

# Guidance on the README file for archaeological datasets

This document is part of the [Protocol for the management of research data of the Institut Català d'Arqueologia Clàssica \(ICAC\)](#) and aims to guide researchers in writing a README file for their datasets, provide best practices and an example template.

Guide based on:

- CSUC [2021]. Readme template. Available: <https://confluence.csuc.cat/pages/viewpage.action?pageId=104529933>
- Varrato, Francesco; Borel, Alain; Gabella, Chiara (2022). README file for datasets - Best practices and template. EPFL. Available: <https://infoscience.epfl.ch/record/298249>

Adapted: Lydia Gil and Arnau Estivill (Institut Català d'Arqueologia Clàssica)

Date: October 18, 2023

Digital version: <http://hdl.handle.net/2072/534226>

*This work is licensed under a Creative Commons Attribution 4.0 International License.*



For any questions regarding this Guide, please contact the ICAC's Documentation Centre and Library.  
PDF/A-1 (ISO 19005-1), created in MS Word (Version 2019) from the docx format.

## Document version control

---

<b>Version</b>	<b>Date</b>	<b>Description</b>	<b>Authors</b>
1.0	2023-10-18	First published version	Lydia Gil and Arnau Estivill

## What is a Readme File?

The README file provides information about the research data in order to help ensure that the data can be interpreted and reused correctly, both by yourself and by other people in the future.

## Best practices<sup>1</sup>

There is not one right way to structure a good README file, but some best practices exist. For example:

### When writing the README file:

- If you plan to deposit the data in a repository, you will need to provide a README file. Creating this archive at the beginning of your research and constantly updating it will make this task easier for you.

### Structure of the README file:

- Name the file simply "README".
- Locate the README file at the root of the project.
- If needed, create one README file per dataset / folder and use the same format for all.
- Save it as a plain text file (txt) and avoid proprietary formats such as MS Word.
- The information must be in English, in addition to the original language, to favour the reusability of the dataset.
- Use standardized date formats: YYYY-MM-DD or YYYY-MM-DDThh:mm:ss.
- Follow the conventions used in archaeology for names, keywords and geospatial data. Whenever possible, use terms from standardized taxonomies and vocabularies.
- Use of blank lines or dashes to separate the document into paragraphs.

---

<sup>1</sup> These lists are non-exhaustive.

- Use bullet points or ordered lists instead of long paragraphs.

### **Content to include in the README file:**

- General information: Title, abstract, keywords, authors, ORCIDs, institutions, dates, funding, contact information and temporary and geographic data coverage.
- Dataset persistent identifier.
- Access information: licenses, terms of use and recommended citation.
- Related items: links to related publications, datasets, websites or other research output.
- Methodology: methods for processing and collecting data, software or information about instruments and standards.
- File overview: files and folders structure description, file formats, files naming convention(s), original source if any, dataset version, updates.
- Data dictionary for tabular data or databases with variables, units of measures, codes, symbols, abbreviations, etc.

### **Template**

The document is organized by blocks and each block contains information between square brackets and their level of obligation.

Download the template in TXT format: <http://hdl.handle.net/2072/534226>

(Available in English, Catalan and Spanish).

-----

This file has been created on [date in format AAAA-MM-DD] by [Name].

-----

#### GENERAL INFORMATION

-----

##### 1. Dataset title:

[Mandatory]

2. Authorship:

[Mandatory | Repeat the structure, one for each author.]

Name:

Institution:

Email:

ORCID:

-----  
DESCRIPTION  
-----

1. Dataset language:

[Recommended]

2. Abstract:

[Mandatory | Explains the content of the dataset, its purpose, what data will be found, the types of files, their formats, etc.]

3. Keywords:

[Mandatory | Minimum 3 keywords from a thesaurus such as [Art and Architecture Thesaurus](#), [Getty Thesaurus of Geographic Names](#), [PeriodO](#), etc. | Repeat the structure, one for each term.]

Term:

Vocabulary:

Vocabulary URL:

4. Date of data collection:

[Mandatory | Single date or date range.]

From: YYYY-MM-DD or YYYY

To: YYYY-MM-DD or YYYY

5. Historical period covered by data:

[Mandatory | Single date or date range.]

Start: YYYY-MM-DD or YYYY

End: YYYY-MM-DD or YYYY

6. Publication Date:

[Mandatory | Date was made available for distribution/presentation.]

YYYY-MM-DD

7. Grant Information

[Mandatory | Repeat the scheme for each agency if necessary.]

Grant Agency:

Grant Number:

8. Geographical location/s of data collection:

[Mandatory | Use Geonames: <https://www.geonames.org/>]

Archaeological site:

Country:

City:

Geographic Unit:

Latitude:

Longitude:

9. Kind of data

[Recommended | Select one or more options.]

- Observation data
- Experimental data
- Simulation data
- Compiled data

- Geospatial data
- Textual data
- Program source code

#### 10. Alternative URL

[Recommended | Link where the dataset can also be accessed. They can also be added to personal or project websites.]

-----  
ACCESS INFORMATION

1. License of the dataset:

[Mandatory | Choose a license for data or software: <https://ufal.github.io/public-license-selector/> ]

2. Citation:

[Mandatory]

2. Dataset DOI:

[Mandatory]

3. Related publications:

[Mandatory | Bibliographic citation]

4. Links to other research outputs and datasets:

[Recommended]

-----  
VERSIONING AND PROVENANCE

1. Specify the origin of the data:

[Recommended]

2. Last modification date:

[Recommended]

YYYY-MM-DD

3. Version dataset:

[Mandatory]

-----  
METHODOLOGICAL INFORMATION  
-----

1. Methods used to collect and generate the data:

[Recommended | Select one or more options.]

- Archaeological excavation
- Archival research
- Archaeological prospections
- Analytical Investigations: Inorganic Materials Study, Dating (Bioarchaeology, Ancient DNA, Dendrochronology, C14)
- Other (explain)

2. Data processing methods:

[Mandatory | Describes how the published dataset was generated.]

3. Software used to generate the dataset:

[Mandatory | Include name and version.]

4. Software or instruments needed to interpret the data:

[Recommended]



5. Information about instruments, calibration, and standards:

[Recommended]

6. Quality-assurance procedures performed on the data:

[Recommended | Select one or more options.]

- Set up of scientific and technical committee.
- Use of tools for automatic checks.
- Data conform to format specification.
- Consistency verified with data models and standards.
- Other (explain)

7. People involved in experiments, surveys, processing, analysis, etc.

[Recommended]

-----  
FILE OVERVIEW  
-----

1. Explain the file naming conversion:

[Mandatory]

*Ex. ArchaeologicalSite\_CaptureYear\_TypeFile\_Number.jpg*

Attributes

-----  
Archaeological site: A very short title representing the archaeological site.

Year: The year of the excavation (YYYY).

Type of file: A very short title of the kind of file.

Number: Three-digit unique sequential identifier such as 001, 002, 003.

2. File list:

[Mandatory | Repeat the scheme for each file.]

File name:

Description:

3. Relationship between files:

[Recommended]

4. File(s) format(s):

[Mandatory]

5. Directory structure:

[Recommended]

-----  
SPECIFIC INFORMATION FOR TABULAR DATA  
-----

1. Data dictionary (variables, units of measures, codes or symbols for missing data, special formats, or abbreviations).

[Mandatory | Check out ICAC's template: <http://hdl.handle.net/2072/536963> ]