Data Management Plan

Emma Lazzeri
ORCID iD: https://orcid.org/0000-0003-0506-046X

Istituto di Scienza e Tecnologie dell'Informazione
Consiglio Nazionale delle Ricerche

Corso di formazione
Praticare l'Open Science nelle scienze della Terra e dell'ambiente
03 Dicembre 2020 | Modulo 4

A summary from the previous modules: How to practice OS?

What are your options?

Your community is already part of a RI

Get engaged, discover and use the RI services, developed tools, infrastructures, repositories, standards, best practices, etc

Your community could be part of a RI

Get engaged with your reference RI, get your community together and liaise with other communities to understand how to apply existing best practices in your specific domain

Your Institution could set up a data strategy

At the level of your Institution, several actions can be taken. You can learn from others' experience: set up a working group, adopt a data policy, design and develop a data registry/repository, give access to your contents

None of the above

You can still apply the Open Science generic best practices that you will learn in this course to your projects and research workflow

Do not reinvent the whee

Build on others' experienc

Liase with your and other communiti

OpenAIRE

European Commission

EOSC-Pillar
Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

ICDI

Ask for Support

OpenAIRE

Get support on general open science practices

noad-it@openaire.eu
www.openaire.eu

EPOS ENVRIFAIR

Get support from your specific domain Research Infrastructure

https://www.epos-ip.org/
https://envri.eu/home-envri-fair/

Competence Center

Be supported at national level from a team of experts in Open Science and EOSC related fields

Coming soon:
https://www.icdi.it/it/attivita/tf-cc

Your Institution

In case your institution has a structured open science strategy and dedicated support!

Data Managemet Plan

OpenAIRE

European Commission

EOSC-Pillar
Coordination and Harmonisation of National &Thematic initiatives to support EOSC

ICDI

@openaire_eu

Data Management Plan

- It is a plan to manage the data within your project (for every kind of project)
- It is a living document (so it needs updates)
- Since it is a plan, you shoud think it from the very beginning (ideally it is included in your project planning)
- Clear rules, less mistakes from the beginning
- It is a structured way to think of your data

What is normally a DMP about?

- Identify the data you are working with in your project
- Decide the strategy to organise your data and the standards you will use
- Daily data management
- What is your plan for sharing your data?
- Will you have issues sharing your data?
- Will you need more resources/budget than expected?

The costs

- The cost related to open access to research data are **eligible as part of the Horizon 2020 grant** - refer to your Grant Agreement conditions.
- For further information, see our guide on cost in research data management.

Data Management Plan Checklist

1. Administrative data

- Basic information about your project: title, acronym, ID, reference numbers
- An abstract of your project highlighting the scope of data collection/creation
- Details related to procedures and policies

OpenAIRE

European Commission

EOSC-Pillar
Coordination and Harmonisation of National &Thematic initiatives to support EOSC

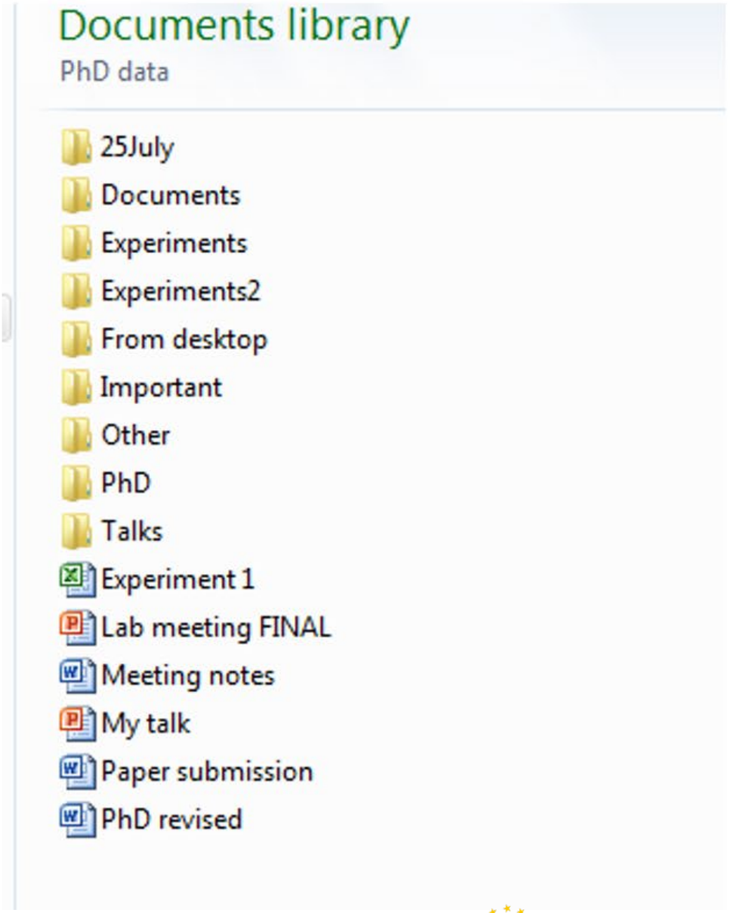
ICDI

2. Data Collection

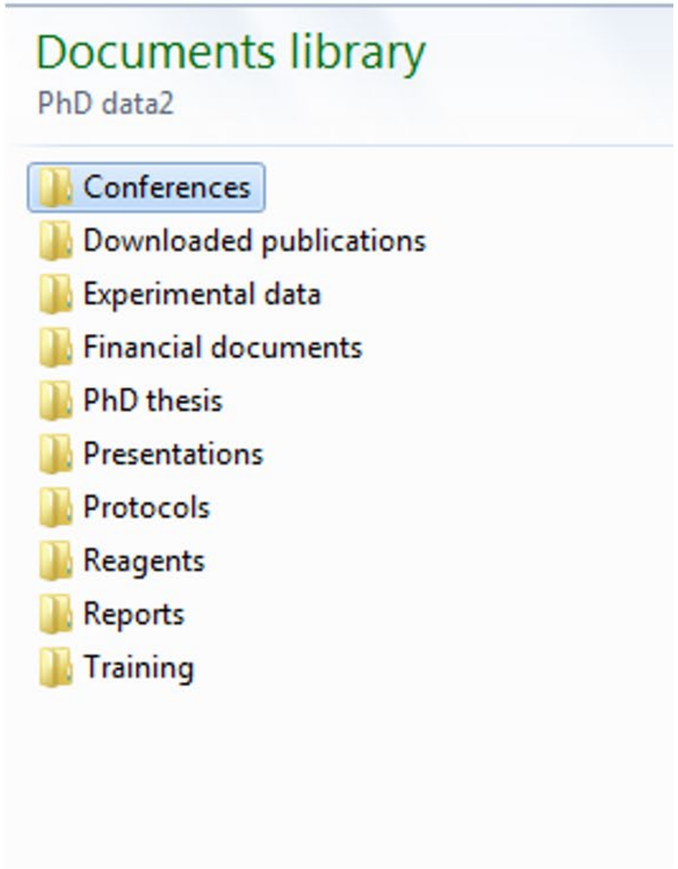
- Are you using existing data?
- What are the standards or methodologies that you will use to collect your data?
- Data Formats and Software
- How will you structure and name your files and folders?

What is your strategy?

Example·A¶



Example·B¶



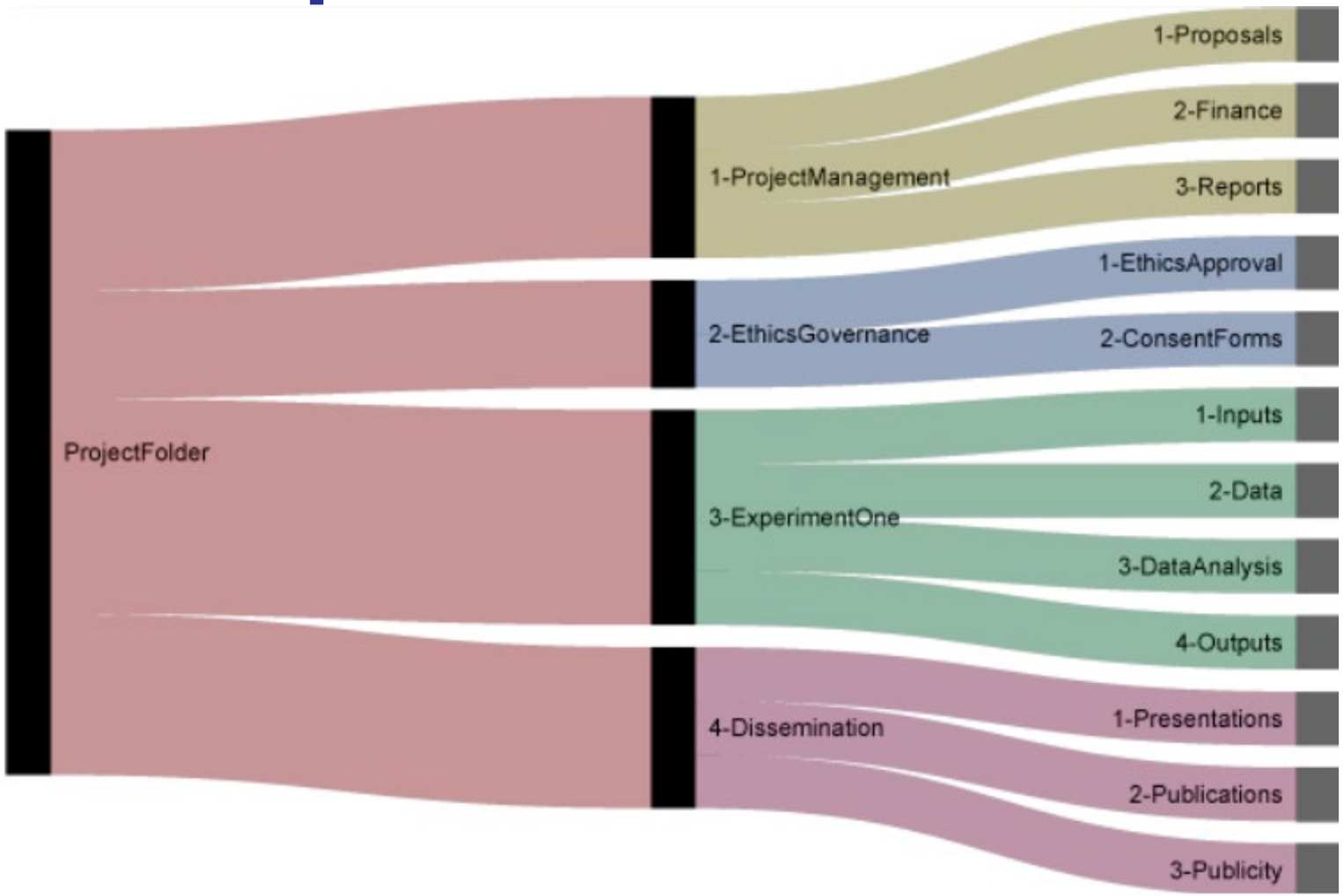
OpenAIRE

European Commission

EOSC-Pillar
Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

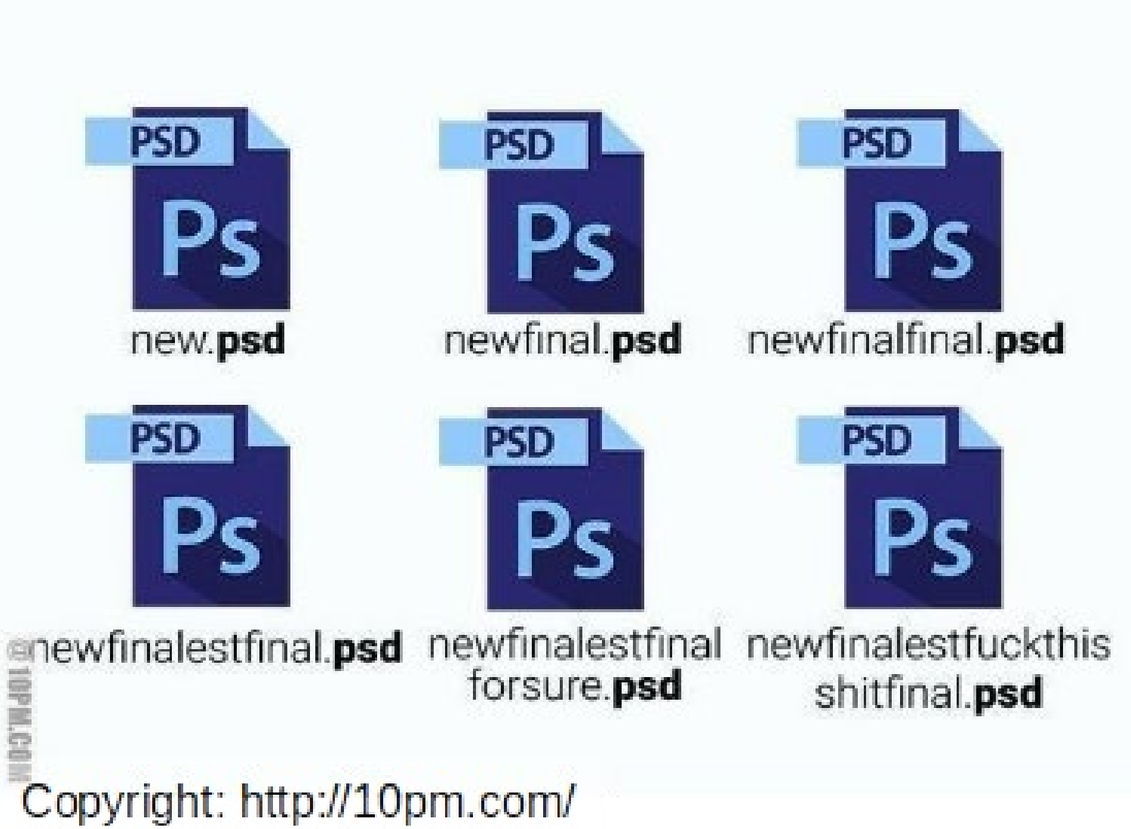
ICDI

A good example



http://nikola.me/folder_structure.html

File naming


new.psd
newfinal.psd
newfinalfinal.psd
newfinalestfinal.psd
newfinalestfinal forsure.psd
newfinalestfuckthis shitfinal.psd
Copyright: http://10pm.com/

A good example: TILS document naming convention

http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf

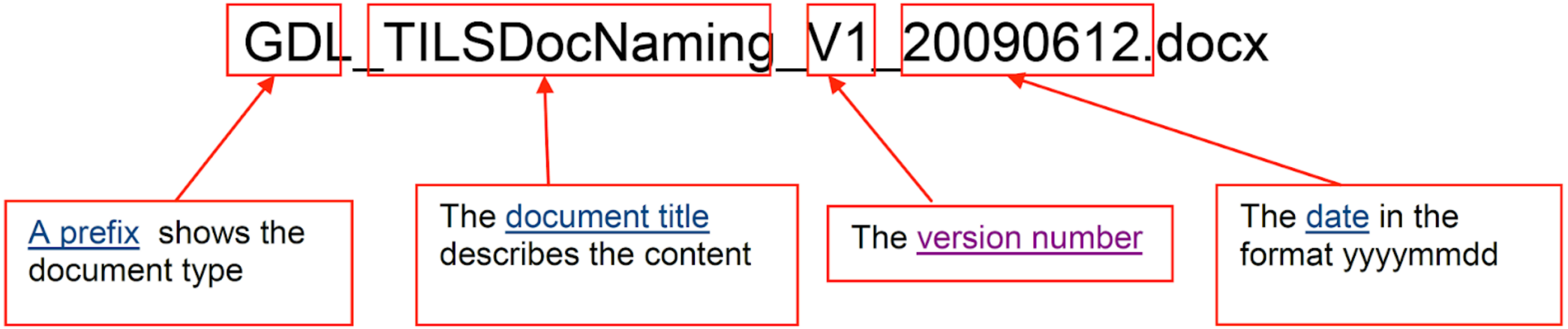
3. Version

(upper case, max 4 chars, optional)

For documents that will continue in various versions use V followed by the version number. Use an underscore to indicate a decimal point if necessary.
Eg. PMF_PRP_ZenMonkeyProject_V2_20090607.docx

New versions should not be created for each iteration of the document, but rather at significant changes or when it has been reviewed or changed by another author.

Document naming for the TILS Division should follow this convention:



Prefix	Meaning
AGD	Agenda
AGR	Agreement
GDL	Guideline
MEM	Memorandum
MIN	Minutes and Notes
PRE	Presentation
PRO	Procedure
PRP	Proposal
REP	Report
TEM	Template

2. Document title/ Description

(mixed case, max 30 chars, no spaces)

- Describes the purpose or "business" of the document. Acronyms, capitalisations, abbreviations can be used, keep in mind that descriptions should be **meaningful** to anyone reading the file name.
- In the case of project documentation use the **project name** or its usual abbreviation
- If possible Departmental Branch and/or Section should be integrated into this field to indicate origin / ownership of document.
- Use only alpha-numeric characters, plus the hyphen and underscore.
- **Do not use spaces.**

3. Metadata and supporting documentation

- Which documentation and metadata will support/describe your data?
- How will you create the supporting documentation and metadata?
- Which metadata standards will you use?

Soo many ways to describe your data

How to create useful README files: https://data.research.cornell.edu/content/readme

Cornell

AUTHOR_DATASET_ReadmeTemplate.txt

```
This DATSETNAMEreadme.txt file was generated on [YYYYMMDD] by [Name]


-------------------
GENERAL INFORMATION
-------------------


1. Title of Dataset


2. Author Information


  Principal Investigator Contact Information
       Name:
          Institution:
          Address:
          Email:
```

README files template:
https://cornell.app.box.com/v/ReadmeTemplate

OpenAIRE

European Commission

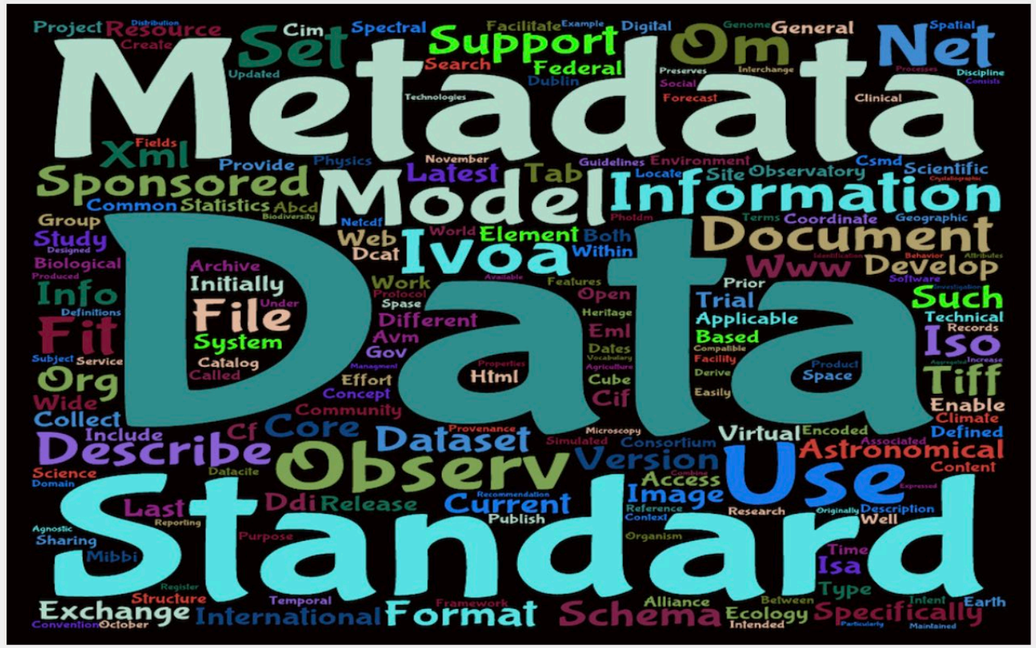
EOSC-Pillar
Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

ICDI

Discipline specific standards

FAIRsharing.org
standards, databases, policies

https://fairsharing.org/standards/



Metadata Standards Directory Working Group

The RDA Metadata Standards Directory Working Group is supported by individuals and organizations involved in the development, implementation, and use of metadata for scientific data. The overriding goal is to develop a collaborative, open directory of metadata standards applicable to scientific data can help address infrastructure challenges.

http://rd-alliance.github.io/metadata-directory/

OpenAIRE
European Commission
EOSC-Pillar
Coordination and Harmonisation of National &Thematic Initiatives to support EOSC
ICDI

4. Legal and ethical aspects

- Did you asked for an informed consensus to share the data and preserve them?
- How will you protect personal data?
- How about data licencing?

5. Data storage and backup

- Do you have enough space to store your data or should you include costs for additional services? Will the services be reliable/trusted?
- How will you share your storage/backup with your collaborators?
- Will you use cloud solutions?
- Will you back your data up? How?

OpenAIRE

European Commission

EOSC-Pillar
Coordination and Harmonisation of National &Thematic initiatives to support EOSC

ICDI

Do not leave it all to Google

Google services Terms of Use:

When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. The rights you grant in this license are for the limited purpose of operating, promoting, and improving our Services, and to develop new ones. This license continues even if you stop using our Services (for example, for a business listing you have added to

https://policies.google.com/terms?hl=en

You have better alternatives…

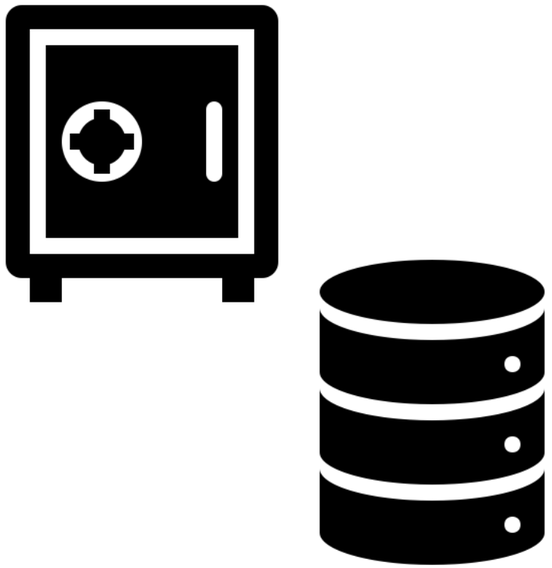
https://www.switch.ch/services/drive/

https://www.surf.nl/en

Your institution probably provides better alternatives:
Ask your IT for support

6. Select and preserve

- Which data shall be preserved or destroyed due to contractual, legal or administrative reason?
- What are the envisaged uses of your data for research purposes?
- Which data shall be preserved and potentially shared?
- What is your long term preservation strategy?
- Did you consider the effort and costs to prepare your data for sharing and preservation?



OpenAIRE

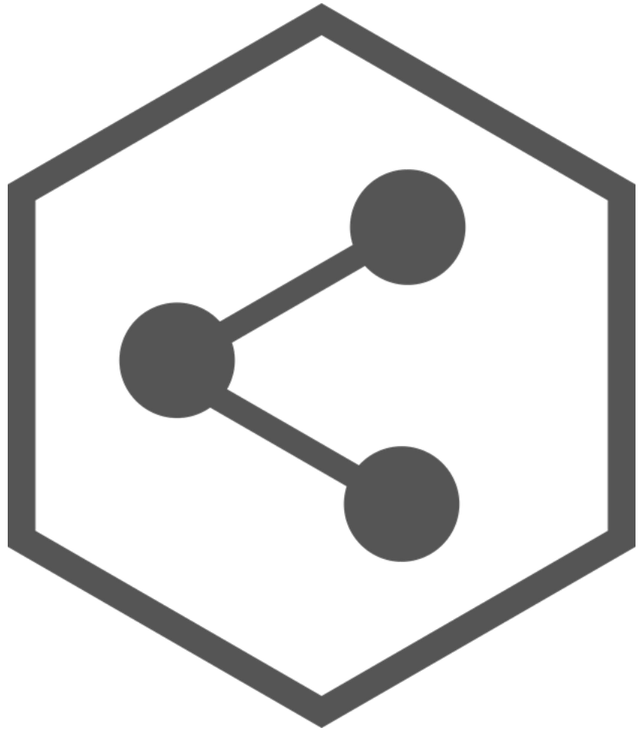
European Commission

EOSC-Pillar
Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

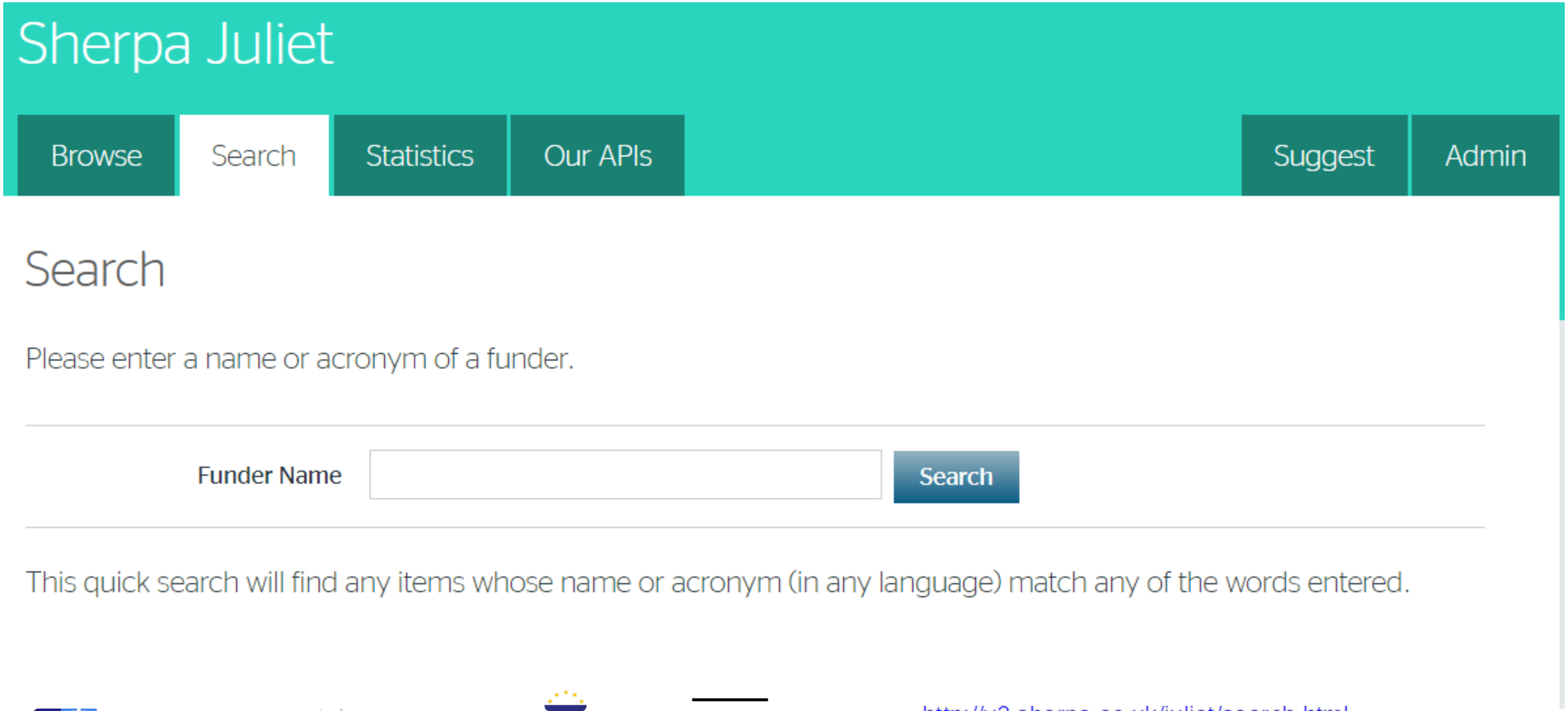
ICDI

7. Data sharing

- Who will you share your data with? Under which conditioons?
- When will you share your data?
- Will you need to apply any access restriction?
- Which actions do you foresee to avoid or reduce access restrictions?
- How will your potential users find your data?



Funder policies on Data sharing



http://v2.sherpa.ac.uk/juliet/search.html

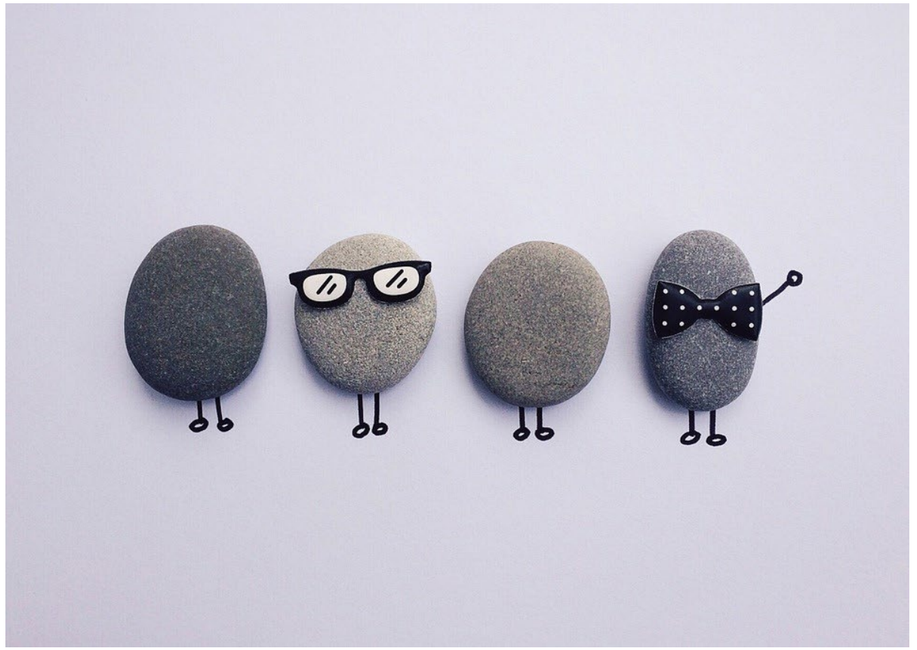
Software

+

https://guides.github.com/activities/citable-code/

8. Responsabilities and resources

- Who is responsible to implement and revise the DMP?
- How will you share responsabilities among partners in collaborative projects?
- Which resources will you need to implement your DMP?
- Will you need any specific external experties or tools?



European Commission and H2020

A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a **DMP should include information** on:

- the handling of research data during & after the end of the project
- what data will be collected, processed and/or generated
- which methodology & standards will be applied
- whether data will be shared/made open access and
- how data will be curated & preserved (including after the end of the project).
- A DMP is required for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. However, projects that opt out are still encouraged to submit a DMP on a voluntary basis.

OpenAIRE

European Commission

EOSC-Pillar
Coordination and Harmonisation of National &Thematic initiatives to support EOSC

ICDI

There are so many tools and resources for DMP out there

Let’s go through some of them

Core Requirements

CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. Data description and collection or re-use of existing data

a. How will new data be collected or produced and/or how will existing data be re-used?

b. What data (for example the kinds, formats, and volumes) will be collected or produced?

2. Documentation and data quality

a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?

b. What data quality control measures will be used?

3. Storage and backup during the research process

a. How will data and metadata be stored and backed up during the research process?

b. How will data security and protection of sensitive data be taken care of during the research?

4. Legal and ethical requirements, codes of conduct

a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

c. How will possible ethical issues be taken into account, and codes of conduct followed?

5. Data sharing and long-term preservation

a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

c. What methods or software tools will be needed to access and use the data?

d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

6. Data management responsibilities and resources

a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

CESSDA DMP Expert guide

PLAN

Overview

Title of the project

Date of this plan

Description of the project
- What is the nature of the project?
- What is the research question?
- What is the project time line?

Origin of Data
- What kind of data will be used during the project?
- If you are reusing existing data: What is the scope, volume and format? How are different data sources integrated?
- If you are collecting new data can you clarify why this is necessary?

Principal researchers
- Who are the main researchers involved?
- What are their contact details?

Collaborating researchers (if applicable)
- What are their contact details and their roles in the project?

Funder (if applicable)
- If funding is granted, what is the reference number of the funding granted?

Data producer
- Which organisation has the administrative responsibility for the data?

Project data contact
- Who can be contacted about the project after it has finished?

Data owner(s)
- Which organisation(s) own(s) the data?
- If several organisations are involved, which organisation owns what data?

Roles
- Who is responsible for updating the DMP and making sure that it's followed?
- Do project participants have any specific roles?
- What is the project time line?

Costs
- Are there costs you need to consider to buy specific software or hardware?
- Are there costs you need to consider for storage and backup?
- Are potential expenses for (preparing the data for) archiving covered?

ORGANISE & DOCUMENT

Organising and documenting your data

Data collection
- How will the data be collected?
- Is specific software or hardware or staff required?
- Who will be responsible for the data collection?
- During which period will the data be collected?
- Where will the data be collected?

Data organisation
- How will you organise your data?
- Will the data be organised in simple files or more complex databases?
- How will the data quality during the project be ensured?
- If data consists of many different file types (e.g. videos, text, photos), is it possible to structure the data in a logical way?

Data type and size
- What type(s) of data will be collected?
- What is the scope, quantity and format of the material?
- After the project: What is the total amount of data collected (in MB/GB)?

File format
- In what format will your data be?
- Does the format change from the original to the processed/final data?
- Will your (final) data be available in an open format?

Folder structure and names
- How will you structure and name your folders?

File structure and names
- How will you structure and name your files?

Documentation
- What documentation will be created during the different phases of the project?
- How will the documentation be structured?

Metadata
- What metadata will be provided with the collected/ generated/ reused data?
- How will metadata for each object be created?
- Is there any program that can be used to document the data?
- Can metadata be added directly into the files or will the metadata be produced in another program or document?

Metadata standard (if applicable)
- What metadata standard(s) will you use?

https://www.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.2.pdf

Adapt your Data Management Plan

A list of Data Management Questions based on the Expert Tour Guide on Data Management

DCC guides

DCC because good research needs good data

Home | Digital curation | About us | News | Events | Resources | Training | Projects

Home > Resources > How Guides > How Develop Rdm Services

In this section

How to Develop RDM Services - a guide for HEIs

https://www.dcc.ac.uk/guidance/how-guides

https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep

Establishing criteria for selection decisions

You should establish criteria to guide selection decisions. The DCC's How to Select and Appraise Research Data for Curation[56] proposes seven criteria as outlined below:

1. **Relevance to mission**: the resource content fits any priorities stated in the institution's mission, or funding body policy including any legal requirement to retain the data beyond its immediate use.
2. **Scientific or historical value**: is the data scientifically, socially, or culturally significant? Assessing this involves inferring anticipated future use, from evidence of current research and educational value.
3. **Uniqueness**: the extent to which the resource is the only or most complete source of the information that can be derived from it, and whether it is at risk of loss if not accepted, or may be preserved elsewhere.
4. **Potential for redistribution**: the reliability, integrity, and usability of the data files may be determined; these are received in formats that meet designated technical criteria; and Intellectual Property or human subjects issues are addressed.
5. **Non-replicability**: it would not be feasible to replicate the data/resource or doing so would not be financially viable.
6. **Economic case**: costs may be estimated for managing and preserving the resource, and are justifiable when assessed against evidence of potential future benefits; funding has been secured where appropriate.
7. **Full documentation**: the information necessary to facilitate future discovery, access, and reuse is comprehensive and correct; including metadata on the resource's provenance and the context of its creation

Citing data

Citing data is important in order to:

- Give the data producer appropriate credit
- Allow easier access to the data for repurposing or reuse
- Enable readers to verify your results

Citation Elements

A dataset should be cited formally in an article's reference list, not just informally in the text. Many data repositories and publishers provide explicit instructions for citing their contents. If no citation information is provided, you can still construct a citation following generally agreed-upon guidelines from sources such as the Force 11 Joint Declaration of Data Citation Principles and the current DataCite Metadata Schema.

Core elements

- There are 5 core elements usually included in a dataset citation, with additional elements added as appropriate.
  - **Creator(s)** – may be individuals or organizations
  - **Title**
  - **Publication year** when the dataset was released (may be different from the Access date)
  - **Publisher** – the data center, archive, or repository
  - **Identifier** – a unique public identifier (e.g., an ARK or DOI)
- Creator names in non-Roman scripts should be transliterated using the ALA-LC Romanization Tables.

Common additional elements

- Although the core elements are sufficient in the simplest case – citation to the entirety of a static dataset – additional elements may be needed if you wish to cite a dynamic dataset or a subset of a larger dataset.
  - **Version** of the dataset analyzed in the citing paper
  - **Access date** when the data was accessed for analysis in the citing paper
  - **Subset** of the dataset analyzed (e.g., a range of dates or record numbers, a list of variables)
  - **Verifier** that the dataset or subset accessed by a reader is identical to the one analyzed by the author (e.g., a Checksum)
  - **Location** of the dataset on the internet, needed if the identifier is not "actionable" (convertable to a web address)

Example citations

- Kumar, Sujai (2012): 20 Nematode Proteomes. figshare. https://doi.org/10.6084/m9.figshare.96035.v2 (Accessed 2016-09-06).
- Morran LT, Parrish II RC, Gelarden IA, Lively CM (2012) Data from: Temporal dynamics of outcrossing and host mortality rates in host-pathogen experimental coevolution. Dryad Digital Repository. https://doi.org/10.5061/dryad.c3gh6
- Donna Strahan. "08-B-1 from Jordan/Petra Great Temple/Upper Temenos/Trench 94/Locus 41". (2009) In Petra Great Temple Excavations. Martha Sharp Joukowsky (Ed.) Releases: 2009-10-26. Open Context. https://opencontext.org/subjects/30C3F340-5D14-497A-B9D0-7A0DA2C019F1 ARK (Archive): http://n2t.net/ark:/28722/k2125xk7p
- OECD (2008), Social Expenditures aggregates, OECD Social Expenditure Statistics (database). https://doi.org/10.1787/000530172303 (Accessed on 2008-12-02).
- Denhard, Michael (2009): dphase_mpeps: MicroPEPS LAF-Ensemble run by DWD for the MAP D-PHASE project. World Data Center for Climate. https://doi.org/10.1594/WDCC/dphase_mpeps
- Manoug, J L (1882): Useful data on the rise of the Nile. Alexandria : Printing-Office V Penasson. http://n2t.net/ark:/13960/t44q88124

https://dmptool.org/general_guidance#metadata-data-documentation

Useful Links

- Zenodo - CERN-OpenAIRE OA repository - catch all

www.zenodo.org

- Choose a license - Creative Commons

https://creativecommons.org/choose/?lang=en

https://chooser-beta.creativecommons.org/

- DMP examples by subject - LIBER

https://libereurope.eu/dmpcatalogue/

- Tools to create your DMP

https://www.openaire.eu/argos/

https://dmponline.dcc.ac.uk/

- Re3Data

https://www.re3data.org/

- Metadata standard Directory - Research Data Alliance

https://rd-alliance.github.io/metadata-directory/

OpenAIRE

European Commission

EOSC-Pillar

Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

ICDI

Thank you!

Emma Lazzeri

emma.lazzeri@isti.cnr.it

This work was partially supported by European Union's Horizon 2020 under projects grant Agreement numbers 857650, and 777541 , and by ICDI (Italian Computing and Data Infrastructure)