Sharing Research Data

where to deposit & publish Research Data

Pedro Principe, University of Minho

CONTEXT OF THE TRAINING IS ESSENTIAL – taking into account the context, the audience and the type of training (we all know is critical for the success)

BUT

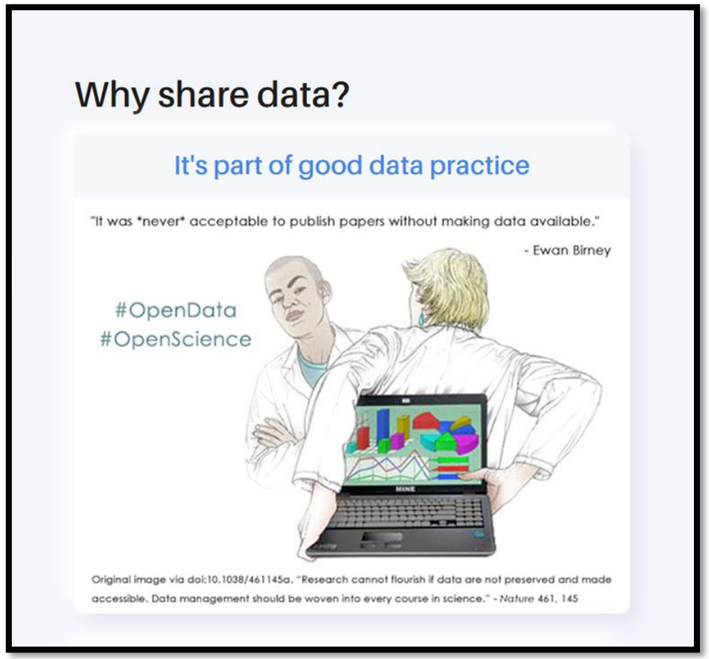
training on RDM in Open Science setting…

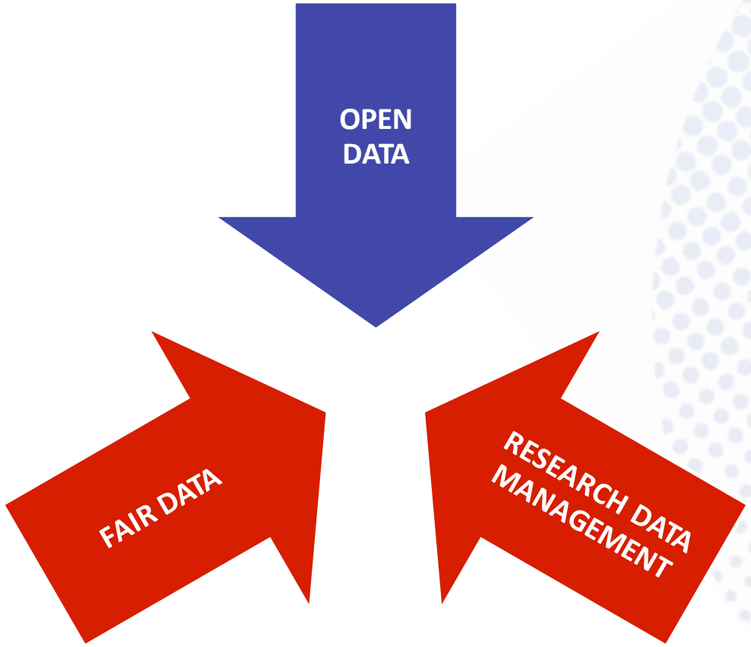
is critical that we focus on SHARING research data;

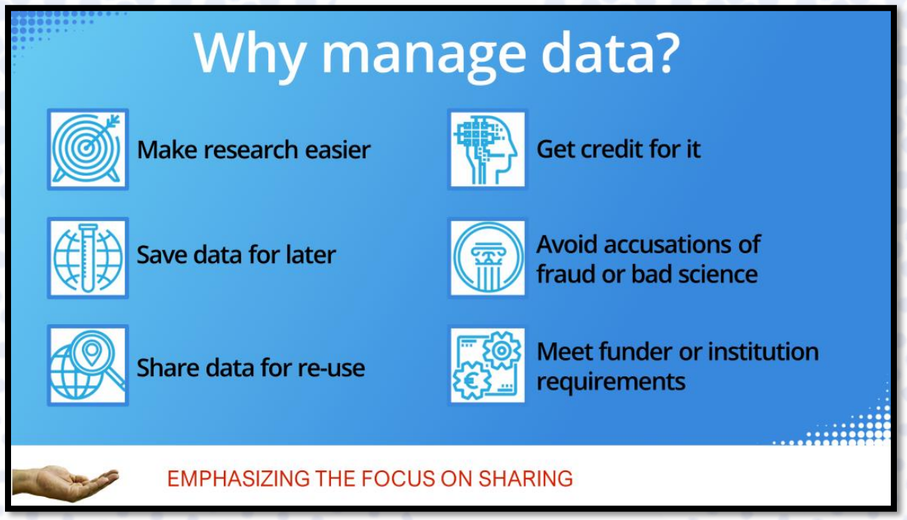
clarifying the meaning and implications of openness;

identifying the VENUES TO PUBLISH and deposit research data;

trying to minimize the complexity of the topics...

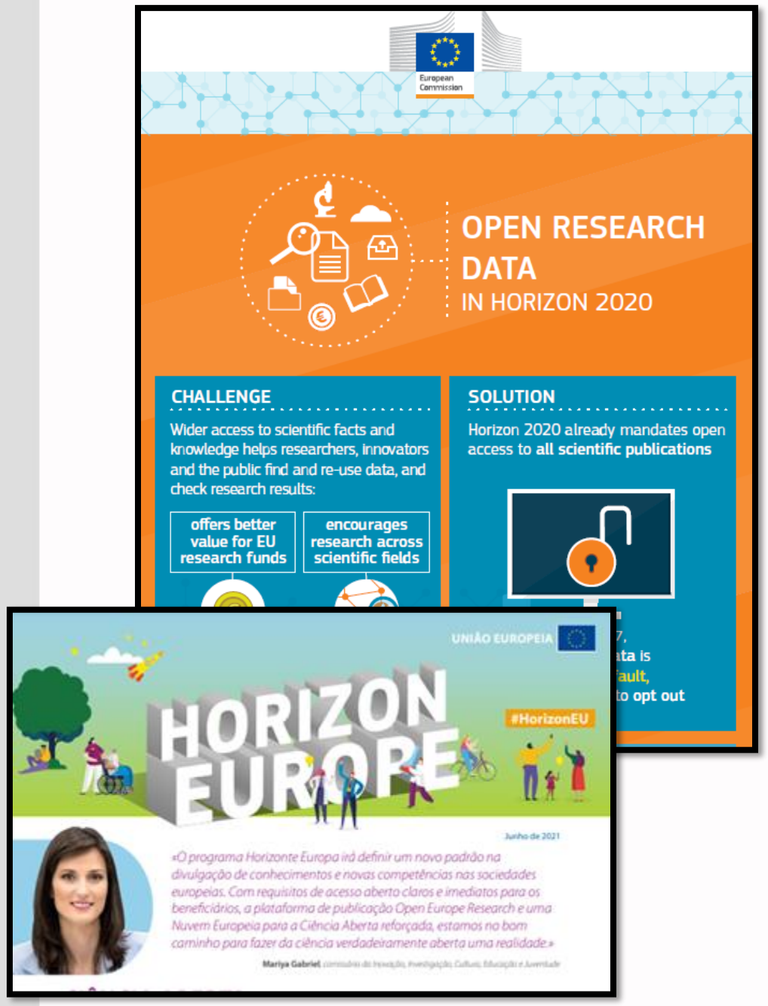






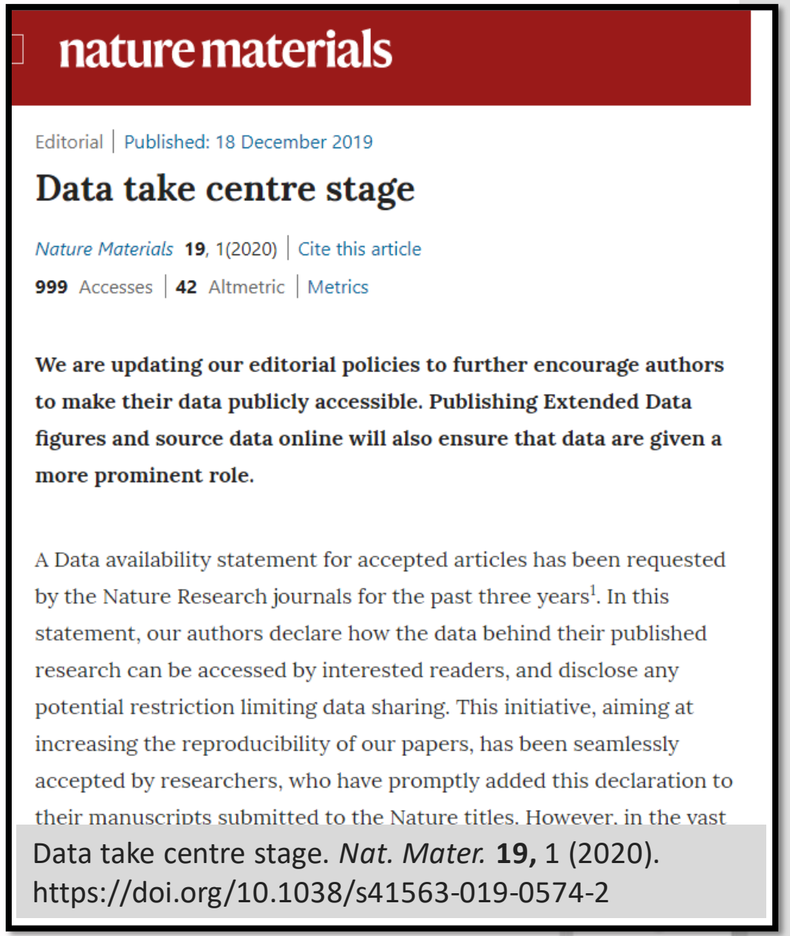
PUBLISH DATA

Universidade do Minho
Serviço de Documentação e Bibliotecas



ENSURE THAT DATA HAVE A MORE PROMINENT ROLE

- I ensure good data management practices (storage and documentation)
- I deposit data in trusted repositories with persistent identifiers.
- I associate data with publications.
- I publish the results of my projects in organized collections and for reuse.



nature materials

Editorial | Published: 18 December 2019

Data take centre stage

Nature Materials **19**, 1(2020) | Cite this article

999 Accesses | **42** Altmetric | Metrics

We are updating our editorial policies to further encourage authors to make their data publicly accessible. Publishing Extended Data figures and source data online will also ensure that data are given a more prominent role.

A Data availability statement for accepted articles has been requested by the Nature Research journals for the past three years[1]. In this statement, our authors declare how the data behind their published research can be accessed by interested readers, and disclose any potential restriction limiting data sharing. This initiative, aiming at increasing the reproducibility of our papers, has been seamlessly accepted by researchers, who have promptly added this declaration to their manuscripts submitted to the Nature titles. However, in the vast

Data take centre stage. *Nat. Mater.* **19,** 1 (2020). https://doi.org/10.1038/s41563-019-0574-2

PUBLISH DATA – Clarify concepts & diferences

AN IMPORTANT DIFFERENCE

- **Deposit**: upload a digital object (data, articles, …) on a platform that allows to correctly describe the object through medatada and that implements long-term preservation.
- **Give access**: once the object has been deposited, the authors can choose the type of access that can be granted (open, restricted, closed, embargoed,…) and assigns a licence to reuse the contents (Creative Commons)

WHAT IS THE DIFFERENCE BETWEEN SHARING, PUBLISHING & ARCHIVING?

SHARED: any way of sharing information, could mean I emailed it to you.

PUBLISH: citable artifact, discoverable.

ARCHIVE: long-term preservation.

Reference:
https://datacarpentry.org/rr-publication/01-publication/

DEPOSIT DATA

- **What to deposit?**
  - → everything needed to find, assess, understand, & reuse data

DATA

- open/common file formats
- use relevant standards for interoperability

METADATA ('data about data')

- highly structured, machine-readable
- fixed set of attributes (schema)
- use existing (domain-specific) standards

ANY OTHER DOCUMENTATION

- codebooks explaining variables
- study context, protocol, methods
- dataset structure
- notes/annotations
- software code
- machine configurations
- consent information
- …

Open Research Data policy requirements

DATA, including metadata, needed to validate the results in scientific publications.

Other data, including metadata, as specified in the Data Management Plan.

OpenAIRE

Universidade do Minho

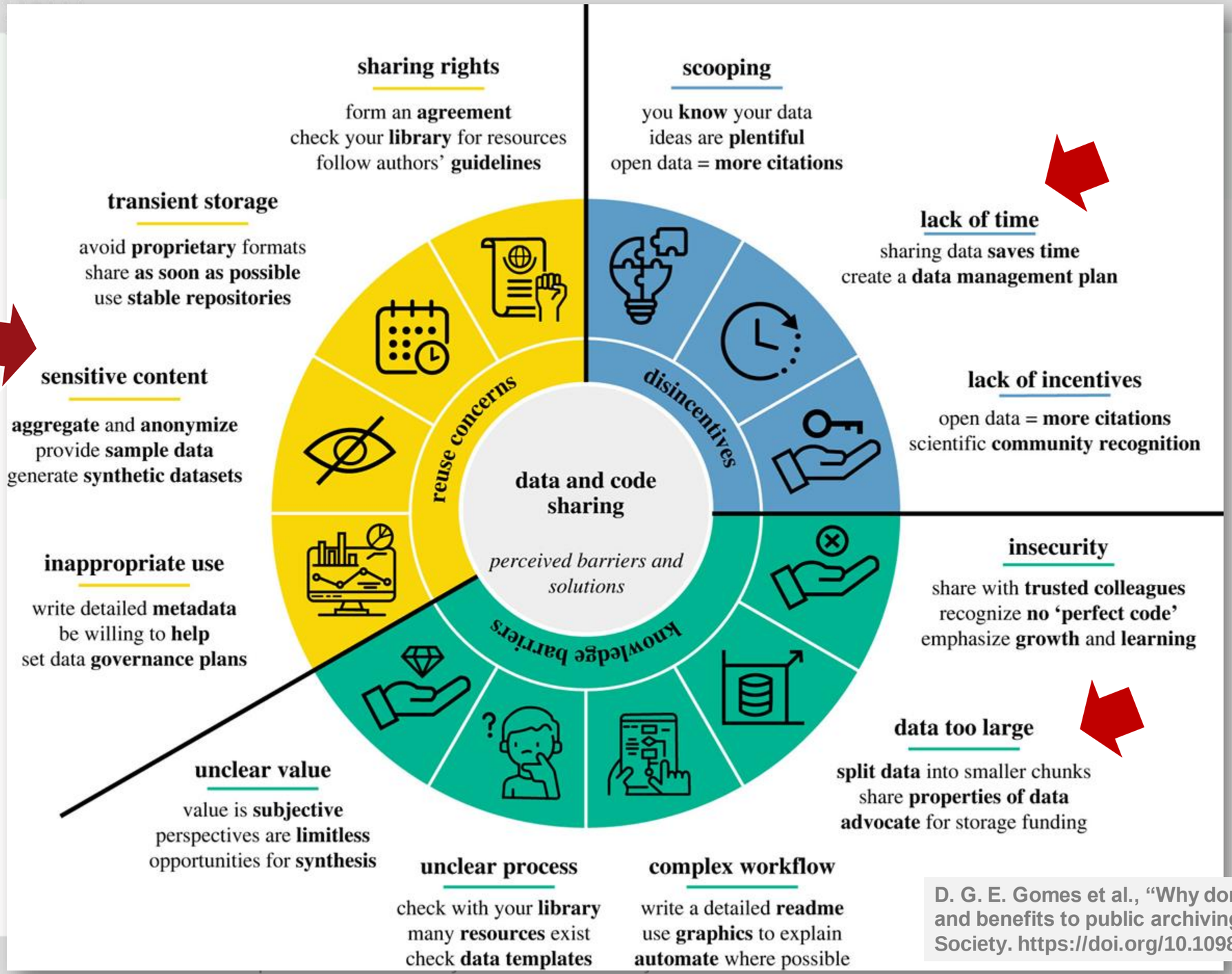
Open Research Data policy requirements

Write, and keep up-to-date, a Data Management Plan.

Deposit the data in a research data repository.

OpenAIRE

Universidade do Minho



Lack of incentives

Concerns

Several Barriers

D. G. E. Gomes et al., "Why don't we share data and code? Perceived barriers and benefits to public archiving practices," 2022, vol. 289, no. 1987. Royal Society. https://doi.org/10.1098/rspb.2022.1113

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

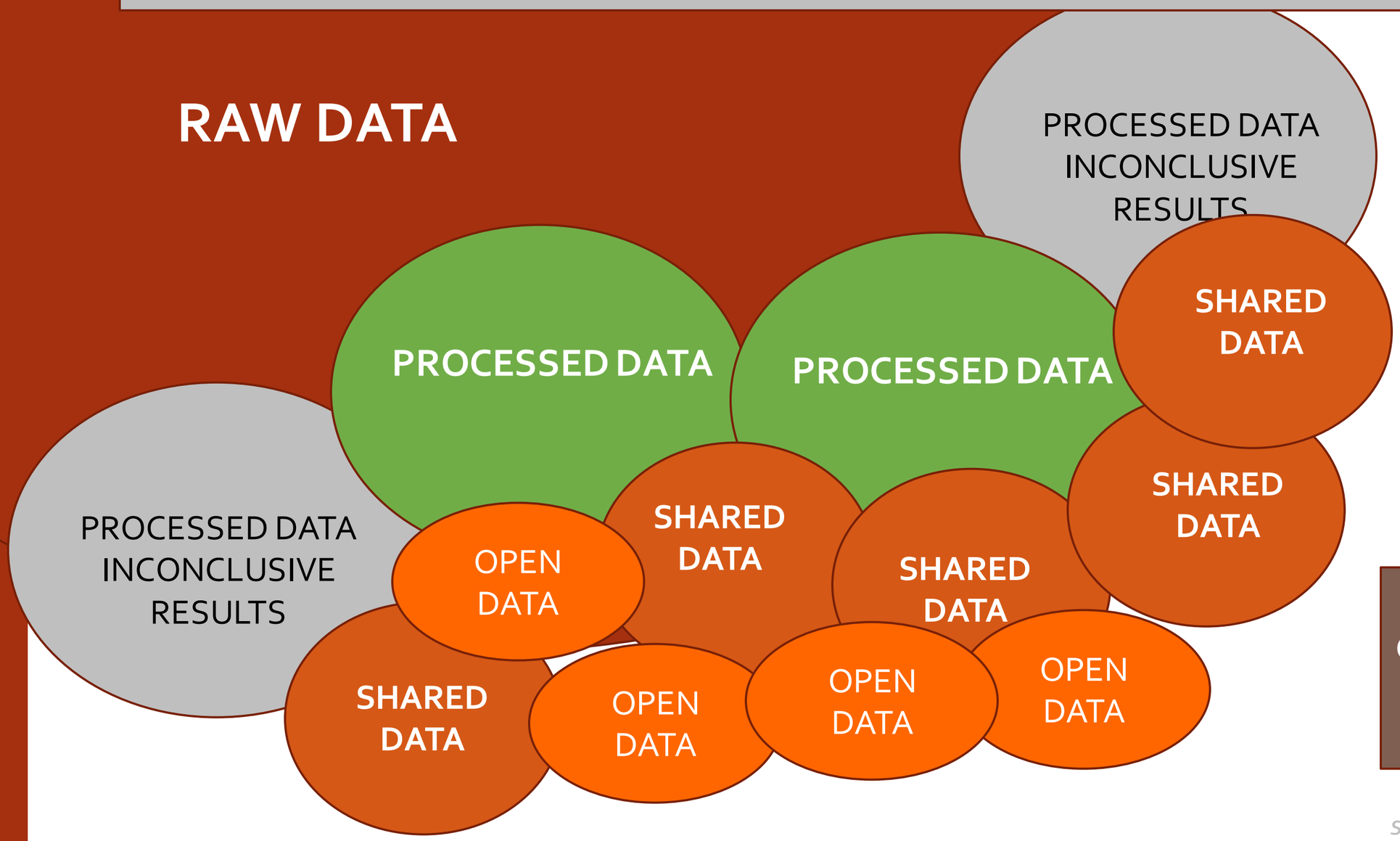
Grantees have the right to **opt-out**, but need to say **why**

intellectual property rights

privacy

might jeopardise project's main objective

Different levels of processing and sharing data



DATA:

Closed

Restricted

Published

Embargoed

Open

Compliance with legal and ethical issues

Slide based on Learn RDM project presentation

PUBLISH DATA – WHERE?

Journal service for supplementary material

- Meet publisher requirements
- Data available from published results
- It can be costly and risky with data rights
- Closed and unlikely access to ensure preservation

Institutional data repository

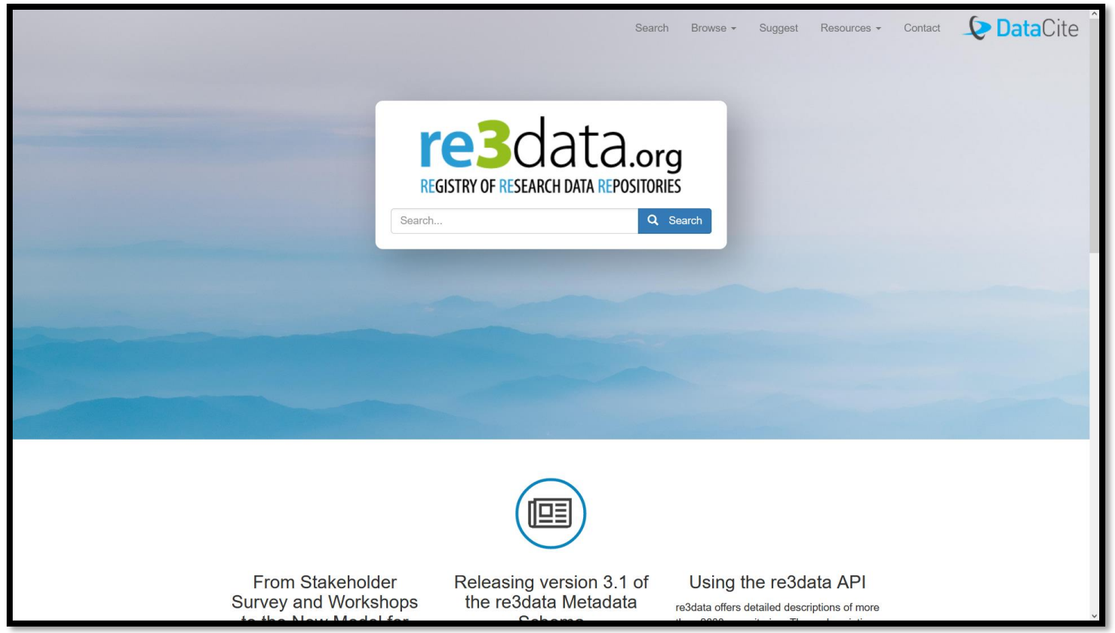
- Accept various types of data, ensure long-term access
- More reliable and there will be no costs
- May not offer long-term sustainable access
- May not have disciplinary metadata

Generic repository

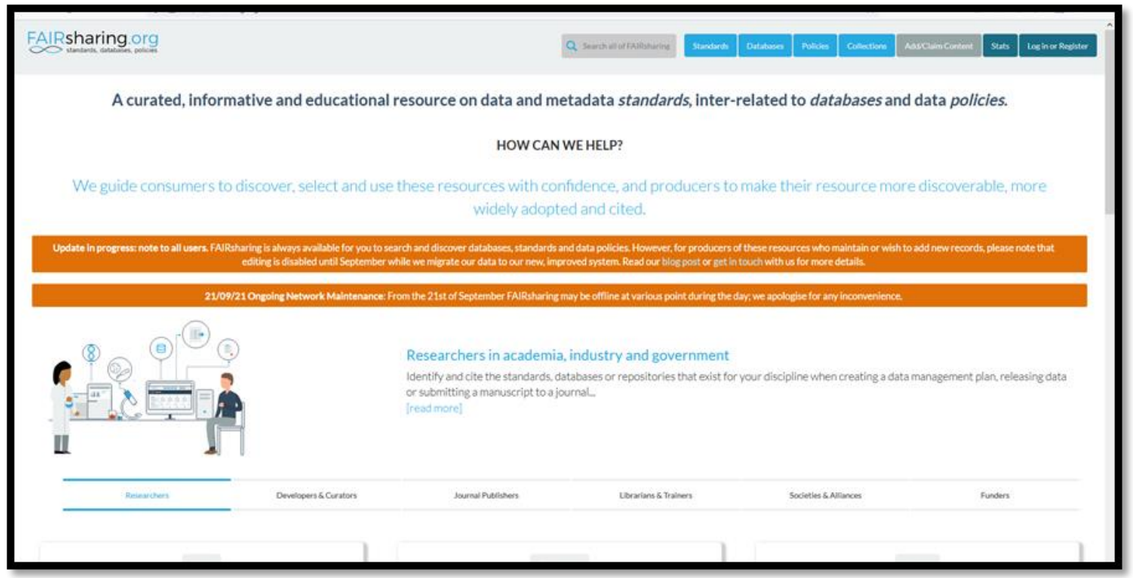
- Reach a wider audience.
- Accepts several types, suitable for interdisciplinary data
- Usually only simple metadata is available
- No editorial control over the quality of deposited materials

Disciplinary repository

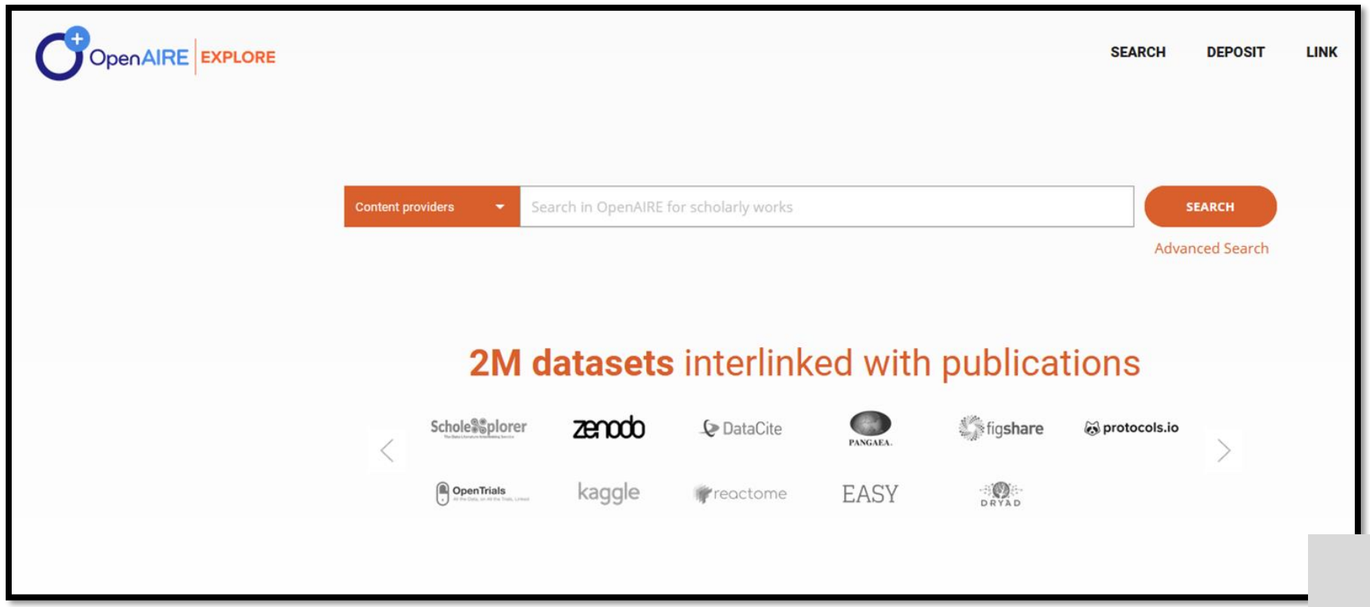
- Offers expertise and experience in data management
- Likely to accept complete data sets
- Selective in the type of data they accept
- Requires planning and high standards, may incur costs



https://www.re3data.org/



https://fairsharing.org/



https://explore.openaire.eu/

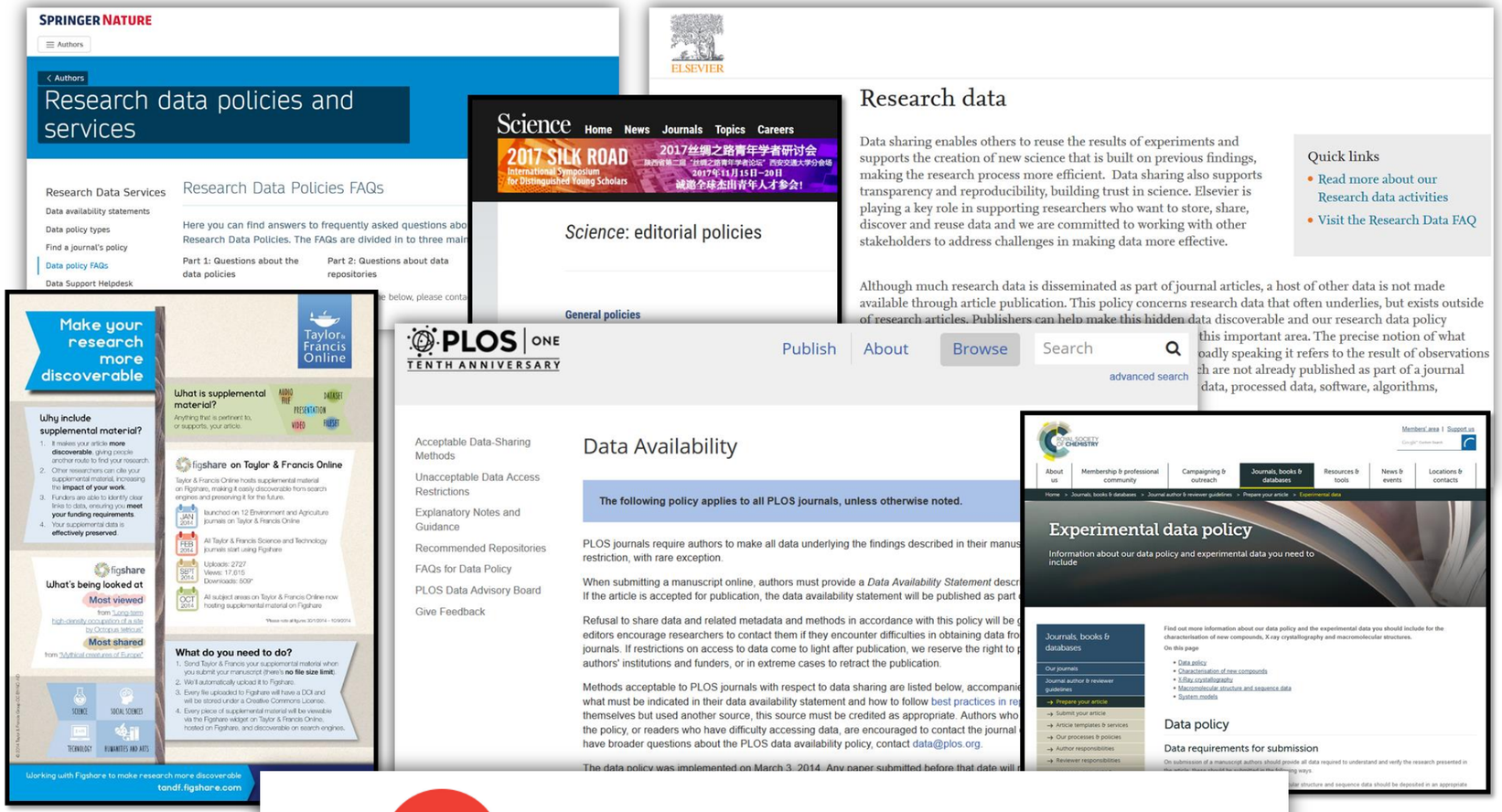
Journal service for supplementary material

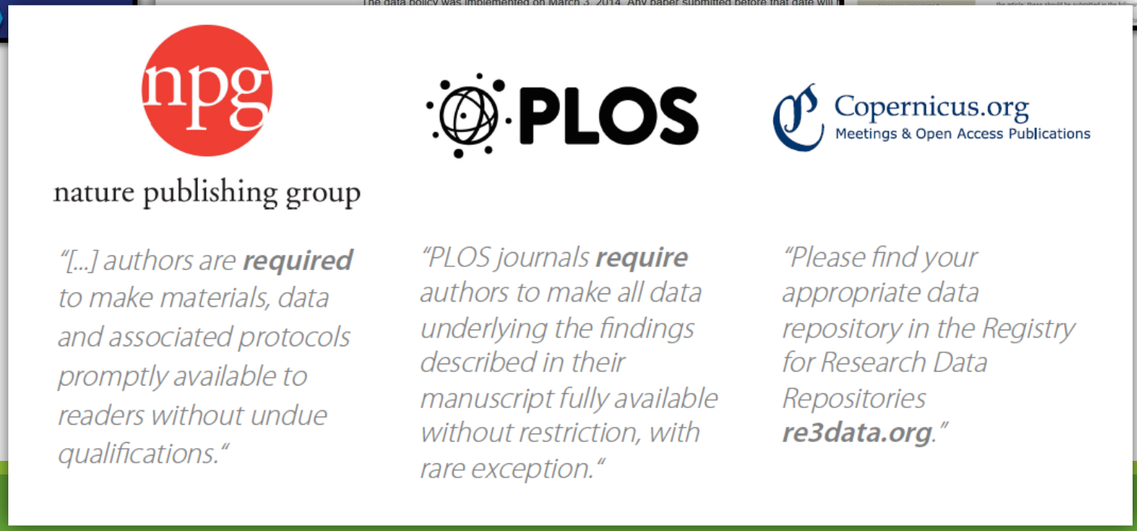
Meet publisher requirements

Data available from published results

It can be costly and risky with data rights

Closed and unlikely access to ensure preservation





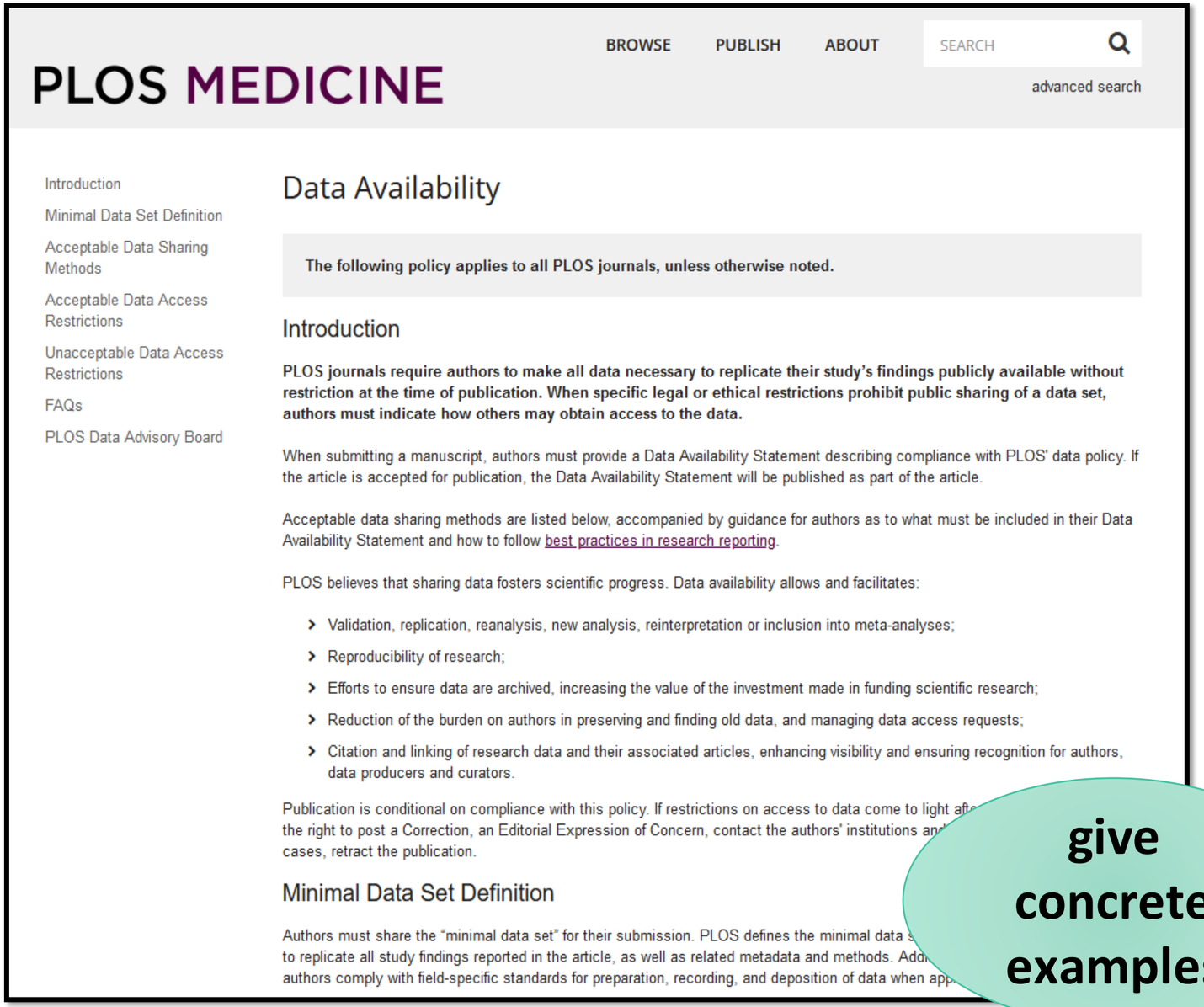
Journal service for supplementary material

Meet publisher requirements

Data available from published results

It can be costly and risky with data rights

Closed and unlikely access to ensure preservation



PLOS MEDICINE

BROWSE PUBLISH ABOUT SEARCH

advanced search

Introduction
Minimal Data Set Definition
Acceptable Data Sharing Methods
Acceptable Data Access Restrictions
Unacceptable Data Access Restrictions
FAQs
PLOS Data Advisory Board

Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

Introduction

PLOS journals require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data.

When submitting a manuscript, authors must provide a Data Availability Statement describing compliance with PLOS' data policy. If the article is accepted for publication, the Data Availability Statement will be published as part of the article.

Acceptable data sharing methods are listed below, accompanied by guidance for authors as to what must be included in their Data Availability Statement and how to follow best practices in research reporting.

PLOS believes that sharing data fosters scientific progress. Data availability allows and facilitates:

- Validation, replication, reanalysis, new analysis, reinterpretation or inclusion into meta-analyses;
- Reproducibility of research;
- Efforts to ensure data are archived, increasing the value of the investment made in funding scientific research;
- Reduction of the burden on authors in preserving and finding old data, and managing data access requests;
- Citation and linking of research data and their associated articles, enhancing visibility and ensuring recognition for authors, data producers and curators.

Publication is conditional on compliance with this policy. If restrictions on access to data come to light aft
the right to post a Correction, an Editorial Expression of Concern, contact the authors' institutions an
cases, retract the publication.

Minimal Data Set Definition

Authors must share the "minimal data set" for their submission. PLOS defines the minimal data
to replicate all study findings reported in the article, as well as related metadata and methods. Add
authors comply with field-specific standards for preparation, recording, and deposition of data when ap

give concrete examples

Data availability policy - publishers

Scenarios:

- **send the dataset to the publisher** and the publisher publishes the dataset online.
- the publisher asks the author to **deposit the dataset in a trusted repository** and to notify the publisher.
- the publisher asks the author to **give contact information** for those who wish to have access to the data.

The requirements are generally found on the journal's website.

A number of journals have a specific Data Availability or Data Archiving Policy

Data paper Journals

- Scientific Data (Nature)
  - https://www.nature.com/sdata
- Data in brief (Elsevier)
  - https://www.journals.elsevier.com/data-in-brief/
- Data (MDPI)
  - https://www.mdpi.com/journal/data
- Patterns (bio data intensive science)
  - https://www.cell.com/patterns

Data in Brief

› Open Access

OpenAIRE

Universidade do Minho
Serviço de Documentação e Bibliotecas

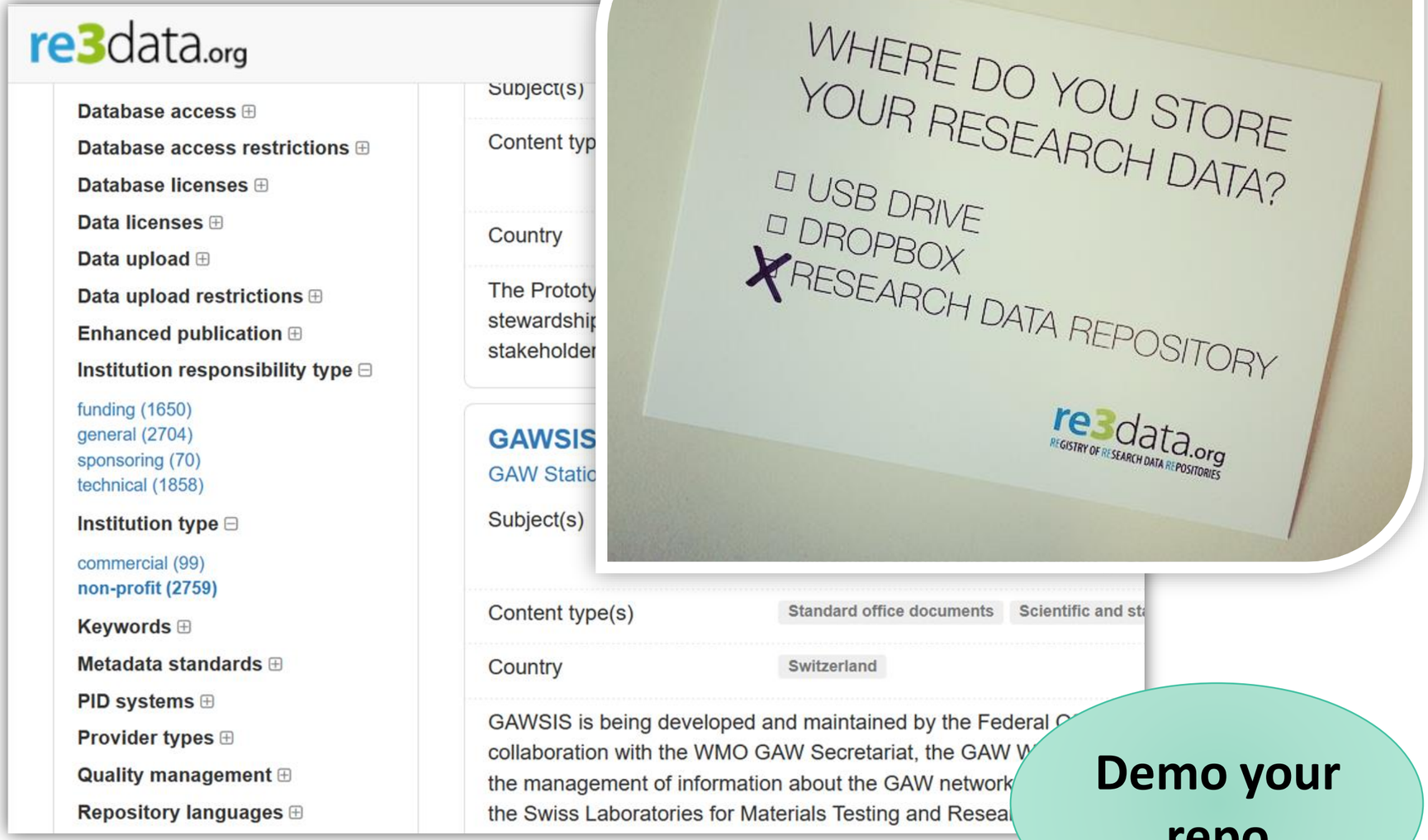
Institutional data repository

Accept various types of data, ensure long-term access

More reliable and there will be no costs

May not offer long-term sustainable access

May not have disciplinary metadata



Demo your repo

data RepositóriUM

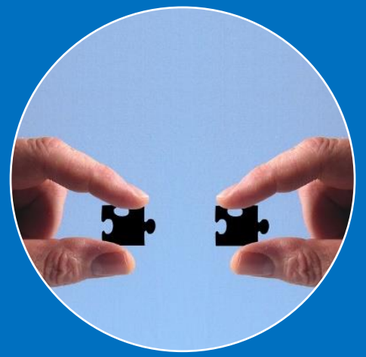
The Dataverse® Project

SHARE

Promote data sharing and the practice of Open Science.

PUBLISH

Make data available for reuse and generate greater impact.

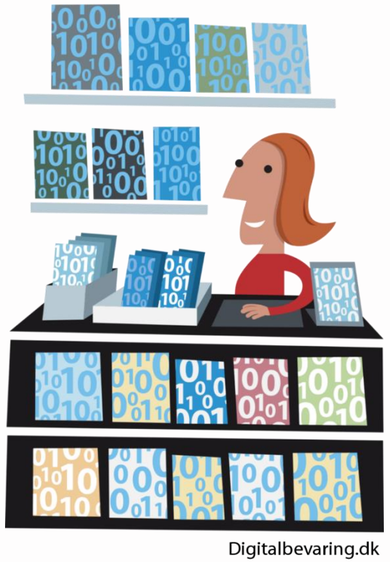


MANAGE

Ensure good documentation and data management practices

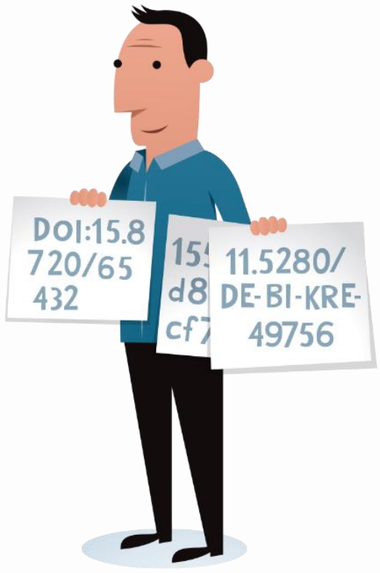
DataRepositóriUM / Dataverse highlights

SECURE DATA MANAGEMENT.



EFFECTIVE SHARING.

TRACK CHANGES.



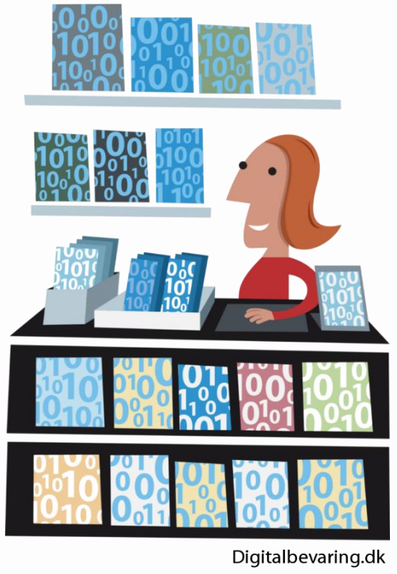
LONG-TERM ACCESS AND PRESERVATION.

DataRepositóriUM / Dataverse highlights

ORGANIZATION AND COMPATIBILITY.

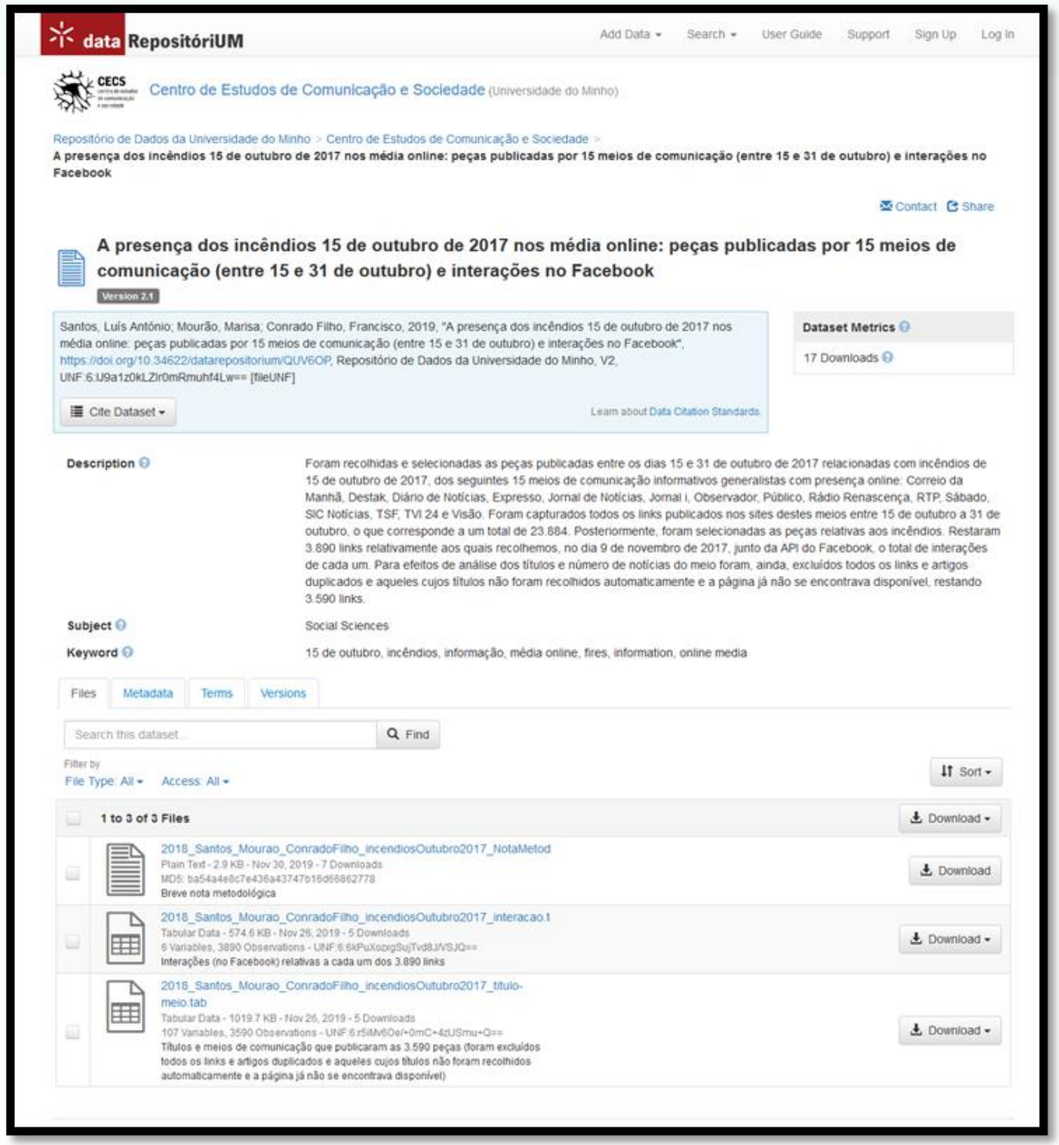
SAVE TIME.

INCREASE RESEARCH VISIBILITY.



MEET GRANT REQUIREMENTS.

DataRepositóriUM in practice



DOI

10.34622/datarepositorium/

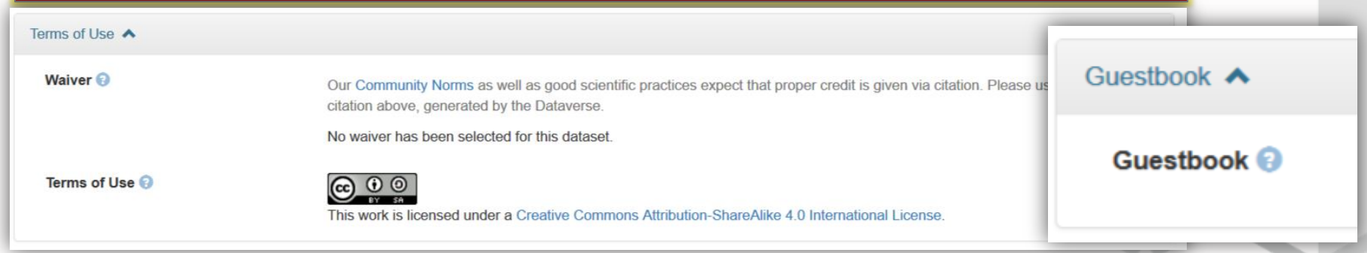
CITATION

Santos, Luís António; Mourão, Marisa; Conrado Filho, Francisco, 2019, "A presença dos incêndios 15 de outubro de 2017 nos média online: peças publicadas por 15 meios de comunicação (entre 15 e 31 de outubro) e interações no Facebook", https://doi.org/10.34622/datarepositorium/QUV6OP, Repositório de Dados da Universidade do Minho, V2, UNF:6:lJ9a1z0kLZlr0mRmuhf4Lw== [fileUNF]

Cite Dataset

Learn about Data Citation Standards.

TERMS OF USE & CONTROLED ACCESS



Generic repository

Reach a wider audience.

Accepts several types, suitable for interdisciplinary data

Usually only simple metadata is available

No editorial control over the quality of deposited materials

zenodo

DRYAD

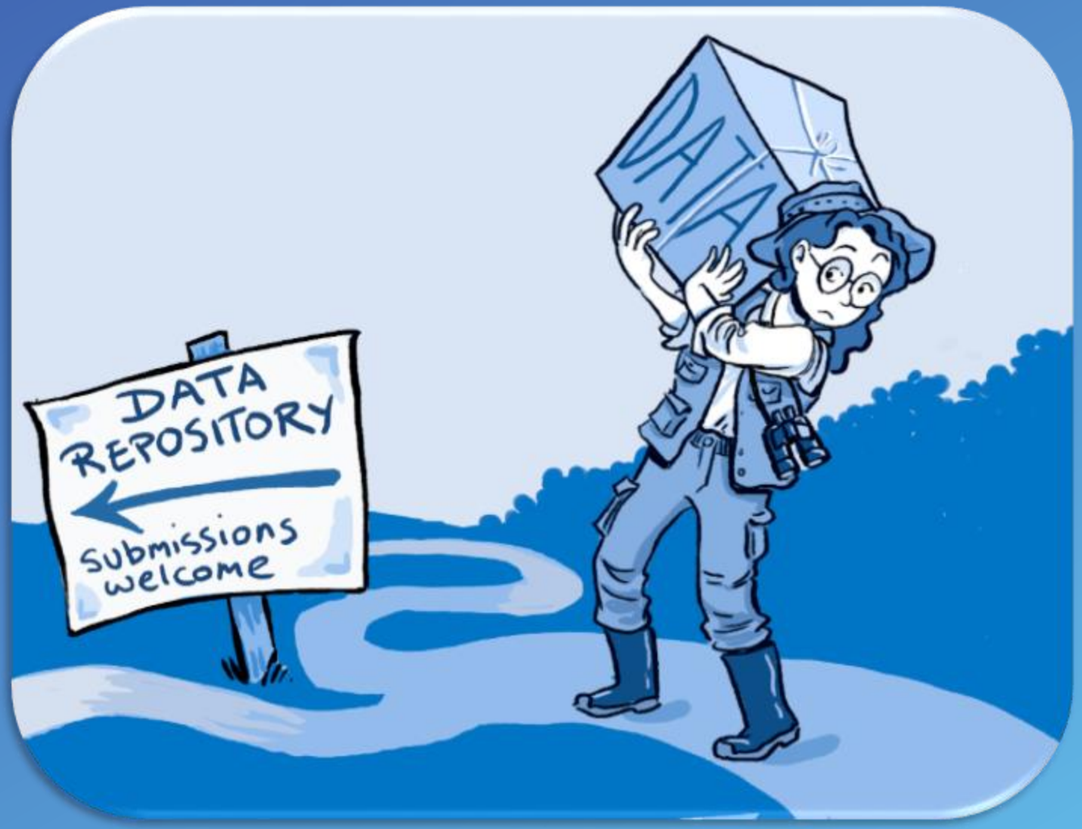
figshare

The Dataverse® Project

Cross-disciplinary repositories

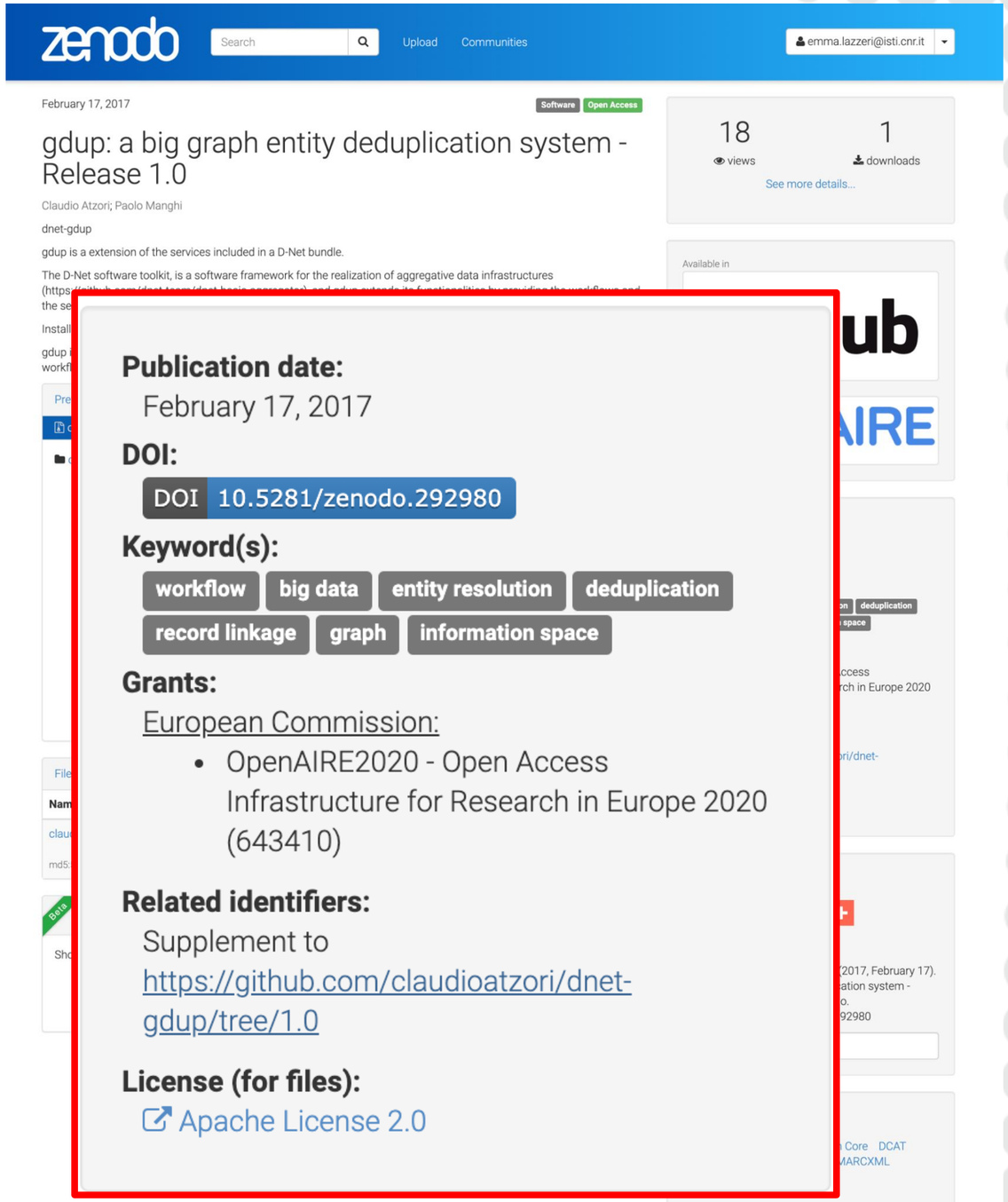
Long tail of research data

zenodo



Zenodo

- **Catch all repository** (CERN and OpenAIRE)
- Free **DOIs** for citation and enable **credit** mechanisms
- Enables tracking of usage (downloads and views) for **credit**
- Citable products are **findable**, reusable, therefore enable degrees of **R* ness**



Disciplinary repository

Offers expertise and experience in data management

Likely to accept complete data sets

Selective in the type of data they accept

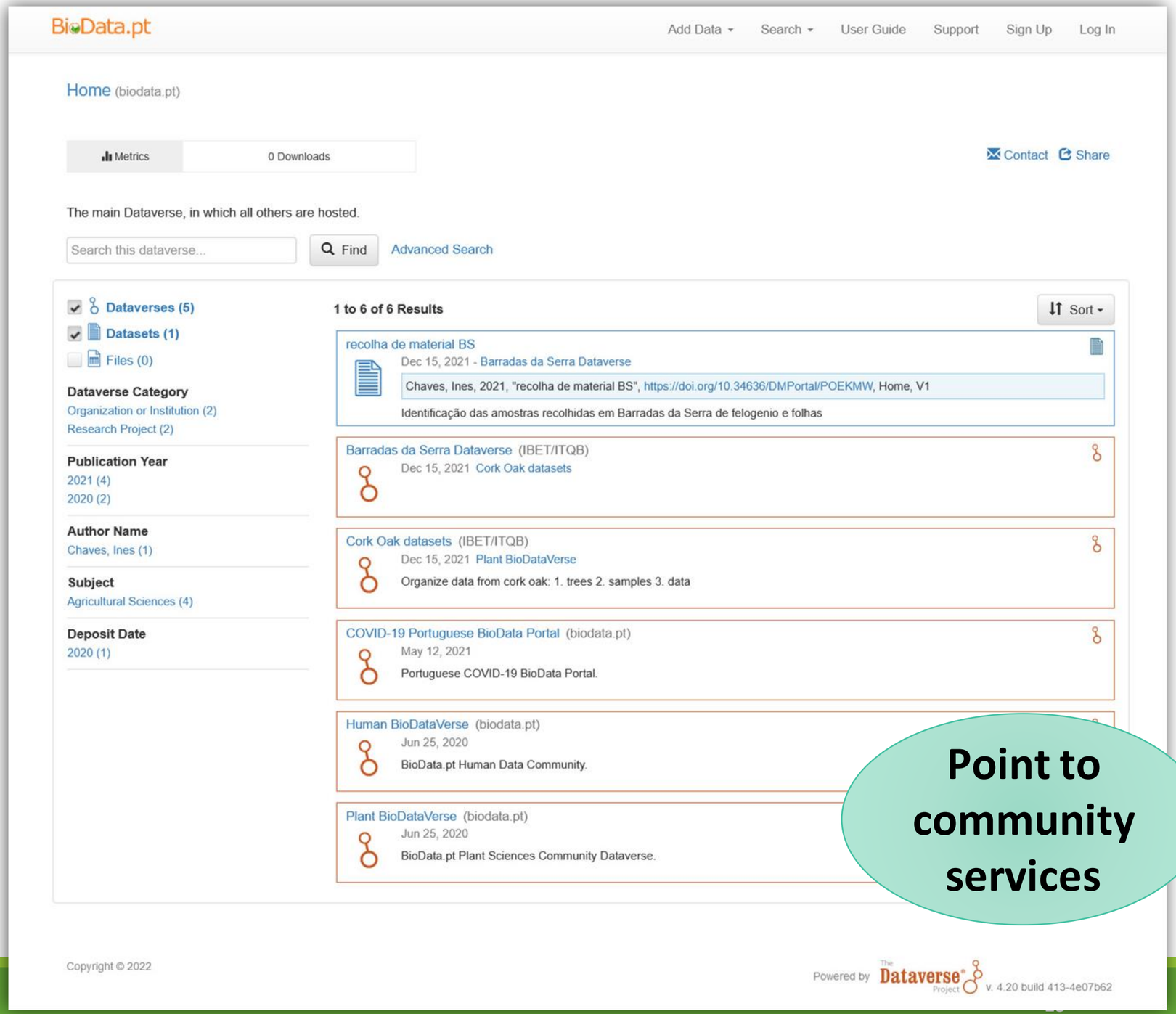
Requires planning and high standards, may incur costs

elixir-europe.org/platforms/data/core-data-resources
elixir-europe.org/platforms/data/elixir-deposition-databases

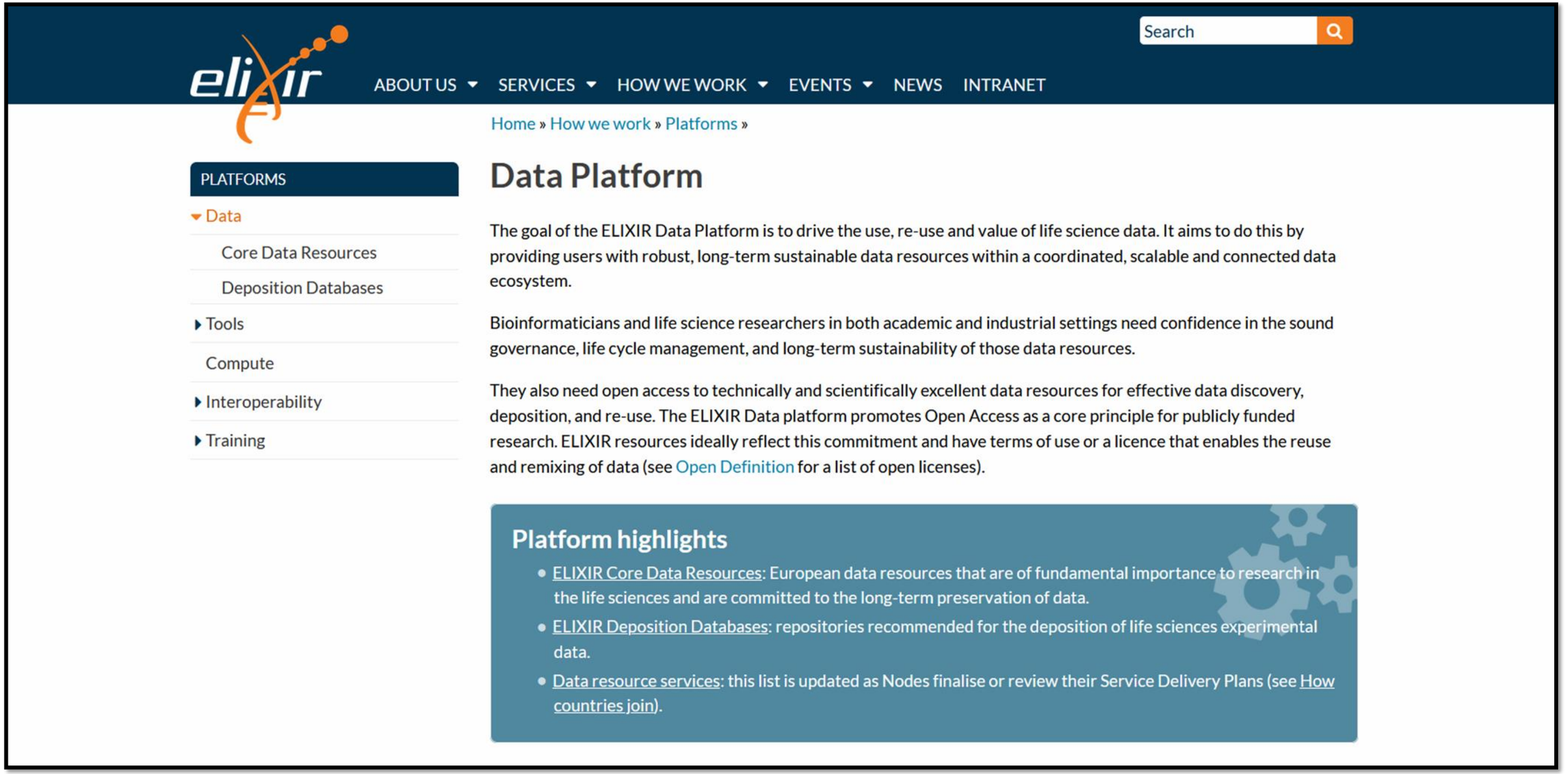
Disciplinary repository

- Offers expertise and experience in data management
- Likely to accept complete data sets
- Selective in the type of data they accept
- Requires planning and high standards, may incur costs

https://dmportal.biodata.pt/



Point to community services



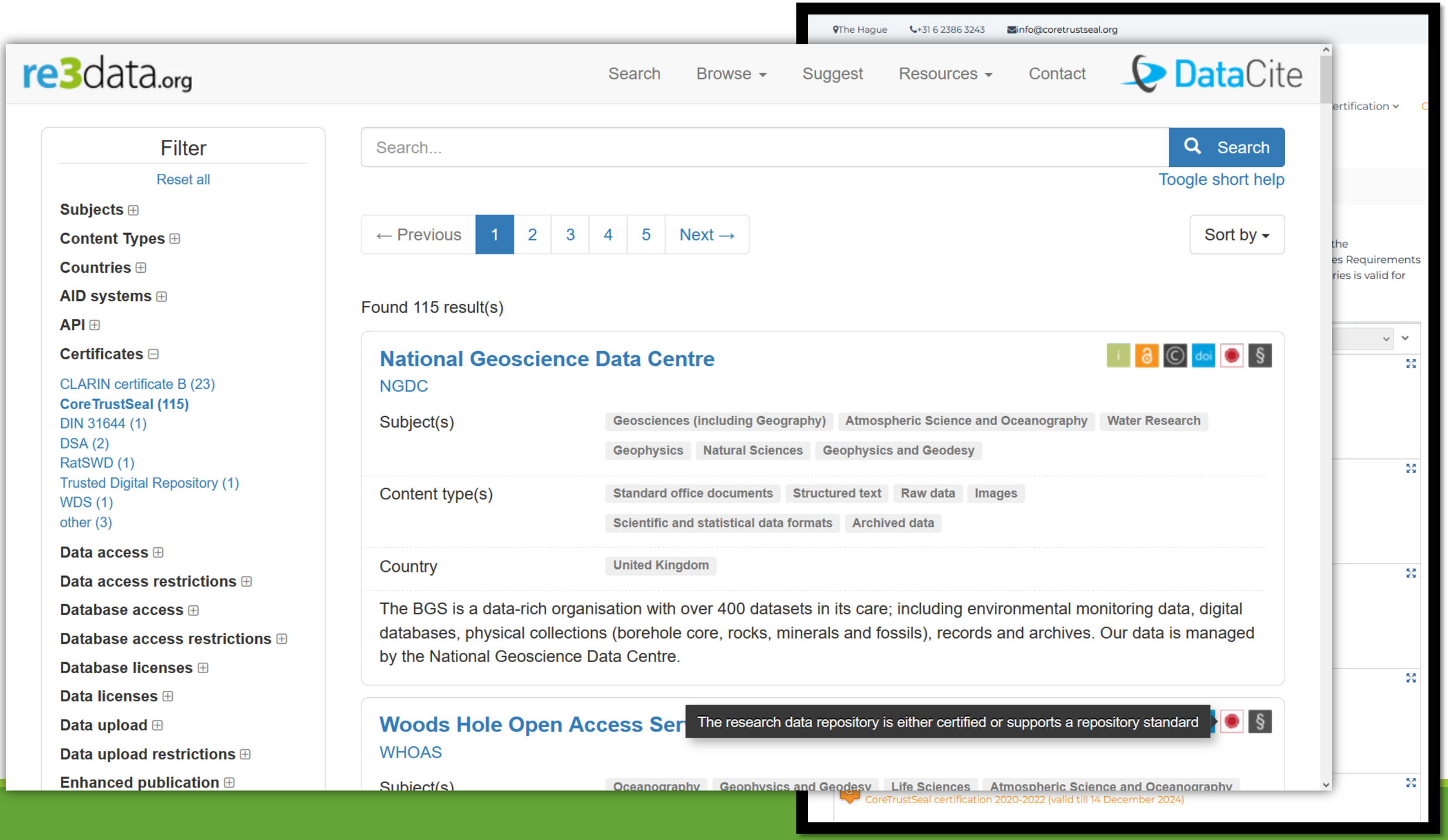
https://elixir-europe.org/platforms/data

TRUSTWORTHY REPOSITORIES

Certification Tools for repository auditing & certification.

- CoreTrustSeal (CTS)
- Nestor Seal;
- ISO 16363: 2013

Digitalbevaring.dk


re3data.org
Search
Browse
Suggest
Resources
Contact
DataCite
Filter
Reset all
Subjects
Content Types
Countries
AID systems
API
Certificates
CLARIN certificate B (23)
CoreTrustSeal (115)
DIN 31644 (1)
DSA (2)
RatSWD (1)
Trusted Digital Repository (1)
WDS (1)
other (3)
Data access
Data access restrictions
Database access
Database access restrictions
Database licenses
Data licenses
Data upload
Data upload restrictions
Enhanced publication
Search...
Search
Toogle short help
← Previous
1
2
3
4
5
Next →
Sort by
Found 115 result(s)
National Geoscience Data Centre
NGDC
Subject(s)
Geosciences (including Geography)
Atmospheric Science and Oceanography
Water Research
Geophysics
Natural Sciences
Geophysics and Geodesy
Content type(s)
Standard office documents
Structured text
Raw data
Images
Scientific and statistical data formats
Archived data
Country
United Kingdom
The BGS is a data-rich organisation with over 400 datasets in its care; including environmental monitoring data, digital databases, physical collections (borehole core, rocks, minerals and fossils), records and archives. Our data is managed by the National Geoscience Data Centre.
Woods Hole Open Access Ser
The research data repository is either certified or supports a repository standard
WHOAS
Subject(s)
Oceanography
Geophysics and Geodesy
Life Sciences
Atmospheric Science and Oceanography
CoreTrustSeal certification 2020-2022 (valid till 14 December 2024)
The Hague
+31 6 2386 3243
info@coretrustseal.org

THANKS

Web
www.openaire.eu

Email
pedro.principe@usdb.uminho.pt

Twitter
@openaire_eu
@pedroprincipe

Services and tools across the Data Curation Lifecycle

HIGHLIGHT TOOLS FOR THE STAGES OF THE RESEARCH DATA LIFE CYCLE

https://dmeg.cessda.eu/Data-Management-Expert-Guide

https://rdmkit.elixir-europe.org/data_life_cycle

My own practices, worries, challenges… and the tools I use for data management in the research life cycle

Showcase tools, ask colleagues to demo…

A world of tools... supporting RDM

https://rdmkit.elixir-europe.org/researcher#national-resources-button
https://dmeg.cessda.eu/
https://datamanagement.hms.harvard.edu/analyze/electronic-lab-notebooks

…

Planning research

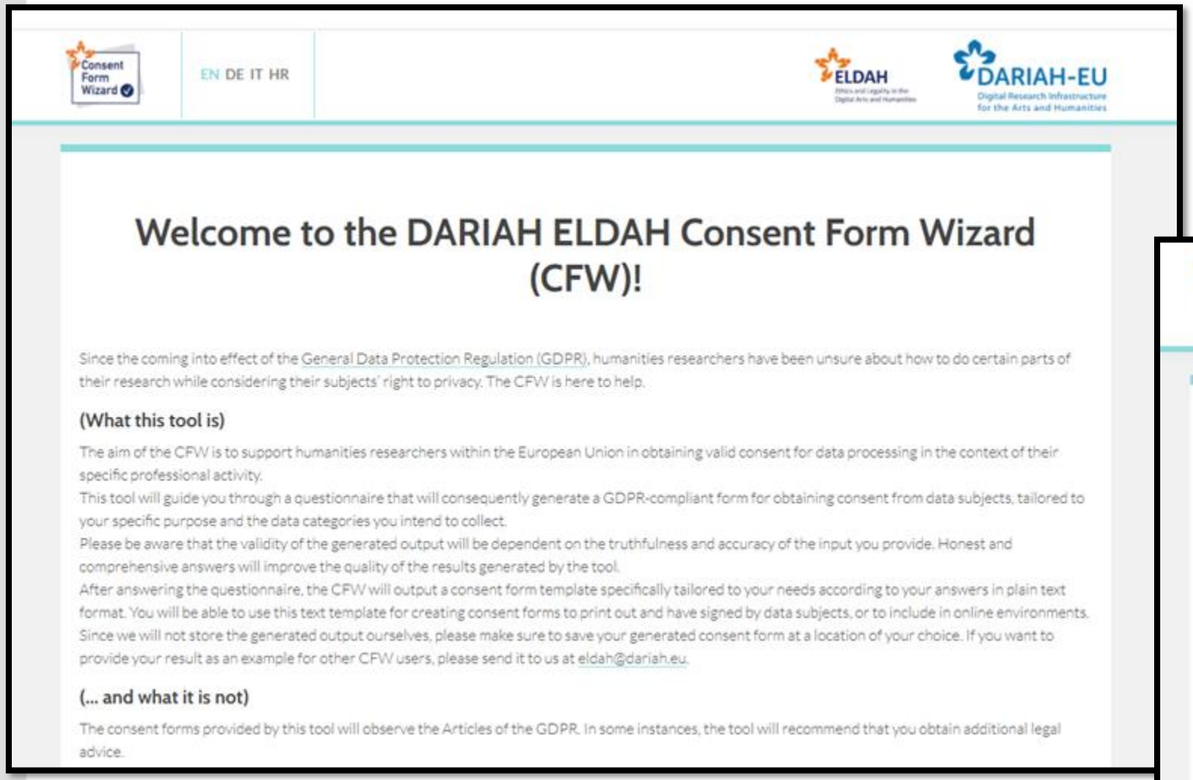
https://ukdataservice.ac.uk/learning-hub/research-data-management/

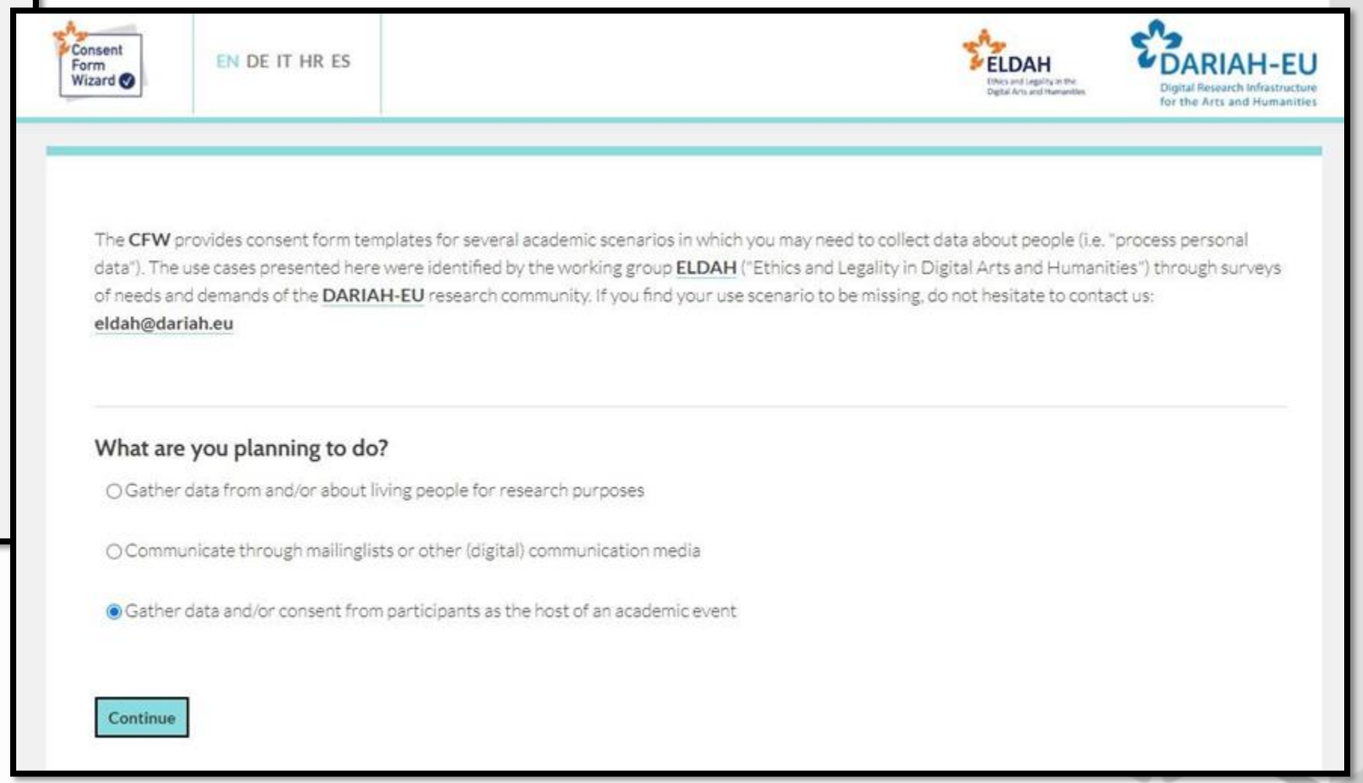
https://youtu.be/-wjFMMQD3UA

DMPs

- DMPonline https://dmponline.dcc.ac.uk/
- ARGOS https://argos.openaire.eu/
- OpenDMP https://gitlab.eudat.eu/dmp/OpenAIRE-EUDAT-DMP-service-pilot
- Data Stewardship Wizard https://ds-wizard.org
- RDMO - RDM Organiser https://rdmorganiser.github.io/en/
- DMPTool https://dmptool.org
- ezDMP https://ezdmp.org
- easyDMP https://easydmp.eudat.eu
- DMP ARIADNEplus https://vast-lab.org/dmp/ariadneplus/form/

Suggestion for preparing Consent forms





https://consent.dariah.eu/

COLLECTING, PROCESSING… DATA

Collecting data:
. Collect data
. Capture data with metadata
. Acquire existing third party data

Processing & analysing data:
. Enter, digitize, transcribe and translate data
. Check, validate, clean, anonymize
. Derive data
. Describe and document data
. Manage and store data
. Analyse and interpret data
. Produce research outputs
. Cite data sources

ELabNotebooks

https://doi.org/10.5281/zenodo.4723752

Research Data Management Working Group
Harvard Longwood Medical Area

Yes
No
* Additional Information available on the ELN subpage

Page last updated 2021-04-19

Features	Specifications									
	Arxspan	Benchling	BIOVIA	Chemotion	Confluence	Docollab	ecLabNote	eLabFTW	eLabJournal	ELOG
Interactivity										
Search functions can search across file formats and beyond typos	✓	*	*	*	*	*	No response received	*	*	*
Ability to manipulate files and images	✓	*	No response received	*	*	*	No response received	✓	✓	*
Support for multiple open windows	✓	✓	*	✓	✓	✓	No response received	✓	✓	✓
Ability to link out	✓	✓	No response received	✓	*	✓	✓	✓	✓	✓
Support for Researcher Documentation										
Hyperlink support	✓	✓	No response received	✓	✓	✓	✓	✓	✓	✓
Metadata Creation Prompts	✗	✓	No response received	✓	✗	✓	No response received	✗	✗	✓
Rights Management (licensing)	✓	✗	No response received	*	*	✓	No response received	✗	✓	*
Protocol Integration	✓	✓	✓	✓	✓	✓	No response received	✓	✓	✓
Adaptability to Lab workflows										
Accounts/Permissions Levels	✓	✓	No response received	✓	*	✓	✓	✓	✓	✓
Internal Data Sharing	✓	✓	*	✓	*	✓	No response received	✓	✓	✓
Adaptable to a Variety of Workflows	*	*	No response received	No response received	*	*	No response received	✓	No response received	No response received
Compatibility with authoring tools	*	✓	No response received	*	*	✓	No response received	✗	✓	✗
Windows Compatible	✓	✓	No response received	✓	✓	✓	✓	✓	✓	✓
Macintosh Compatible	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Linux Compatible	✓	✓	✗	✓	✓	✓	No response received	✓	✓	✓
Android Compatible	✓	✓	✓	✓	✓	✓	No response received	✓	✓	✓

HARVARD MEDICAL SCHOOL | DATA MANAGEMENT WORKING GROUP | DATAMANAGEMENT.HMS.HARVARD.EDU

HARVARD MEDICAL SCHOOL

Spreading the Word on Going Digital in the Lab

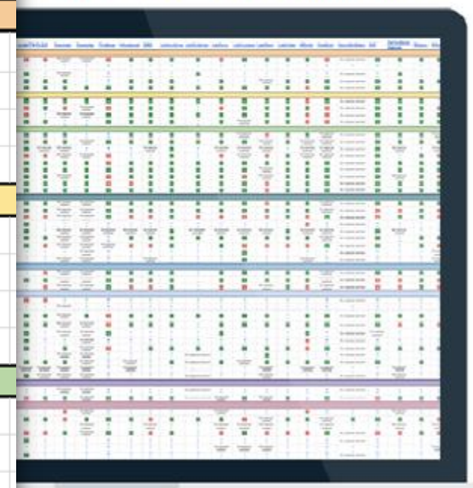
The Data Management Working Group

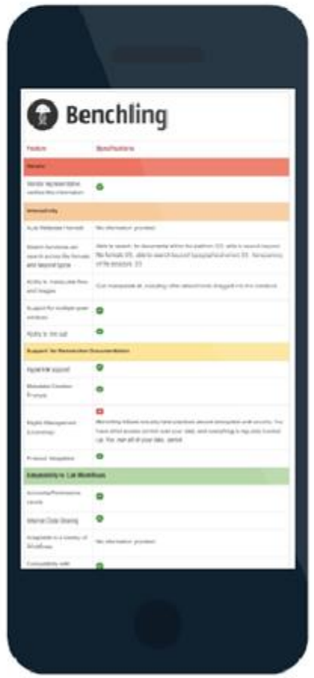
The Harvard Medical School (HMS) Data Management Working Group (DMWG) was established in response to identified unmet needs in biomedical research in the HMS community. One subgroup was created to study and support researchers' interest in and use of Electronic Lab Notebooks
mine use cases for ELNs
the needs of researchers
ailable with current
oduced a web-based guide
G to act as a consultant
their ELN practices.

The Electronic Lab Notebook Matrix

The matrix table compares Electronic Lab Notebook (ELN) options, helping researchers identify appropriate solutions. We sent a survey to 26 vendors, and created a matrix based on the responses. This matrix does not recommend any particular solution(s), but serves as an educational tool and decision map for librarians and researchers alike to navigate the growing ELN market. The group continuously processes requests for updates to current matrix information, as well as inquiries to add additional ELNs. Mentions of this matrix on Twitter, in Nature, and Wikipedia show that the wider community welcomes resources that make research data management solutions more findable, accessible, interoperable, and reusable (FAIR).

.HARVARD.EDU/ELECTRONIC-LAB-NOTEBOOKS





ectronic laboratory notebook."
d41586-018-05895-3

RDM Support webpages in institutional websites

LONGWOOD
RESEARCH DATA MANAGEMENT

Search

About | Plan & Design | Collect & Analyze | Store & Evaluate | Share & Publish | Training & Events

Store & Evaluate

Store & Evaluate

- Storage Options
- Data Safety
- Data Security
- Data Retention
- Archives and Records Management
- Data Destruction

Storage Overview

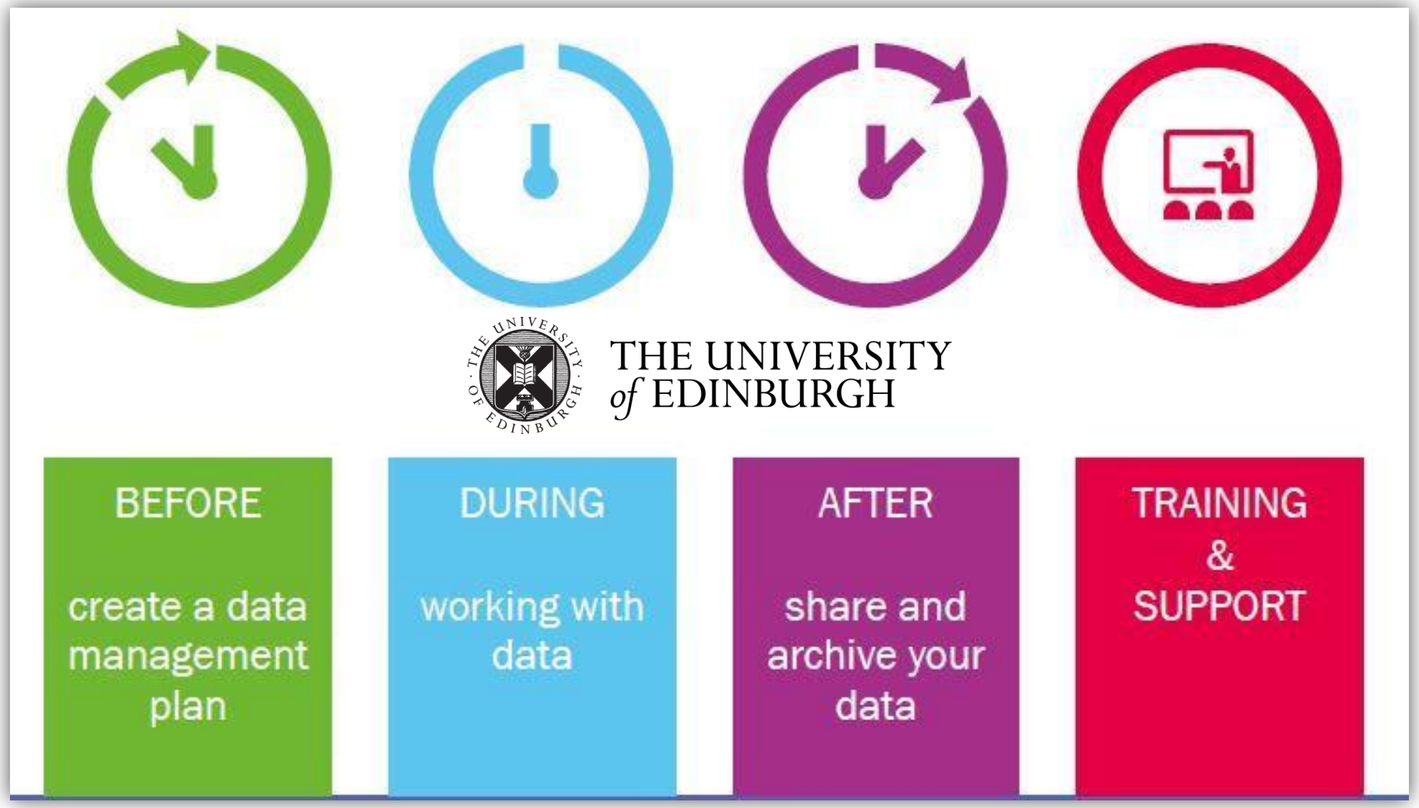
Every stage of the Biomedical Data Lifecycle centers around the management of data storage. Proper storage maintenance throughout the lifecycle is imperative to ensure data remains secure and adheres to recommended safety protocols.

Research requires increasingly complex arrangements for the storage and transmission of research data. Robust data privacy and security planning is needed to protect the privacy of research subjects and to secure sensitive, personally identifiable information.

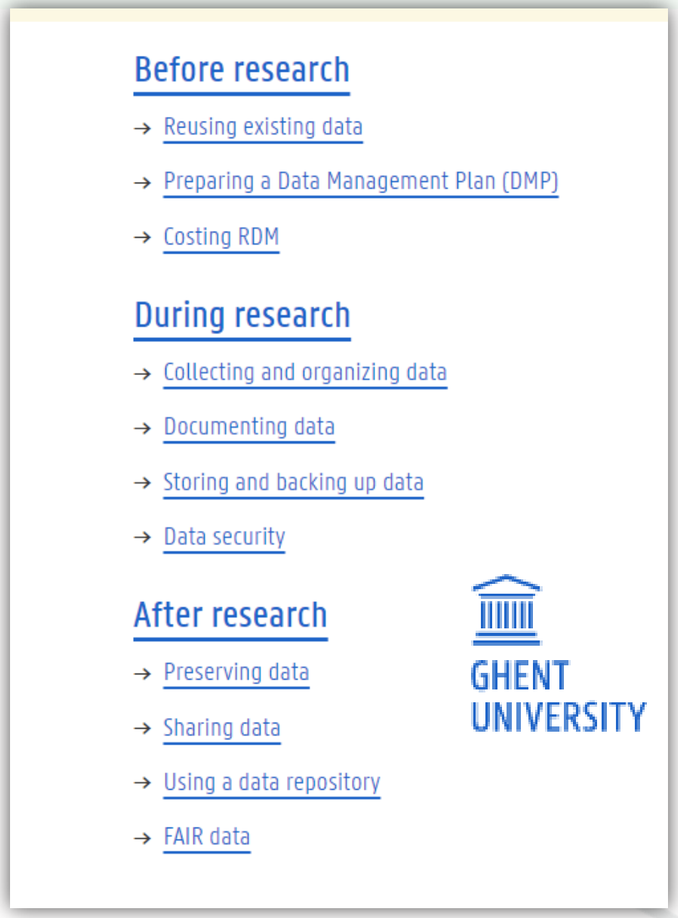
- In the Plan & Design stage, you may be asked to

https://datamanagement.hms.harvard.edu/store-evaluate

Research work life cycle (simplified)

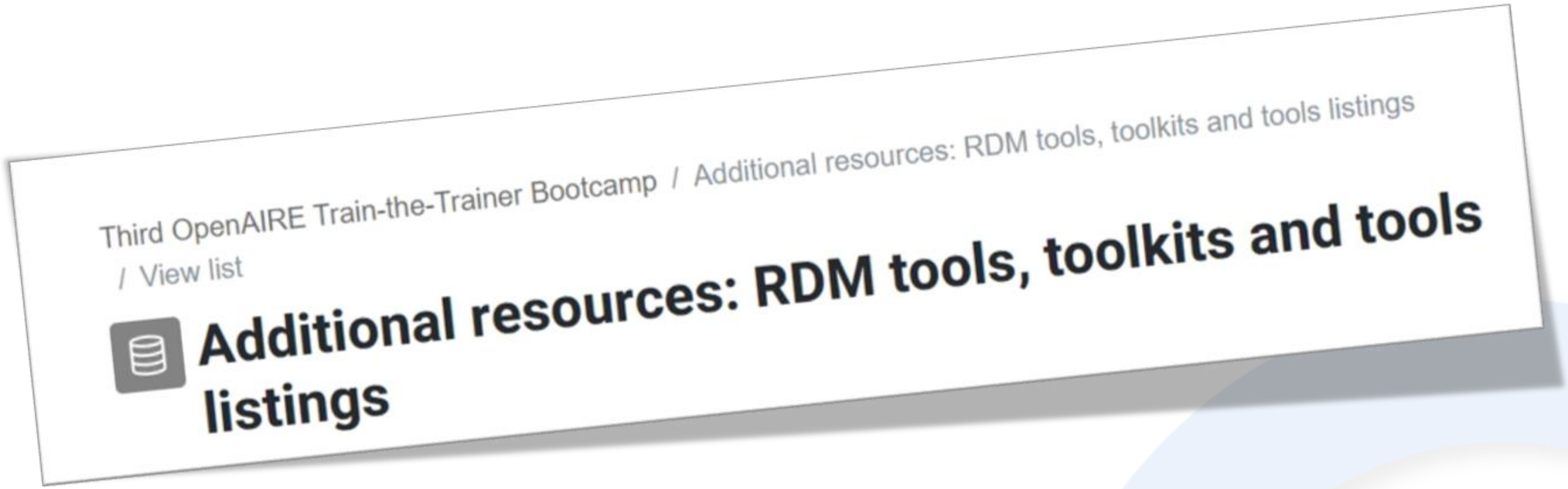


https://www.ed.ac.uk/information-services/research-support/research-data-service



https://www.ugent.be/en/research/datamanagement

SELECTION OF RESOURCES THAT RECOMMEND RDM TOOLS



Feel free to add resources (or individual tools) you are using.

THANKS

Web
www.openaire.eu

Email
pedro.principe@usdb.uminho.pt

Twitter
@openaire_eu
@pedroprincipe