Institut Català d'Arqueologia Clàssica

January 2024

RESEARCH DATA MANAGEMENT TRAINING IN ARCHAEOLOGY

NOTE: This document is part of the Protocol for the management of research data of the Catalan Institute of Classical Archaeology (ICAC-CERCA) and contains an asynchronous training workshop to acquire basic notions about the concepts, methods and tools associated with data management of research. It is also available on the Canvas platform.

Acknowledgements

Training based on Muilenburg, Jennifer [2020]. Research Data Management Class Template. University of Washington. Available at: https://lor.instructure.com/resources/410b649b56324c449d89112868855622

Adapted by: Lydia Gil and Arnau Estivill

Documentation Centre, Library and Open Science

Catalan Institute of Classical Archaeology (ICAC-CERCA)

Plaça d'en Rovellat, s/n - 43003 Tarragona

Phone number: 977 24 91 33 / https://openscience.icac.cat / lgil@icac.cat

Date: 2024-01-31 Version: 1.0

https://doi.org/10.51417/rdm_001

PDF/A-1 (ISO 19005-1) created using MS Word (2019 version) of docx format.

Index

Introduction

Welcome

Welcome to the **Research Data Management training in Archaeology**, a workshop for the **Catalan Institute of Classical Archaeology (ICAC-CERCA)** community which was created to make the content **openly available and as a self-learning course.**

The course is aimed at PhD candidates at the Institute who require a hands-on introduction to Research Data Management (RDM). However, the course can also be useful for researchers and master students interested in learning the basics of RDM.

It is structured in four successive modules that can be completed at your own pace. Each module contains informative material and self-assessment quizzes to reinforce the knowledge acquired on certain topics. This self-learning course also includes suggested exercises to apply what you will be learning throughout the content.

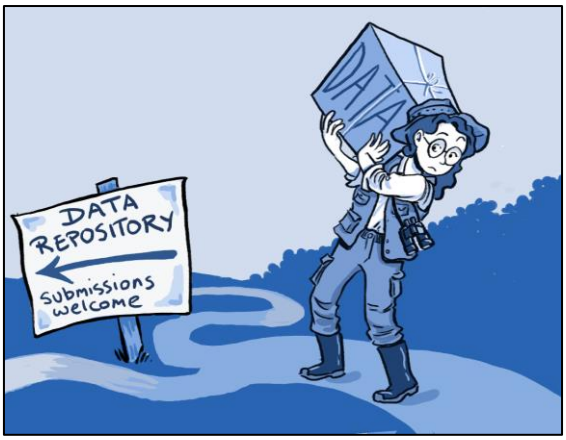


Table of Contents

- **Module 4**: The Big Picture
- **Bibliographical references**

Connect with us

If you would like support and advice specific to your project, write us to schedule a consultation.

Have some feedback? If you notice any issue, or have suggestions or requests, please contact us.

Why bother with data management?

Data is the center of most research, and this course will teach you how to give your future self a true gift: data that is where you need it, when you need it, and how you need it. Good data management practices will let you:

Image credit: Genpact

- **Protect your data from loss.** Whether it's spilling your coffee on your laptop or accidentally hitting 'delete' when you meant 'copy,' you'll be safe in the knowledge you have backup copies that will prevent you from having to do everything over again.
- **Find it when you need it**. So, you ran that analysis three times with three different initial variables and ended up with three output files. Which one were you supposed to work with, again? A logical and concise folder organization and file-naming convention will make it obvious.
- **Secure your data**. Will you automatically remember whether you shared that Dropbox folder with a collaborator from three projects ago? No worries - you'll have decided to use your institution server instead, and you'll know precisely who has access.
- **Reuse your old data**. You get a flash of inspiration in the shower: what if I compare that one dataset I created last year with the one I'm creating now? Consistent metadata standards and thoroughly recorded protocols will make it easy.
- **Share your data**. You want to join the Open Science movement, or you just want to be able to share your data with a collaborator. Piece of cake - you'll have planned for the level you want to share and will have found a distribution mechanism that works for you.

If you have time, please read this supplemental article on the 10 Simple Rules for the Care and Feeding of Scientific Data.

Module 1: Planning for Data Management

- **Overview: Before you start your project**

Taking the time to think through your plan for data management *prior* to beginning data collection can help you refine the **systems you will put in place** and the **strategies you will follow** over the course of your project to make sure you reap the benefits of good data management.

By the end of this module, you will be able to:

- Identify the characteristics of your data that will have an impact on your data management strategies.
- Locate data management requirements specified by funders, journals or other third parties.
- Describe the roles and responsibilities that will need to be delegated in order to execute your data management plan.

Resources 1A: Data Characteristics

- **What do we have here?**

The first step in the data management planning process is to assess the **type of data** you will collect or produce:

- Observational (questionnaires, interviews, photos...)
- Experimental (Gas Chromatography-Mass Spectrometry, Isotope Analysis, C14 Results...)
- Computational (sensor data, GIS, program source code...)
- Compiled data (compiled databases, 3D models, text and data mining...)
- Reference data (old image databases, archive collections...)

	Your research methodology will have important effects on how and what **metadata** you collect, how much of your **process** needs to be documented, and how easy it will be to **keep track** of that additional information.
	Assessing your data will also give you clues as to the **value** of your data for **sharing and long-term archiving**.

- **Technical characteristics**

And then there are some nitty-gritty issues to consider:

	**Format** How accessible is your data? Are you going to need to worry about software compatibility for your collaborators - or your future self? Do you use open formats?
	**Size** How much of it will there be? Does your computer have enough disk space to work with it? Or are you going to need to figure out how to get access to a server or cloud storage?

Having a good handle on your data's general characteristics will help you set up your processes to make life as easy as possible over the course of the project.

Resources 1B: FAIR data and software

- **FAIR ≠ Open**

FAIR ensures data can be found, accessed, understood and reused.

Data can be shared under restrictions & still be FAIR: "As open as possible, as restricted as necessary".

- **FAIR data principles**

	**Findable** Data and metadata are easy to find by both humans and computers.
	**Accessible** Both humans and computers can readily access or download datasets.
	**Interoperable** Data from different datasets are prepared to be combined or exchanged.
	**Reusable** Published data can be easily combined/replicated in future research.

- **Steps toward FAIRer data**

To achieve FAIRness, data objects should at least have:

- **A persistent identifier (like DOIs) for the data object as a whole.** Without a persistent identifier, the data object simply will not be findable let alone reusable in the long-run.
- **A sufficient set of metadata.** A sufficient and standardised set of metadata (elements which describe the data) will enhance findability, interoperability, and reusability. The quality of the descriptive information regarding the data has a profound impact on their reusability. So the more documentation of the data's context, the better.

- **A clear licence.** Researchers (and computers) who find a dataset should immediately know what they are allowed to do with it. Stating clear re-use rights is like having a warm 'Welcome' on the doormat of your dataset.

One of the ways to make sure your data will not become useless in the long-run is to choose a (trusted) data repository which has these attributes built into its infrastructure for dataset submission.

How FAIR are your data? Have a look at the checklist by Jones and Grootveld.

- **Software can be made FAIR as well**

FAIR principles can be applied to the development of research software to improve data management and stewardship. Here are five recommendations for FAIR software.

Image source: Netherlands eScience Center and DANS

Resources 1C: Open formats

- **Formats for long-term accessibility**

An open format is a file that can be read by multiple software. Using open formats serves two main objectives:

- Ensuring access to your documents regardless of the software used by other people.
- Ensuring the longevity of your documents so that they remain accessible even if the software used to create them is not available or no longer exists.

Examples of open file formats are PDF/A, CSV, TIFF, OpenDocument Format (ODF), ASCII, tab-delimited format, geoJSON and KML.

- **Preservable digital objects are vital for access by future researchers and for data interoperability.**

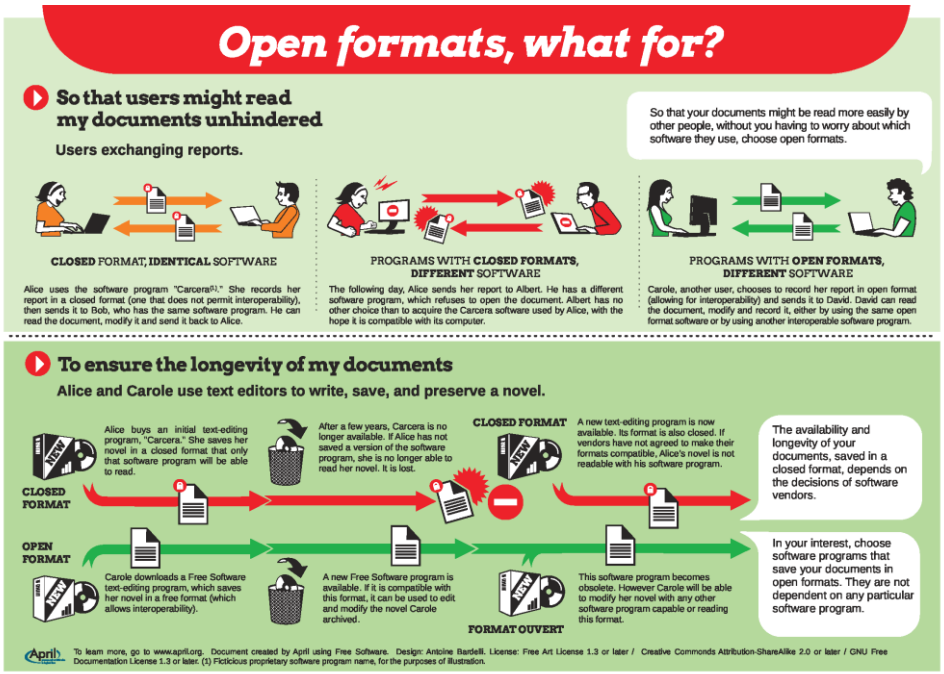


Image credit: April

Follow this link for information on the file formats recommended by the ICAC-CERCA for long-term preservation.

Resources 1D: External Data Requirements

- **Not just a good idea...**

We'll talk about motivations for data management as we go along (**Reproducibility**, **Open Science & Scholarship** **Not losing all your thesis work**), but you also need to be aware that your funder - and many of the journals you want to publish in - have requirements for how you handle your data during the project and what to do with it after you're finished. So it's important to **research their data management and data sharing requirements** before you start collecting.

Most funding agencies have a public access requirement attached to their grants -for publications and also for data. Make sure to check the current requirements on your funder's website.

- 8 ambitions of the EU's open science policy
- Spain finally has its open science strategy (2023-2027)
- Implementation of a Catalan Open Science strategy objectives and measures

At the beginning of a project, you may not know where you'll end up publishing the results, but it's still a good idea to familiarize yourself with the data policies of the major journals in your field. Information about academic journal requirements is typically provided on the publisher's website.

Here's Elsevier list of requirements about data.

- **Legal issues**

The management and sharing of research data creates questions of data ownership, data protection and ethical approval. It is advisable to address these questions before the project starts, and to make sure that all parties are in agreement.

	Archaeological and Historic Environment archives may sometimes include personal, confidential and sensitive data.
	This may raise issues of confidentiality and privacy covered by institutional ethics policies, and that data may fall within the scope of the General Data Protection Regulation (GDPR), the Spain's law on data protection and other legislation.
	Most data, however, can be shared ethically and legally if researchers employ strategies of informed consent and anonymisation.

Funders expect your data management plan to demonstrate that you have considered these issues.

	Check out this website which contains a collection of questions with guiding information on intellectual property and personal data.

Resources 1E: Roles and Responsibilities

- **Who's doing what?**

If you are working with a team, it can be easy to assume someone else is handling the data management. Teams should **clearly define the roles** of individual members as part of a data management plan.

The roles and responsibilities your team designates will depend on the number and location of researchers involved and the nature of the project. Regardless, **every task** you identify in your data management plan needs to have a **specific person** assigned to it.

Browse a sample **list of roles** and responsibilities from Data One.

To assign roles and responsibilities you can consider the following questions:

- Who is the project director?
- Who collects the data?
- Who knows where all of the data is stored and makes sure it's organized correctly?
- Who generates the metadata?
- Who analyzes the data?
- Who is responsible for the backup schedule?

It is very important for the ICAC-CERCA that researchers follow best practices on Research Data Management (RDM). That is why since 2021 the Institute has published its protocol which provide a clear division of roles and responsibilities around RDM. Click here to find out.

Quiz 1

Question 1

- When following the FAIR principles, access to data ensures the _____ of data.

Question 2

- FAIR data is _____ equivalent to _____ data.

Question 3

- If data contains confidential/sensitive/personal information, the data should remain closed, but you need to provide access to the metadata to comply with the _____ principles.

Question 4

- If you want to make data _____ a data set should have a persistent identifier.

Suggested Exercise 1: Exploring the FAIRness of datasets

Find a dataset

In this activity, you will **select** a publicly available data set, **examine** its contents, and **answer questions** about your selected data set.

1. Choose a dataset that interests you just by reading the title.

- *Images and GIS Data from an Archaeological Evaluation at Medebridge Solar, Fen Lane, and Medebridge Road, South Ockendon, Essex, 2022*
- *Amphoric Epigraphy: Data and graphs for the elaboration of food supply networks in Roman Empire*
- *Catalogue Female Monasticism in Medieval Ireland An Archaeology TCollins*

Search for the dataset with the provided title. Feel free to use your usual search methods - no restrictions there

If none of the titles we provide are interesting for you, you can search for datasets in the following data repositories using keywords: ADS, CORA.RDR or Zenodo.

2. Once you find the dataset reflect on how Findable, Accessible, Interoperable, Re-usable (FAIR) you think this dataset is. You can download the following template to guide you with your assessment in this link.

Do you have any question about this activity? Feel free to share your questions or concerns by email!

Module 2: Active Data Management

- **Overview: During the course of your project**

In Module 1 we examined milestones in the planning process for data management. By carefully developing a plan before your research begins, you will avoid time-consuming problems.

In this module, we will discuss Data Management **concepts** you will encounter **during the research project**.

By the end of this module, you will be able to:

- Define metadata.
- Recognize the importance of procedural and data documentation.
- Plan for storage and backup.
- Identify how to appropriately name, version, and organize your data.

Resources 2A: What is Metadata?

- **What is metadata?**

Simply put, metadata is data about data. This is the information that is necessary to interpret the data; researchers often refer to the process of compiling metadata as **data description** or **data documentation**. Complete metadata also includes information about the **processes** used for collecting, analyzing, and archiving your data.

	To see an example of data documentation, view this detailed README file. This file includes a description of the project, a data dictionary, as well as information about licensing and citation.
	Have a look at the ICAC-CERCA guide to writing **README file for archaeological datasets**.
	Take a look at the ICAC-CERCA **data dictionary template**.

- **What is a metadata standard?**

Many disciplines have come together to agree on the basic set of metadata that is necessary to understand the data collected in their field. This is often accompanied by a standard for the format the metadata uses, which allows for comparison between datasets and easy electronic transfer.

	There are a large number of metadata initiatives. They can be developed by a wider community or specifically for archaeology. These initiatives range from extremely detailed and specific metadata systems such as the Federal Geographic Data Committee's (FGDC) Content Guidelines for Digital Geospatial Metadata to the much simpler and more generalised Dublin Core.

As an example, the USGS has excellent information about the metadata standards and formats that have developed around geospatial data, including Federal Geographic Data Committee (FGDC) standards like Content Standard for Digital Geospatial Metadata (CSDGM).

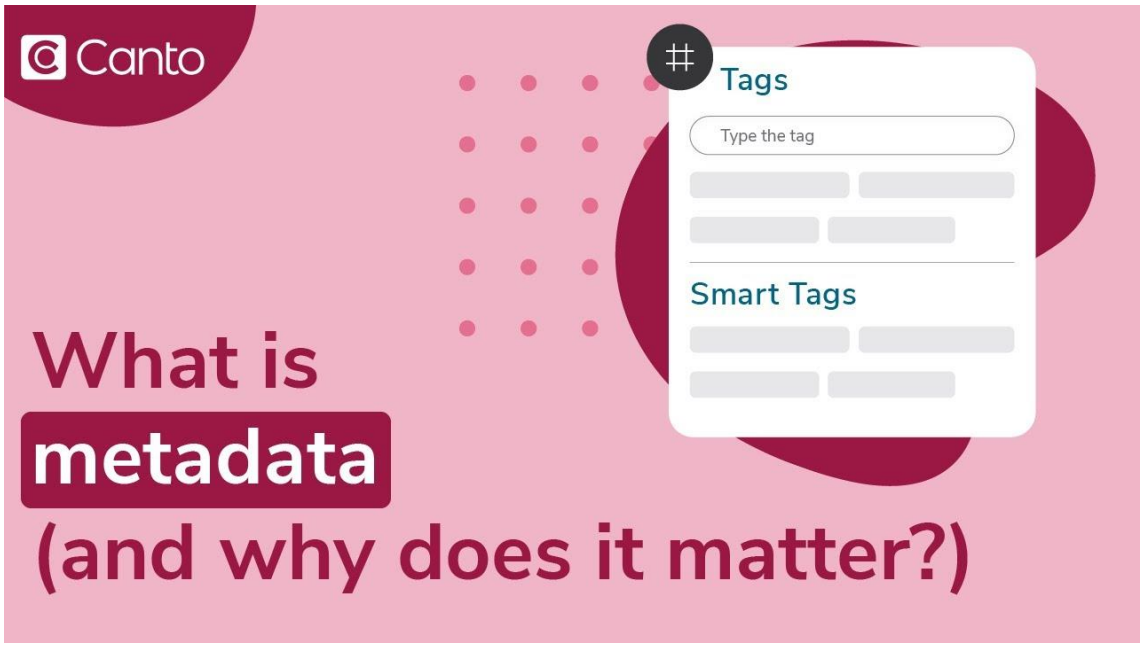
- **Why is metadata important?**

Documenting what you did during your research is important for several reasons. It will not only help you remember what steps you took to reach a particular result, it also facilitates future use of your data by yourself as well as others. Good metadata increases the visibility of your research through search engines, for example, and allows other researchers to understand what your research contains.

What is metadata (and why does it matter)?

Please watch the entire video. Length: 03:57

https://www.youtube.com/watch?v=fZWg0ClQkYQ



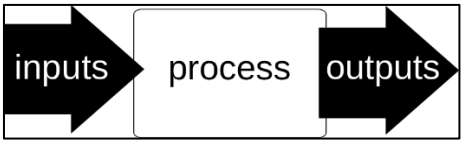
Resources 2B: Documenting Your Procedures

- **Documentation**

As we said in Resources 2A: What is Metadata?, documenting your process is an integral part of metadata. This is something that science has always valued - hence the long tradition of keeping a field notebooks or excavation diaries. Keep in mind that there are a number of important process issues that may seem secondary to your data documentation but are actually critical to record for reproducibility.

Things like:

- The **software** used to create and store the data. It will often be important, for example, to know what version of the software you were using.
- The **code you wrote** to manipulate or transform your data**.**
- And of course, the **protocols and workflows** you used throughout the project.



- **Software and services to the rescue**

Thankfully, there are a number of places to turn when you need help with this kind of documentation.

- **Versioning software** like Git can be a life-saver when working with code and other complex, collaborative projects.
- **Hosting services** like GitHub can help you collaborate with others and keep track of that code you so carefully versioned, but its "Issues" feature actually makes it a fantastic general project management tool, as well. For code-intensive projects, Jupyter offers a platform to integrate explanatory text with live code and visualizations with over 40 programming languages.

Information about **Journals and Electronic research notebooks in archaeology** is described here.

Didn't find what you were looking for? You can email us with your requests for help!

Resources 2C: Storage, Backup and Security

- **Storage and backup**

Planning for storage and backup of data costs a small investment in time, but pays off in the event of a disaster or unforeseen occurrence. Complicated research projects sometimes require years to gather data. Losing that research can be catastrophic.

When choosing a suitable storage solution to fit your project's needs, a lot of questions need answering. For example:

- How much storage space do I need?
- Who needs access?
- What precautions should I take to protect my data against loss?
- Which storage solutions are suitable for personal data?

It is an important aspect of data management planning to determine what your storage needs are and select solutions accordingly.

Portable device (*Laptops, tablets and external hard-drives*)

- Use for temporary, short-term storage for non-sensitive data.
- Conduct regular checks to ensure your device is working and that files are accessible.
- Not recommended for long-term storage, as their longevity is uncertain and they are easily damaged.

Cloud storage (*OneDrive, SharePoint, Institution's OwnCloud)*

- Use cloud services for granting shared, remote and easy access to data and other files to all involved in the project.
- Read the terms of service. Especially focus on rights to use content given to the service provider.
- Opt for European, national, or institutional cloud services which store data in Europe if possible.
- Do not make this your only storage and backup solution.
- Do not use for unencrypted (sensitive) personal data.

Local storage (*desktop computers and personal laptops)*

Using desktop computers and personal laptops as the primary way of storing and accessing data and files is only suitable for projects involving very few people (ideally: only yourself) and where data and files will not have to be moved back and forth between personal computers frequently.

Networked drives (*shared drives on institution server)*

- Use for distributed collaborative projects involving many people who need access to data and files.
- Use in combination with a suitable security strategy to protect data and files against unauthorised access.
- Think about long-term archival solutions for data that is complete and has been analysed. Valuable storage space might be released in this way.

Tip to avoid data loss:

You should regularly back up copies of your work using the **3-2-1 Rule**: keep at least three copies of your data, on two kinds of media, at least one of which is in a geographically separate place.

Don't let this be you!

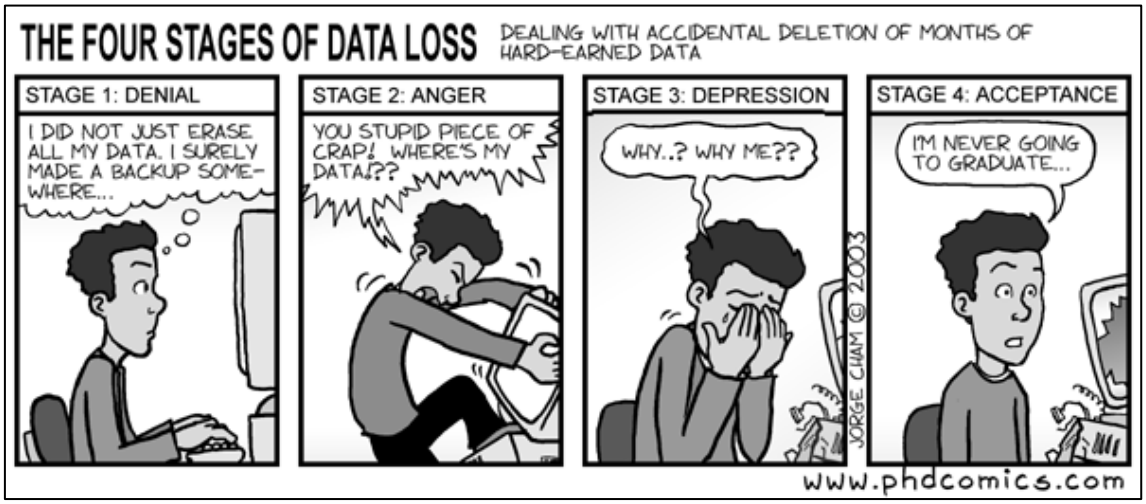


Image credit: PhD Comics

- **Security**

As we learned in Module 1, there are times when the data you're producing and working with is so **sensitive** that you have strict **legal requirements** around its handling. But we **all** should be concerned about the security of our technical infrastructure.

11 practical ways to keep your IT systems safe and secure.

Resources 2D: Data Organization

- **Data Organization**

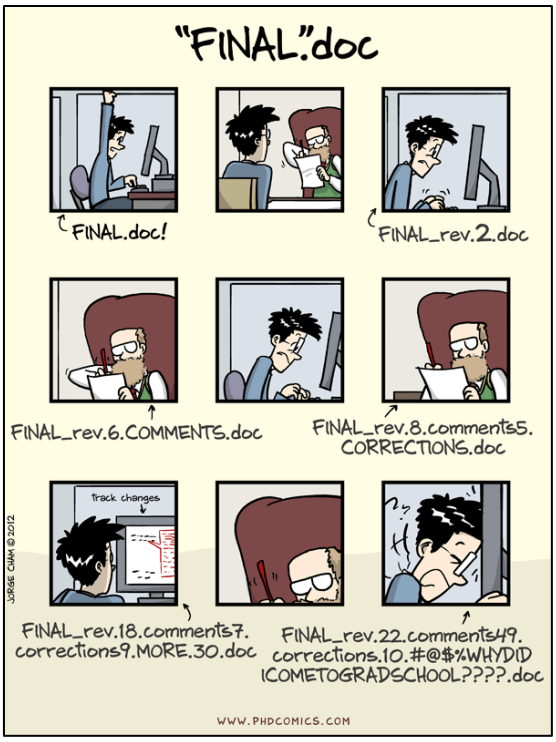


Image credit: PhD Comics

Creating a **well-organized hierarchy of files** with **clear naming conventions** is an important part of improving your research process. This is especially important if you are working with large data sets and complex output files or coordinating with multiple people at multiple institutions.

There are many ways to structure your folders, and multiple naming conventions you can use. The key is **consistency**. Make your file names **brief but descriptive** and consider including information about **dates and versioning**. **Best practice** is to consult with your group or with your co-workers to develop one schema that everyone is willing to follow consistently.

- **Best practices**

Some universal best practices for file naming are:

- Create meaningful but brief names.
- Use file names to classify types of files.
- Avoid using spaces, dots and special characters (& or ? or !).
- Use hyphens (-) or underscores (_) to separate elements in a file name.
- Use three letter file extensions to ensure backwards compatibility (ex: .doc, .tif, .txt).
- Include versioning of file names where appropriate.
- Do not use letter case to identify different files (ex: datasetA.txt vs. dataseta.txt)

Here you can learn more tips for naming your files.

Some tips for organizing directories

- Use a system that is logical to you, but simple and self-explanatory to others.
- Organize data hierarchically, and identify ways to divide your data into categories or attributes such as project, time, location, file Type (text, images, models, etc.)
- Include basic information such as the project title, dates, and some kind of unique identifier, such as a grant number.
- Document your file directory structure and describe the types of records that should be maintained in those folders in a README file.

Example folder structure

In this example, data and documentation files are held in separate folders. Data files are further organised according to data type and then according to research activity. Documentation files are organised also according to type of documentation file and research activity.

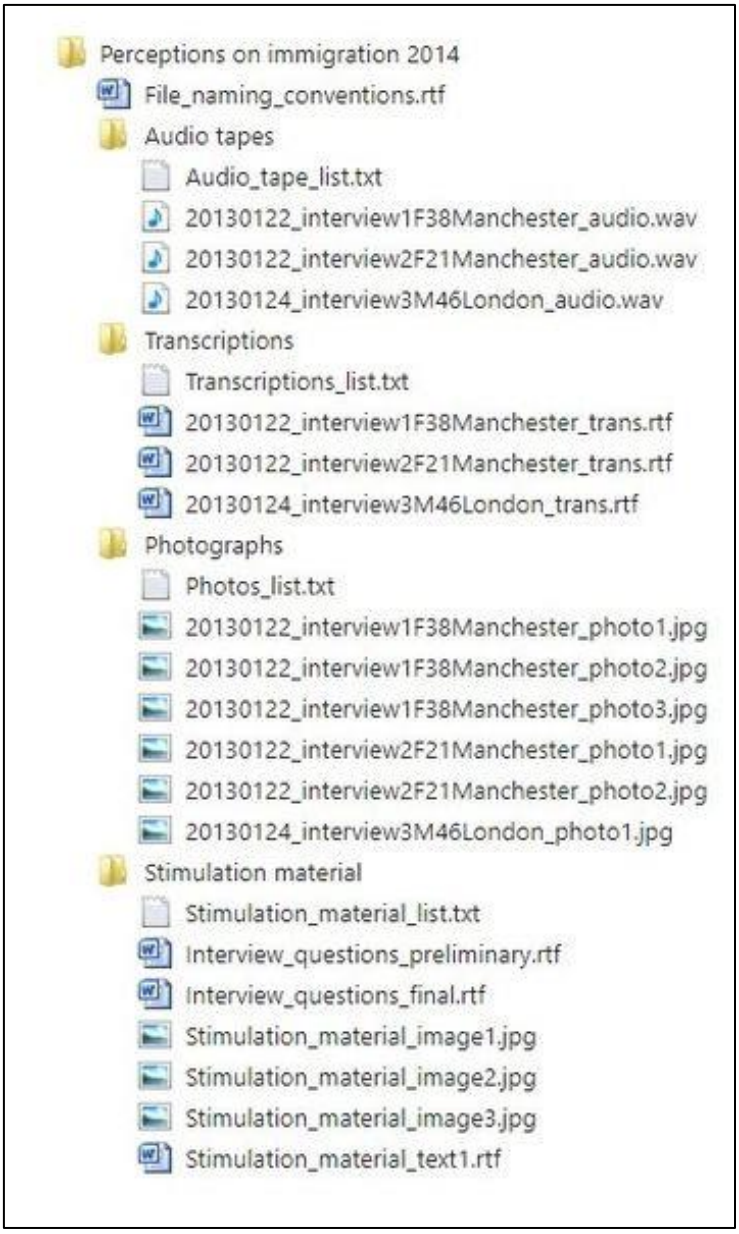


Image credit: UK Data Service (2017). Organising data.

If you have many files already named, consider using a file renaming application such as ReNamer (Mac/Windows).

Quiz 2

Question 1. Consider the following file names. Decide which of the three options uses the most effective file naming convention.

- ❑ 20190324_Attachment
- ❑ 240319attch
- ❑ 24 March 2019 Attachment

Question 2. Consider the following file names. Decide which of the three options uses the most effective file naming convention.

- ❑ 2019-08-11_bioassay_toxicity_V1.sps
- ❑ labtox_recent_110820_old_version.sps
- ❑ FFTX_Méta_3776438656.sps

Question 3. Good data organisation makes the data Findable for yourself and your collaborators.

- ❑ True
- ❑ False

Question 4. Thinking about data organisation at the beginning of your project can save you a lot of time later on.

- ❑ True
- ❑ False

Question 5. Consistency is not relevant in data organisation.

- ❑ True
- ❑ False

Suggested Exercise 2: Project File Structure and Naming

The next exercise asks you to describe the structure of your project folders and define and describe the file naming system for each different data types.

1. Briefly list the data generated and used in the course of your project.

- ❑ Documents and reports
- ❑ Spreadsheets
- ❑ Databases
- ❑ Photographs and images
- ❑ 3D Models, Visualisation and Virtual Reality
- ❑ Laser Scanning
- ❑ LIDAR and Satellite Imagery
- ❑ Maps and plans
- ❑ GIS
- ❑ Geophysics
- ❑ Software
- ❑ Other

2. What formats might be involved in the project?

Kind of data	Format	Is open format?	Software used to create file

3. What folder structure will you use for your project? (and how does the data fit in there)

4. What naming convention are you going to use? (show us an example based on your expected dataset).

Example: VMF2013_photo_001.jpg (site name, year of the archaeological intervention, kind of data, number and format).

Have questions about this activity? Feel free to share your questions or concerns by email!

Module 3: Retaining, Sharing, and Archiving

Overview 3: After the project has ended

In Module 2 we learned about data management challenges you will confront during your research project. But responsible stewardship of data does not end with publication.

More and more frequently, grant funders and journals impose requirements about how data is handled *after* the completion of the project. In this module, we will discuss **best practices for retaining, sharing, and archiving** your data. Although these concepts are frequently conflated, each requires a distinct set of considerations and action.

By the end of this module, you will be able to:

- Distinguish between retention, sharing, and archiving.
- Locate resources to retain, share, and archive data.
- Identify what questions to ask when preparing your data for sharing and archiving.

Resources 3A: Data Preparation

- **Housecleaning**

Did you get behind on your metadata and procedural documentation? Now's the time to go back and make sure it's **up to date**. Did you end up using some data from one file and another piece from a second? You need to make sure your data set includes **only the most relevant version of all the data** you used in your publication. You're now putting together a data set that you can be proud to send out into the world!

Image credit: Allie Brosh, Hyperbole and a Half

- **Other considerations**

As we discussed in Module 1, there may be additional issues that you need to address before you share your data. Establish the impact of any **privacy concerns or limitations.** This may mean you need to **anonymize your data.**

As an example, see this ICPSR's guidance on respecting confidentiality was written for social science research, many of the concepts are relevant to any data that includes personally identifiable information.

You should also determine **the best format** to use for sharing and archiving your data.

The ICAC-CERCA has recommendations for formats to use, but the gist is that you should aim for **open** formats that are **widely used** and, if possible, **self-documenting.** They should have **as few legal and technical hurdles,** like external dependencies or patent issues, as possible.

This optional reading, Nine simple ways to make it easier to (re)use your data explains further steps to take to prepare your data for sharing - Step 4 covers file formats to better ensure longevity and access to your data.

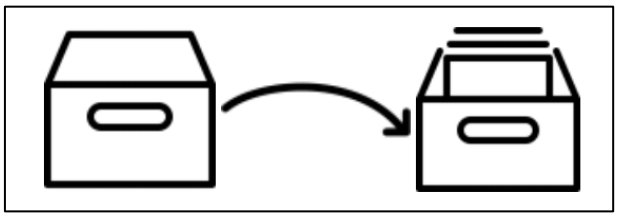
Resources 3B: Retention

- **To keep or not to keep**

In a data management plan, retention is defined as the storing of data after the completion of a research project. Being thoughtful about how and where you store your data is important for a number of reasons:

- You or one of your collaborators may want to use that data again.
- Funders, journals, and sponsoring institutions may all have requirements for whether and how data is kept or destroyed at the end of a project.

Proper retention is an extension of the storage and backup and organizing habits you established during the active data collection phase. After the project is over you must once again determine what to keep, where, and for how long.



Many funders and sponsoring institutions instruct researchers to retain data for a particular length of time. For example, the ICAC-CERCA's own policy on the management of research data stipulates a minimum retention period for research data of ten years after publication or public release of the work of the research.

To do this, you must consider both what you think you may want to re-use, and also the legal and policy requirements that may be attached to the data. For example, data containing personally identifiable information may be required to be stored using a particular protocol and destroyed after a certain amount of time.

For any questions and clarifications, please do not hesitate to reach out to us.

Resources 3C: Sharing

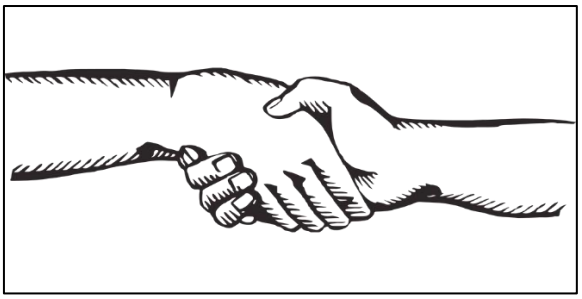
- **Why share?**

There are many benefits to sharing data. First, **sharing improves transparency** by clearly showing the underlying research that supports your findings. Second, sharing data leads to an increase in citations and allows others the opportunity **to use your data for further research.**

In Module 1 you learned how to determine whether there were external **data sharing requirements**, and whether there were **licensing, copyright, or sensitive data** issues you needed to take into account about your decision to share. At this point, you will want to once again consider these requirements and issues. If you decide to share your data, you should attach a license such as the ones developed by Creative Commons.

As for data, it is important to explain how your code can be used or cited by others (considering related restrictions). For the code it is recommended to use specific software licenses, such as MIT or GPL.

Journals often **require** supplemental materials, including research data, to be made available to readers. PLOS, for example, requires authors to provide research data to the public upon publication of the article. The easiest way to do this is by depositing data into a repository. See PLOS's policies for more on this example.



- **Choosing the mechanism for sharing**

The best way to make sure your data is available to others is to publish your data to an appropriate sharing site. Some examples are:

- Whenever possible, deposit your data in a discipline specific repository. Some consolidated disciplinary repositories are: Archaeology Data Service (ADS), Open Context or The Digital Archaeological Record (tDAR).
- Archive your data in our institutional repository: CORA. Research Data Repository (CORA.RDR).

Data can also be published in journals specialized in data (data publications):

- Journal of Open Archaeology Data (JOAD).
- Journal of Open Humanities Data (JOHD).

If you want to start writing your README file, explore ICAC-CERCA'S guidance. Be sure to share your README file in the same location as the data!

Resources 3D: Archiving

- **Didn't we just talk about this?**

Pay close attention to whether your funder requires **sharing, archiving, or both**! Because as it turns out, what a librarian or a funder thinks about when we say 'archive' has a very different meaning than if you were to talk to a computer scientist. Archiving in this sense is about **long-term preservation and accessibility**, and it requires a specialized infrastructure and commitment to ongoing maintenance that many organizations just can't support. Data is actually very fragile!

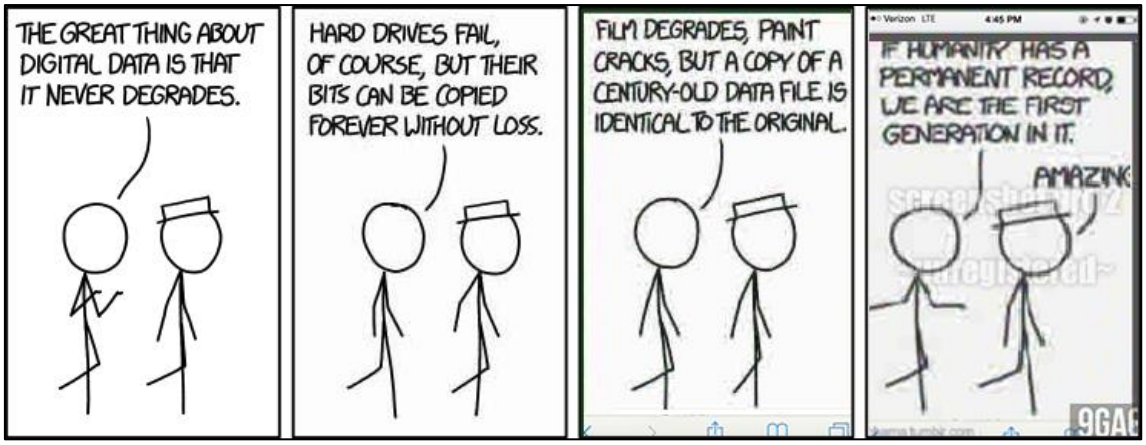


Image credit: xkcd

- **So, what should I archive?**

This may depend on whether your data are **observational** (is it unique?), **experimental** (is it reproducible?), or **computational** (will you archive the outputs, or just the algorithms and initial conditions?). Try to envision who might be able to re-use the data in the short term, and whether it could still be interesting to others ten years from now.

- **Well, then where should I archive?**

Not every repository or community website is equipped to provide long-term preservation. Make sure to investigate how the repository is funded and what preservation actions they take to ensure your data will continue to be found and used. For example, the Archaeology Data Service (ADS) provides a guide on how it actively preserves data.

The ICAC-CERCA participates in the CORA.Research Data Repository, which is the trusted repository specifically for research data from Catalan universities and CERCA centres.

There are other many repositories that are good for both sharing and archiving. Funder-affiliated repositories are usually a good bet, but for those who don't have access to that kind of resource you can always deposit in Zenodo is definitely here for the long-term!

For more options you can browse Re3data.org, the Registry of Research Data Repositories. But be sure to check whether an individual repository's funding and preservation strategies leave you confident they'll be able to take care of your data in the long-term.

Check out this infographic with characteristics and benefits of publishing research data in CORA.Research Data Repository.

Quiz 3

Question 1. Metadata (data about data) is not an important type of documentation.

- ❑ True
- ❑ False

Question 2. Data repositories is an excellent way to publish data.

- ❑ True
- ❑ False

Question 3. GitHub/GitLab and Jupyter notebooks are good tools for documenting code.

- ❑ True
- ❑ False

Question 4. Publishing research data makes research transparent and it facilitates reproducibility.

- ❑ True
- ❑ False

Question 5. Data can be FAIR and not open.

- ❑ True
- ❑ False

Suggested Exercise 3: Data Requirements

1. What data will / could be archived?

- ❑ Raw and processed data
- ❑ Relevance of content to others
- ❑ Ease of reuse of the format by others
- ❑ Data linked to a publication
- ❑ Investigation verification

2. Do you have authority to deposit these data?

- ❑ Yes
- ❑ No
- ❑ I don't know

3. Do you have any restrictions that may affect your data?

- ❑ Contractual obligations
- ❑ Legal obligations: protection of personal data (LOPDGDD, RGPD,)
- ❑ Legal obligations: copyright, intellectual property
- ❑ Ethical restrictions
- ❑ Commercial aspects (p. ex. patentability)
- ❑ No obligations
- ❑ I don't know

4. In which repository would you like to store your data?

- ❑ CORA.Research Data Repository (CORA.RDR)
- ❑ Thematic Research Data Repository
- ❑ Multidisciplinary repository (e.g., Zenodo, Figshare, Dryad)

Module 4: The Big Picture

Overview: Conclusion

Welcome to the last module of the **ICAC-CERCA training on Research Data Management**! In Module 3 we covered the steps required to retain, share, and archive your data. In this module, we will discuss the **research lifecycle and data management plans** in a larger context.

By the end of this module, you will be able to:

- Recognize the research data lifecycle.
- Start writing your own data management plan.
- Learn about support services for data management at the ICAC-CERCA.

Resources 4A: Research Lifecycle

- **Data lifecycle**

Photo credits: Disney

In the previous modules, we've engaged with some specific data management-related issues. Here, we'd like to take a step back and look at the big picture. To wit: ***Data has a lifecycle.*** Choices that you make at one stage of your project will affect how you are able to interact with your data at a later stage. So, when you make decisions, try to have the **whole** lifecycle in mind!

- **What comes around, goes around**

There are many different graphic representations of this lifecycle, but most generally follow the same formula. Take a look at this lifecycle from the Archaeology Data Service. How will it be reflected in your own research process?

Image credit: Archaeology Data Service adaptat per l'ICAC-CERCA.

- **Project Planning:** At this step, you design your research and actively collect data. You also plan for data management and sharing, as well as capture and create metadata.
- **Data collection:** This is the stage in which you do the "dirty work" of working with your data (entering, digitizing, transcribing, checking, validating, cleaning data...). You also anonymize where necessary, describe, and manage and store data.
- **Data analysis:** You begin to interpret data, derive data, and produce research outputs, and write publications. This is also the step at which you begin to prepare data for preservation.
- **Data archiving:** At this step you migrate data to its best format and suitable medium, refine the metadata and documentation you produced, as well as archive data for long-term preservation.
- **Data discovery:** This step is when you focus on how you will distribute and share your data, including controlling access, assigning licenses where appropriate, promoting your data, and providing proper citation.

- **Reuse:** In a cyclical fashion, track your research, start new research, make revisions, examine your findings and those of others, and teach and learn from other researchers.

Resources 4B: Data Management Plans

- **Keeping your funders informed**

Many funding agencies require **Data Management Plans** (DMPs) to be included in all grant proposals. If you've thought about how the lessons, we've covered in this course apply to your own data: congratulations, you have all of the conceptual information you need to write your DMP!

Different funders have different specifications for what should be included in their data management plans, but in general this document **describes your data**, the **metadata standards** you will use, your **storage and backup proce**dures, your plans for **retention**, **sharing**, and **archiving**, and your **policies for re-use**.

Browse the Public DMPs plans created using the CORA. eiNa DMP.

- **eiNaDMP can make it easier**

Figuring out exactly what your funder thinks a DMP should look like can be difficult. Thankfully, the CSUC's eiNa DMP has a number of templates that are tailored to different funder requirements.

When you log in to eiNa DMP you will be directed to the 'My Dashboard' page. From here you can edit, share, download, copy or remove any of your plans. You will also see plans that have been shared with you by others.

If you choose to create a new DMP, you will be brought to a list of DMP templates based on funding agency. Fill in the template, and at the end the website will format your answers into a completed PDF or DOCX.

Check out **ICAC-CERCA's customize a DMP templates**.

- Data Management Plan (DMP) for grant proposals
- Data Management Plan for Horizon Europe
- Data Management Plan (DMP) for the Agencia Estatal de Investigación
- Data Management Plan (DMP) for PHD researchers
- Data Management Plan (DMP) for master's degree Final Project
- Software Management Plan (SMP)

Resources 4C: Next Steps

- **What's next?**

Here are some first steps you can take to start applying what we've covered in the course to your own work.

1. Start writing your data management plan (DMP) and permanently go back to evaluate what is working and what needs to be improved and update the plan accordingly.
2. It is well worth the effort, or even mandatory, to add a README file to each new project. So, the sooner you start writing it, the better.
3. Talk about research data and software with your peers, with your supervisor. Ask others for their experiences, tips and tricks. Maybe discussing research data and software in your team/project meetings can help you and others.
4. Check out the ICAC-CERCA's Open Science website. There are many more resources available to you there and can be a starting point for finding them.
5. Take a look at the infrastructure provided at the Institute for the Research Data Management: Research Portal of Catalonia (CORA.PRC), Research Data Repository (CORA.RDR), eiNa DMP and RECERCAT.

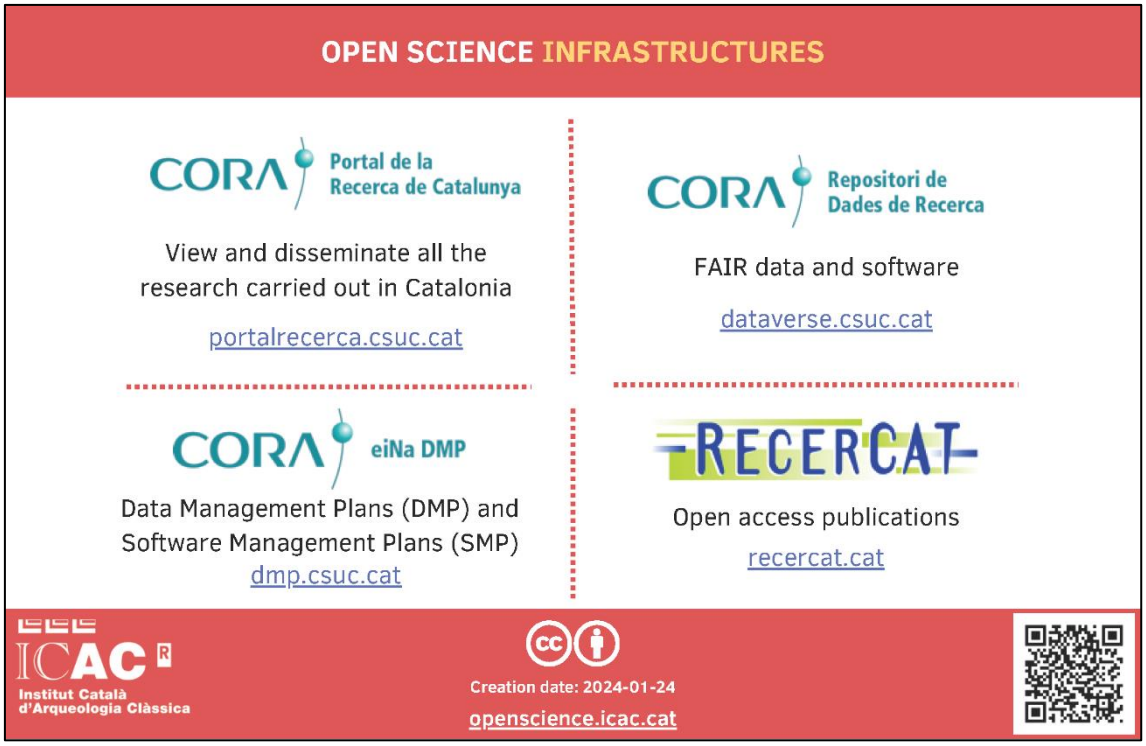


Image credit: Catalan Institute of Classical Archaeology

- **We can give you a hand!**

There are a variety of different resources at the ICAC-CERCA that can assist researchers with their data before, during, and at the end of their project. We would love to help you find them!

Please feel free to contact us if you have any questions or concerns.

Quiz 4

Question 1. What is the management of research data?

- ❑ It refers to organization of metadata standards.
- ❑ It refers to actions such as planning, collecting and organizing research data.
- ❑ It divides the personal data from the study data.

Question 2. Which of these options is research data?

- ❑ Interviews
- ❑ Photos
- ❑ Simulations
- ❑ Software

Question 3. What is research data?

- ❑ Research data is any information of metadata and primary research data.
- ❑ Research data is any information collected or generated for the purpose of analysis, in order to generate or validate scientific claims.
- ❑ Research data is any data, that is generated during a scientific work process.

Question 4. Who profits from research data management?

- ❑ The team members who are involved in the project
- ❑ The research community
- ❑ Society
- ❑ Someone who loves bureaucracy

Question 5. Which activities are included in research data management?

- ❑ Collecting and organizing research data
- ❑ Documenting and describing research data
- ❑ Sharing and controlling access to research data
- ❑ Writing a DMP
- ❑ Storing, backing up and preserving research data

Conclusion

- **Conclusion & Survey**

Our final activity is an easy one: take the **survey on ICAC-CERCA training on Research Data Management**!

https://forms.office.com/e/XTe6twgpmE

We value your feedback and appreciate your interest and enthusiasm for improving your research skills. This survey provides us with the information we need to improve this workshop.

Thank you again for your participation, and here's wishing you success in your future research efforts!

Bibliographical references

ADS [2024]. *Data Management & Sharing Plans*. Archaeology Data Service. https://archaeologydataservice.ac.uk/help-guidance/how-to-prepare-data/data-management-sharing-plans/

ADS [2024]. *Preservation Policy and Procedures.* Archaeology Data Service. https://archaeologydataservice.ac.uk/about/policies/ads-policies-and-procedures/

ADS [2024]. *Sensitive data*. Archaeology Data Service. https://archaeologydataservice.ac.uk/help-guidance/how-to-prepare-data/sensitive-data/

April (2013). *Open formats, what for?* Promoting and Protecting Free Software. https://www.april.org/en/open-formats

Belarte Franco, M. C.; Pastor Quiles, M.; Cuscó Badia, O.; Camañes Villagrasa, M. P. (2023). *Protohistoric Combustion Structures from the Western Mediterranean*, https://doi.org/10.34810/data775, CORA.Repositori de Dades de Recerca, V1; readme.txt

Briney, K. (2013). *Save your thesis (and back it up too).* Data Ab initio. http://dataabinitio.com/?p=341

Canto (2021). *What is metadata (and why does it matter)?* [Vídeo]. YouTube. https://www.youtube.com/watch?v=fZWg0ClQkYQ

CESSDA Training Team (2017 - 2022). *Storage*. CESSDA Data Management Expert Guide. https://dmeg.cessda.eu/Data-Management-Expert-Guide/4.-Store/Storage

Creative Commons (2023). *Choose a License*. Creative Commons. https://chooser-beta.creativecommons.org/

CSUC (2023). *What is CORA.RDR and what does it offer?* Recercat. http://hdl.handle.net/2072/536911

European Commission (2023). *Open Science*. Research and innovation. https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

FGDC (s. d.). *Content Standard for Digital Geospatial Metadata (CSDGM).* Federal Geographic Data Committee. https://www.fgdc.gov/metadata/csdgm-standard

Generalitat de Catalunya. Departament d'Empresa i Coneixement (2019). Implementation of a catalan open science strategy Objectives and Measures. *Pacte Nacional per la Societat del Coneixement*. https://recercaiuniversitats.gencat.cat/web/.content/23_PNSC/document/annex_f_ciencia_oberta_en.pdf

Gil, L. (2022). *Guia per elaborar un "Pla de gestió de dades (DMP)" per l'alumnat de Màster*. Recercat. http://hdl.handle.net/2072/522870

Gil, L.; Estivill, A. (2023). *Data dictionary template*. Recercat. http://hdl.handle.net/2072/534226

Gil, L.; Estivill, A. (2023). *Guidance on a "Data Management Plan (DMP)" for PhD researchers*. Recercat. http://hdl.handle.net/2072/534262

Gil, L.; Estivill, A. (2023). *Guidance on the Data Management Plan (DMP) for grant proposals*. Recercat. http://hdl.handle.net/2072/533371

Gil, L.; Estivill, A. (2023). *Guidance on the Data Management Plan (DMP) for Horizon Europe*. Recercat. http://hdl.handle.net/2072/535959

Gil, L.; Estivill, A. (2023). *Guidance on the Data Management Plan (DMP) for the Agencia Estatal de Investigación*. Recercat. http://hdl.handle.net/2072/536906

Gil, L.; Estivill, A. (2023). *Guidance on the Software Management Plans (SMP)*. Recercat. http://hdl.handle.net/2072/536964

Gil, L.; Estivill, A. (2023). *Open science infrastructures at the Catalan Institute of Classical Archaeology (ICAC)*. Recercat. http://hdl.handle.net/2072/537187

Gil, L.; Estivill, A. (2023). *Readme file for archaeological datasets*. Recercat. http://hdl.handle.net/2072/534226

Goodman, A., Pepe, A., Blocker, A. W., (2014). *Ten Simple Rules for the Care and Feeding of Scientific Data*. PLoS Comput Biol 10(4): e1003542. https://doi.org/10.1371/journal.pcbi.1003542

HES-SO Valais-Wallis (2023). *Research Data Management: Type of data.* Médiathèques HES-SO Valais-Wallais. https://hevs-ch.libguides.com/RDM/Research-Data#s-lg-box-16072466

ICAC (2022). *Protocol per a la gestió de dades de recerca de l'Institut Català d'Arqueologia Clàssica (ICAC)*. Recercat. http://hdl.handle.net/2072/521684

ICAC (2023). *File formats.* Open Science ICAC. https://openscience.icac.cat/file-formats/

ICAC (2023). *File naming*. Open science ICAC. https://openscience.icac.cat/file-naming/

ICAC (2023). *Open Research Tools*. Open Science ICAC. https://openscience.icac.cat/open-research-tools/

ICO (s. d.). *11 Practical ways to keep your IT systems safe and secure*. Information Commissioner's Office. https://ico.org.uk/for-organisations/advice-for-small-organisations/whats-new/blogs/11-practical-ways-to-keep-your-it-systems-safe-and-secure/

ICPSR [2024]. *Data Confidentiality.* Data Management & Curation. https://www.icpsr.umich.edu/web/pages/datamanagement/confidentiality/index.html

Jones, S., & Grootveld, M. (2017). How FAIR are your data? (Version v3). Zenodo. https://doi.org/10.5281/zenodo.5111307

Karoune, E., & Plomp, E. (2022). Removing Barriers to Reproducible Research in Archaeology (Version v5). Zenodo. https://doi.org/10.5281/zenodo.7320029

OHSU Library (2024). *Research Data and Reproducibility: Organizing Data*. The Oregon Health & Science University (OHSU) Library. https://libguides.ohsu.edu/research-data-services/organizing-data

Ollé, C. (2023). *El Estado español dispone finalmente de su estrategia en ciencia abierta (2023-2027)*. Open Science Spain. https://opensciencespain.org/es/el-estado-espanol-dispone-finalmente-de-su-estrategia-en-ciencia-abierta-2023-2027/

OSI (2023). *GNU General Public Licenses*. Open Source Initiative. https://opensource.org/licenses-old/gpl-license-html/

OSI (2023). *The MIT license*. Open Source Initiative https://opensource.org/license/mit/

Piwowar, H.A.; Vision, T. J. (2013). *Data reuse and the open data citation advantage*. PeerJ 1:e175 https://doi.org/10.7717/peerj.175

PLOS ONE (2019). *Data Availability*. PLOS ONE. https://journals.plos.org/plosone/s/data-availability

UAB [2024]. *Informació sobre la protecció de dades.* Propietat intel·lectual i accés obert. https://webs.uab.cat/dretsautor/category/dades-personals/

University College London (2021). *Defining open science & scholarship*. UCL Library Services. https://www.ucl.ac.uk/library/open-science-research-support/open-science/defining-open-science-scholarship

USGS (s. d.). *Describe (Metadata/Documentation).* U.S. Geological Survey. https://www.usgs.gov/data-management/describe-metadatadocumentation

Valentine, T.; DataONE (2011). "Best Practice: Define roles and assign responsibilities for data management". Data Management Skillbuilding Hub. https://dataoneorg.github.io/Education/bestpractices/define-roles-and

White, E., Baldridge, E., Brym, Z., Locey, K. J., McGlinn, D. J., & Supp, S. R. (2013). Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution*, *6*(2). https://doi.org/10.4033/iee.2013.6b.6.f

ICAC
Institut Català d'Arqueologia Clàssica

Consortium integrated by:

Generalitat de Catalunya

UNIVERSITAT ROVIRA i VIRGILI

CERCA Center:

CERCA Centres de Recerca de Catalunya

Accredited center:

ICREA

HR EXCELLENCE IN RESEARCH

ceics campus of international excellence southern catalonia

Documentation Centre, Library and Open Science

Catalan Institute of Classical Archaeology (ICAC-CERCA)

Plaça d'en Rovellat, s/n - 43003 Tarragona

Phone number: 977 24 91 33 / https://openscience.icac.cat / lgil@icac.cat