Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Consiglio Nazionale delle Ricerche

Research Data & Open Science

6 Novembre 2023

PART 2

Lottie Provost
Istituto di Linguistica Computazionale
"A. Zampolli" - CNR
lottiemiaprovost@cnr.it
https://orcid.org/0000-0001-5279-797X

FOSSR
Fostering Open Science in Social Science Research
Innovative tools and services to investigate economic and societal change

This presentation is on Zenodo
DOI 10.5281/zenodo.10074214

Some of the slides you will see in this course are taken from or adapted from previous materials produced by my colleagues. Attribution is indicated in the specific slides.

Special thanks to Francesca Di Donato https://orcid.org/0000-0003-0144-8934

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

and Gina Pavone https://orcid.org/0000-0003-0087-2151

What you can do with this presentation

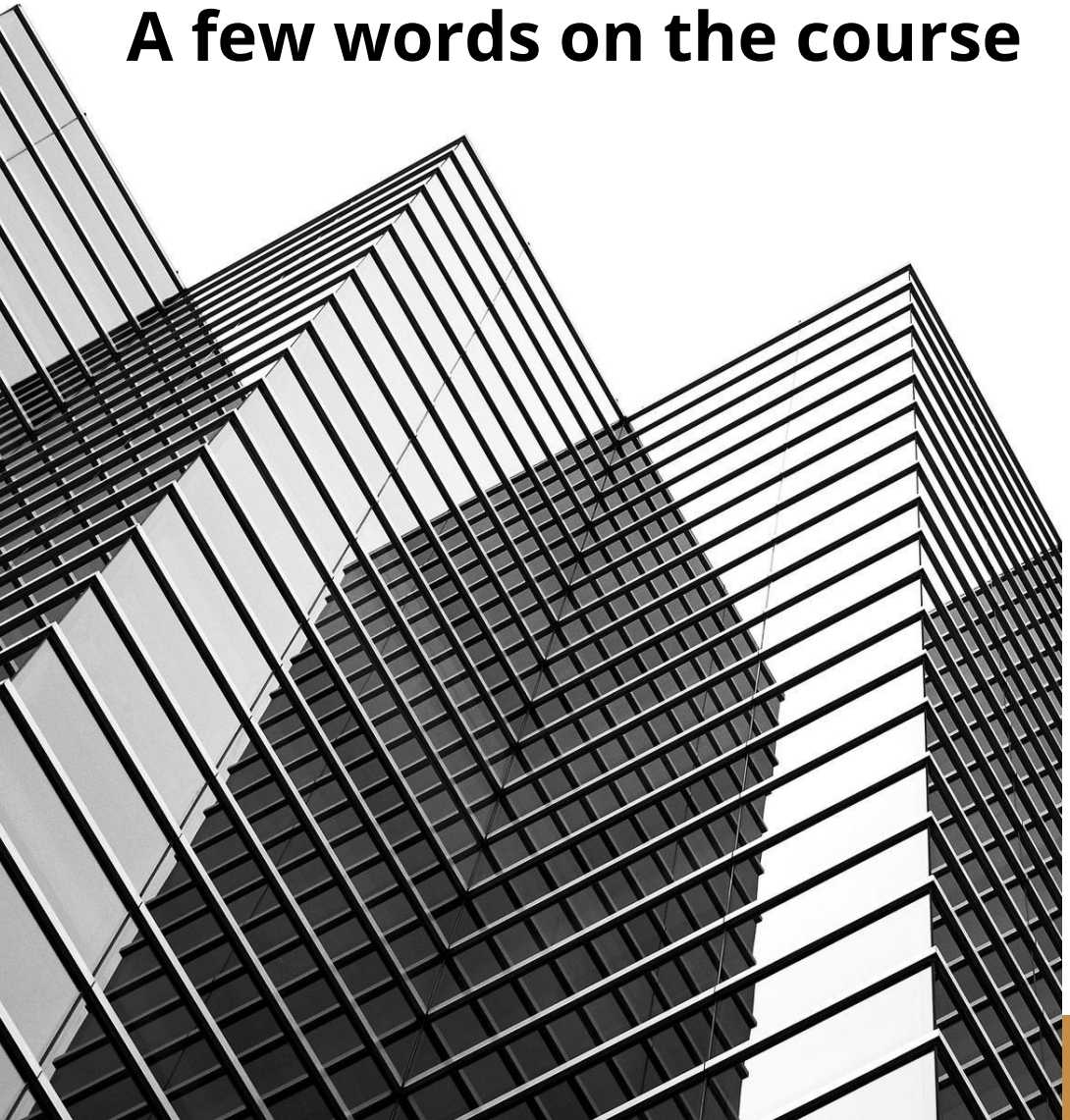
You are free to

- distribute, remix, adapt, and build upon the material in any medium or format

Provided that

- you attribute the work to its author and respect the rights and licenses associated with its components

A few words on the course



The course methodology:

- the aim of the course is to introduce you to the concepts and practices underlying research data and open science, and to provide you with tools to help you get started.
- your involvement is crucial !

Structure of the course

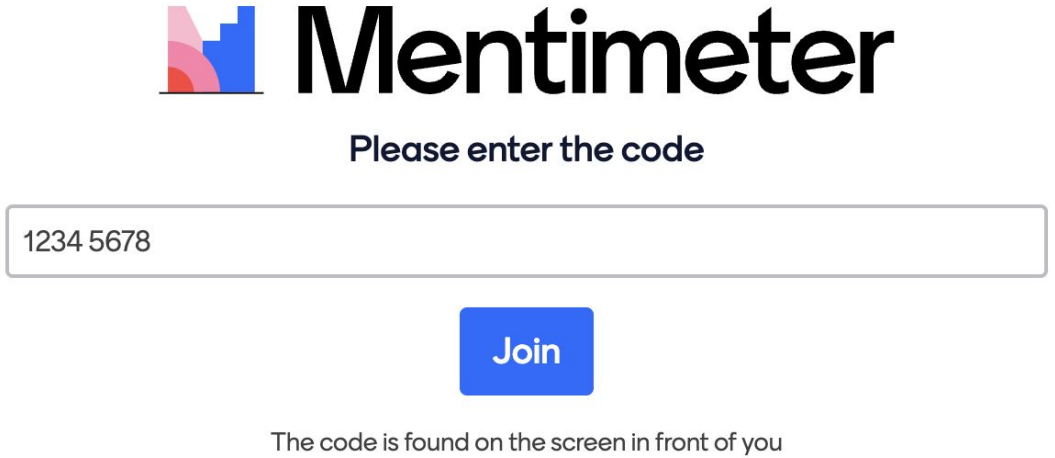
- 3 modules, frontal lesson

In each module:

- Test your understanding as we go (questions to be answered by you, on Mentimeter)
- Final open Q&A

Be ready !

When you see a slide with this image: Go to **Menti.com** with the code on the slide and answer the question(s)



Today's journey PART 2

RESEARCH DATA & RESEARCH DATA MANAGEMENT

Q&A

OPEN SCIENCE AND OPEN ACCESS IN PRACTICE

Q&A

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
Consiglio Nazionale delle Ricerche

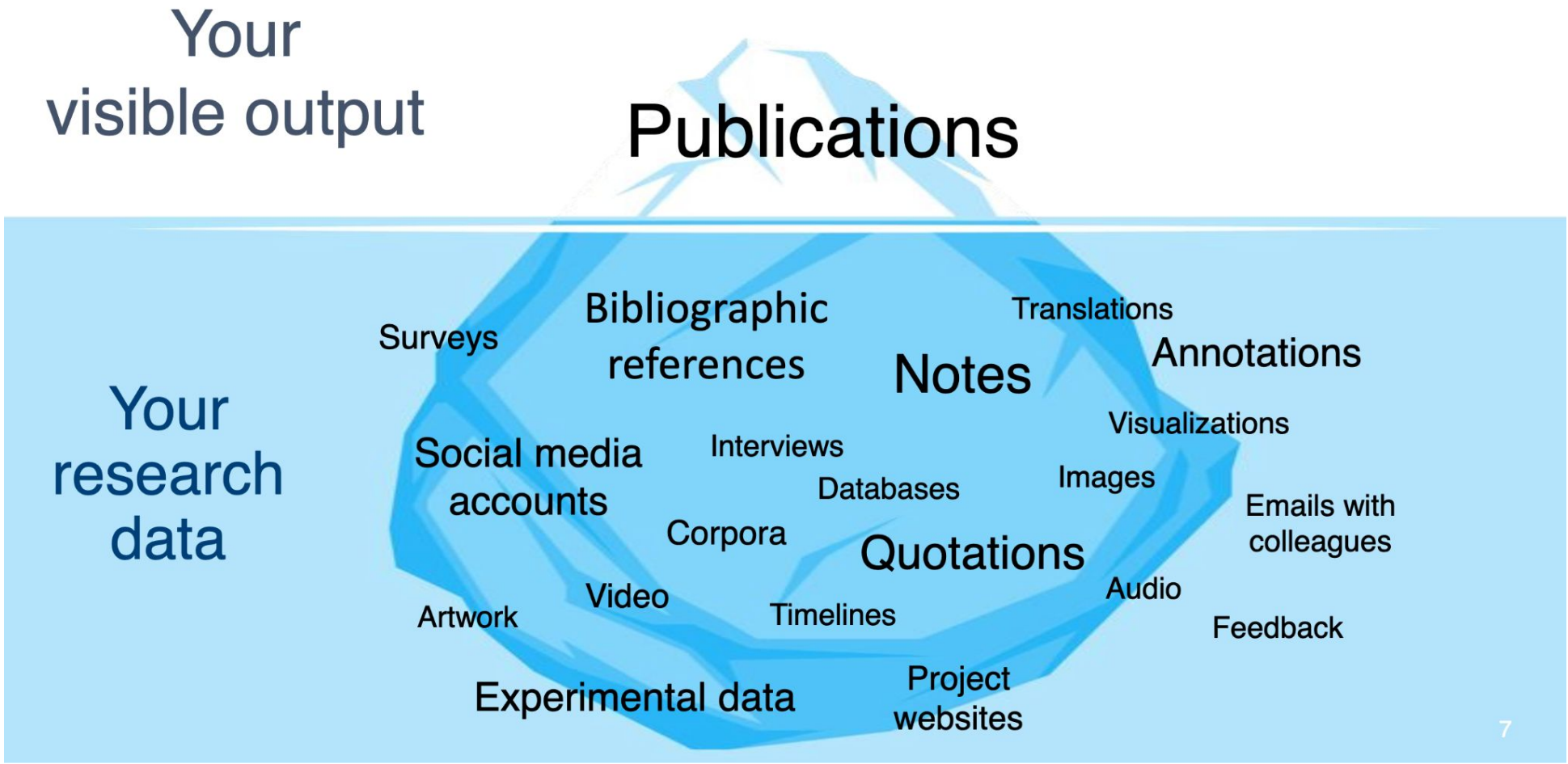
RESEARCH DATA & RDM

Core values
Data Jargon
Data Stewards

Why do we talk about research data?

> "An article [...] in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship."
>
> Buckheit, Jonathan B., and David L. Donoho. "WaveLab and Reproducible Research." In Wavelets and Statistics, edited by Anestis Antoniadis and Georges Oppenheim, 55–81. Lecture Notes in Statistics. New York, NY: Springer, 1995. https://link.springer.com/chapter/10.1007/978-1-4612-2544-7_5

Publications are just the visible output of research. Most of the work of scholarship happens within the vast amount of materials that underlie each publication: the research data.


Your visible output
Publications
Your research data
Surveys
Bibliographic references
Translations
Annotations
Notes
Visualizations
Social media accounts
Interviews
Databases
Images
Emails with colleagues
Corpora
Quotations
Audio
Video
Timelines
Artwork
Feedback
Experimental data
Project websites
7

An introduction to RDM

Ghent University Data Stewards (2020). Knowledge clip: What is Research Data Management (RDM)?

1. Core values

Research integrity

"In our modern knowledge society, scientific and scholarly research has thereby acquired an indispensable role. In providing knowledge and understanding of all aspects of reality, science and scholarship also provide the building blocks for political decision-making and the stimulus for societal development and economic growth (...)

If scientific and scholarly research is to perform this role properly, research integrity is essential. Researchers who are not guided by the principles of honesty, scrupulousness, transparency, independence and responsibility risk harming both the quality and the trustworthiness of research"

| *Netherlands Code of Conduct for Research Integrity*, 2018

The European Code of Conduct for Research Integrity, Revised edition 2023.

Versione Italiana

Reliability

Honesty

Respect

Accountability

"Good research practices are based on fundamental principles of research integrity. They guide individuals, institutions, and organisations in their work as well as in their engagement with the practical, ethical, and intellectual challenges inherent in research."

Reproducibility

"Reproducibility means that research data and code are made available so that others are able to reach the same results as are claimed in scientific outputs"

| The Open Science Training Book, 2018

Image by Auke Herrema – Het Bouwteam (2014). https://www.fosteropenscience.eu/content/cartoonpublication-and-data

2. Data jargon

Research Data

"Research data constitute primary research data (the raw, rough measurements or observation) and secondary research data (the results after the data have been processed by a researcher (recoded, combined, categorised, visualised, etc.))."

Utrecht University, 2016

Photo by Priscilla Du Preez on Unsplash

Research Data

"Research data may be facts, observations, interviews, recordings, measurements, experiments, simulations and software; numerical, descriptive and visual; raw, cleaned up and processed; they may or may not support an actual or intended publication; and may be stored and exchanged in various formats on various storage media".

Van Berchum & Grootveld, 2017

Image by Alexander Schimmeck on Unsplash

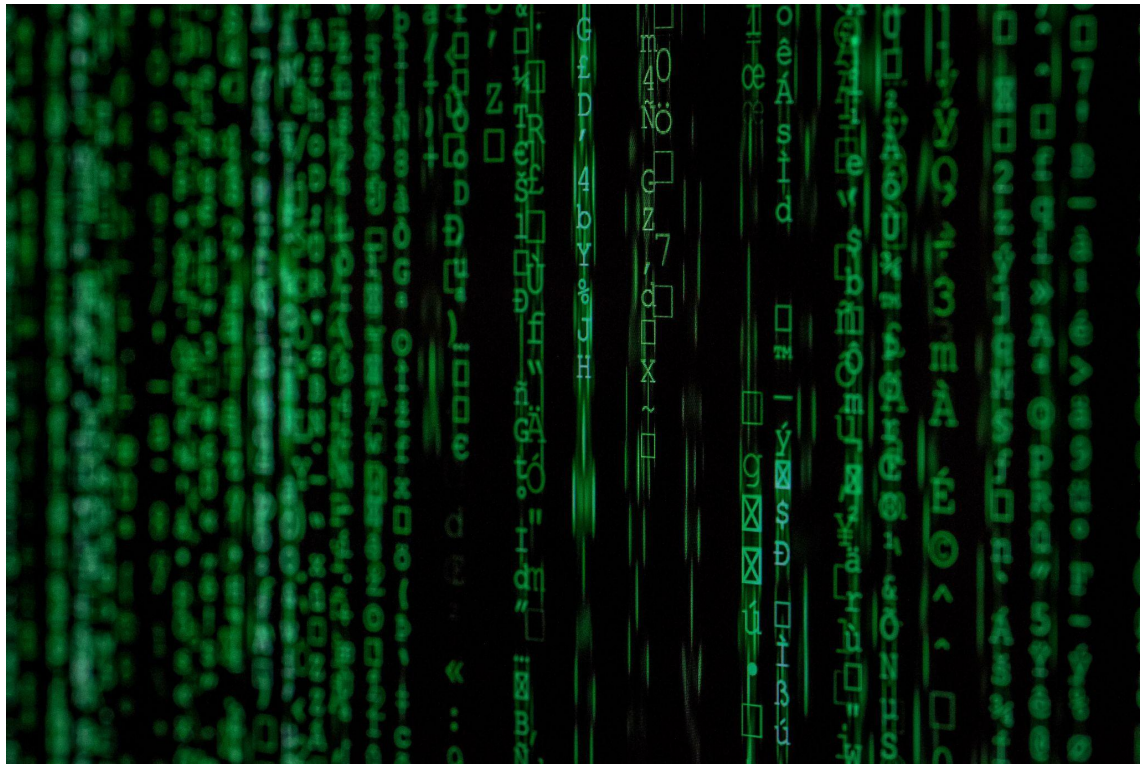


Image by Markus Spiske on Unsplash

Research Data

"Research data are the factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated."

OECD, 2007

Ways to look at research data

How was the data collected/obtained?

What forms does the data take?

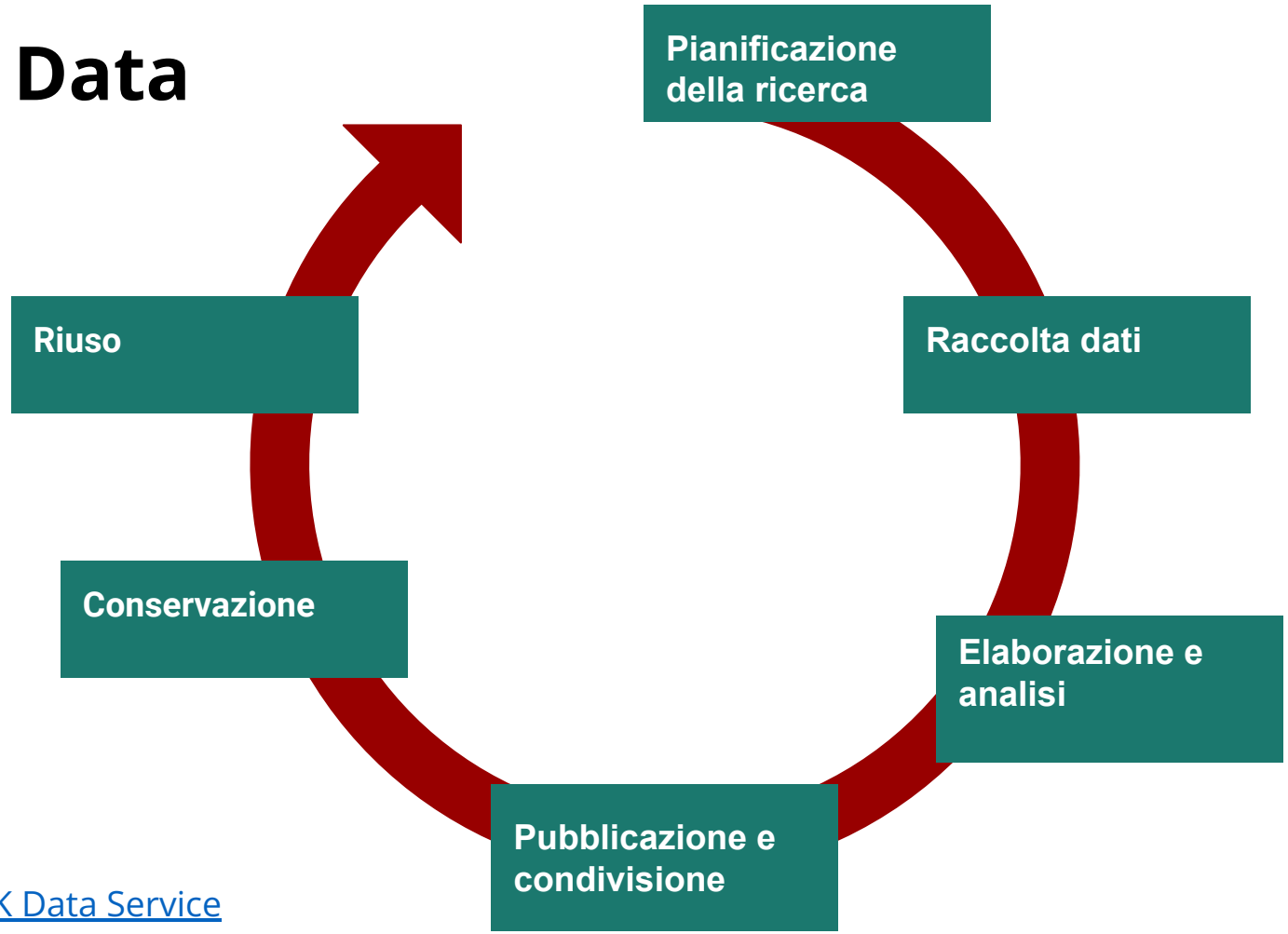
On which formats is the data stored?

What is the size of the data files?

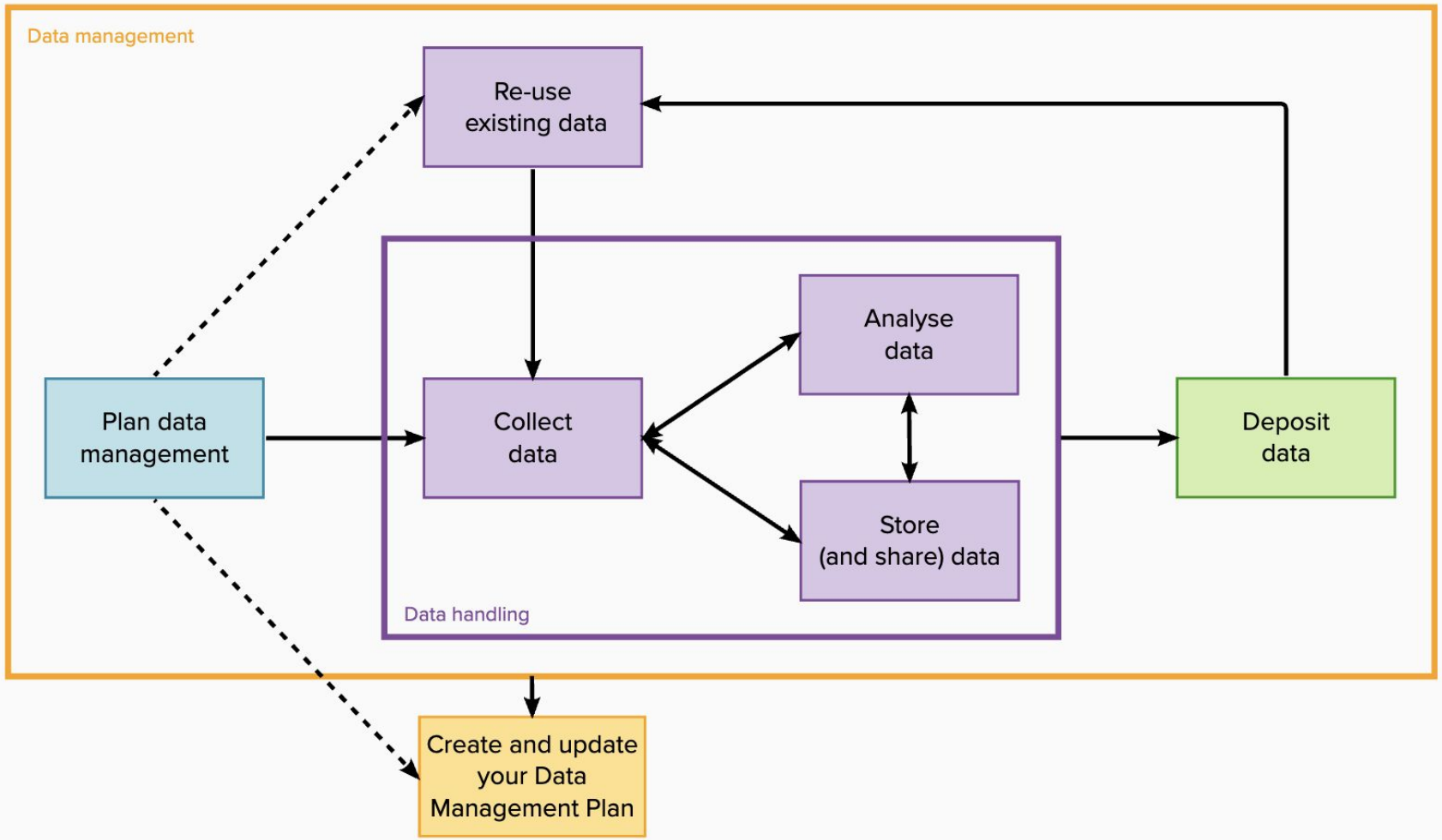
What is the phase of the research lifecycle?

Source image

Research Data lifecycle

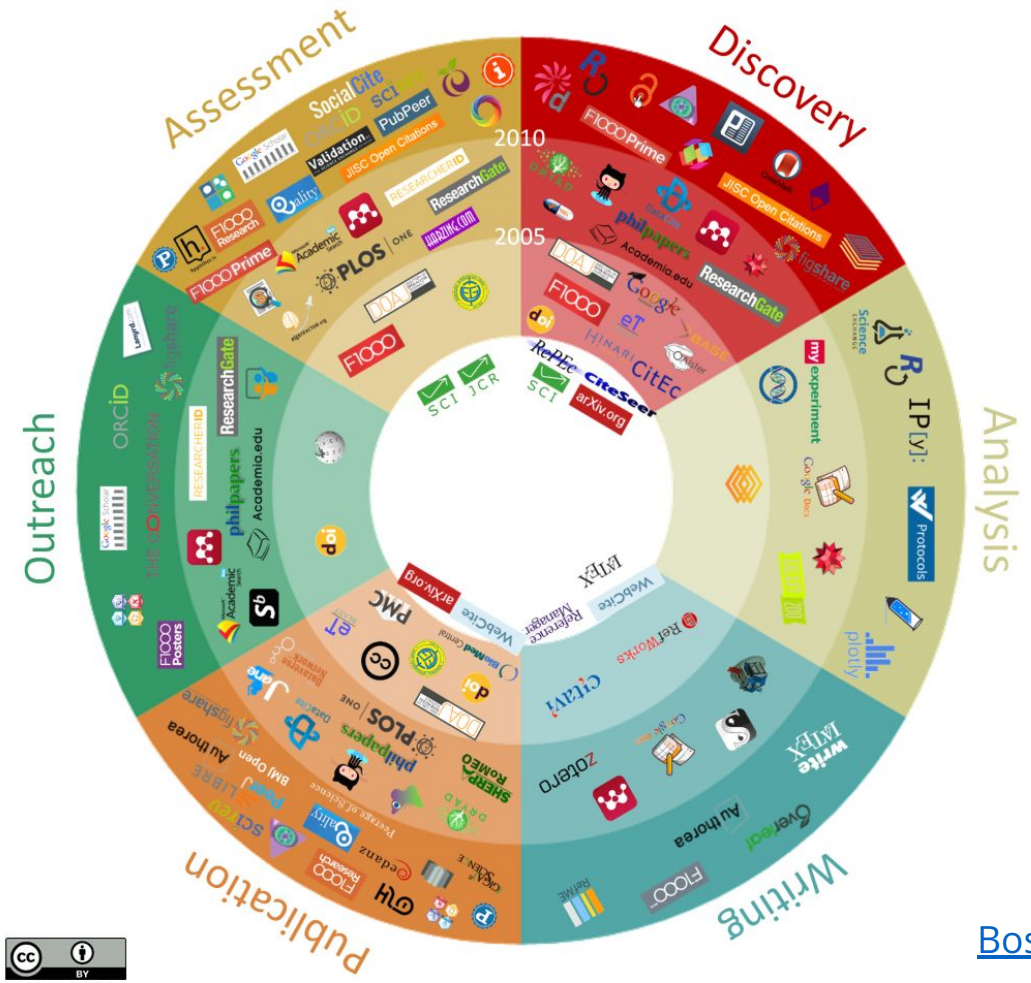


Adapted from Lifecycle UK Data Service



Research Data Management: Data Lifecycle

Gualandi, B., Caldoni, G., & Marino, M. (2022). Research Data Management: Data Lifecycle. Zenodo. https://doi.org/10.5281/zenodo.7249051



The Circle

Bosman & Kramer, 2015

FAIR Principles for Scientific Data Management, 2016

FAIR principles

Findable

- Persistent identifier (e.g. DOI)
- Rich metadata
- Searchable and discoverable online

Accessible

- Deposited on a trusted repository (e.g. Zenodo)
- Data can be restricted and still FAIR – "as open as possible, as closed as necessary"

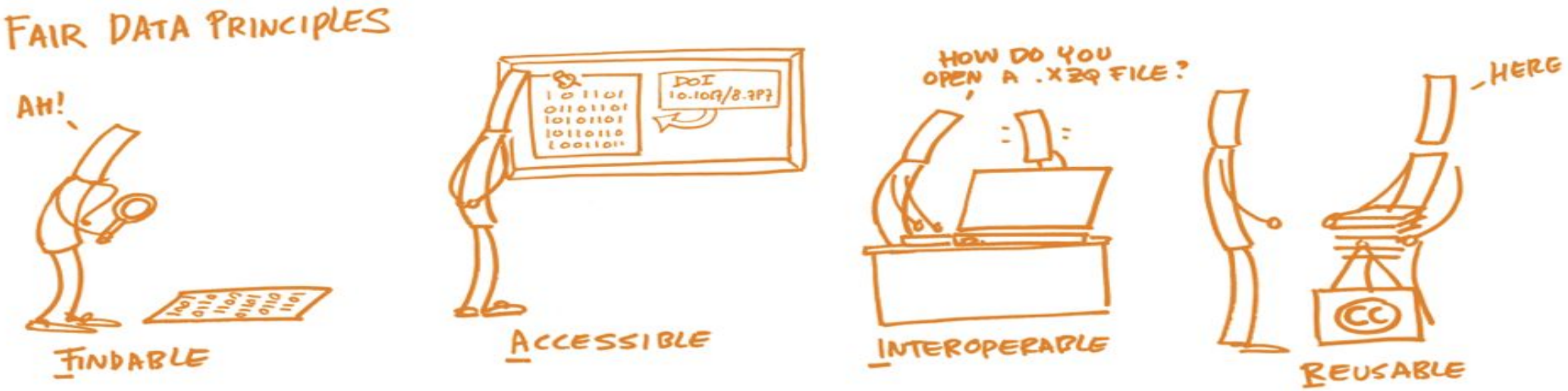
Interoperable

- Open and/or standardised file formats

Reusable

- Well documented (e.g. README files), including provenance and tools / instruments needed to reproduce the results
- Clear licence (e.g. CC BY 4.0, CC0)

https://www.openaire.eu/how-to-make-your-data-fair



Findable:
Others can easily discover your data

Accessible:
It is clear who, when and how can access your data (does not mean open)

Interoperable:
Your data can be integrated with other data and/or they can be easily used and read by machines

Reusable:
Your data can be reused by others in new research

Source: Pavone, Gina. (2022, May 31). FAIR principles and Data Management Plan. Zenodo.

Importantly, it is our intent that **the principles apply not only to 'data'** in the conventional sense, **but also to the algorithms, tools, and workflows** that led to that data.

All scholarly digital research objects—from data to analytical pipelines—**benefit from** application **of these principles,** since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

Wilkinson et al.

Image by IMattSmart on Unsplash

⚠️ **FAIR is not the same as Open!**

Open Data

"Open data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose."

Open Knowledge Foundation, 2005

Image by Calvin Hanson on Unsplash

RDM

* What will be necessary for using or collecting your particular type of data?
* How will you maintain the integrity of the data, make sure that they are not lost due to technical mishaps, and that the right people can access the data at the appropriate time?
* Look forward to the future, provide detailed and structured documentation tso that you can share your data with other colleagues and prepare them for long-term availability.

Image by Fabian Quintero on Unsplash

3. Data Stewards

Che cosa succede quando i dati non sono gestiti?

I dati non sono più riutilizzabili!!!

Valentina Pasquale – Istituto Italiano di Tecnologia

iit ISTITUTO ITALIANO DI TECNOLOGIA

Perché servono i « data steward»?

Per poter produrre dati FAIR i ricercatori hanno bisogno di un supporto professionale adeguato, così come già avviene in altri contesti (TT, comunicazione, etc.)

EOSC SRIA v1.0

Strategic Research and Innovation Agenda (SRIA)
of the
European Open Science Cloud (EOSC)
Version 1.0 15 February 2021

News Blog

We need 500.000 respected data stewards to operate the European Open Science Cloud Interview to Barend Mons, Chairman of the High Level Expert Group on EOSC, May 2016

04/05/16 09:08

At the e-IRG workshop in Amsterdam, we had the opportunity to talk to Barend Mons who is chairing the High Level Expert Group on the European Open Science Cloud, an advisory group to the European Commission. To be successful, the European Science Cloud needs a lot of experts to operate it, Barend Mons told us. Data stewards that have a lot of knowledge about managing and maintaining data. Experts who are well respected with a solid career path. Barend Mons also discussed several other findings of the Expert group, whose report will be published very soon.

We are here at the e-IRG workshop in Amsterdam and we are talking with Barend Mons. Welcome. You just had a presentation here, and you were also part of the panel. One of the things that you focused on, was the European Open Science Cloud, because you are chairman of the Expert Group of the European Commission. So can you tell a little bit about the progress? What is the status of the European Science Cloud?

7.4. Critical success factors

The developments and expected impacts described above will not happen spontaneously. For these benefits to materialise a number of critical success factors (CSFs) must be in place. The following CSFs have been identified for EOSC:

- Researchers performing publicly funded research make relevant results available as openly as possible;
- Professional data stewards are available in research-performing organisations in Europe to help implement FAIR principles and support Open Science;

WORLD VIEW · 25 FEBRUARY 2020

Invest 5% of research funds in ensuring data are reusable

It is irresponsible to support research but not data stewardship, says Barend Mons.

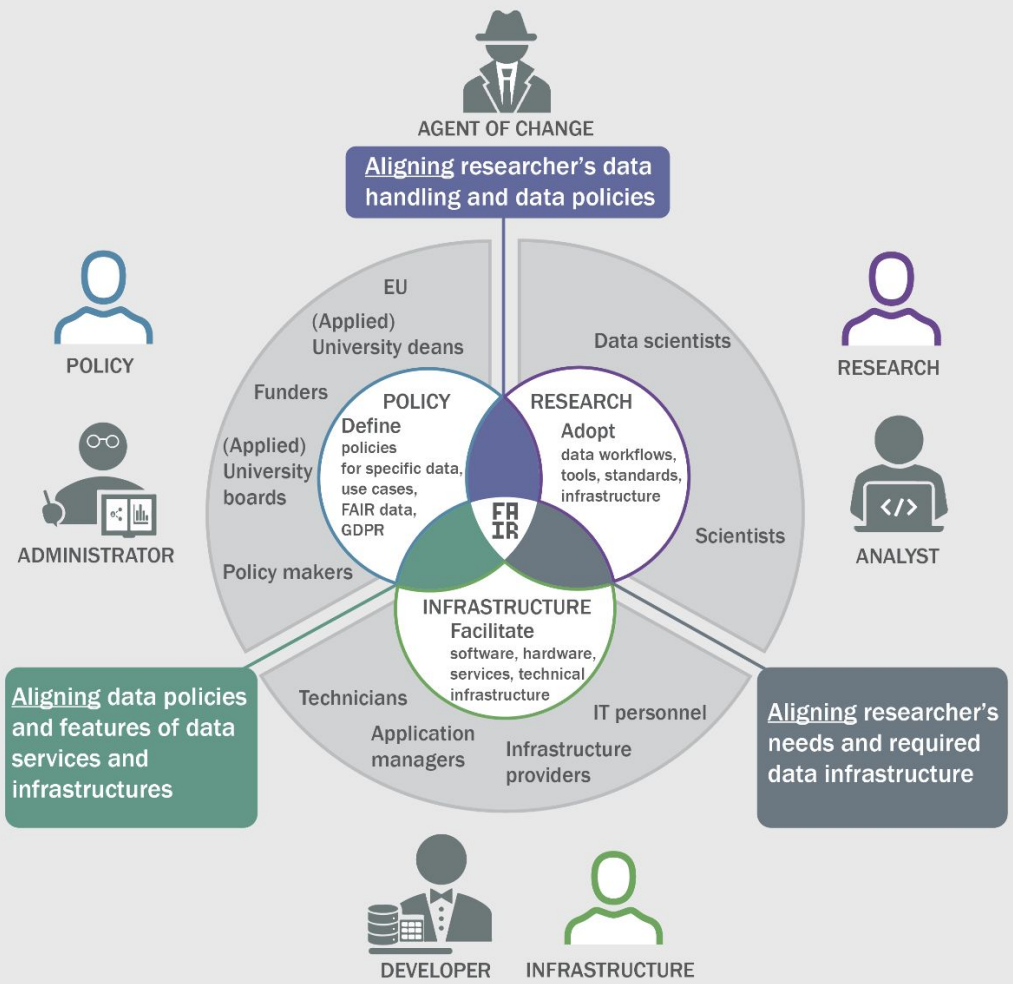
Barend Mons

http://www.doi.org/10.1038/d41586-020-00505-7

to the European Commission. To be successful, the European Science Cloud needs a lot of experts to operate it, Barend Mons told us. Data stewards that have a lot of knowledge about managing and maintaining data. Experts who are well respected with a solid career path. Barend Mons also discussed several other findings of the Expert group, whose report will be published

Valentina Pasquale – Istituto Italiano di Tecnologia

ISTITUTO ITALIANO DI TECNOLOGIA



National Forum for Research Data Management- Denmark

Four roles of a data steward

ZonMw/ELIXIR project- The Netherlands

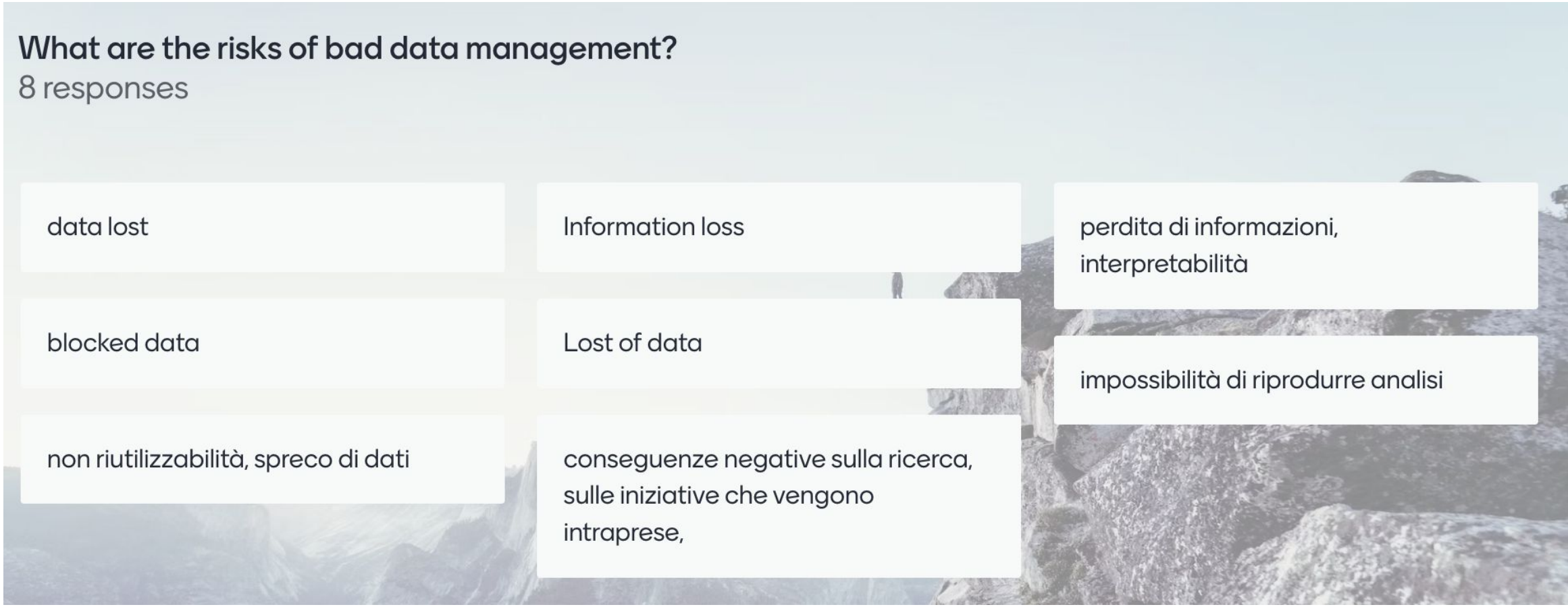
Three roles of a data steward

Frederike Schmitz. (2020). The roles of data stewards in the data stewardship landscape identified in Denmark and the Netherlands. Zenodo. https://doi.org/10.5281/zenodo.4321265

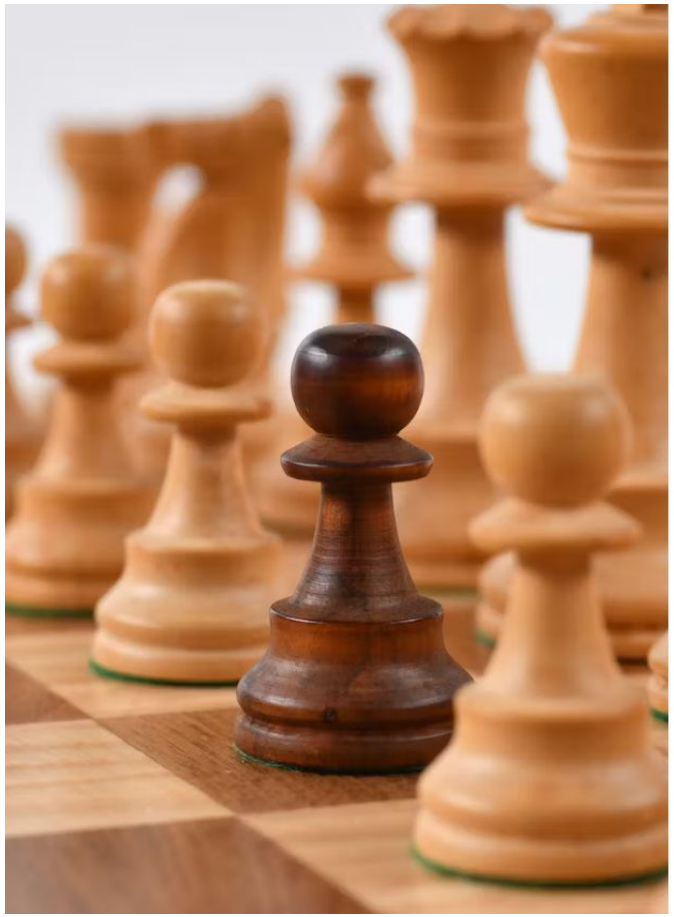
Go to Menti.com

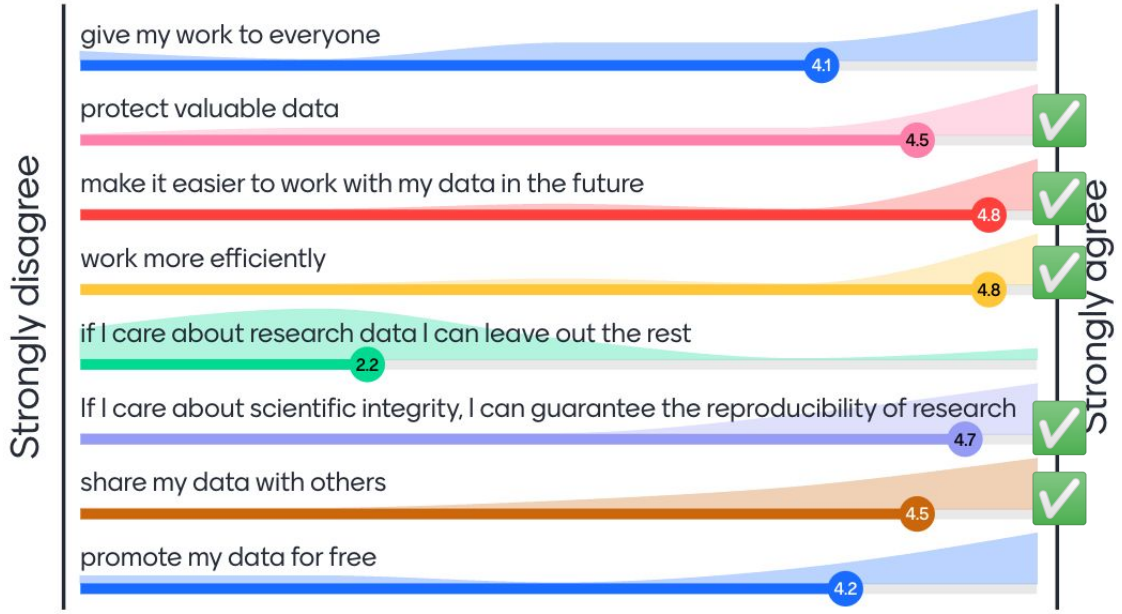
Image by Volodymyr Hryshchenko on Unsplash

Menti results (1/3)

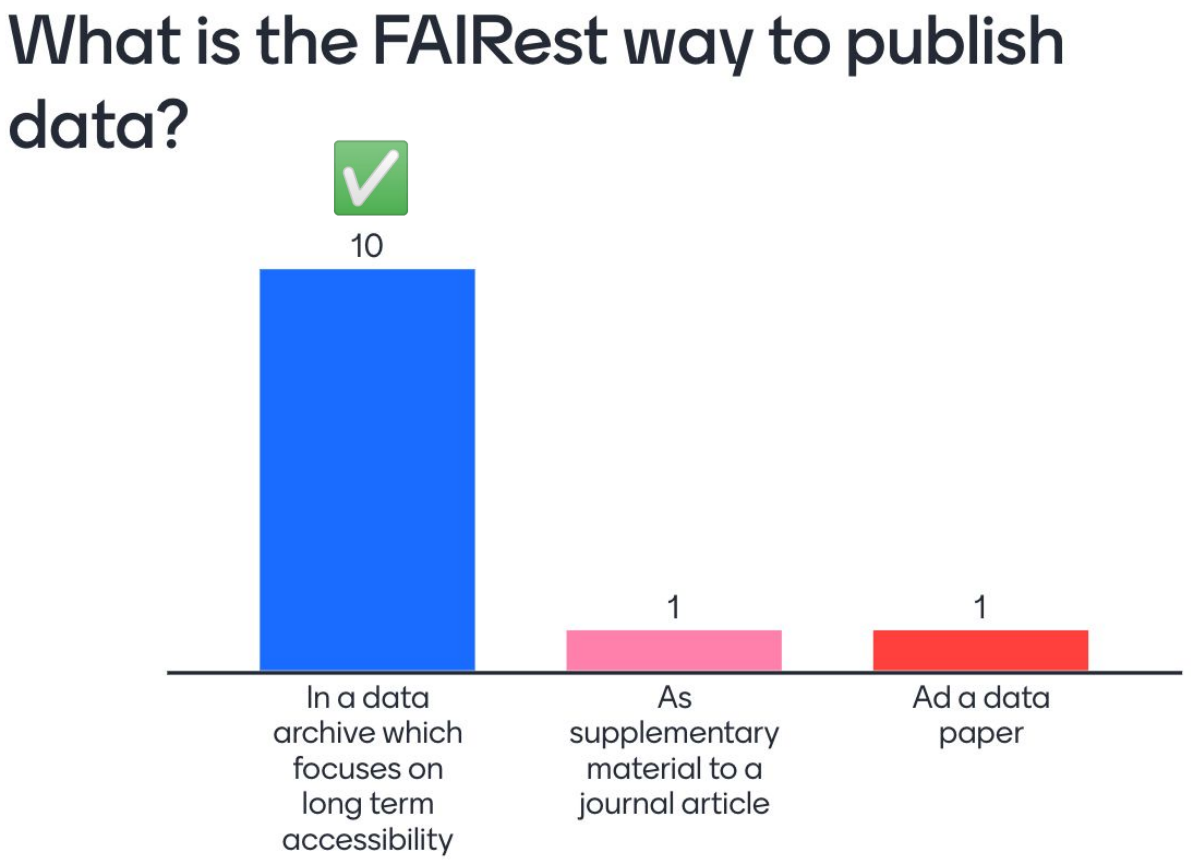


Menti results (2/3)





Menti results (3/3)



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Consiglio Nazionale delle Ricerche

OPEN SCIENCE AND OPEN ACCESS IN PRACTICE

EU-funded projects
Open Access: how?
Get started with Zenodo

1. Open Science and Open Access in EU funded projects

An overview of Horizon Europe

EC priority

Open Science

An approach to the scientific process that focuses on spreading knowledge as soon as it is available using digital and collaborative technology. Expert groups, publications, news and events.

The EU's open science policy

Open science is a policy priority for the European Commission and the standard method of working under its research and innovation funding programmes as it improves the quality, efficiency and responsiveness of research.

When researchers share knowledge and data as early as possible in the research process with all relevant actors it helps diffuse the latest knowledge.

And when partners from across academia, industry, public authorities and citizen groups are invited to participate in the research and innovation process, creativity and trust in science increases.

That is why the Commission requires beneficiaries of research and innovation funding to make their publications available in open access and make their data as open as possible and as closed as necessary. It recognises and rewards the participation of citizens and end users.

Furthermore, the European Open Science Cloud EN will enable researchers across disciplines and countries to store, curate and share data.

The effective linking of open science practices to innovation and business models requires careful consideration of issues such as Intellectual Property Rights (IPR), licensing agreements, interoperability and reuse of data.

EU support for open access

Open access is the practice of providing online access to scientific information that is free of charge to the user and is reusable.

It is now widely recognised that making research results more accessible to all contributes to better and more efficient science, and to innovation in the public and private sectors

The Commission supports open access, specifically in its funding programmes.

Open access to scientific information in research and innovation refers to 2 main categories

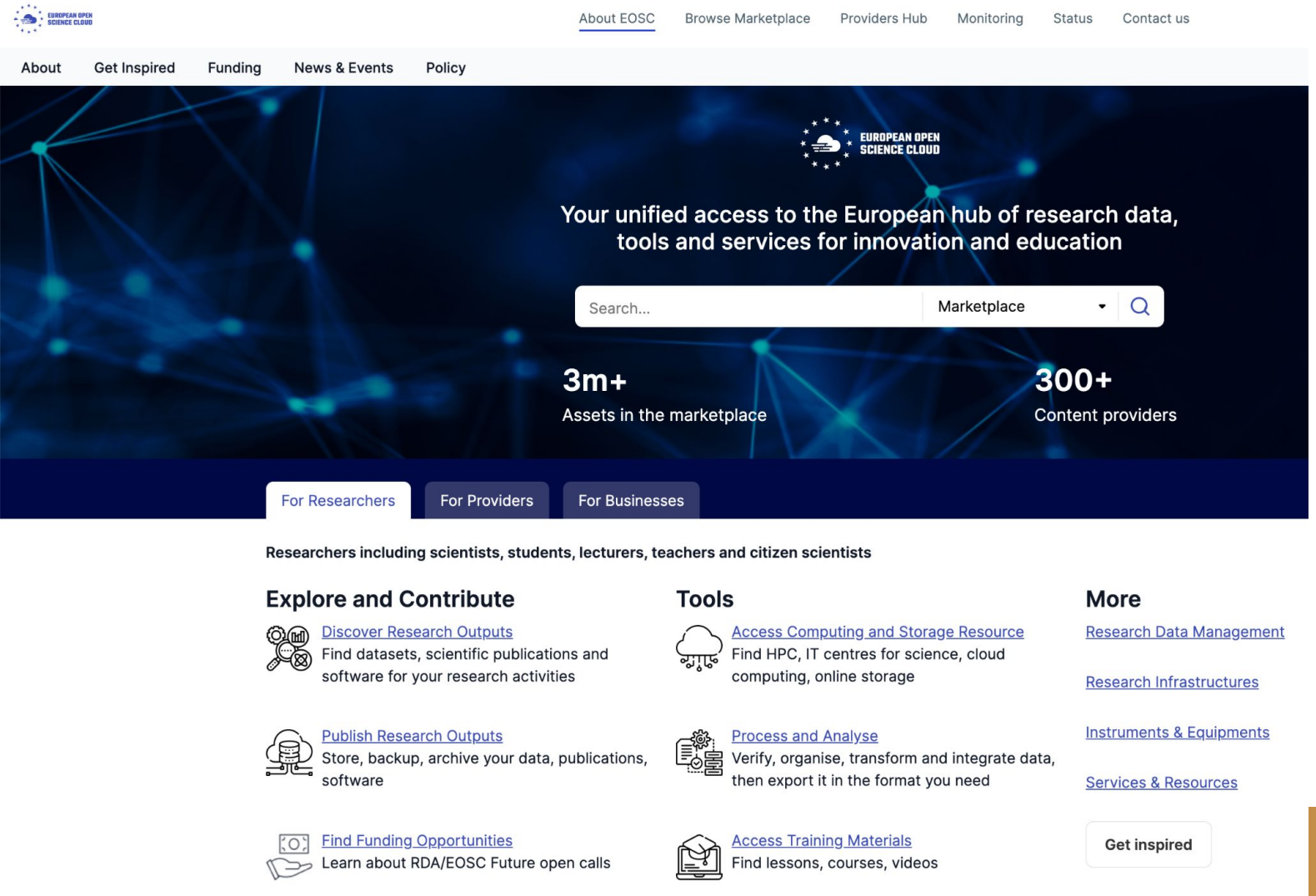
- peer-reviewed scientific publications (primarily research articles published in academic journals)
- scientific research data: data underlying publications and/or other data (such as curated but unpublished datasets or raw data)

Background note EN that provides more information on open access to scientific publications and research data at the EU and national level.

https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en

Image by JJ Ying on Unsplash

EOSC https://eosc-portal.eu/



"EOSC will be a multi-disciplinary environment where researchers can publish, find and re-use data, tools and services, enabling them to better conduct their work."

https://eosc.eu/eosc-about

What can you find ?

- Meta information
- Policies across Europe
- Use cases
- Marketplace
- Information for service providers

Why is Open Science so important?

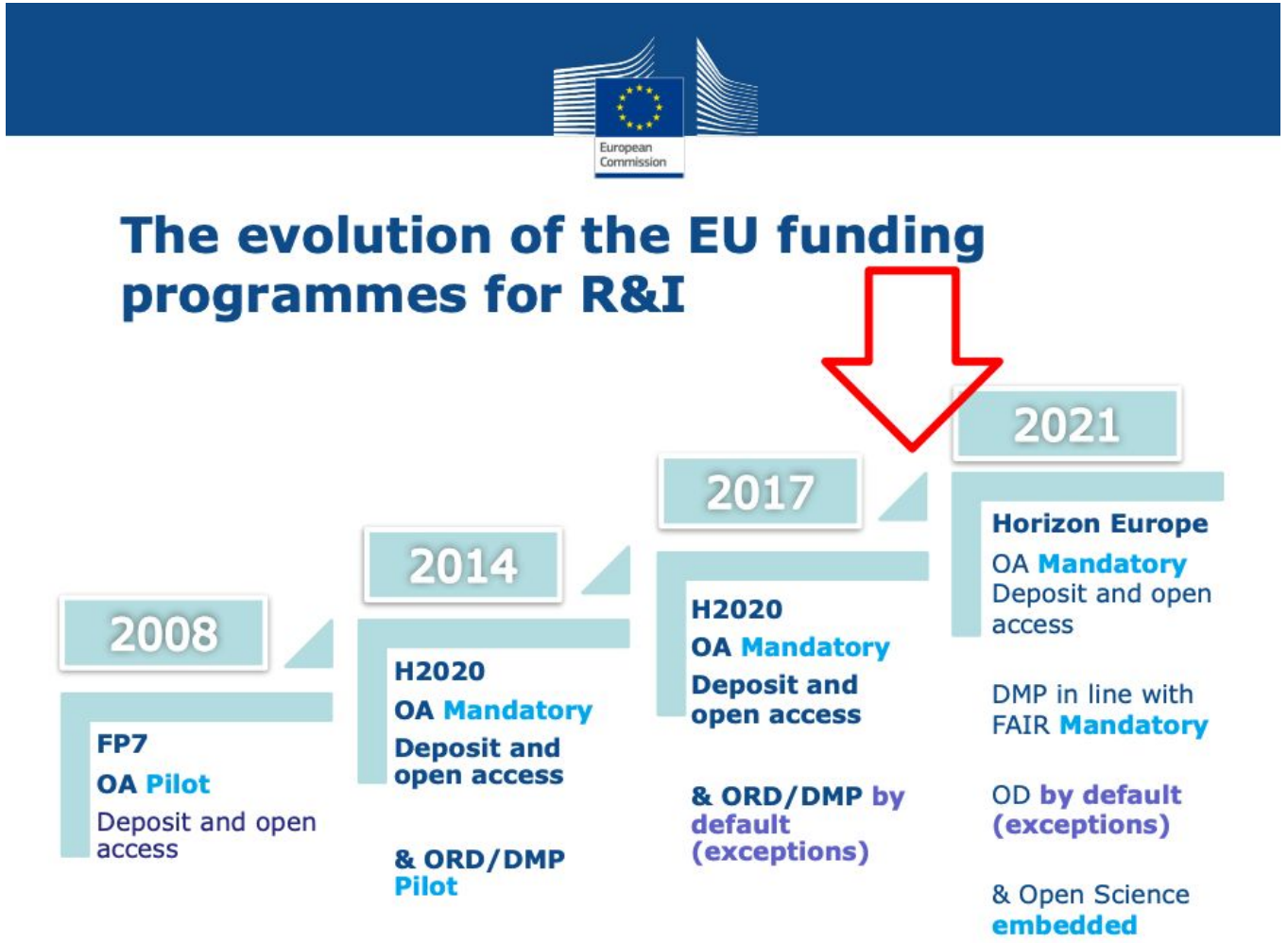
It's good for science: efficiency, verifiability, transparency, interdisciplinarity, trust in science

It's good for the economy: access to and re-use of scientific information by industry, innovation

It's good for society: broader, faster, transparent & equal access for and possibility for participation by citizens, increased societal impact of science and research

****International collaboration necessary**

Source: slide Fit for Purpose! Shaping Open Access and Open Science Policies for Horizon Europe, Victoria Tsoukala, PhD, DG RTD Open Science (Unit G4), PUBMET 2019, Zadar, September 19th, 2019



Source: slide Fit for Purpose! Shaping Open Access and Open Science Policies for Horizon Europe, Victoria Tsoukala, PhD, DG RTD Open Science (Unit G4), PUBMET 2019, Zadar, September 19th, 2019

Source Image

From

Competition

Collaboration

Cambio di prospettiva: la valutazione

Open Science practices

early and open sharing of research (for example through preregistration, registered reports, pre-prints, or crowd-sourcing)

research output management including research data management measures to ensure **reproducibility** of research outputs

providing **open access** to research outputs (e.g. publications, data, software, models, algorithms, and workflows) through deposition in trusted repositories

participation in **open peer-review**

involving all relevant knowledge actors including citizens, civil society and end users in the co-creation of R&I agendas and contents (such as citizen science)

Recap: mandatory/ recommended Open Science practices in Horizon Europe

Open Science practices

What?	How?	Mandatory in all calls/recommended
Early and open sharing of research	Preregistration, registered reports, preprints, etc.	Recommended
Research output management	Data management plan (DMP)	Mandatory
Measures to ensure reproduciblity of research outputs	Information on outputs/tools/instruments and access to data/results for validation of publications	Mandatory
Open access to research outputs through deposition in trusted repositories	• Open access to publications • Open access to data • Open access to software, models, algorithms, workflows etc.	• Mandatory for peer-reviewed publications • Mandatory for research data but with exceptions ('as open as possible…') • Recommended for other research outputs
Participation in open peer-review	Publishing in open peer-reviewed journals or platforms	Recommended
Involving all relevant knowledge actors	Involvement of citizens, civil society and end-users in co-creation of content (e.g. crowd-sourcing, etc.)	Recommended

Slide courtesy Victoria Tsoukala, EC

2. Open Access: how to?

What does Open Access mean ?

Open access is online access at no cost for the end user of research outputs such as scientific publications, data or other engineered outcomes and processes *(e.g. software, models, algorithms, protocols and electronic notebooks)*. Open access often carries less restrictive copyright and licensing barriers than traditionally published works, for both the users and the authors.

Open access: **Offer specific information on how you will meet the open access requirements**, that is **deposition and immediate open access** to publications and open access to data (the latter with some exceptions and within the deadlines set in the DMP) through a **trusted repository,** and under **open licenses**.

Horizon Europe Programme Guide, 2023

Slide from: Giglia, Elena. (2021, November 4). Open Science cafè - Open Access alle pubblicazioni in Horizon Europe. Zenodo.

Open Access to Scientific publications in Horizon Europe

Model Grant Agreement Art.17

Open science: open access to scientific publications

The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:

1 - at the latest at the time of publication, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications

2 - immediate open access is provided to the deposited publication via the repository, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights; for monographs and other long-text formats, the licence may exclude commercial uses and derivative works (e.g. CC BY-NC, CC BY-ND) and

3 - information is given via the repository about any research output or any other tools and instruments needed to validate the conclusions of the scientific publication.

Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the open access requirements.

Metadata of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine-actionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication.

Horizon Europe Model Grant Agreement, 2021

2.1 Choosing a repository and licensing your work

What's a repository?

A repository is an online archive, where researchers can deposit digital research outputs and provide (open) access to them.
Repositories help manage and provide access to scientific outputs and contribute to the long term preservation of digital assets.

They can be institutional, operating with the purpose to collect, disseminate and preserve digital research outputs of individual research organisations (institutional repositories, e.g. the repository of University X) or domain-specific, operating to support specific research communities and supported/endorsed by them (e.g. Europe PMC for life sciences including biomedicine and health or arXiv for physics, mathematics, computer science, quantitative biology, quantitative finance and statistics; Phonogrammarchiv for audiovisual recordings; the CLARIN-DK-UCPH Repository for digital language data or the European Nucleotide Archive or databases of astronomical observations operated by the European Southern Observatory, among others).

There are also general-purpose repositories, such as for example Zenodo, developed by CERN.

⚠️ Personal websites and databases, publisher websites, as well as cloud storage services (Dropbox, Google drive, etc) are NOT considered repositories.

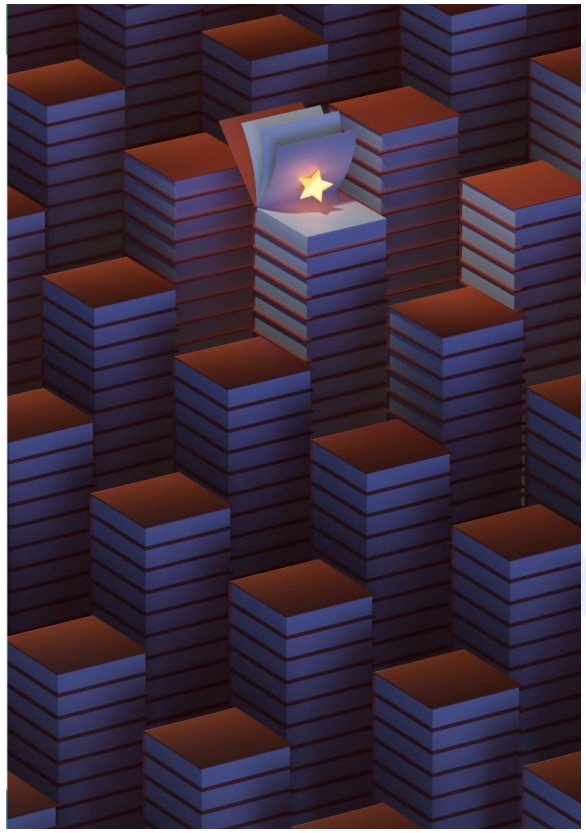
⚠️Academia.edu, ResearchGate and similar platforms do not allow open access under the terms required and therefore are also NOT considered repositories.

Image by Jiawei Zhao on Unsplash

What's a trusted repository?

Trusted repositories can be grouped into three categories which may overlap:

- certified repositories, such as those certified by international organisations or government-authorised certification bodies *(e.g. CoreTrustSeal, nestor Seal DIN31644, ISO16363)*
- disciplinary or domain repositories commonly used and endorsed by the research communities, and which are recognised internationally
- general-purpose repositories, institutional repositories or any other repositories that present the essential characteristics of trusted repositories, i.e.:
  - display specific characteristics of organisational, technical and procedural quality, *such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents*, thus facilitating their use and re-use in the short- and long-term. Trusted repositories have specific provisions in place and offer explicit information online about their policies, which define their services *(e.g. acquisition, access, security of content, long-term sustainability of service including funding, etc)*
  - provide broad, equitable and ideally open access to content free at the point of use, as appropriate, and respect applicable legal and ethical limitations. They assign persistent unique identifiers to contents *(e.g. DOIs, handles, etc)*, such that the contents *(publications, data and other research outputs)* are unequivocally referenced and thus citeable. They ensure that contents are accompanied by metadata sufficiently detailed and of sufficiently high quality to enable discovery, reuse and citation and contain information about provenance and licensing. Their metadata is machine-actionable and standardized *(e.g. Dublin Core, Data Cite, etc)* preferably using common non-proprietary formats and following the standards of the respective community the repository serves, where applicable
  - facilitate mid- and long-term preservation of the deposited material. They have mechanisms or provisions for expert curation and quality assurance for the accuracy and integrity of datasets and metadata, as well as procedures to liaise with depositors where issues are detected. They meet generally accepted international and national criteria for security to prevent unauthorized access and release of content and have different levels of security, depending on the sensitivity of the data being deposited, to maintain privacy and confidentiality.



Repository Open Access

A repository stores Open Access digital objects and makes them available and downloadable. It's accessible and interoperable through a OAI-PMH protocol and it deploys a **long-term archiving policy**

- Institutional
- Thematic/Disciplinary
- Literature
- Data
- Catch All

How do you choose a Repository?

Literature Repositories: Open Access Repository Directory
www.opendoar.org
Data Repositories: Registry of Research Data Repository
www.re3data.org

Source: slide n. 66, E. Lazzeri, F. Di Donato, Open Science: Why it is important, 10.5281/zenodo.4317277

Why do you need to deposit in a Repository?

Preserve

Repositories are managed by institutions, countries, transnational infrastructures or solid scientific communities that implement long-term curation and preservation of contents

A repository provides a public interface that allows anyone to access the metadata of digital objects. The author can assign different access rights for attachments (open, restricted, closed, embargoed, ...)

Share

Source: slide n. 67, E. Lazzeri, F. Di Donato, Open Science: Why it is important, 10.5281/zenodo.4317277

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Consiglio Nazionale delle Ricerche

CC Licenses

6 types of CC licenses

From most to least permissive

CC BY: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

CC BY includes the following elements:
BY – Credit must be given to the creator

CC BY-SA: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.

CC BY-SA includes the following elements:
BY – Credit must be given to the creator
SA – Adaptations must be shared under the same terms

CC BY-NC: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.

It includes the following elements:
BY – Credit must be given to the creator
NC – Only noncommercial uses of the work are permitted

CC BY-NC-SA: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.

CC BY-NC-SA includes the following elements:
BY – Credit must be given to the creator
NC – Only noncommercial uses of the work are permitted
SA – Adaptations must be shared under the same terms

CC BY-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, and only so long as attribution is given to the creator. The license allows for commercial use.

CC BY-ND includes the following elements:
BY – Credit must be given to the creator
ND – No derivatives or adaptations of the work are permitted

CC BY-NC-ND: This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.

CC BY-NC-ND includes the following elements:
BY – Credit must be given to the creator
NC – Only noncommercial uses of the work are permitted
ND – No derivatives or adaptations of the work are permitted

The Creative Commons Public Domain Dedication

CC0 (aka CC Zero) is a public dedication tool, which allows creators to give up their copyright and put their works into the worldwide public domain. CC0 allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, with no conditions.

https://creativecommons.org/about/cclicenses/

Before applying a CC License or CC0 to your work, be aware:

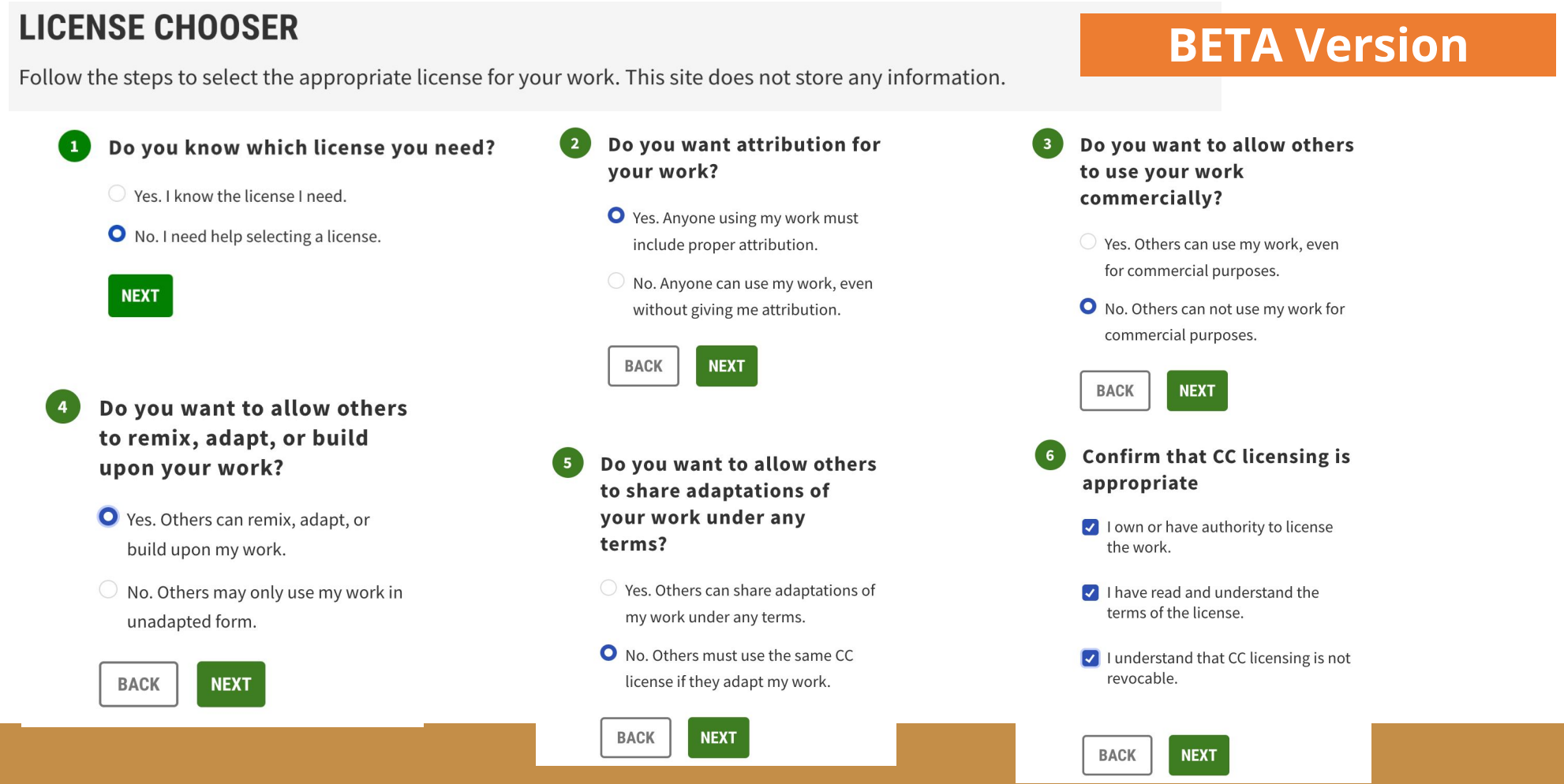
⚠️ licenses and CC0 cannot be revoked

⚠️ You must own or control copyright in the work

Photo by Justin Chrn on Unsplash

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

How to choose the right CC License with the Creative Commons License Chooser



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

Another example with the License chooser

1 **License Expertise**
I need help selecting a license.

2 **Attribution**
Anyone using my work must include proper attribution

3 **Commercial Use**
Others can not use my work for commercial purposes.

4 **Derivative Works**
Others can remix, adapt, or build upon my work.

5 **Sharing Requirements**
Others must use the same CC license if they adapt my work.

6 **Confirm that CC licensing is appropriate**
I confirmed the appropriateness of CC licensing.

RECOMMENDED LICENSE

CC BY-NC-SA 4.0

Attribution-NonCommercial-ShareAlike 4.0 International

This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.

BY: Credit must be given to you, the creator.

NC: Only noncommercial use of your work is permitted.
Noncommercial means not primarily intended for or directed towards commercial advantage or monetary compensation.

SA: Adaptations must be shared under the same terms.

See the License Deed

MARK YOUR WORK

Choose the kind of work to get appropriate license code or public domain marking.

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

To recap

	YOU CAN			YOU MUST	YOU MAINTAIN
	Share (copy and redistribute the material in any medium or format)	**Use for commercial purposes**	**Adapt** (remix, transform and build upon the material)	**Attribute** Give appropriate credit, provide a link to the license, and indicate if changes were made.	**Copyright, database rights**
CC BY	Yes	Yes	Yes	Yes	Yes
CC BY NC	Yes	No	Yes	Yes	Yes
CC BY ND	Yes	Yes	No If you remix, transform, or build upon the material, you may not distribute the modified material.	Yes	Yes
CC BY NC ND	Yes	No	No	Yes	Yes
CC0	Yes	Yes	Yes	No	No: Waived

2.2 Metadata and Persistent Identifiers

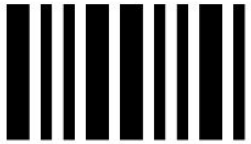
Metadata

- Data describing data
- Very important for:
  - Access
  - Comprehension
  - Process
- Use your discipline specific standards: you will spend less time **curating and interpreting data** and more time to actually make science!

Source: slide n. 32 , E. Lazzeri, F. Di Donato, FAIR principles and Open data, 10.5281/zenodo.4450515, slightly modified

Persistent Identifiers

- A **persistent identifier** (PI or PID) is a long-lasting reference to a document, file, web page, or other object.
- The term persistent identifier is usually used in the context of **digital objects** that are accessible over the Internet.
- Typically, such an identifier is not only persistent but **actionable**: you can plug it into a web browser and be taken to the identified source.
- It is like the barcode used on products…

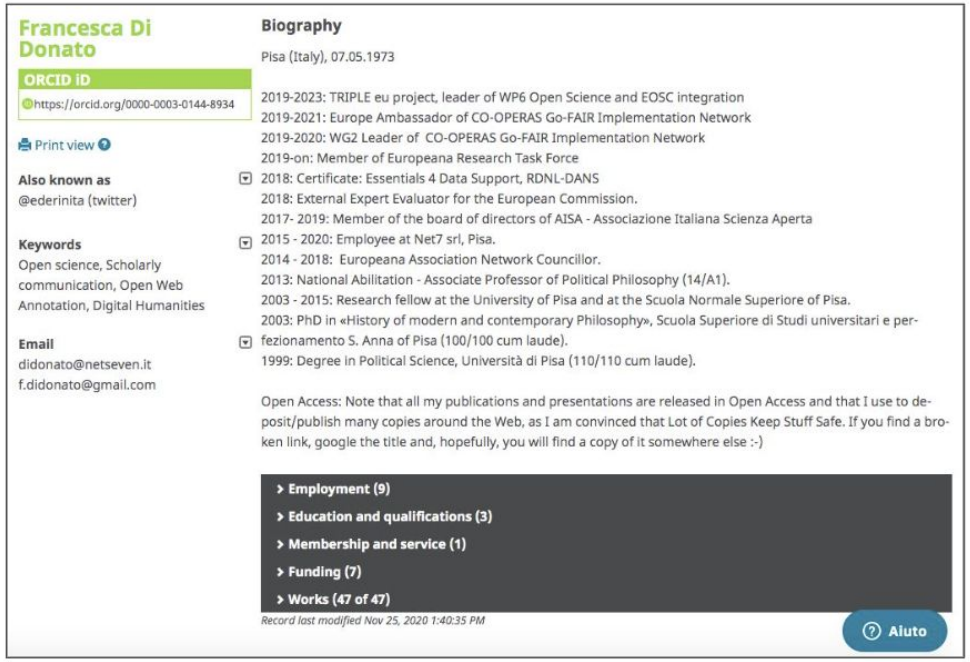


Source: slide n. 38 , E. Lazzeri, F. Di Donato, FAIR principles and Open data, 10.5281/zenodo.4450515

Orcid

The Open Researcher and Contributor ID (ORCID) is a nonproprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors

Do you have one? **You should…**



Definition: Wikipedia
https://orcid.org/

Source: slide n. 40 , E. Lazzeri, F. Di Donato, FAIR principles and Open data, 10.5281/zenodo.4450515, slightly modified

DOI – Digital Object Identifier

- In computing, a **digital object identifier** (DOI) is a persistent identifier or handle used to identify objects uniquely, standardized by the International Organization for Standardization (ISO).
- A DOI aims to be **resolvable**, usually to some form of access to the information object to which the DOI refers.
- This is achieved by **binding the DOI to metadata** about the object, such as a URL, **indicating where** the object can be found
- a DOI differs from identifiers such as ISBNs and ISRCs which aim only to identify their referents uniquely

Definition: Wikipedia

Source: slide n. 41 , E. Lazzeri, F. Di Donato, FAIR principles and Open data, 10.5281/zenodo.4450515

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Aspects to think about to make your research outputs FAIR

- **Documentation**

  Give the context to make your data understandable by others

- **Metadata**

  Make your data easy to find

- **Data formats**

  Make your data simple to combine to other data and machine readable.

- **Access to data**

  It means to decide who will have access to your data and how

- **Persistent identifiers**

  Persistent links to data that allows other to find and cite (give credit to) your data.

- **Licenses**

  Are used to tell others how they can reuse your data.

Image by Joanna Kosinska

Where to find support?

"S-LÉGAMI! Open Access - Manuale d'uso per ricercatori", a cura di Matteo Di Rosa, Claudia Iasillo et al., 2022, APREquaderni

Any Questions?

Photo by Camylla Battani on Unsplash

3. Get started with Zenodo

How to use Zenodo - a video tutorial

Zenodo Factsheet
OpenAIRE

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

Zenodo upload guide (mandatory and recommended)

1. Files	choose the file you want to upload
2. Communities* (recommended)	choose the community
3. Upload type	select the type of document (poster, deliverable, lesson…)
4. Basic information	DOI , publication date, title, authors (and ORCID). In description add the abstract. Version, language used, key words are recommended.
5. License	Choose if you want to deposit in OA and select the corresponding license under which you wish to share your work
6. Funding* (recommended)	Funder and Grant agreement number
7. Related/ alternate identifiers* (recommended)	Specify if your work is related to other kinds of outputs (PID)

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Consiglio Nazionale delle Ricerche

zenodo Search Upload Communities provostlottie@gmail.com

Search uploads... New Upload

New upload

Instructions: (i) Upload minimum one file and fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.

Files — Choose files — Start upload

Drag and drop files here

— or —

Choose files

(minimum 1 file required, max 50 GB per dataset - contact us for larger datasets)

If you're experiencing issues with uploading larger files, read our FAQ section on file upload issues.

Communities — recommended

Specify communities which you wish your upload to appear in. The owner of the community will be notified, and can either accept or reject your request.Please make sure your record complies with the content policy of the communities you add; reported abuse will be followed by account inactivation.

Start typing a community name...

Upload type — required

Publication | Poster | Presentation | Dataset | Image | Video/Audio | Software | Lesson | Physical object | Workflow | Other

Publication type Journal article

Publication type.

https://zenodo.org/

formation — required

Digital Object Identifier e.g. 10.1234/foo.bar

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.

Reserve DOI

Publication date * 2023-04-21

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

Title *

Required.

Authors * Family name, given names | Affiliation | ORCID (e.g.: 0000-0002-1825-0097)

Optional.

+ Add another author

Description * Source

Required.

Version

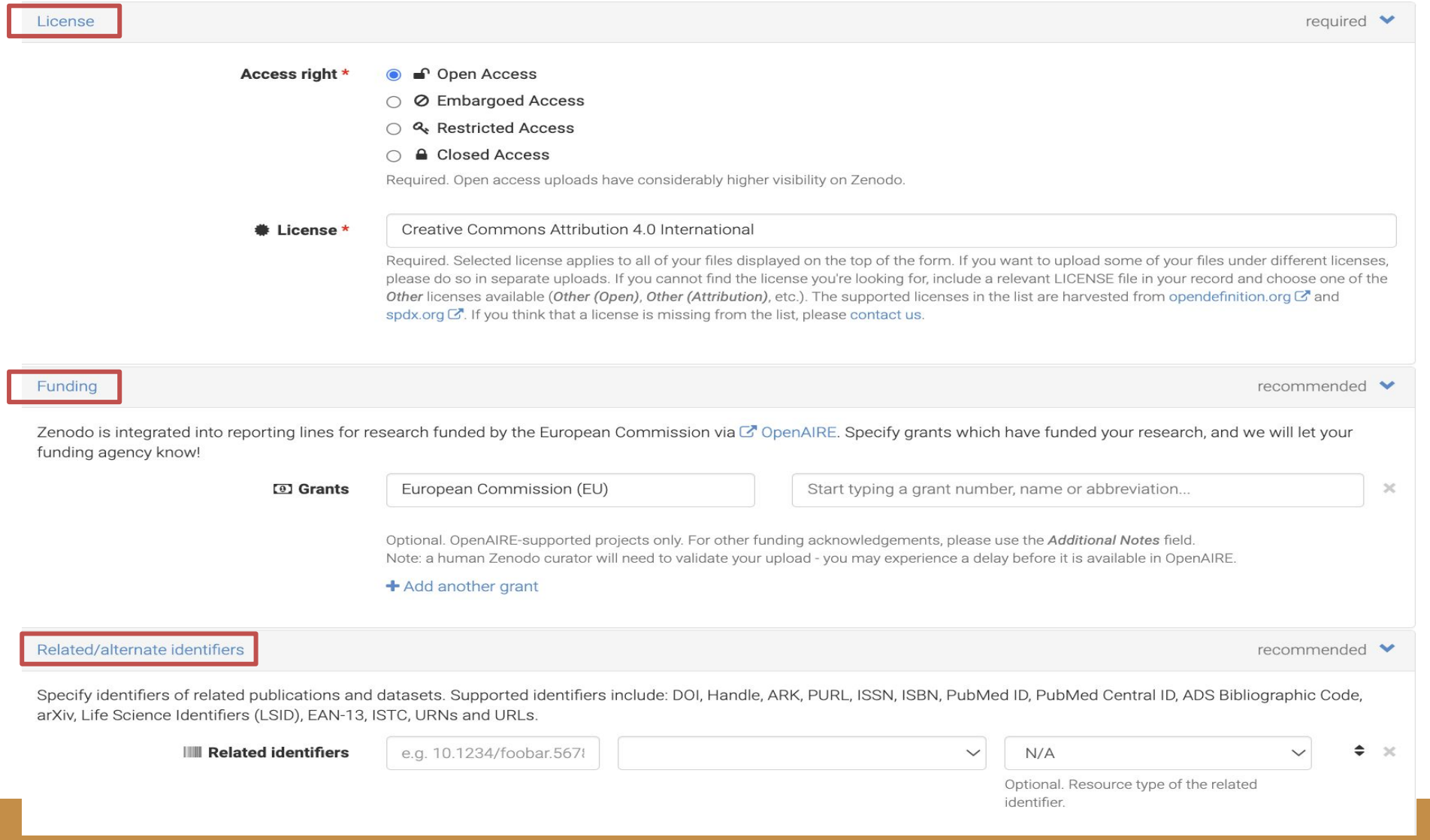
Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended.
See semver.org for more information on semantic versioning.

Language e.g.: 'eng', 'fr' or 'Polish'

Optional. Primary language of the record. Start by typing the language's common name in English, or its ISO 639 code (two or three-letter code).
See ISO 639 language codes list for more information.

Keywords

+ Add another keyword



Step by step 2/2

https://zenodo.org/

Consiglio Nazionale delle Ricerche

zenodo | Search | Upload | Communities | provostlottie@gmail.com

November 10, 2022 Workflow Open Access

TRIPLE Training Toolkit

Francesca Di Donato; Lottie Provost; Tiziana Lombardo; Michela Vignoli; Stefanie Pohle; Erzsébet Tóth-Czifra; Yin Chen; Emilie Blotière

Abstract →

The TRIPLE Training Toolkit is part of the work performed by Work Package 6 (WP6) under Task 6.3 in the TRIPLE Project (Transforming Research through Linked Interdisciplinary Exploration). The project is funded by the European Commission, under Grant Agreement No. 863420 and will run for 42 months starting from October 2019.

In light of the need for a common understanding of European Open Science advancements and to support the uptake of Open Science practices within SSH research and training communities, Task 6.3 produced two kinds of outputs.

- The TRIPLE Open Science Training Series is a series of 12 open and reusable training events specifically designed to upskill researchers in FAIR and Open Science. The organisation of the training series enabled a reflection on current challenges trainers face in making FAIR-by-design training resources and how to overcome them.
- The TRIPLE Training Toolkit is an open workflow for trainers to reproduce and adapt to organise training events following a FAIR-by-design method. It was created following the delivery of the TRIPLE Open Science Training Series.

The purpose of the TRIPLE Training Toolkit is to provide effective support to the research community in the uptake and application of Open Science and FAIR Data management practices within training activities and to address the frequent findability and reusability issues related to the management of digital training materials.

The Toolkit shows how the digital training materials created within the project are in line with the FAIR principles and enables for the experiment to be reproduced. It includes 11 reference documents referred to as reproducible templates that trainers can use and adapt to their needs along with illustrations of the process to facilitate the uptake of the method.

The following files are deposited in Zenodo to serve as a reference for those wishing to reproduce this experiment within their own institution or for their own training activities.

Please note: A first version (0.1) was deposited on Zenodo when the training series was still ongoing. Versions 2.0 and 2.1 contain inaccuracies, they are available on Zenodo but we recommend using version 3.0.

The first document to read is the README.

Edit

New version

Communities

Italian Computing and Data Infrastructure - ICDI ✖ Remove

OPERAS: open scholarly communication in the european research area for social sciences and humanities ✖ Remove

Indexed in OpenAIRE

Preview

Files (1.6 MB)

Name	Size	
01 README_TRIPLE_Training_Toolkit.docx md5:4b0ece11a2809c4a49624dd89ac7a47d	9.0 kB	Download
02 TRIPLE_Training_Toolkit_Workflow1.png md5:70aabd8b27d819e5ace23ec3224660d7	192.5 kB	Preview Download
03 TRIPLE_Training_Toolkit_Workflow2.png md5:20bf982a0e7d35a81200631260145ddd	263.0 kB	Preview Download
04 Guidelines_Organisation_TRIPLE_Training_Toolkit.docx md5:715eaf210cb0bb070859c78d51a7534f	19.4 kB	Download
05 To_Do_TRIPLE_Training_Toolkit.xlsx md5:bf84027705216553a38db6e55e727bcf	18.9 kB	Download
06 List_Past_Events_TRIPLE_Training_Toolkit.docx md5:13066646ee70649f1803f15aaac63a3f	159.0 kB	Download
07 Training_Objectives_Learning_Outcomes_TRIPLE_Training_Toolkit.docx md5:c8e7e77d2dee9d6c9e5de66712e065b2	227.4 kB	Download
08 Internal_Training_Needs_Survey_TRIPLE_Training_Toolkit.docx md5:faaf601c2ebdffc04cdb7c3a4398ee33	8.2 kB	Download
09a Internal_Training_Needs_Results1_TRIPLE_Training_Toolkit.png md5:d039f3de5d0e5fa1e9733bd1d338ffde	242.5 kB	Preview Download
09b Internal_Training_Needs_Results2_TRIPLE_Training_Toolkit.png md5:0214a2a908766357a9730b7c41d80829	234.8 kB	Preview Download
10 Post_Training_Survey_TRIPLE_Training_Toolkit.docx md5:476d8cf77b184cd85fba8554ef2ca4ff	8.6 kB	Download
11 Post_Training_Survey_Results_TRIPLE_Training_Toolkit .xlsx md5:e2143fc322d0f29bc86f04407bd76c98	44.9 kB	Download
11 Post_Training_Survey_Results_TRIPLE_Training_Toolkit.xlsx md5:95cd6c9a80119e58985626d9bc03d736	44.9 kB	Download
12 Promotion_Dissemination_Template_TRIPLE_Training_Toolkit.xlsx md5:0eedbf3c972fdfd94d230a99090ba58f	39.1 kB	Download
13 Enlarged_Audience_Template_TRIPLE_Training_Toolkit.xlsx md5:240a0cf51576dfaf16e9758f6a2d4620	70.1 kB	Download

← **Files**

Publication date:
November 10, 2022

DOI:
DOI 10.5281/zenodo.7309919 ←

Keyword(s):
Open Science | Training | TRIPLE | Workflow | FAIR principles | Reusability ←

Grants:
European Commission:
- TRIPLE - Transforming Research through Innovative Practices for Linked interdisciplinary Exploration (863420) ←

Related identifiers:

Cites
10.5281/zenodo.6207720 (Lesson)
10.5281/zenodo.5795376 (Lesson)
10.5281/zenodo.5566572 (Lesson) ←
10.5281/zenodo.5510388 (Lesson)
10.5281/zenodo.5045044 (Lesson)
10.5281/zenodo.5036684 (Lesson)
10.5281/zenodo.4707975 (Lesson)

Supplementary material
10.5281/zenodo.5544033 (Presentation)

Communities:
Italian Computing and Data Infrastructure - ICDI
OPERAS: open scholarly communication in the european research area for social sciences and humanities

Versions

Version	Date
Version 3.0 10.5281/zenodo.7309919	Nov 10, 2022
Version 2.1 10.5281/zenodo.7308477	Nov 9, 2022
Version 2.0 10.5281/zenodo.7243778	Nov 9, 2022
Version 0.1 10.5281/zenodo.6256198	Feb 25, 2022

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.6256197. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share

Cite as

Francesca Di Donato, Lottie Provost, Tiziana Lombardo, Michela Vignoli, Stefanie Pohle, Erzsébet Tóth-Czifra, Yin Chen, & Emilie Blotière. (2022). TRIPLE Training Toolkit (3.0). Zenodo. https://doi.org/10.5281/zenodo.7309919

Data Management tools to help you get started

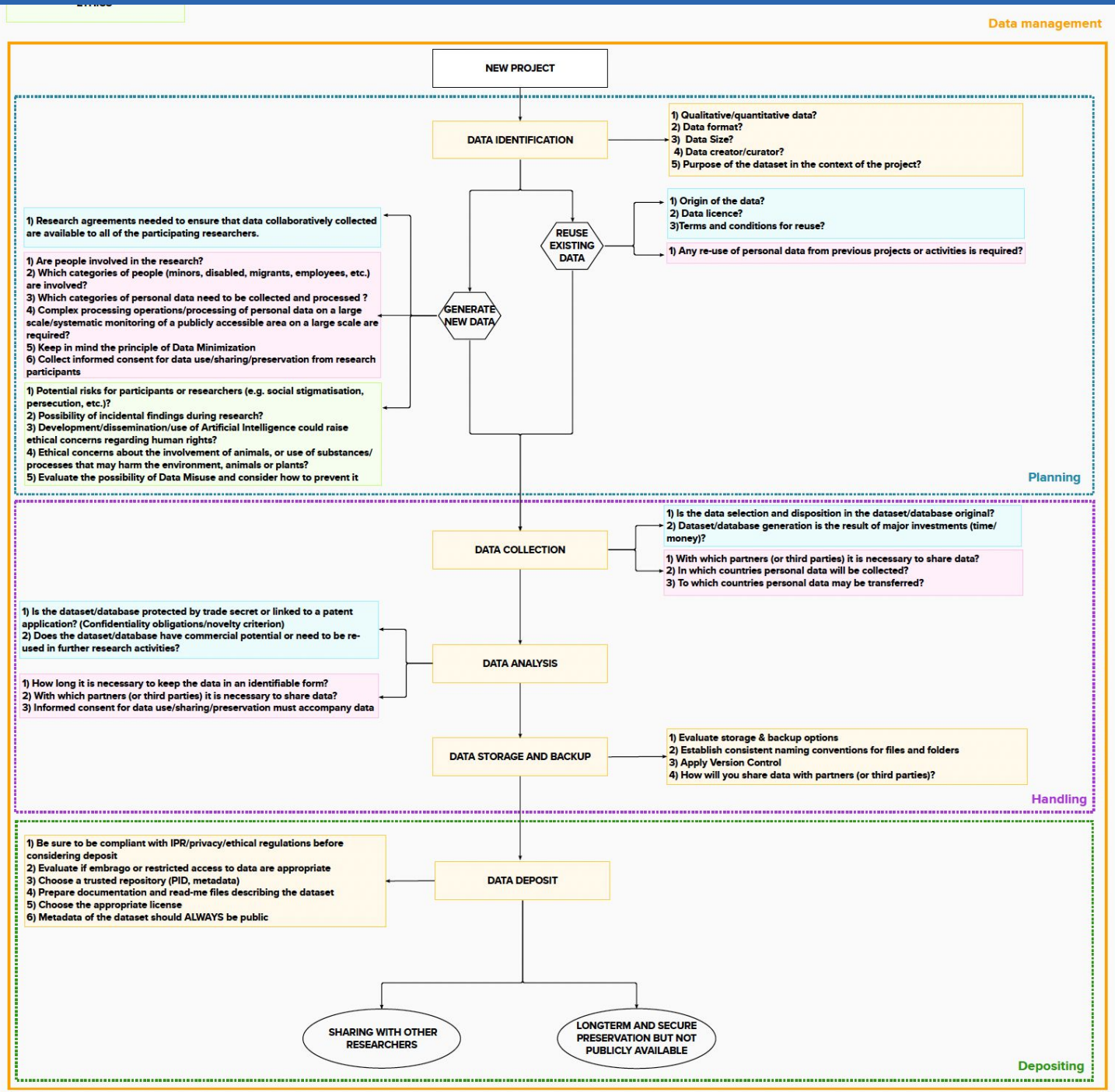
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Univers e della Rice

Decision Tree for Data Management

Legend:

- DATA MANAGEMENT
- INTELLECTUAL PROPERTY RIGHTS
- PRIVACY
- ETHICS



Caldoni, Giulia, Gualandi, Bianca, & Marino, Mario. (2022). Research Data Management Decision Tree. Zenodo.

DCC Checklist for a Data Management Plan

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	**Questions to consider:** - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	**Questions to consider:** - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) policy? - Does your funder have a Research Data Management policy? - Are there any formal standards that you will adopt? **Guidance:** List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.
Data Collection	
What data will you collect or create?	**Questions to consider:** - What type, format and volume of data? - Do your chosen formats and software enable sharing and long-term access to the data? - Are there any existing data that you can reuse? **Guidance:** Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.
How will the data be collected or created?	**Questions to Consider:** - What standards or methodologies will you use? - How will you structure and name your folders and files? - How will you handle versioning? - What quality assurance processes will you adopt? **Guidance:** Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning
	for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.
Documentation and Metadata	
What documentation and metadata will accompany the data?	**Questions to consider:** - What information is needed for the data to be to be read and interpreted in the future? - How will you capture / create this documentation and metadata? - What metadata standards will you use and why? **Guidance:** Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Data Management Best Practices Evaluation Checklist

File Formats

Retain the original, unedited outputs from software and hardware to preserve source data.

- ☐ Do not edit or alter the raw data file. Keep it in its native format and create a copy for editing or further manipulation.

Ensure future access to your data files by using standard, stable, commonly-used file formats.

- ☐ Non-proprietary formats are preferred (particularly for final versions).
- ☐ Be aware of what software is required to view and process data files, and be wary of software lifespans.

File Organization

A logical and organized folder structure can make it easier to keep track of project information.

- ☐ Avoid complex directory hierarchies and consider that folder names will sort alphabetically.
- ☐ Avoid keeping duplicate working copies of files (backup copies are not considered duplicates in this context).

Develop a file and folder naming convention and document it so all team members can follow it.

Good practices in choosing file and folder names:

- ☐ Uniquely name each file.
- ☐ Be consistent and include similar information in all file names of the same file type.
- ☐ Consider sorting order (usually lexicographic) and logical hierarchies in file directories.
- ☐ Avoid ambiguous and confusing names, such as 'MyData' or 'sample'
- ☐ Derivatives and versions should have similar (but differentiated) names to keep them co-located but still uniquely identified.
- ☐ Names should reflect the contents of the file and/or the stage of development.
  - When using dates, if you want the files to sort chronologically, put the year first and use numerical two-digit months and days (YYYY-MM-DD). (Example: March 7, 2004 would be written '2004-03-07'.)
- ☐ Use only alphanumeric characters but use dashes (-) or underscores (_) instead of spaces; avoid special characters such as colons (:) and slashes (/).
- ☐ Avoid using case differences to distinguish between files: 'Record', 'record', and 'RECORD' may be three different file names or the same file name, depending on the operating system.

Documentation

If possible, document data characteristics and workflows in a digital format at the time that data files are created or altered.

- ☐ Create readme files and data dictionaries to provide digital documentation of data characteristics, workflows, progress, results, software, etc.
- ☐ Document any database data field (variable/column) characteristics for later interpretation. Possible elements are field name, field description, and permitted range of values.
- ☐ Document data file and collection characteristics that are relevant for later interpretation. Possible elements are file name and path, relevant dates, creation method, and status.
- ☐ Digitize (scan) relevant paper laboratory or field notes so that they can be more easily shared along with the data.

Data files are more easily and accurately interpreted when they are documented using a formal, standardized metadata format.

- ☐ Consider using discipline-specific metadata specifications and schemas. (see a list of schemas at http://rd-alliance.github.io/metadata-directory/standards/)
- ☐ Develop strategies for streamlining the metadata data entry process, such as through the use of templates to input information that is consistent across all project metadata.
- ☐ Store the metadata file close to the data (or embedded if possible) to ensure discovery.

Storage and Security

Ensure data redundancy and replication, and avoid single points of failure.

- ☐ Never rely on a single copy of data. Have at least two backup copies (with at least one in a remote location) in addition to the working copy.
- ☐ Document your data storage system and data backup policy.
- ☐ Back up data regularly. Backups are particularly important if using portable media, such as laptops and flash drives.
- ☐ Use managed, networked storage whenever possible (Example: departmental network drive with system administrator).

Access and Use Restrictions

Data files may be protected by ownership rights or licenses.

- ☐ Ensure that you have proper permissions to use and share data, considering any license agreements or ownership issues.
- ☐ Document any access or use restrictions in the metadata, readme file, or data dictionary.

Protect sensitive and confidential information.

- ☐ Datasets that include confidential information should have that information de-identified or suppressed before being shared.

Source: https://www.library.ucsb.edu/sites/default/files/attachments/data-curation/resources/DataEvaluationChecklist.pdf

Any Questions?

Photo by Camylla Battani on Unsplash

THANK YOU!

Please write to lottiemiaprovost@cnr.it for any questions

fossr.dissemination@ircres.cnr.it

@fossrproject

fossr.eu

fossr-eu

@fossr

zenodo.org/communities/fossr