Data sharing & publishing

Obrad Vučkovac
University of Belgrade
Vinča Institute of Nuclear Sciences – Library

EIFL Open Science bootcamp

Training concepts

Training concept

- Get to know concepts of RDM, FAIR, DMP first - you can leave Sharing and publishing at the end;
- You train RDM in Open Science setting - when designing the trainings, emphasize the focus on sharing;
- Be clear with Open Data - what can and cannot be open;
- Identify the venues to publish and deposit research data.

Data with more prominent role

Simplify this complex topic by concentrating on these four points that researchers need to adhere:

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

Your audience should recognize topics that you were talking about earlier (RDM, FAIR, DMP).

Clarify concepts

Deposit data: upload a digital object (research data) on a platform that enables correct description with metadata and implements long-term preservation.

Give access: authors choose the access type (open, closed, restricted, embargoed) of deposited data and assign a licence for reuse (Creative Commons)

Photo by Sigmund on Unsplash

Clarify concepts

Non-synomymous terms

- *shared*: any way of sharing information, could mean I emailed it to you.
- *publish* : citable artifact, discoverable.
- *archive* : long-term preservation.

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

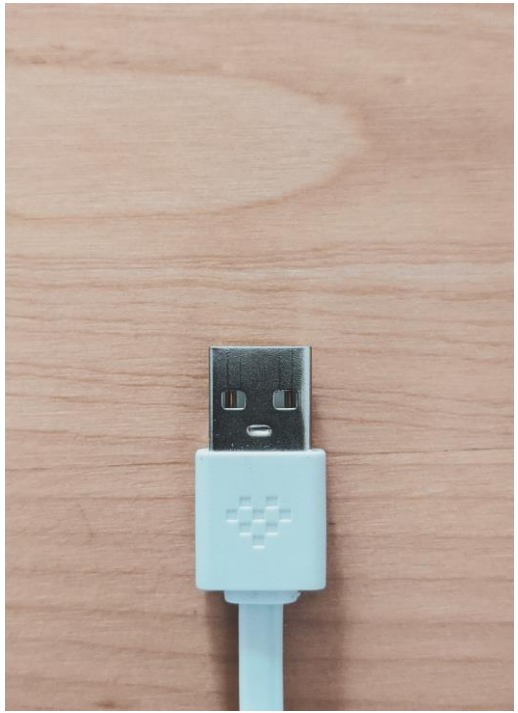


Photo by Sigmund on Unsplash

What to deposit?

Follow and comply with FAIR principles

DATA	METADATA	DOCUMENTATION
• open / common formats • relevant standards for reusability	• structured, machine-readable • metadata schema - fixed set of attributes • use existing (domain-specific) standards	• software code; • protocols, methods; • consent information; • data quality (instrument calibrations...); • machine configurations; • etc.

Open data

Funder requirements

European Commission - Horizon Europe

- "**as soon as possible** and within the deadlines set out in the DMP, **ensure open access** — via the repository — **to the deposited data**, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC0) or a licence with equivalent rights, following the principle 'as open as possible as closed as necessary', unless providing open access would in particular:
  - be against the beneficiary's legitimate interests, including regarding commercial exploitation, or
  - be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP

- "Metadata of deposited data must be open under a CC Public Domain (CC 0) or equivalent, in line with the FAIR principles (in particular machine-actionable)…"

EU Grants - AGA - Annotated Model Grant Agreement

Funder requirements

European Commission - Horizon Europe

- “as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, **under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC0) or a licence with equivalent rights**, following the principle ‘as open as possible as closed as necessary’, unless providing open access would in particular:
  - be against the beneficiary’s legitimate interests, including regarding commercial exploitation, or
  - be contrary to any other constraints, in particular the EU competitive interests or the beneficiary’s obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP

- “**Metadata** of deposited data must be open **under a CC Public Domain (CC 0) or equivalent**, in line with the FAIR principles (in particular machine-actionable)…”

EU Grants - AGA - Annotated Model Grant Agreement

Funder requirements

Wellcome Trust

"Published outputs that arise from our funding must be **open and accessible** to everyone."

Wellcome - Open Access policy

Funder requirements

Bill & Melinda Gates Foundation

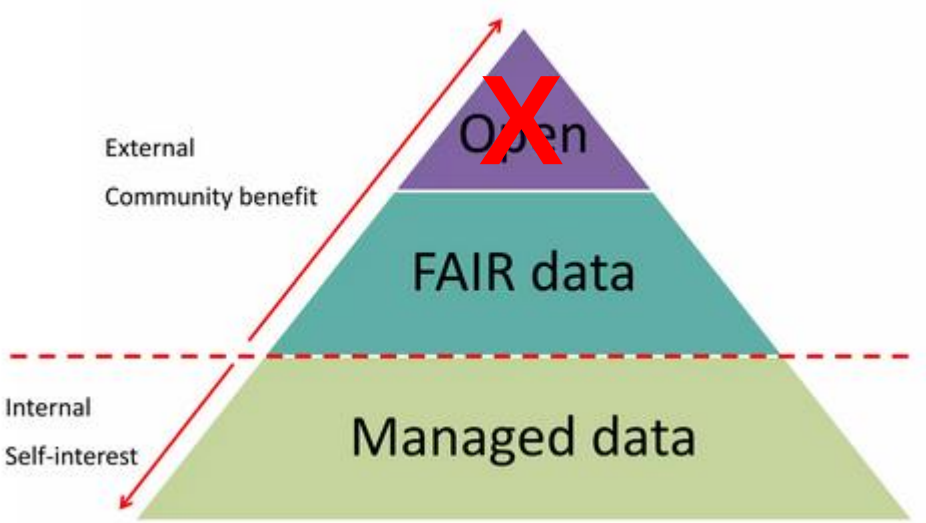
“ **Publications and Underlying Data Will Be Accessible and Open Immediately.** All Funded Research including articles accepted for publication shall be available immediately at publication, without any embargo period. Each accepted article must be accompanied by a Data Availability Statement that describes where any primary data, associated metadata, original software, and any additional relevant materials necessary to understand, assess, and replicate the reported study findings in totality can be found.”

“Note: We do not require sharing of data that is ethically unsound or legally encumbered.”

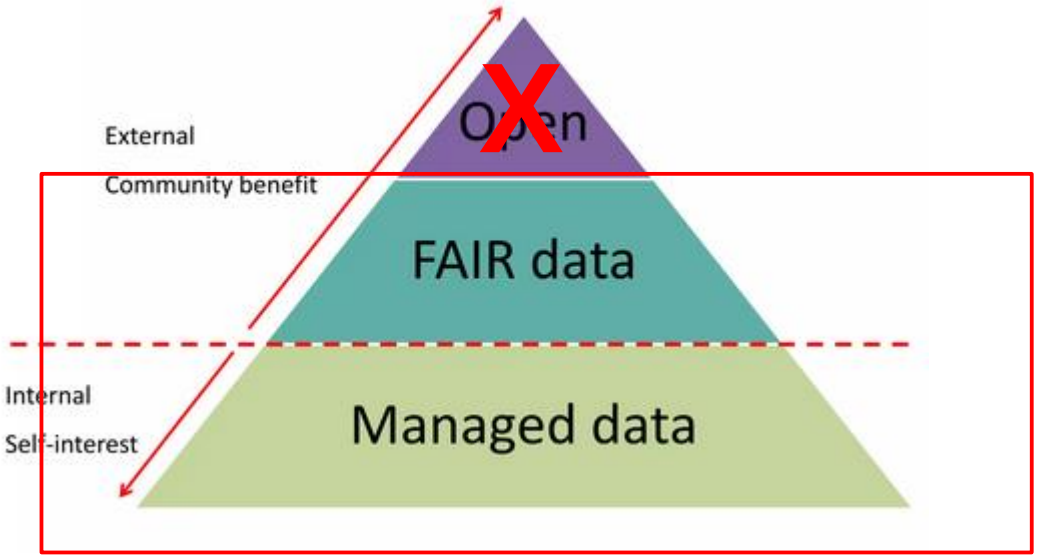
Gates foundation - Open Access policy

“as open as possible,
as closed as necessary”

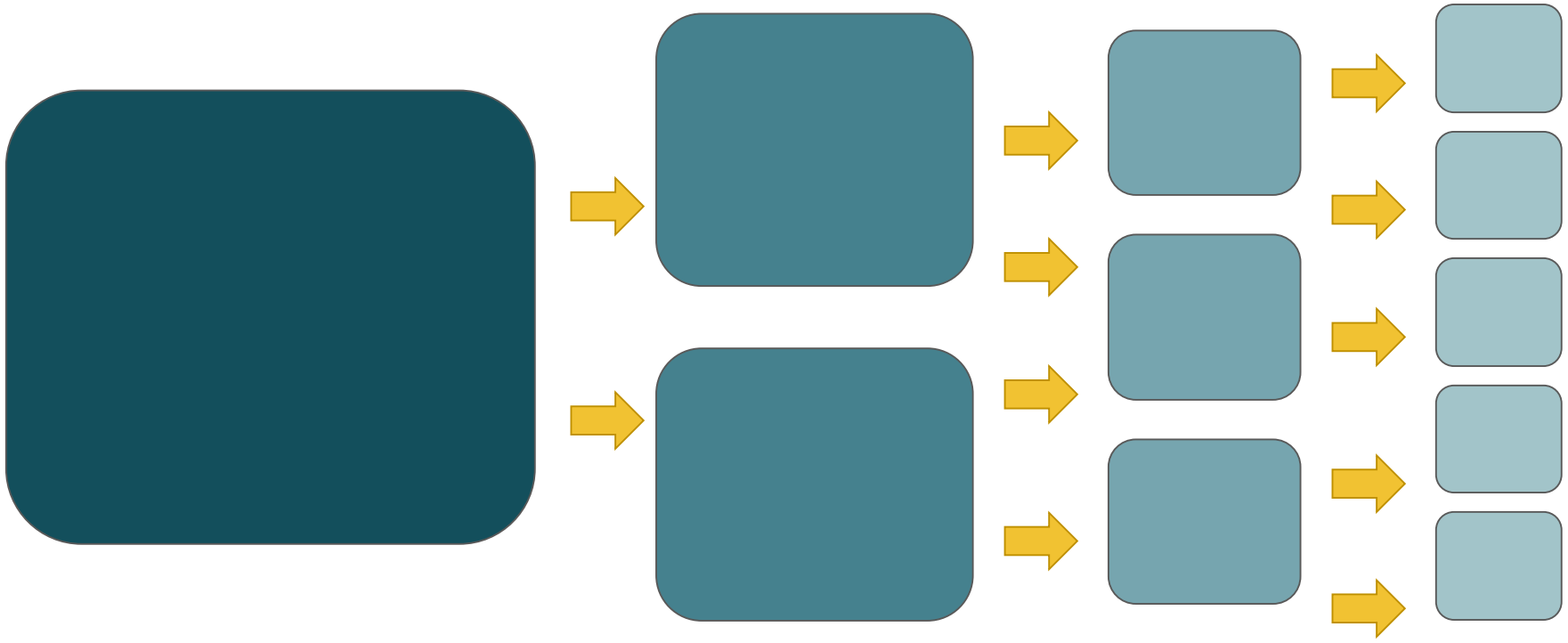
FAIR - even if data is not open



FAIR - even if data is not open



Different levels of data



Publishing and sharing data

Where to publish data?

- Journals & supplementary materials
- Institutional data repository
- General purpose repository
- Thematic/disciplinary repository

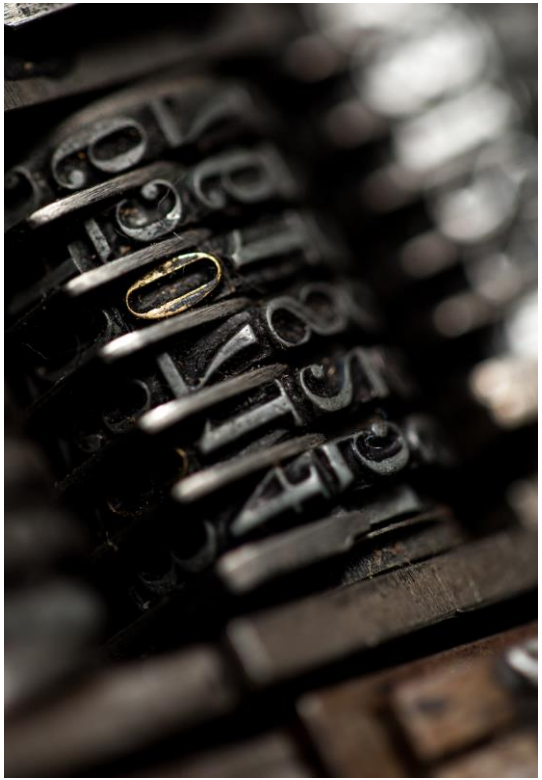


Photo by Alex Zaj on Unsplash

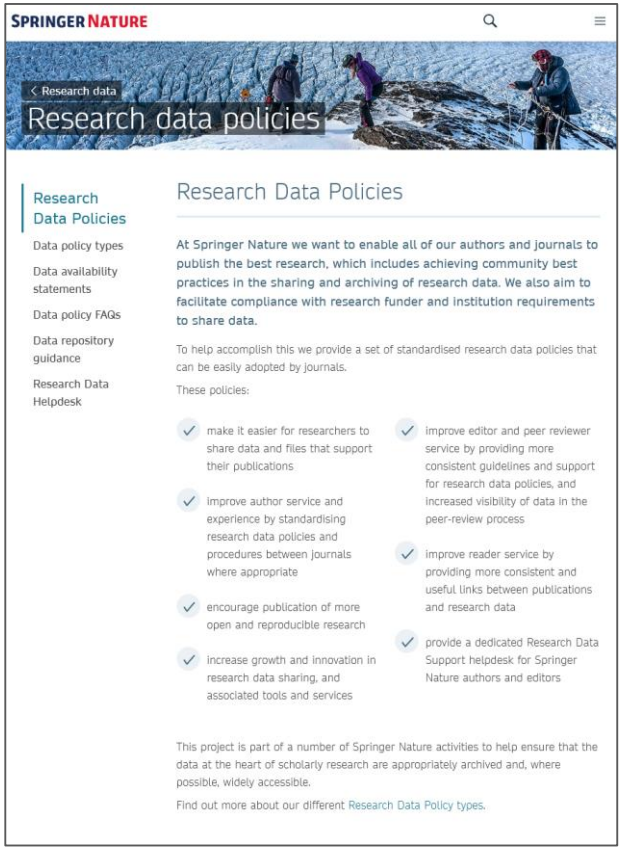
Journals & supplementary materials

PROS:

- publisher requirements
- data is available from articles
- (data papers)

CONS:

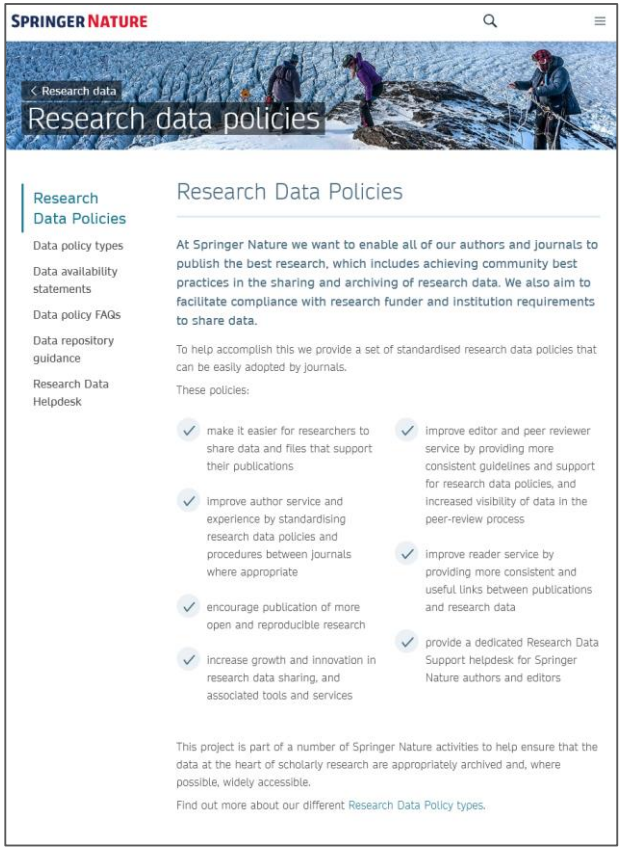
- risky with data rights
- may cost
- may be closed
- long-term preservation?!



Journals & supplementary materials

The ways journals publish data:

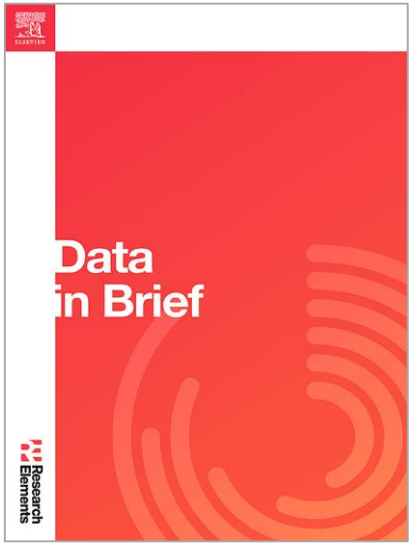
- **send the dataset to the publisher** to publish online;
- the publisher asks **authors to deposit** the dataset in a **repository** and to provide a link in a article;
- author give **contact information** to anyone who wants to have access to the research data.



Data papers journals

data

A peer-reviewed open-access journal
Biodiversity Data Journal
Making your data count! ISSN 1314-2828 (online)



scientific **data**

Institutional data repository

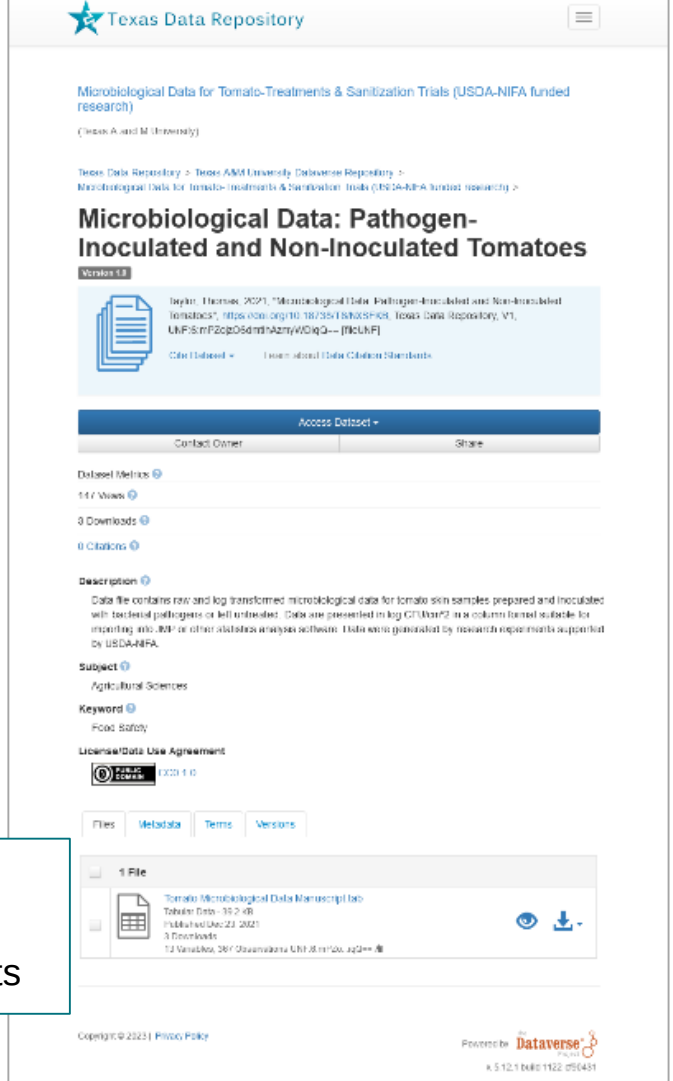
PROS:

- long-term preservation and access
- accept various data types
- no costs

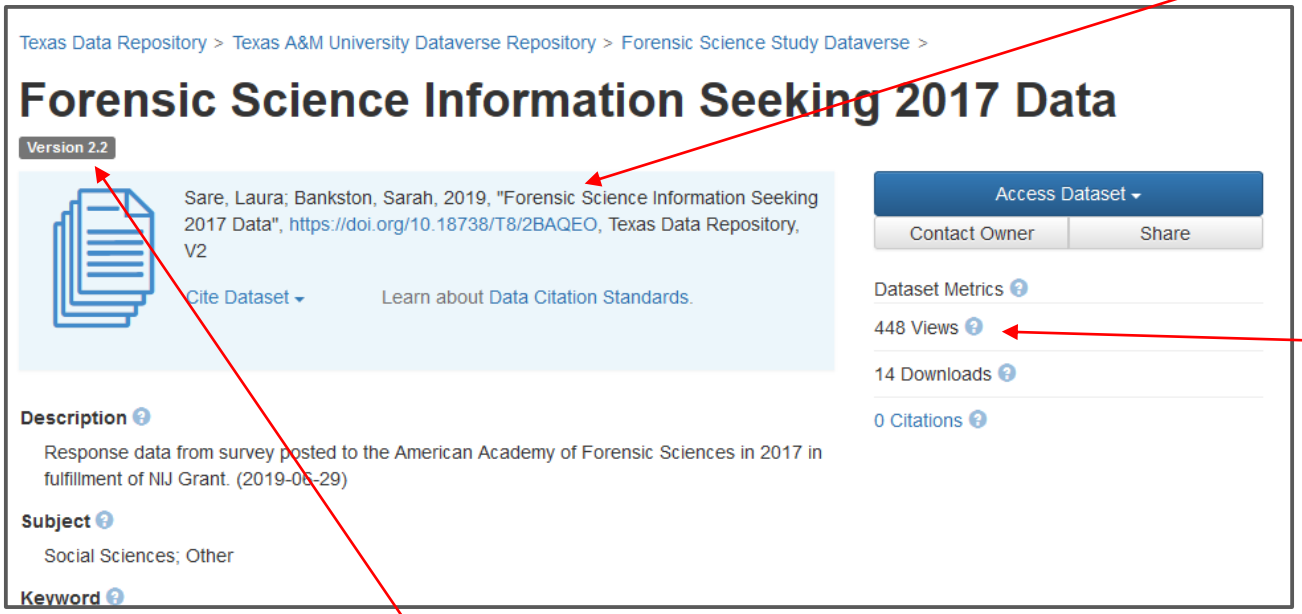
CONS:

- may not have disciplinary metadata
- less visible than thematic repos

If your institution have a data repository, demonstrate how to deposit and all the positive aspects



Institutional data repository

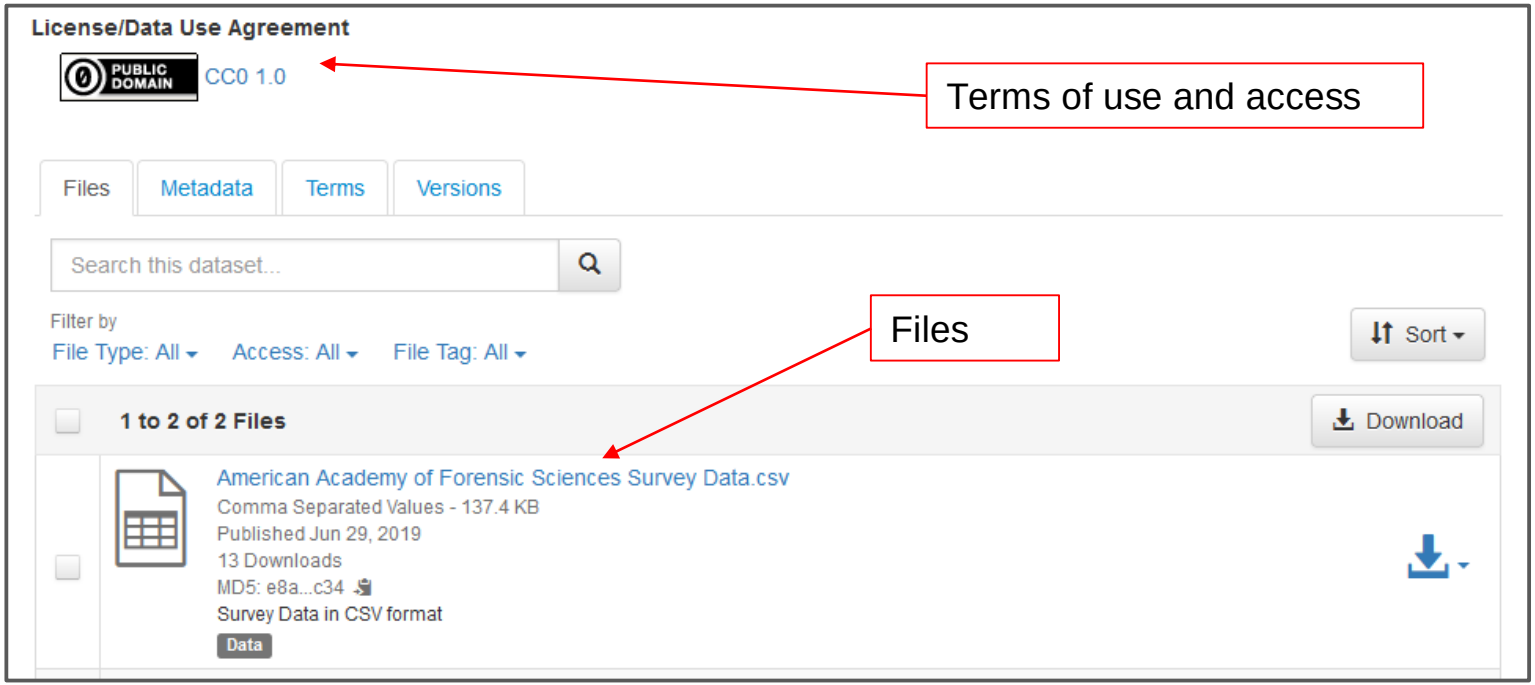


DOI and citation

Usage statistics

Version control

Institutional data repository



General purpose repository

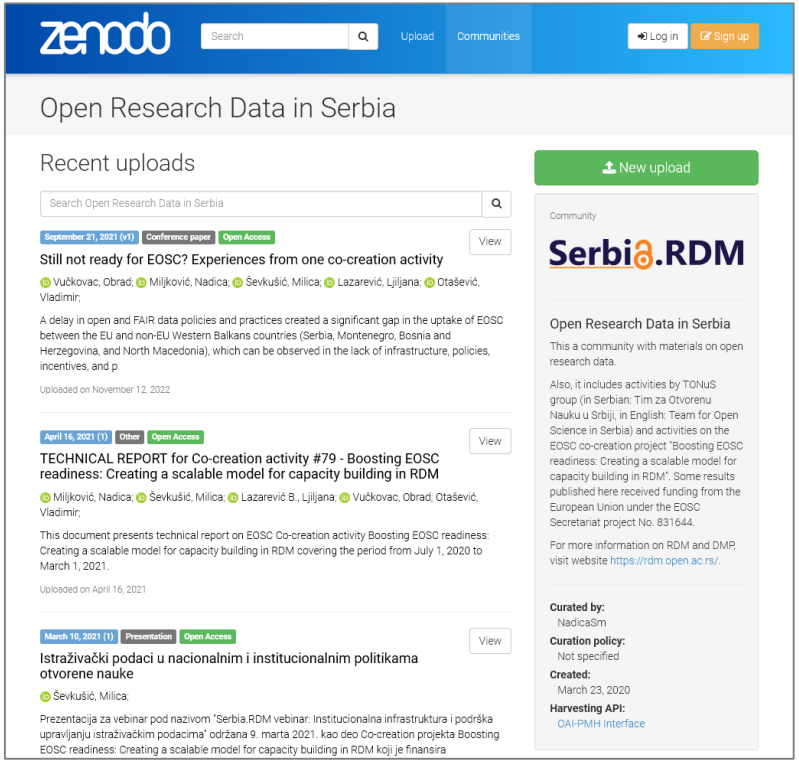
PROS:

- better visibility, wider audience
- accept various data types
- suitable for interdisciplinary data

CONS:

- (usually) only simple metadata
- no quality control over deposited data

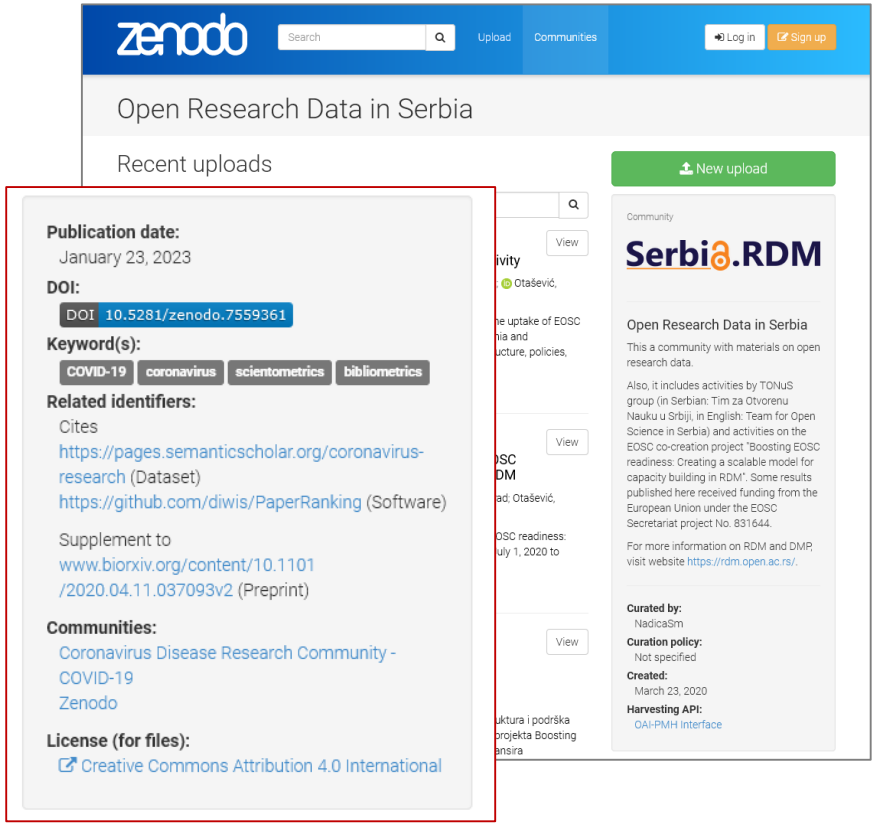
Zenodo, Figshare, OSF, Dryad



General purpose repository

Advantages:

- free persistent identifier (DOI) - excellent for tracking citations;
- usage track - downloads and views;
- supports different data types;
- different types of licenses available (not just CC, but Apache, GNU, etc.);
- research communities - join existing or create yours



Disciplinary repository

PROS:

- offer data management services
- likely to accept complete datasets
- excellent visibility within domain

CONS:

- may include costs
- require high standards in (meta)data quality

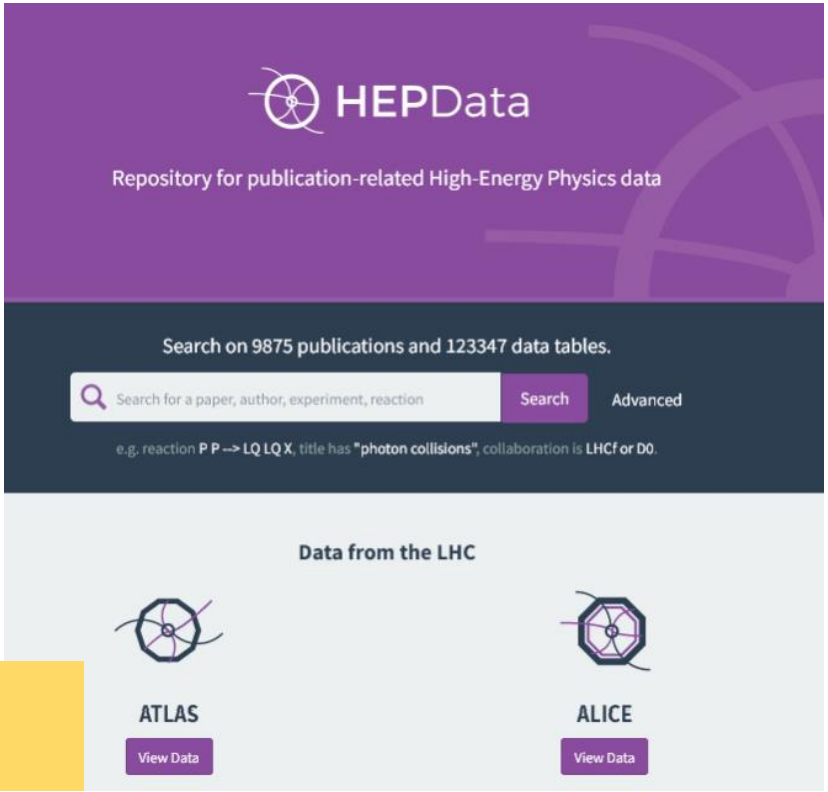
Disciplinary repository

PROS:

- offer data management services
- likely to accept complete datasets
- excellent visibility within domain

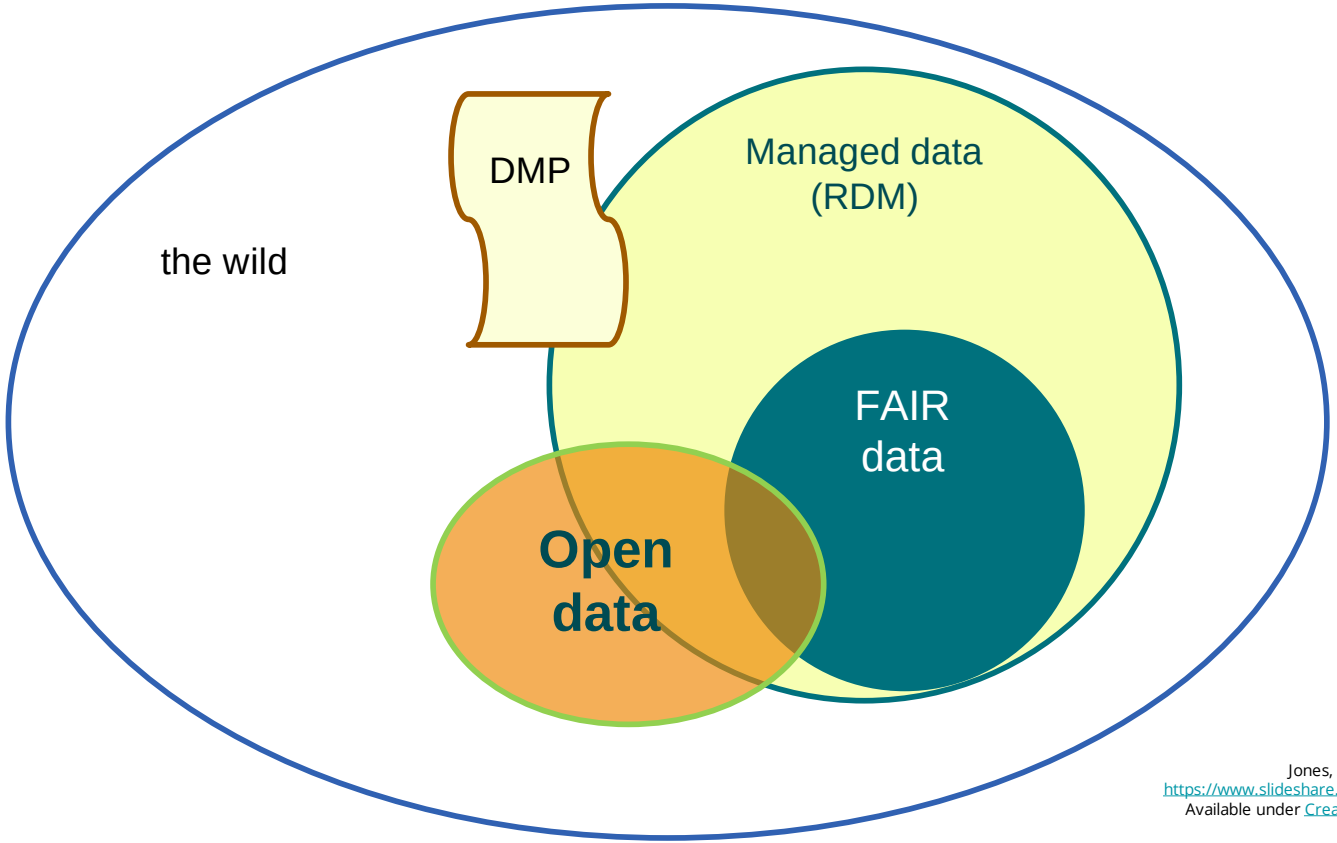
CONS:

- may include costs
- require high standards in (meta)data quality



Point to the community services

Summary



Questions?

Obrad Vučkovac

University of Belgrade

Vinča Institute of Nuclear Sciences - Library

ORCID: 0000-0001-5616-2680