ClimDivE

A Data Management Plan created using DMPonline.be

Creators: Judith Biernaux ⓘ https://orcid.org/0000-0002-0307-5847, n.n. n.n., Anaïs Vignoles

Affiliation: Université de Liège

Funder: European Commission (Horizon)

Template: Horizon Europe DMP +

Principal Investigator: n.n. n.n., Anaïs Vignoles

Grant number / URL: 10110788

ID: 204402

Start date: 01-01-2024

End date: 31-12-2025

Project abstract:

Currently facing major environmental changes due to global warming, a growing concern of modern societies is related to how we will be able to cope with these rapid changes to survive. This kind of preoccupation is probably not new: over the course of its history, the human species has been able to adapt to very different environmental conditions thanks to diversified cultural adaptations. The objective of the project "ClimDivE" is to understand the role of environmental change in the diversity and evolution of cultural adaptations at large geographic and chronological scales, using the archaeological record of the Gravettian (32-26 000 years before present) as an example framework. At this period, archaeologists have shown that two neighboring regions, Western and Central Europe, exhibit very different cultural trajectories, without being able to understand the large-scale mechanism that influenced this pattern. The ClimDivE project aims to explore the factors composing this mechanism by integrating archaeological data and paleoclimate simulations into a unified approach: "Eco-cultural niche modeling". This approach derived from the field of distributional ecology allows one to identify the environmental conditions associated with specific cultural traits. To do so, Dr. Vignoles will gather pertinent archaeological occurrences and environmental variables for the Central European Gravettian. These data will feed a mathematical model that represents the niche. These results will be compared with her previous work on the Western European Gravettian. Based at Liège with shorter stays in Kansas University and Paris-Saclay, the research will be supervised by three internationally recognized researchers specialized in the Gravettian, ecology and paleoclimatology. The proposal will result in a model underlining different types of factors – environmental, cultural, demographical – probably implicated in the separate evolution of neighboring regions during the Gravettian

Last modified: 06-06-2024

Created using DMPonline.be. Last modified 06 June 2024

1 of 9

ClimDivE
Full DMP

Version information

Action number

10110788

Action acronym

ClimDivE

Action title

Influence of Climate variability on the Gravettian cultural Diversity and Evolution

DMP version number

v1

Date

Question not answered.

1. Data summary

1.1 Will you re-use any existing data and what will you re-use it for?

The data that will feed the database will be secondary (bibliographic sources or direct contact with researchers)
Primary data will consist of climatic simulations parameters and results.

1.2 What types and formats of data and other research outputs will the project generate or re-use?

The format of the data base will be a LibreOffice base file. It will be accompanied by csv files for each table, a code for the data base structure and queries in SQL.
The data that will feed the database are textual. They will be referenced from bibliography or from researchers who agreed on sharing their unpublished data.
The simulation outputs format will be raster files (probably GeoTiff ; to be defined).
Other research outputs will be the code for niche modeling (R script files) and output models (either .jpg or GeoTiff rasters).

1.3 What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The purpose of data generation and reuse (data base and simulations) is to be combined in the niche modeling process. The models will be used to answer the two specific objectives of the project.

1.4 What is the expected size of the data that you intend to generate or re-use?

Given the relatively light formats (text- and table-based) of the files, an estimate of a few hundreds of Kb can be proposed. However, simulations files (high-res spatialised data) can add up to a higher volume. A finer estimate will be provided along the project but sufficient storage is made available by host institutions.

1.5 What is the origin/provenance of the data, either generated or re-used?

For the database input :

- For the secondary data collected by contacting researchers:
  - Data Transfer Agreements (DTAs) will be put in place with partners if needed that will document the data origin.
  - If no DTA is needed, sufficient provenance documentation will be requested upon contacting the researcher responsible for the dataset
- For the secondary data that is available in the literature, provenance is documented in the database itself with the dataset reference and metadata.

The simulation parameters will be collected from published material that will be documented through its metadata.

1.6 To whom might your data be useful ('data utility'), outside your project?

The sites data base will be useful for the broad archaeological community working on the Gravettian in Europe.
The simulations will be useful for any person needing to estimate the climate around 32 000 to 26 000 years ago.

2.1 FAIR data: Making data findable, including provisions for metadata

2.1.1 Will data and other research outputs be identified by a persistent identifier?

- Yes: describe below

The data base of sites, the simulations and the scripts will be published on an open repository (DataVerse ULiège) and will thus be associated to a DOI.

2.1.2 Will rich metadata be provided to allow discovery?
What metadata will be created?
What disciplinary or general standards will be followed?
In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

We will include rich metadata to ensure findability and reusability of the data.
The metadata will consist of fields associated with a Dataverse ULiege repository : citation metadata (Persistent identifier, publication date, title, author, point of contact, description, subject and distribution date), project metadata (project type, project category, PI, location, period), and other descriptors if needed, such as sites coordinates.
Additionnally, the fields of the database will be associated with a glossary indicating the definition of specific terms. Terms will be linked to open Thesaurus entries (ex. PACTOLS https://pactols.frantiq.fr/opentheso/). The glossary will be published together with the database.
The simulations will be associated with model parametrizations described in a datapaper or alternatively in a READ.ME file.
Code will be thoroughly commented and be associated with a READ.ME file.

2.1.3 Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

- Yes: describe below

Keywords will be associated with the data through Dataverse metadata.They will describe the nature, geographic and temporal scope of the data.

ex for the database : data base, Upper Palaeolithic, Gravettian, Central Europe, lithic industries

ex for the simulations : iLOVECLIM, palaeoclimatic simulations, Late Pleniglacial, Europe

2.1.4 Will metadata be offered in such a way that it can be harvested and indexed?

- Yes: describe below

The metadata associated with ULiege DataVerse is indexed and harvested by the main web research engine through OAI-PMH protocol.

2.2 FAIR data: Making data accessible

2.2.1 Will the data and other research outputs be deposited in a trusted repository?

- Yes: describe below

The outputs will be deposited in intitutionnal repositories : ULiege DataVerse for the datasets, simulations, as well as GitHub for codes and scripts ; ORBi for the publications.

2.2.2 Have you explored appropriate arrangements with the identified repository where your data and other research outputs will be deposited?

- No

Dataverse, GitHub and ORBi are institutional tools at ULiège and are planned to be used under normal conditions. No specific arrangement is needed given the nature of the datasets, and data will be made available openly.

2.2.3 Does the repository ensure that the data and other research outputs are assigned an identifier? Will the repository resolve the identifier to a digital object?

The ULiege data verse automatically assigns a DOI to the outputs.

2.2.4 Will all data and other research outputs be made openly available?

- Yes

The data will be made available under an open data licence such as CC-BY 4.0.

2.2.5 Is an embargo applied to give time to publish or seek protection of the intellectual property (e.g. patents)?

- No

2.2.6 If an embargo is applied (see question 2.2.5), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Not applicable

2.2.7 Will the data and other research outputs be accessible through a free and standardized access protocol?

- Yes: describe below

The data and outputs will be accessible in the ULiege Dataverse, where it can be freely downloaded by anyone. The chosen licence will ensure proper citation of datasets.

2.2.8 If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

Not applicable

2.2.9 How will the identity of the person accessing the data be ascertained?

Not applicable.

2.2.10 Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

- No

Not applicable

2.2.11 Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why.

- Yes

The datasets will be made available under an open-data licence such as CC-BY 4.0. The metadata will be discoverable by main search engines.

2.2.12 Will metadata contain information to enable the user to access the data?

- Yes

The metadata will contain information so that the user can access the data, such as the DOI (that serves as a URL to access the datasets through simple download).

2.2.13 How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Dataverse ensures a sustainable availability of the data and metadata, through a DOI, as long as institutions maintain the repositories.

2.2.14 Will documentation or reference about any software needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

The outputs will be shared in open formats (such as csv, GeoTiff). The data base will be developped under a open source database software (LibreOfficeBase), which makes it accessible using most open, free software. The code will be written with R, ensuring its openness. The metadata will specify the versions of the software and R language used.

2.3 FAIR data: Making data interoperable

2.3.1
What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines?
Will you follow community-endorsed interoperability best practices? Which ones?

The data and metadata will follow the DublinCore specifications to ensure different levels of interoperability (https://www.dublincore.org/specifications/dublin-core/interoperability-levels/).

Additionnally, the fields of the database will be associated with a glossary indicating the definition of specific terms. Terms will be linked to open Thesaurus entries (ex. PACTOLS https://pactols.frantiq.fr/opentheso/). The glossary will be published together with the database.

2.3.2 In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies:
Will you provide mappings to more commonly used ontologies?
Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

A Glossary for the database will be generated with links to published thesaurus; in case further details need to be specified, we will add them as supplementary documents in the publication of data.

The generated vocabularies will be published in the data paper, or will be available together with the data in the public repository (as a separate file or a READ.ME file)

2.3.3 Will your data and other research outputs include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

- Yes

For the data base, the datasets will be extracted from publications, that will systematically be referenced in each table.
The climate simulations parametrization will be based on published knowledge from the modeled time periods, that will be referenced in the READ.ME.
Some of the niche models will be based on a previous data base that I created in my previous work and is currently in press.

2.4 FAIR data: Increase data re-use

2.4.1 How will you provide documentation needed to validate data analysis and facilitate data re-use?

All the shared materials will be sufficiently documented for autonomous reuse through README files, commented codes and scripts, structured vocabularies and any other relevant process. All the methodologies will be transparently shared in assorted publications.

2.4.2
Will your data and other research outputs be made freely available in the public domain to permit the widest re-use possible?
Will your data and other research outputs be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

The data will be made available under an open data licence such as CC-BY 4.0.

2.4.3 Will the data and other research output produced in the project be useable by third parties, in particular after the end of the project?

- Yes

Yes, given the properties of the CC-BY 4.0 licence, there will be no restrictions to reuse even after the project, apart from the obligation of

citation.

2.4.4 Will the provenance of the data and other research outputs be thoroughly documented using the appropriate standards?

- Yes

See sections 1 and 2.1

2.4.5 Describe all relevant data quality assurance processes.

Even though all the collected datasets will go through the analyses processes, they may be of various levels of quality. For each analysis, a quality threshold will be applied a posteriori and strictly documented in the publications. For example, the peer-reviewed characteristics of the harvested datasets may come into consideration.

For simulation-based datasets, quality of outputs will be discussed with project experts.

3. Other research outputs

3.1 Do you have any additional information, that was not addressed in the previous sections, which you wish to provide regarding other research outputs that are generated or re-used throughout the project?

No - this management plan has been designed both for datasets and other research outputs (see sections 1-2)

4. Allocation of resources

4.1 What will the costs be for making data and other research outputs FAIR in your project?

Datasets and other outputs will be made FAIR using Dataverse, ORBi and GitHub, all made available by ULiège at no additionnal costs.

4.2 How will these be covered?

Not applicable

4.3 Who will be responsible for data management in your project?

The person responsible for data management will be Anaïs Vignoles.

4.4 How will long term preservation be ensured?

A sustainable version of the published material will remain on Dataverse.
For internal archival of the datasets, a strategy will be designed with the help of ULiège IT service, where long-term preservation datacenters are made available.

5. Data security

5.1 What provisions are or will be in place for data security?

All datasets will be stored on the ULiège installation of OneDrive, where sufficient storage space is available. A mix of OneDrive and DoX (the ULiège IT-operated cloud platform) will be used for storage and backups. The ULiège IT service is ISO27001-certified.

Working datasets will be temporarily stored on mobile storage solutions when needed (laptops and mobile hard-drives). A mobile hard-drive will also be used for additionnal backup.

For security reasons, portable data storage supports will not be used for travelling but their content will be backed-up to a cloud-based platform (OneDrive or DoX) and used remotely, using the ULiège VPN service.

The ongoing documents, codes, datasets that are stored in open repositories are protected via "private view" functionalities and restrictive use licenses.

5.2 Will the data be safely stored in trusted repositories for long term preservation and curation?

- Yes

The data will be sustainably stored on ULiège Dataverse and the ULiège IT datacenters for a long-term preservation (especially after the project ends).

6. Ethics

6.1 Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?

- Yes

Unpublished secondary data will be shared according to the Data Transfert Agreements when applicable.

6.2 Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

- Not applicable

Not applicable

7. Other issues

7.1 Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

- No

Not applicable

ClimDivE
Grant proposal

---

1. Data summary

1.1 Types of data/other research outputs

Question not answered.

2. FAIR principles

2.1 Findability of data/research outputs

Question not answered.

2.2 Accessibility of data/research outputs

Question not answered.

2.3 Interoperability of data/research outputs

Question not answered.

2.4 Reusability of data/research outputs

Question not answered.

3. Resources and responsibilities

3.1 Curation and storage/preservation costs

Question not answered.

3.2 Person/team responsible for data management and quality assurance

Question not answered.