Open Science and Data Management Plan

Module 3.1 - Horizon Europe and EOSC
15/05/2023, Area Science Park

Gina Pavone, CNR-ISTI
ORCID 0000-0003-0087-2151

Today's agenda

18/05/2023 AREA SCIENCE PARK

Leftovers

The clear distinction between Open and FAIR

DMP and FAIR

Where to find information; FAQs; FAIR principles quiz

Hands-on 1

Revisions of existing DMPs

Hands-on 2

Sketching the main contents of a DMP

Interaction

Go to:
https://www.menti.com

Voting code: 1728 5770

https://www.menti.com/alsq6nioemuk

Leftovers…

Nel maggio 2017 un gruppo di lavoro informale sui dati della ricerca (costituito da Politecnico di Milano, Università di Milano, Università di Torino, Università di Trento, Università Ca' Foscari Venezia) ha redatto una checklist con una griglia in lingua italiana per l'elaborazione di un Data Management Plan.

DATA MANAGEMENT PLAN CHECKLIST /Griglia per il piano di gestione dei dati

ADMINISTRATIVE PLAN DETAILS	Informazioni generali sul progetto di ricerca
Project Name	*Inserire il nome del progetto*
Acronimo	*Inserire l'acronimo del progetto, se applicabile*
Grant Reference Number	*Inserire il riferimento alla call (es: call Horizon2020 ID:……) e al numero di identificazione del progetto presentato, se disponibile*
Persistent Identifier	*handle o DOI del DMP, ricavabile dopo l'inserimento nel repository*
Funder	*Inserire il nome del finanziatore/dei finanziatori* *Es: European Commission (H2020)*
Principal Investigator/Researcher	*Inserire il nome del ricercatore autore del documento* *Es: Laura Rossi*
Principal Researcher ID ORCID	*Inserire l'identificativo ORCID del ricercatore* *Es: 0000-0003-4170-6345*

http://wikimedia.sp.unipi.it/images/Grigliapianodigestionedatiricerca.pdf

Exercise on Zenodo sandbox

1.Create an account on Zenodo sandbox: https://sandbox.zenodo.org/

2. Prepare a file to be uploaded there: a presentation, a dataset...whatever

Did you use the Orcid for the author(s)?

Did you notice that you can use communities to enhance findability?

Did you fill in the field “subject”?

- Look at this example on the use of subjects

Did you used keywords?

- Look at how they used keywords to express a taxonomy in this case

What field would you use if you are uploading a contribution to a conference?

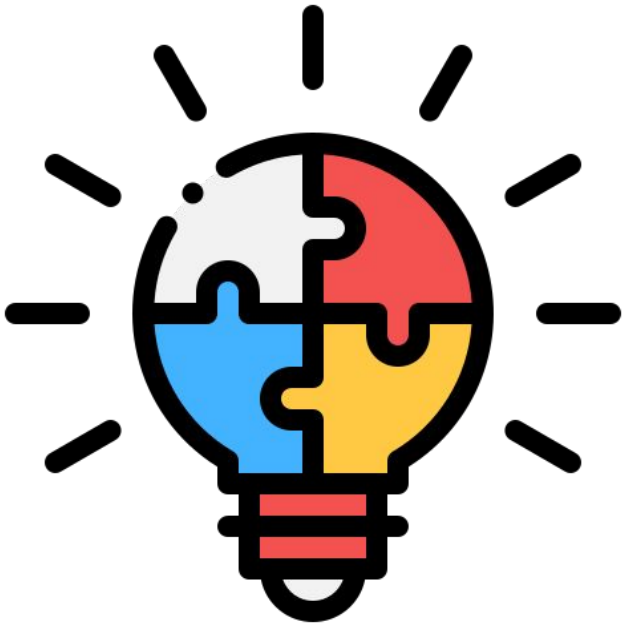
Did you linked to other resources (ie. some sort of supplementary material or whatever)?

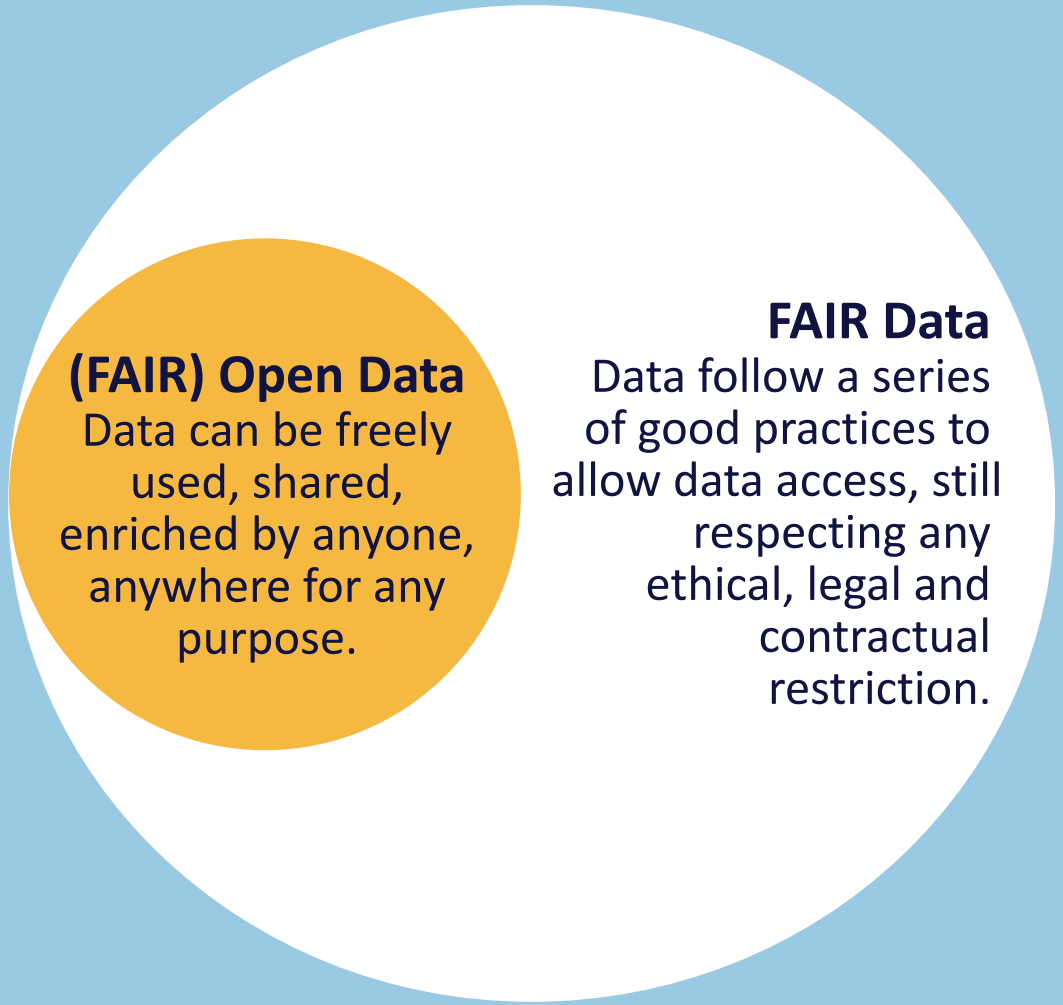
What if you want to add editors (or other contributors) to your upload?

Look at this example of GitHub+Zenodo for software

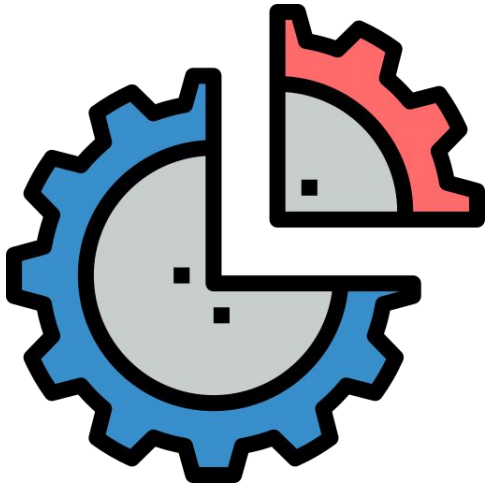
Zenodo Search guide

Open Data
and
FAIR Data
are different
concepts




(FAIR) Open Data
Data can be freely used, shared, enriched by anyone, anywhere for any purpose.
FAIR Data
Data follow a series of good practices to allow data access, still respecting any ethical, legal and contractual restriction.

Why do we need a distinction?



Research data could:

- Contain personal information (privacy e GDPR)
- Fall under copyright (in the case of a database with creative structure)
- Fall under the Sui Generis right (database obtained thanks to a substantial investment)
- Be protected by patent or industrial secret

Data sharing needs to respect the specific law.
Data needs to be protected against non authorised access.

Data are not yours!

Cartoon credit: Foster project

Data are **not** intellectual work, they are facts and information.

Copyright protection covers expressions and not ideas, procedures, operating methods or mathematical concepts as such.

Protection is on databases and not on data. Data are protected only and especially when they are collected and organized in a database.

The sui generis property right (only in Europe) covers not only the reproduction and dissemination of the database, but also the extraction and reuse of substantial parts of the database.

A valid resource (in Italian)

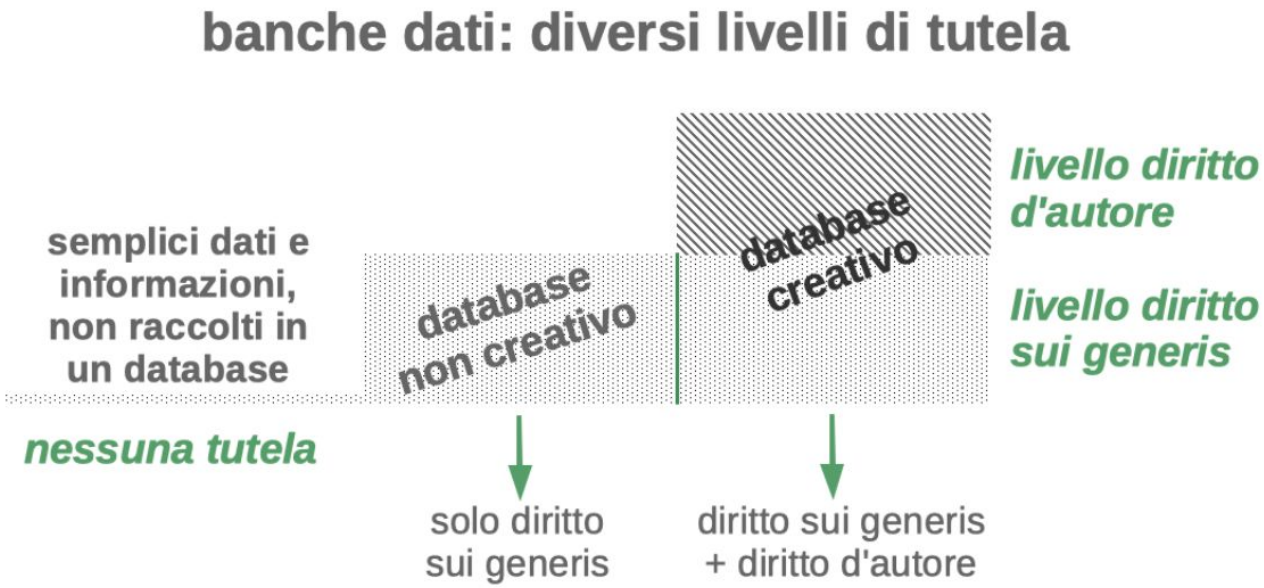


Figura 1: I due diversi livelli di tutela per le banche dati basati sul requisito del carattere creativo

Aliprandi, Simone. (2022). Aspetti legali degli open data: la guida definitiva (1.0 (maggio 2022)). Zenodo. https://doi.org/10.5281/zenodo.6575822

Data and law protection

- **Raw data** are not protected by copyright
- **Database** is defined as a collection of independent works, data or other materials arranged in a systematic or methodical way
- **Copyright** protects the structure, selection or arrangement of the database contents, not the data
- **Sui generis database right**: protects the substantial effort in obtaining data (not creating). Note: the right owner is often the institution.

Data to be handled with great care:

- **Personal data**: any information about an identified or identifiable natural person (directly or indirectly)
- Personal **sensitive data** (i.e. revealing racial or ethnic origin, political views, religious or philosophical beliefs, membership of a trade union, genetic data, biometric data, data about health or someone's sexual behavior or sexual orientation)
- Data protected by **IPR** (Intellectual Property Rights) agreements
- **Confidential data** (i.e. commercial agreements)

This means that access to the data must be managed and restricted.

They still can be FAIR

Security

Measures to keep data safe

- Data relating to people or containing commercial information requires extra care to be taken including additional security measures such as **encryption, anonymisation and storing data in a more secure location.**
- Consider anonymization softwares and pseudonymization techniques.
- Personal data is any information that relates to an identifiable, living individual and requires careful handling under the **GDPR legislation**.
- Data management should be aligned with **ethics applications,** and care taken to secure your research data from loss or theft.

FAIR RDM for researchers

FINDABILITY

- Standardized practices related to organizing files (e.g., standardized file naming conventions)
- uploaded to a repository that assigns a persistent identifier (e.g., DOIs, RRIDs, etc.) and describes data sets with standardized metadata.
- high quality metadata

ACCESSIBILITY

- Data is available through a clearly defined process for members of the research team.
- When data and other materials are made available to others, there should be a clear path to gaining access.

INTEROPERABILITY

Data should be structured in a standard way so it can be easily combined with other similarly structured data sets. For example describing and organizing data (e.g., applying appropriate metadata, maintaining data dictionaries) and saving files in open or nonproprietary file formats.

REUSABILITY

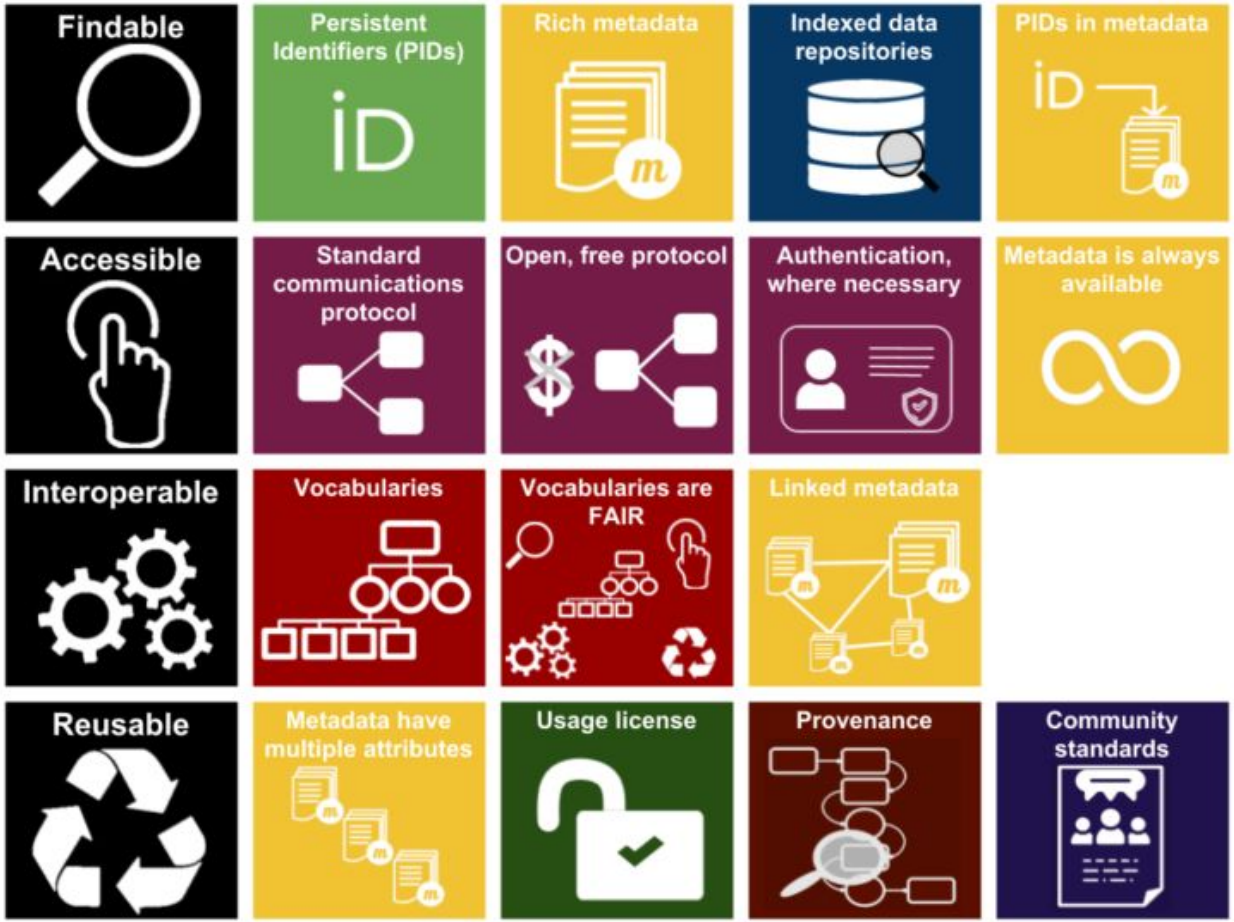
Data should be saved, organized, and described with its future (re)use in mind.

A future user may be a member of the research team coming back later on the same data, or another researcher with other research purposes.

Freely adapted from Borghi, J., & Van Gulick, A. (2022). Promoting Open Science Through Research Data Management. *Harvard Data Science Review*, *4*(3). https://doi.org/10.1162/99608f92.9497f68e

FAIR must haves

Resources to support FAIR data



Source: [ARDC FAIR data training]

Exercice 1: take the quiz!

Go to the following link and answer the questions!

https://docs.google.com/forms/d/e/1FAIpQLSe6Srqsx77yPibAOB0W0CMbDBAPlWcI8oRTmG52X1tJd4vF-g/viewform

TIME: 20 minutes

DMP –
Where to find information?
Who to ask?

1. Funding Agency/Institutional policies and Guidelines

Carefully check the requirements of your grant and of your institution.

They often have specific instructions and templates for creating a DMP tailored to their funding program or institutional policies.

2. Research office/Grant support office/librarians

Reach out to research support staff at your institution.

They can provide examples of already written DMPs, offer templates, or answer questions regarding data management.

3. Ethics committee or legal departiment

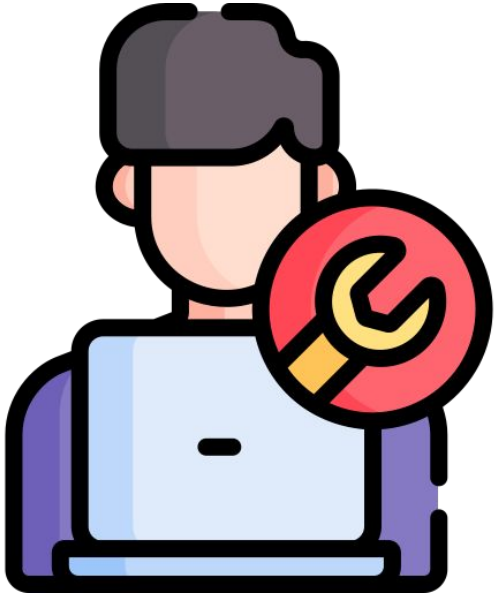
You can refer to specific deliverables or data management and data protection policies.

Are there legal aspects to consider (non-disclosure agreements, patent possibilities, etc.)?

You can ask for support from the legal department, e.g. the data protection office

4. IT staff

Ask the IT staff, e.g. to get info on the institution's cloud space and the respective back-up and disaster recovery strategies; in case add your own measures.



5. Check the policy of the chosen repository

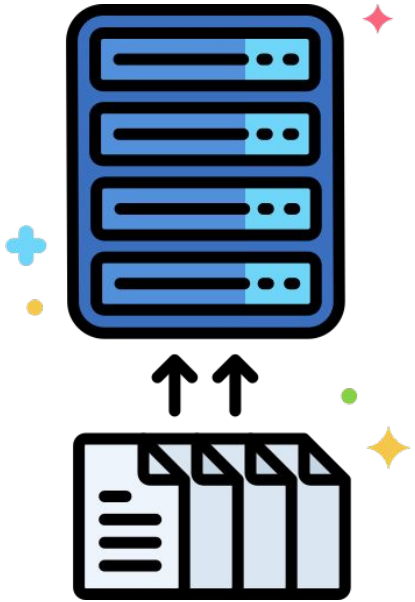
How long is storage guaranteed?

Are the results indexed somewhere?

What kind of metadata schema do they use?

Do you need to complement them with something more specific?

Etc.



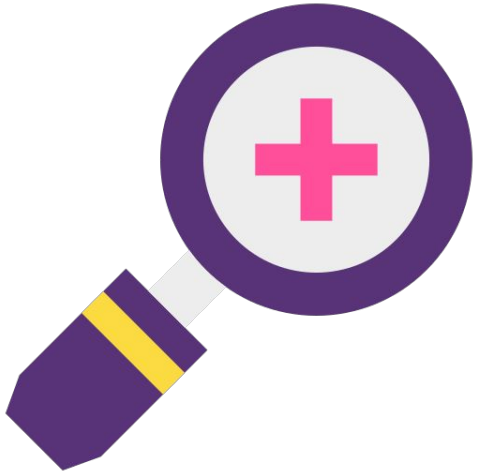
6. Furthermore…

- **Research Peers and Colleagues**. Consult with experienced researchers or colleagues who have previously written DMPs or have expertise in data management. They may be able to provide insights, tips, and suggestions based on their own experiences.

- **Project Partners and Collaborators**. If you are part of a collaborative project, engage with your project partners and collaborators. Discuss data management practices, share ideas, and ensure alignment in DMP preparation to ensure consistency and compliance across the project.

- **Online Resources and Communities**. There are various online resources, websites, and forums dedicated to data management and DMPs. Websites of research organizations, associations, and research blogs often provide helpful information and resources.

- **Explore relevant data repositories or platforms specific to your research field**. They often provide documentation or guidelines on data management, including recommended practices for creating a DMP.



DMP –
Frequently Asked Questions

Is a DMP required for all research projects?

DMP requirements vary depending on funding agencies, institutions, and specific research contexts. While some projects may have mandatory DMP requirements, **it is generally good practice to create a DMP regardless of formal requirements** to enhance data management and research integrity.

Under HE, does a DMP need to be included as part of the proposal?

At proposal stage, applicants that intend to generate and/or reuse data and/or other research outputs (except for publications) during the project must provide an outline of how the data and/or other research outputs (e.g. workflows, protocols, software, samples) will be managed in line with the FAIR (Findable, Accessible, Interoperable, Reusable) principles. This obligation does not apply to European Research Council (ERC) calls.

Beneficiaries must submit a data management plan to the granting authority in accordance with the grant agreement (normally by month 6). In cases of a public emergency, and if the work programme requires so, proposers should submit a Data Management Plan (DMP) already when submitting the proposal or at the latest by the signature of the grant agreement.

https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq;type=0,1;categories=;tenders=;programme=null;keyword=;freeTextSearchKeyword=;matchWholeText=true;period=null;status=0;sortQuery=publicationDate;faqListKey=faqSearchTablePageState

Who should be involved in creating and reviewing a DMP?

Researchers, principal investigators, data managers, and relevant stakeholders should be involved in creating and reviewing a DMP to ensure that all aspects of data management and compliance are adequately addressed.

Are the requirements for open science and exploitation conflicting in Horizon Europe?

Open science requirements do not impact the decision to commercially exploit research results, such as through patenting.

Regarding research data, open access is the **default requirement**, but it should follow the principle of being "as open as possible, as closed as necessary." This means that beneficiaries may choose to **restrict** access to their data, either **partially or completely, also on a temporary basis**, considering their legitimate interests such as c**ommercial exploitation** or other constraints like data **protection rules, privacy, confidentiality, or trade secrets**. **Beneficiaries need to provide justification in the Data Management Plan (DMP) for any exceptions made to restrict access to the research data.**

https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq;type=0,1;categories=;tenders=;programme=null;keyword=;freeTextSearchKeyword=;matchWholeText=true;period=null;status=0;sortQuery=publicationDate;faqListKey=faqSearchTablePageState

Under Horizon Europe, can the data be kept closed if **patenting** is foreseen? Is this a legitimate reason to keep the data closed?

Yes. Open access to data is required under the principle ‘as open as possible, as closed as necessary’. This means that, as an exception, beneficiaries may decide to close (all or some of) their data, maybe temporarily, taking into consideration their legitimate interests, including **commercial exploitation** e.g. through patenting. **Beneficiaries must explain in the Data Management Plan (DMP)** the exception(s) under which they choose to restrict access to some or all of the research data.

https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq;type=0,1;categories=;tenders=;programme=null;keyword=;freeTextSearchKeyword=;matchWholeText=true;period=null;status=0;sortQuery=publicationDate;faqListKey=faqSearchTablePageState

Under HE, for how long after the end of a project should the data be made available for re-use and validation purposes?

Beneficiaries need to provide digital or physical access to data or other results needed for the validation of the conclusions of peer-reviewed scientific publications, provided that their legitimate interests or constraints are safeguarded (this does not apply under the ERC Work Programmes 2021-2023).

It is recommended to keep **data available for a substantial period (at least 5 years, and preferably 10 years), or longer according to disciplinary deposition practices.** This is in line with the recommendations of the European Code of Conduct for Research Integrity. The Data Management Plan is expected to include information about long-term preservation of data.

https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq;type=0,1;categories=;tenders=;programme=null;keyword=;freeTextSearchKeyword=;matchWholeText=true;period=null;status=0;sortQuery=publicationDate;faqListKey=faqSearchTablePageState

Are there any Data Management Plan (DMP) examples available?

A DMP template is provided under the reporting templates in the reference documents of the Funding & Tenders Portal. The use of such a template is recommended for Horizon Europe beneficiaries. The Horizon Europe Programme Guide contains useful references that can be used to find DMP examples in different disciplines, and points to specialised journals that publish DMPs. For actions supported by the European Research Council (ERC), a separate template is provided.

Also look at this.

Have a look at the Liber DMP catalogue on Zenodo

https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq;type=0,1;categories=;tenders=;programme=null;keyword=;freeTextSearchKeyword=;matchWholeText=true;period=null;status=0;sortQuery=publicationDate;faqListKey=faqSearchTablePageState

Interaction

Go to:
https://www.menti.com

Voting code: 1728 5770

https://www.menti.com/alsq6nioemuk

Exercise 2: revision of existing DMPs

You will be divided into **groups**.

Observe these examples:

- https://www.lib.umn.edu/services/data/dmp-examples
- https://zenodo.org/record/1251693#.ZFtpsOxBwq0

Identify the key elements of the DMP: is there anything missing?

Identify possible areas for improvement.

Is there any example of a virtuous element?

Discuss in your group one of the given examples and note down the main aspects.

A spokesperson from each group reports the main points discussed (you can write on mentimeter)

Interaction

Go to:
https://www.menti.com

Voting code: 1728 5770

https://www.menti.com/alsq6nioemuk

Some observations…

In the first DMP here https://www.lib.umn.edu/services/data/dmp-examples

- No administrative data (or very few)
- There is confusion on the concept of metadata documentation (internal RDM concepts are present and not data description elements for externals)
- There is no description of data types, only variables are mentioned, but it is not explained which data and what they are used for
- There is no mention of roles and responsibility, nor costs
- There is a description of data utility (good!)
- There is a mentions of the main software (and thus formats)

Some observations...

In the second DMP here https://www.lib.umn.edu/services/data/dmp-examples

- no administrative data
- no mentions of metadata nor metadata standard
- no mention of costs
- no mention of standardised resources (eg. controlled vocabulary)
- details on how personal data will be managed - confidentiality (good!)
- quite detailed on documentation

Some observations…

In this DMP:

https://zenodo.org/record/1251693#.ZF4Rr-xBzxp

- Very few information about the project
- Dataset structure only for primary data
- Data collection is detailed on the individual dataset
- Distinction made between data sharing (inside the consortium) and archiving and preservation
- Confusion between archiving and storage
- Consent form and information sheet included (it is ok also to refer to specific deliverables on ethics, if any)

Exercice 3: hands-on session

Draft your own DMP

Go to https://jamboard.google.com/d/1tGszx6CWSlJnrDGKrQ5ngyr4bnCROcS2mEw2UtCgs_8/edit?usp=sharing

Duplicate a jamboard (only one!) for each group.

Within each group, choose a research project on which to do the draft DMP. Or draft it on the basis of the mock case research on page 5 of the jamboard.

Divide the parts to be filled in among yourselves (write on the same jam board as the group)

Discuss the aspects you have doubts about.

Finally, return to the main room and a spokesperson presents the draft.

Let's look at some examples
follow the link to the Liber DMP catalogue:

https://zenodo.org/communities/liber-dmp-cat/?page=1&size=20

Interaction

Go to:
https://www.menti.com

Voting code: 1728 5770

https://www.menti.com/alsq6nioemuk

THANK YOU

GET IN TOUCH

gina.pavone@isti.cnr.it

VISIT
open-science.it

Consiglio Nazionale delle Ricerche

infrascience

OpenAIRE

Co-funded by the European Union