJECT	
Project number:	101073543
Project acronym:	AntCom
Project name:	From Antiquity to Community: Rethinking Classical Heritage through Citizen Humanities

DATA MANAGEMENT PLAN	
Date:	[1/3/2025]
Version:	[2]

0. Executive Summary

This document represents the second version of the Data Management Plan (DMP) for the project *AntCom - From Antiquity to Community. rethinking classical heritage through citizen humanities*.

AntCom aims to train a new generation of highly skilled cultural heritage researchers to harness the potential of the digital transition and address the professional and societal challenges it entails. The focus is on "classical" heritage from Europe's Graeco-Roman past, with training transferable to other cultural heritage domains. This choice is motivated by the transnational and identity-shaping nature of classical heritage across Europe, raising urgent issues relevant to education and identity.

AntCom's research is divided into three work packages addressing three different typologies of heritage: manuscript, linguistic and narrative.

The project has three aims, that is to:

- digitize and analyze obliterated manuscript heritage using advanced imaging techniques;
- document the transformation of classical heritage in oral, ritual, and monumental traditions;
- digitally archive collected data and engage communities in rethinking their past.

AntCom is characterized by the integration of humanities, STEM and citizen engagement. This approach ensures that the research is grounded in both latest scientific advancements **and** the lived experiences and traditions of the communities.

By combining the analytical power of STEM with the contextual insights of the humanities, the project not only enhances the understanding of cultural heritage but also empowers communities to actively participate in preserving their heritage. This interdisciplinary approach translates into a innovative training program, aiming to create new professional figures of heritage practitioners.

Given the strong project's emphasis on digitization and broadening access, **this plan details how all generated and collected data will remain findable, accessible, interoperable, and reusable, while ensuring compliance with data protection regulations.** It also describes which typologies of data the project will generate, whether and how they will be made accessible for verification and re-use, and how they will be curated and preserved.

This DMP has been compiled based on the consolidation of the individual DMPs drafted by the DCs and will be updated periodically. New versions of the DMP will be created should significant changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.

Table of Contents

1. Data Summary

1.1. Introduction

AntCom is a cross-European research and training initiative that rethinks the concept of heritage through an inclusive, transdisciplinary, and citizen-engaged approach. Coordinated by a consortium of four universities—University of Southern Denmark (SDU), University of Salento (UniSalento), University of Verona (UniVr), and University of Santiago de Compostela (USC)—the project explores the layered nature of ancient European heritage, focusing on material, linguistic, and narrative dimensions that have been marginalized or erased over time.

By combining a diverse expertise, drawn from fields such as history, philology, linguistics, oral history, library science, engineering, and digital humanities, AntCom aims to mobilize communitied in order not only to share but also to co-create knowledge, and thus challenge dominant historiographical narratives about the uses of the Graeco-Roman past. Doctoral candidates will be trained in innovative methods, including multispectral imaging, oral history, and citizen humanities, and will engage in placements, workshops, and schools across the consortium. Through this integrated and collaborative framework, AntCom contributes to the development of inclusive memory practices and transformative heritage policies in Europe.

The project is characterized by diverse approaches and hence a diverse pool of data collected/managed/curated. **Therefore it was important in a first phase that priority was given to the DCs working out their individual DMPs, given the importance of data collection from an array of diverse sources and living informants, as well as the multiple actors involved (beneficiaries, associated partners, communities).** At the same time, however, a centralized DMP is crucial to streamline data collection as well as to ensure replicability of the research results.

The following plan consolidates the individual DMPs worked out by each DC, providing an overview of the different typologies of data mobilized by the project and policies for their archiving and conservation.

1.2. Which data

AntCom is expected to generate/gather:

a) *interviews*

b) *fragments' and palimpsests' digitization*

c) *digitization of archival and published material not covered by copyright*

d) *digitization of literary texts of oral origin*

e) *images*

f) *audio files*

g) *video files*

h) *multimedia documents*

i) *bibliographies*

j) *computational models (image optimization, AI)*

k) *softwares*

l) *codes*

1.3. Purpose of the data collection and relation to the objectives of the project

The data collection in the AntCom project aligns with its objectives by directly contributing to its scientific, societal, and educational goals. Here is how this alignment is achieved:

- Scientific Exploration and Preservation: The project involves digitizing and analyzing manuscripts, including palimpsests, to uncover hidden texts and historical insights. This data collection contributes to advancing knowledge in codicology, philology, and historical studies, particularly regarding ancient texts and their cultural significance.

- Community Engagement: A key objective is to actively involve citizens in the scientific process of data collection. This includes initiatives like interviewing citizens on classical narratives related e.g. to the tower of Hercules in Galicia, or integrating high school students into reading and analyzing Virgil's literary works directly from one of their oldest manuscripts, fostering a participatory approach to the study of Graeco-Roman antiquity.

- Training and Development: The collected data supports the training of doctoral fellows by providing them with hands-on experience in interdisciplinary methodologies, such as multispectral imaging and manuscript analysis. This aligns with the project's goal of structuring doctoral training at a European level and enhancing researchers' career prospects.

- Dissemination and Impact: The collected data is used to create resources that are on sharing platforms (e.g. Fragmentarium https://fragmentarium.ms/, Europeana https://www.europeana.eu/en/collections, Terra e Memoria https://www.terraememoria.usc.gal/). Models as well as software will be made available through GitHub and Zenodo. The combination of these platforms will ensure open access to both classical heritage and tools to manage it according to best practices. This aligns with the project's commitment to open science practices and maximizes its societal impact by making research outputs accessible to both academic communities and the public.

In particular it is expected that:

- Interviews are tied to specific work packages (e.g., WP3, WP4) that focus more on social, cultural, and political dimensions. They support research activities such as case study development, and narrative construction. Interviews, audio, and video files (a, f, g, h) will contribute to fulfilling the project's objective on community engagement, through:
    - *direct interaction with communities*: interviews are used to engage directly with local communities, particularly in regions like Galicia and Gríko-speaking areas where the project focuses.
    - *gathering local and first-hand knowledge*: they emphasize the project's aim to encompass diverse voices, including underrepresented or marginalized groups. Interviews help bring forward local and indigenous knowledge, ensuring that scientific discussions are grounded in modern social realities.
    - *Creating a dialogue*: interviews are designed as collaborative spaces for dialogue. This two-way communication helps communities feel heard and included, reinforcing trust and legitimacy in the project's aims.
- Digitalization of fragments, palimpsests and literary texts (b, c, d, e), which will be collected mainly in WP2 and WP3, directly fulfills the project's objective of scientific exploration and preservation through:
    - *Unearthing Hidden Texts*: The digitalization, particularly with multispectral imaging, allows for the recovery of hidden texts in palimpsests. For example, manuscripts like Mss. LXII (60) and XL (38) contain palimpsest portions of Livy and Virgil with hidden Greek marginalia. Digitalization facilitates detailed analysis of the codicological and book-historical aspects of the fragments. This enables a deeper understanding of the texts and their historical context, supporting the scientific exploration of these materials.
    - *Accessibility and Preservation*: By digitizing both manuscripts and fragments, the project ensures that these fragile historical documents are preserved for future research, thus contributing to the dissemination and impact of the new retrieved material.
    - *Citizen Science Integration*: The project actively engages citizens in the scientific process of data collection, specifically involving citizenship in reading Virgil's text integrating high school students in the project.
- Images and multimedia documents (f,g,h) are used throughout the project to ensure e.g. visibility of the projects' results on the social media channels, as well as to provide systematic reports of the project's developments.

1.4. Types and formats of data generated/collected

- Microsoft Office file formats will be the reference for documents, spreadsheets, presentations and transcripts (a, d, i).
- Fragments' and palimpsests' digitisation (b) will generate image files of manuscript fragments (digital facsimiles; hyper- and multispectral imaging) as well as descriptive metadata relating to the fragments and their host volumes. The latter will be compiled as text and or XML files. Palimpsests digitisation (b) will generate image data (including HSI & MSI) in open format (.tiff, .jpeg, .jpeg2000), as well as RAW data in proprietary formats (Phase One IIQ, Nikon NEF). Furthermore, the project will record cataloguing data in two open and reusable formats, .json and .cvs. In .json files context can be added for producing linked data. Moreover, the project is also expected to generate unstructured data (RGB images in .jpg and .png format, as well as hyperspectral data cubes in .bin format)
- Digitisation of of archival/published material and literary texts (c,d) will produce data in .pdf, .txt, and other standard text file format.
- Images will be stored in .cr2/.nef, .tiff, .jpeg high resolution formats.
- Audio files will be stored in .mp3, .wav, m4a formats (f)
- Videos in .mp4, .mkv format; .mov; .mxf (g)
- Multimedia (e) will be stored in.wav (audio) and .mkv, .mp4 (video).
- Metadata will be stored as csv, .json, .pdf.

1.5. Re-use of existing data

Existing data might be used to provide, for example, benchmarks for evaluations of advantages/disadvantages of chosen solutions. Such data will be gathered from sources such as scholarly papers, established industry publications (e.g. existing hyperspectral or rgb data is expected to be used for training machine learning models or for augmenting the dataset that we have), other studies of recognized value and/or clear reliability. The existing hyperspectral or rgb data is expected to be used for training machine learning models or for augmenting the dataset that we have. Wherever pre-existing data will be used, clear reference will be made to it by the user, including either a link to it (if available online) and/or publishing information (if no online reference existed).

Other data reused and digitized will come from printed repositories such as e.g. collections of texts out of copyright (for literature in Griko); from on-line repositories (e.g. collections of sources in Griko from amateur scholars' websites); from National archives. In latter case, if the relevant documents are not digitized, the project will primpt/support digitization through the hosting institutions.

In some cases (e.g. DC 8), the project will also re-use and analyse data produced by AntCom's associated partners such as the "Consorcio de Turismo da Coruña" that manages the UNESCO site of the Tower of Hercules. This data include photographic and architectural documentation, audiovisual material, and publications aimed at the public such as brochures. A communication channel has already started with the partners to identify authorisations and methods of use in case of existing copyrights. Common bibliography on existing data is shared on AntCom open group library on Zotero.

1.6. Origin of the data

New data will be originated by the project activities, for example:

- Interviews will be conducted to people within and outside the project consortium; see *AntCom Oral History Participation and Recording Agreement* (D1.3) for details about consent and treatment.
- Fragments digitization, e.g bibliographic, book historical and codicological information relating to the fragments, bibliographic, book historical and codicological information relating to the host volumes, and digital images (facsimile reproductions and hyperspectral imaging) and corresponding metadata.
- Palimpsests digitization, e.g. the data is collected from original sources (medieval manuscripts of the Biblioteca Capitolare in Verona) via digitization protocols, using state of the art cameras and procedures. Data will be produced based on inspection and existing bibliography regarding the cataloguing of the object.
- Digitization of literary texts, previously published but existing only in hard copies, not easily accessible.
- Bibliographic repositories, with relevant digitization of bibliographic material where copyright allows it.
- Images; for examples posted photos on social media to document AntCom events.
- Audio files include for examples audio recordings of gríko-songs and dialogues between gríko-speakers.
- Video files will include for instance audiovisual recordings of popular narratives and short documentaries, which will be key to preserve local communities' oral history regarding the legendary traditions and their memory practices related to the Tower of Hercules.
- Multimedia documents (these include for instance internal or confidential reports, activities plan, spread sheets, presentations; website contents, social media posts).
- Optimization/AI model for image post-processing.
- Software for the optimiozation of processes of image acquisitions.
- Codes for image post-processing.

1.7. Size of data

The exact size of data is not yet known, however anticipates to generate approximately:

- 2 to 4 TB of images TIFF and approximately 100 MB of datasets (in CSV and JSON format). Photos are stored in JPEG compressed format unless high resolution is required e.g. for publishing purposes; moreover, images will be

included in documents (e.g. reports) in such a manner not to increase significantly the dimension of the files themselves

- 20 GB of models
- 100 GB – 1 TB machine learning modules
- 10 MB fellows personal documents (i.e. self-evaluation, CDP, DMP)
- 5 GB of textual material
- 50 GB of photos
- 20 GB of audio material
- 120 GB (compressed of video material)
- 50 GB of multimedia materials
- 100 MB for confidential and ufficial reports
- 4 MB Project internal documents, such as training and communciation handbooks and plans.
- 20 MB of metadata

Wherever possible, the best compromise between storage efficiency and accessibility of information will be sought.

1.8. Who will make use of the data

The data generated within the project will be used mainly by project members to ensure a correct undertaking of the activities outlined in the DoA. The use of the results, to which data constitute a fundamental part, is regulated by the Consortium Agreement and stored in accordance to the guidelines of the present DMP.

If the data generated are not deemed confidential or instrumental for commercial exploitation, they will be disclosed by publication on:

- the project website, under “Resources” (e.g. Training Handbook, Communication Plan, and deliverables classed as Public within the Grant Agreement of AntCom).
- Fragmentarium (Public Repository) for raw/elaborated data. Amongst such data, we foresee to publish data leading project’s scientific publications.
- GithHub
- Europeana
- Hypotheses

- Terra e Memoria at USC

With respect to data of a personal nature, anonymization is proposed to be used as a main strategy to ensure protection of personal data, and any publication of such elaborated data will be regulated according to the regulations in force (see section 4).

2. FAIR data

To make data FINDABLE, the project team will ensure each file of data is accompanied by a set of metadata to aid discoverability.

2.1. Making data findable, including provisions for metadata

All project's data are collected and stored in a trusted repository, i.e. *Ucloud*, an interactive digital research environment hosted by SDU and built to support the needs of researchers for both computing and data management, throughout all the data life cycle. Public data will be available on the project website, while sensitive documents are stored on Ucloud. In order to access projects' data stored on Ucloud, all project members have undergon online individual recognition carried out by the service support team at UCloud, SDU. For internal members access to Ucloud is unlimited. External reviewers or 3rd party interested in the data will be provided access to SDU drive platform, through an access link which is only valid for a limited period of time, i.e. 1 month.

Moreover, public data will also be stored in GitHub/Europeana /Hypotheses/Terra e Memoria in this case the JSON format will apply; this should ensure that the metadata will be exported in the DataCite Metadata Schema v3.1 standard, according to the OpenAIRE guidelines.[1] The metadata will be generated upon publishing onto the repository.

Bibliographies are stored in a public group on Zotero, which is also the tools used consortium-wide to manage references and literature lists.

For the data stored internally and shared amongst partners on UCloud, the team will strive to adopt the same standards for consistency, abiding by the following list of fields:

- file size,
- file extension,
- when the file was created
- author of the file (name and affiliation)
- a reference to the project (acronym and GA number).

Metadata in Microsoft Office files are stored in the "Properties" form accompanying each file, which can be completed/modified according to the above requirements; for other formats, guidelines from the software provider will be sought. For pictures, metadata can

[1] OpenAIRE guidelines: https://guidelines.openaire.eu/en/latest/

be added with photo editor software (proprietary or open source) or with other tools (like file management tools enabling metadata editing).

Every effort will be made to ensure that the data/publications are made available on the repositories in their definitive version, once the record has been published, it is no longer possible to change the files in the record. If further changes will be required on published versions the new version will entail a versioning of the DOI.

As an aid to identify and make searchable the data files, the naming will be developed to include most of the following information, as applicable:

- Project name or acronym
- Researcher name/initials
- Date or date range
- Type of data
- Version number of file
- Three-letter file extension for application-specific files

Files will be collected in directories containing a readme.txt file that explains the naming format along with any abbreviations or codes used.

Wherever possible, tags or keywords will be associated with the files to make them searchable according to themes. For files in Microsoft Office format, this can be done either upon saving a new file or working on the properties of the files; for other file formats, this will be explored by the beneficiary generating the file.

For published data, the metadata are created upon uploading onto Fragmentarium and GitHub. As for the digitization of SDU's fragments, dince SDU library has no preferences as far as bibliographic metadata is concerned, the project follows Fragmentarium's standard metadata template. Fragmentarium is the trailbraser laboratory for manuscript fragments, containing more than 7000 manuscript fragments, that will eventually also host SDU's fragments. Fragmentarium's template puts the fragment at the centre and it requires the following main categories: library, country of location, text language, dating, material, script type, decoration, musical notation, document type, publication for, language of description, illumination, restauration, reconstruction, curated tags, social tags, range types, projects.

2.2. Making data accessible

AntCom project aims to publish as a minimum:

- Project deliverables classified as Public
- Journal papers in gold/diamond (preferred) open access or on institutional repositories
- CS science project once deemed fit for publication

Because of the potential confidentiality, however, some of the data generated within the project might not be publishable because of GDPR issues. Such data are likely to be still exchanged internally, e.g. as results required to progress towards the objectives of the project, and their internal use is regulated by the Consortium Agreement. In these cases, all data are accessible to all project's members on Ucloud Platform.

The data will be made available through the project website and CORDIS library (public deliverables) and through GitHub /Fragmentarium (other data). According to the degree of access, journal papers might also be published on the project website and on the repository of the Institution/university generating the paper, as well as on the publisher's Journal site. As of January 2025, all the data generated by the project are consistently stored on UCloud. Public data are/will be made available through the AntCom webiste /GitHub/Fragmentarium/Europeana/Hypoitheses/The Danish Royal Archives have been elected as repositories of choice for all open access data generated by this project. For the second reporting period we plan to initiate an AntCom GitHub, as we have reached a critical mass of codes/softwares/model suitable for a consolidated outlet.

All files will be available in the format (extension) most readable/ most used, such as .pdf or .txt or ISO-STEP. If specific formats have to be used, a freely available reader will be identified for viewing.

2.3. Making data interoperable

The degree of interoperability of data generated is expected to be high as for instance, common formats of files and metadata recognized standards will be used. Should specialist metadata be used, a definition will be provided into a "Notes" or "Comments" field or within the readme.txt file within the folder where the files is stored.

English language will be used for all published and internally shared data, considering the international make up of the consortium. If languages other than English are used, e.g. for interviews of local stakeholders, English transcripts/translations will be made available.

2.4. Increasing reuse of data

Both for internally shared data and for published data, every partner generating data will be responsible for the quality of their data in terms of content (e.g. only data from test run in valid conditions, advanced draft versions of documents, etc.) and formalities, for example elimination of outliers from numerical data; formatting, internal referencing, correct use of pictures, etc.

Each publication will be provided with an ORCID number, which provides a permanent ID for researchers, linking them to their published work and datasets. Data for publication will need to meet the requirements set in this document (e.g. metadata, format as identified etc.). The Project Coordinator, in collaboration with the work package leaders, will check the overall quality of the data and their coherence with the requirements of the latest DMP, and will work with the data owner for addressing any shortcomings.

It is expected that non sensitive data will be public will be freely available. Published data are generally intended for reuse and will be left on the AntCom website and UCloud for at least five years unless otherwise specified.

Finally, AntCom makes use of Standardized Data Collection Methods: systematic checks are applied to avoid data entry errors. These include validating input formats (e.g., dates, email addresses) and using dropdowns or auto-completion features to minimize human error, for example, excel sheets, which use built-in functions (e.g., data validation, conditional formatting) for basic validation, or Google Forms (with built-in validation checks) in relation to the self-evaluation questionnaires. Specific measures will be implemented to ensure that data quality does not compromise security and privacy requirements, such as *Anonymization*, which ensure that personally identifiable information (PII) is anonymized where necessary (e.g., using hashing or pseudonymization); Moreover, data stored on Ucloud undergo a strict clearance protocol, which include virtual personal identification. In this perspective, Ucloud also offers a series of *Access Controls* functions, which set up user roles and permissions to control who can access and modify data, ensuring only authorized personnel can make changes.

3. Other research outputs

Project's generated research-outputs (such as articles, posters, presentations, ppts, publications) are stored on UCloud, in a "Outputs" folder. Publications are also stored on institutional repositories and linked to each researcher through their ORCID.

Raw data resulting from CH campaigns are stored on UCloud in a relevant folder.

4. Allocation of resources

Policy of the Consortium is to apply the DMP starting from the data collection phase. As the data gathering/generating is responsibility of each partner as applicable, so it is the inclusion of the DMP requirements. Use of Ucloud/GitHub/ Fragmentarium/Europeana/Hypotheses is free of charge (UCloud is subsidized by SDU), while a budget for scientific paper publication – in relation to Open Access policies – has been set aside for each beneficiary.

The project commits to maintaining Zotero AntCom bibliography alive for 5 years after the end of the project.

A list of available Open Access journals has been added as append to the Training Handbook.

The Project Coordinator will be in charge of ensuring data identified for publication are not jeopardizing and IPR protection through consultation with the workpackage leaders and partners involved.

Preservation of data on Ucloud is expected to be free of charge for the foreseeable future, and for at least five years after the end of the project. After the specified time, such data might be moved back to a private repository/backup drive, and the AntCom project partners notified. Some personnel time might be required to liaise with Ucloud and the AntCom project partners for such operations.

5. Data security

For all data, every partner will be responsible for ensuring the data generated (raw or elaborated from other partner's data) are regularly backed up and stored safely according to

internal safe keeping procedures. If such data is also shared internally at the Consortium, the data owner will be responsible also for the methods used to share information. At SDU data are backed up as they are created/updated on the institutional server OneDrive. At UniLe and UniVr data are backed up and stored on hard drives. At USC copies are made daily in an incremental way, that is, taking into consideration the changes with respect to the previous day and/or the complete copies that are made on one of the days of the weekend. When data include databases it is recommended that a dump of the databases to the disk is made so that the backups copy these dumps. UCloud itself performs backups in accordance with a backup policy that complies with ISO 27001 data security standards.

As previously mentioned, SDU Ucloud virtual platform is en entrusted repository for data recovery and secure storage/archiving: *UCloud is a highly secure platform for interactive high-performance computing suitable for sensitive data analyses in compliance with GDPR regulations. UCloud is built from the ground up following rigorous security and privacy-by-design principles. All the systems at the eScience Center are tested against comprehensive security benchmarks by the Center for Internet Security (CIS)*. In short, Ucloud follows the highest standards of information security management with a formal ISO27001 certification. The SDU eScience is ISO/IEC27001:2013 certified since February 2020, the first Danish university or public institution certified in the country.

6. Ethics

6.1. Responsibilities

According to the definitions of the GDPR , the partners collecting personal data are the processors, while the controller, who is ultimately responsible for the data treatment, remains the Coordinator of the project. Every processor is given by the controller the responsibility for implementing every measure for ensuring the protection of the personal data.

6.2. Consent to collect and treat the data

The consent to film and audio-record the interviews has been asked and obtained before/at the beginning of each interview; similarly, for the questionnaires, the respondents have given their written consent to the treatment of the data. AntCom has compiled with the support of AISO (*Italian Association of Oral Hisotry)* a formal *Participation and Recording Agreement* which needs to be signed prior every interview (see D1.3).

6.3. Protection

The personal data are to be collected as much as possible in an anonymous form, and where that is not possible, they will be anonymized before any sharing. The processor undertaking such anonymization is required to keep separated any files containing trace of the personal details (name, contacts etc) from the anonymized data. This is to avoid that images, opinions, statements etc. can be referred to a specific person. Such files should be kept in a safe location to minimize data breach, and destroyed after the time that will be stipulated within the consortium. Password protection and encrypting might be used to enhance security of data. No personal details shall be shared with other partners of the project.

7. Other

No other national/funder/sectorial/departmental procedures for data management are currently in use.

8. Appendix

a. Information required by the DMP online tool:

Admin Details

Plan Name: AntCom - Data Management Plan

Plan ID: D1.5

Grant number: 101073543

Principal Investigator / Researcher: Aglae Pizzone

Plan Data Contact: czi@sdu.dk

Plan Description: This Initial DMP describe the data expected to be generated by project AntCom, how such data will be made FAIR and how personal data collected will be protected.

Funder: European Commission HORIZON – MSCA – 2021-DN

Institution: SDU

Your ORCID: N/A

b. Zipped files potential content stored on Ucloud:

1. models and results
2. raw data and results
3. software and documentation
4. multimedia (.docx, .pdf, .ppt., . xlsx)
5. images
6. videos

c. Indicative Potential File Types:

Text	.txt; .doc; .docx; .pdf; .odt
Tables / Spreadsheets	.xls; .xlsx
Images	.tiff; .jpg/.jpeg; .png; .cr2/; .nef

xml Files	.xml
Video	.mkv, .mp4; .mov; .mxf
Audio	.mp3, .wav; .m4a
Publications	.indd; .pdf; .epub