

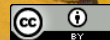
# Data sharing & publishing

Obrad Vučkovic

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)

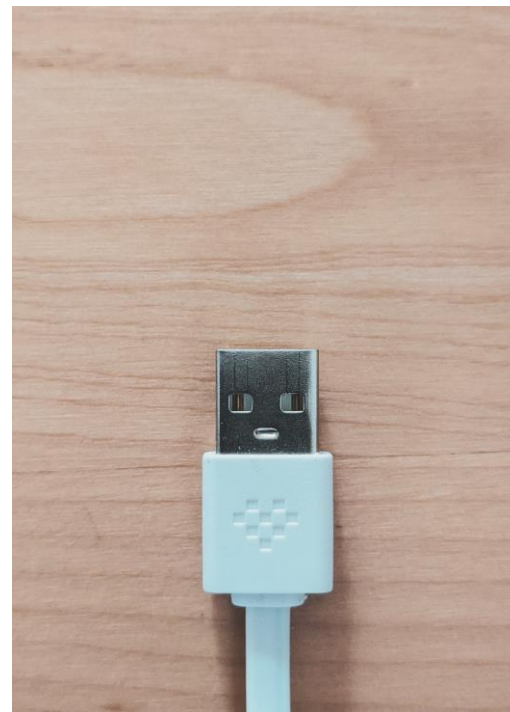


Photo by [Sigmund](#) on [Unsplash](#)

# Clarify concepts

## Non-synonymous 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/>

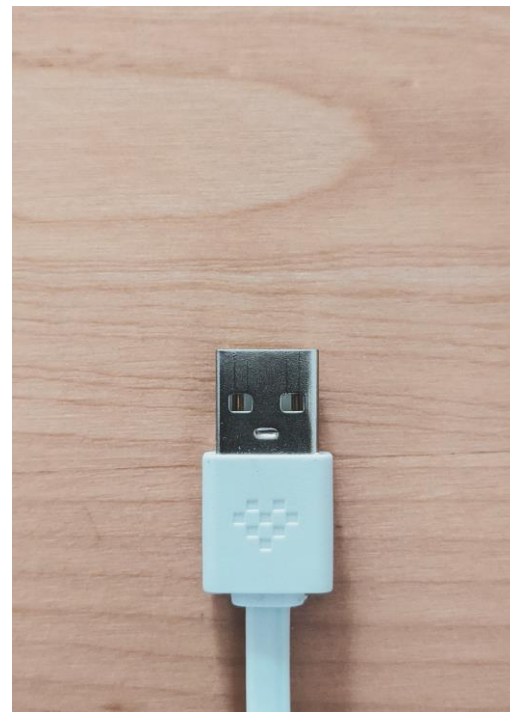


Photo by [Sigmund](#) on [Unsplash](#)

# What to deposit?

Follow and comply with FAIR principles

DATA	METADATA	DOCUMENTATION
<ul style="list-style-type: none"><li>● open / common formats</li><li>● relevant standards for reusability</li></ul>	<ul style="list-style-type: none"><li>● structured, machine-readable</li><li>● metadata schema - fixed set of attributes</li><li>● use existing (domain-specific) standards</li></ul>	<ul style="list-style-type: none"><li>● software code;</li><li>● protocols, methods;</li><li>● consent information;</li><li>● data quality (instrument calibrations...);</li><li>● machine configurations;</li><li>● etc.</li></ul>

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)...”

# 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)..”

# 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

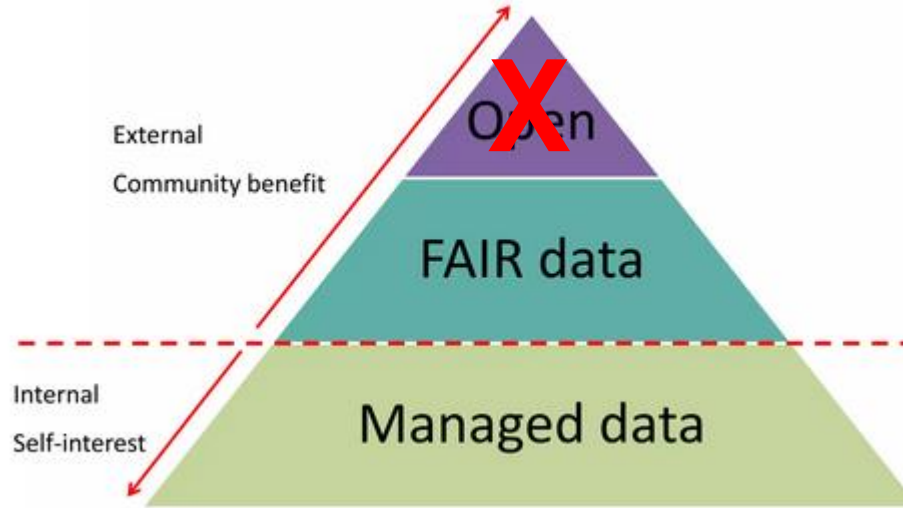
**“ Publications and Underlying Data Will Be Accessible and Open Immediately.** All Funded Research including articles accepted for publication shall be available immediately at publication, without any embargo period. Each accepted article must be accompanied by a Data Availability Statement that describes 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.”

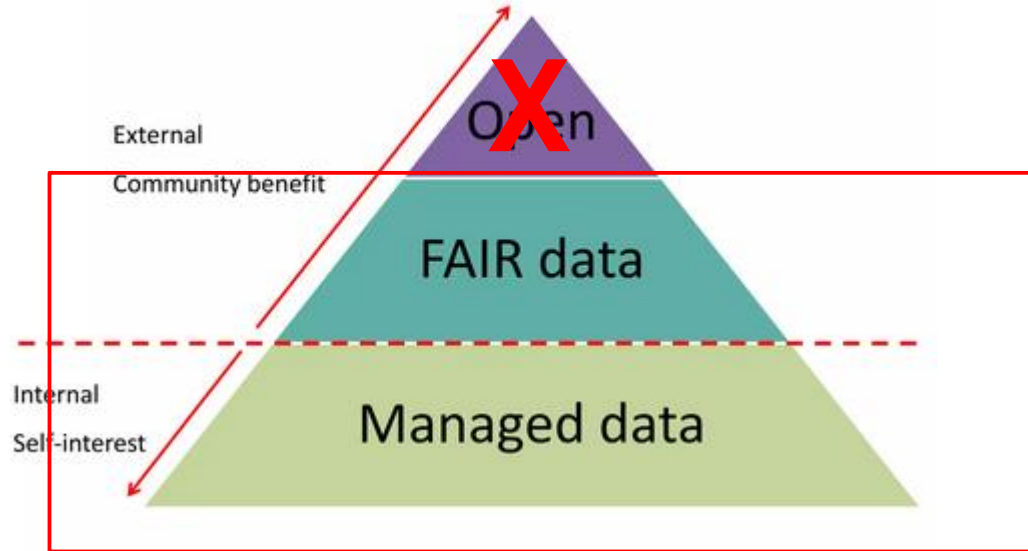
*Gates foundation - Open Access policy*

*“as open as possible,  
as closed as necessary”*

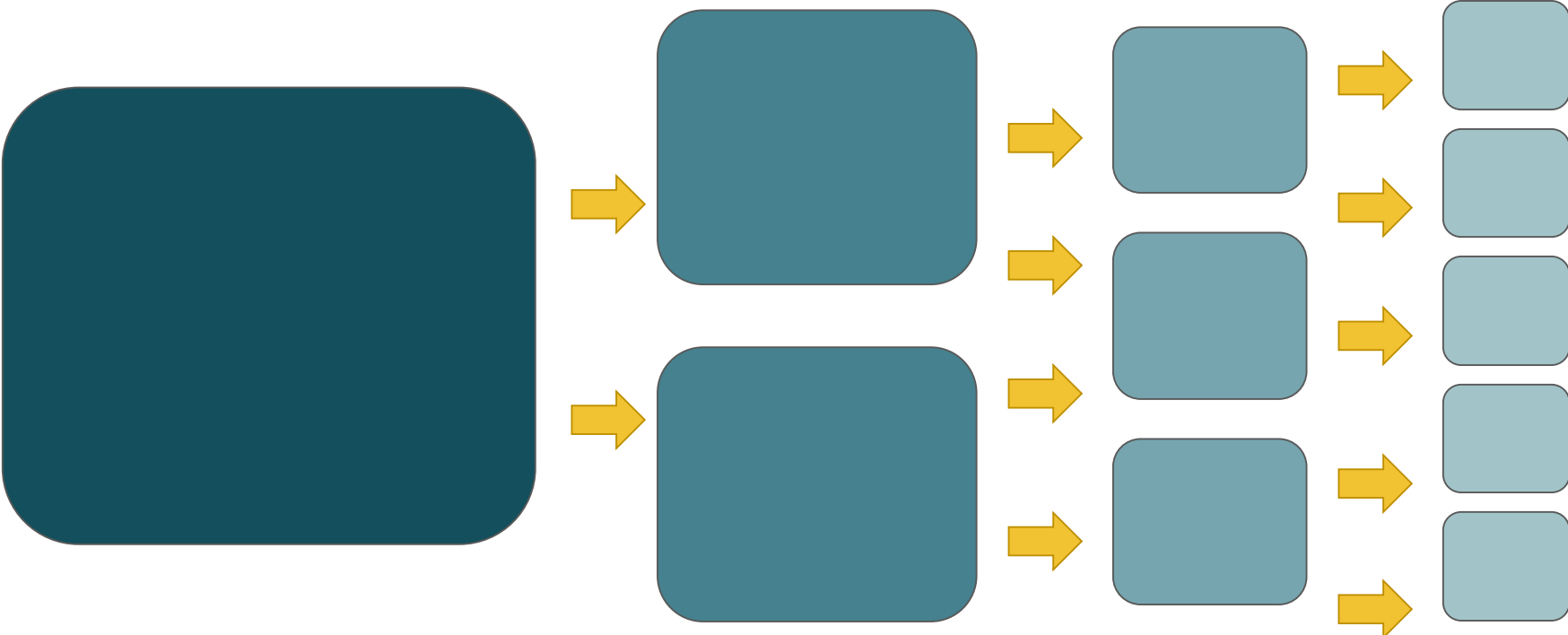
# FAIR - even if data is not open



# FAIR - even if data is not open



# Different levels of data



Raw data

Processed data

Shared data

Open data



# Publishing and sharing data

# Where to publish data?

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

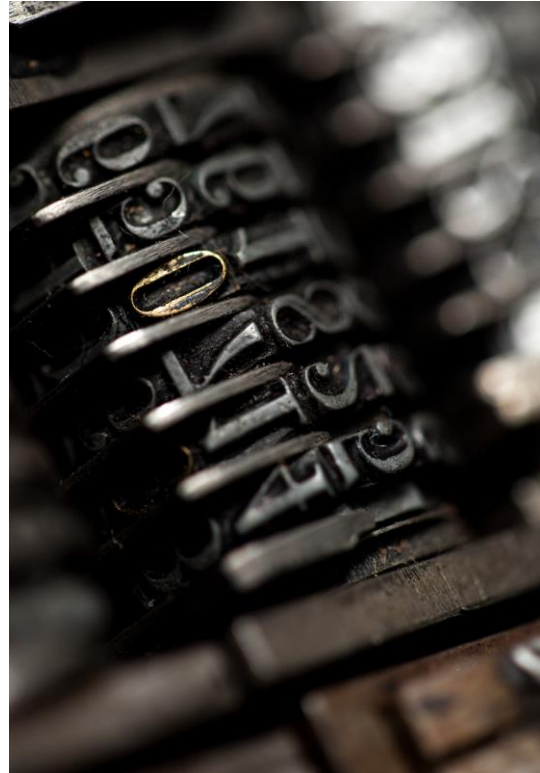


Photo by [Alex Zaj](#) on [Unsplash](#)

# 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?!

**SPRINGER NATURE** Q ☰

[Research data](#)

## Research data policies

### Research Data Policies

At Springer Nature we want to enable all of our authors and journals to publish the best research, which includes achieving community best practices in the sharing and archiving of research data. We also aim to facilitate compliance with research funder and institution requirements to share data.

To help accomplish this we provide a set of standardised research data policies that can be easily adopted by journals.

These policies:

- ✓ make it easier for researchers to share data and files that support their publications
- ✓ improve editor and peer reviewer service by providing more consistent guidelines and support for research data policies, and increased visibility of data in the peer-review process
- ✓ improve author service and experience by standardising research data policies and procedures between journals where appropriate
- ✓ improve reader service by providing more consistent and useful links between publications and research data
- ✓ encourage publication of more open and reproducible research
- ✓ provide a dedicated Research Data Support helpdesk for Springer Nature authors and editors
- ✓ increase growth and innovation in research data sharing, and associated tools and services

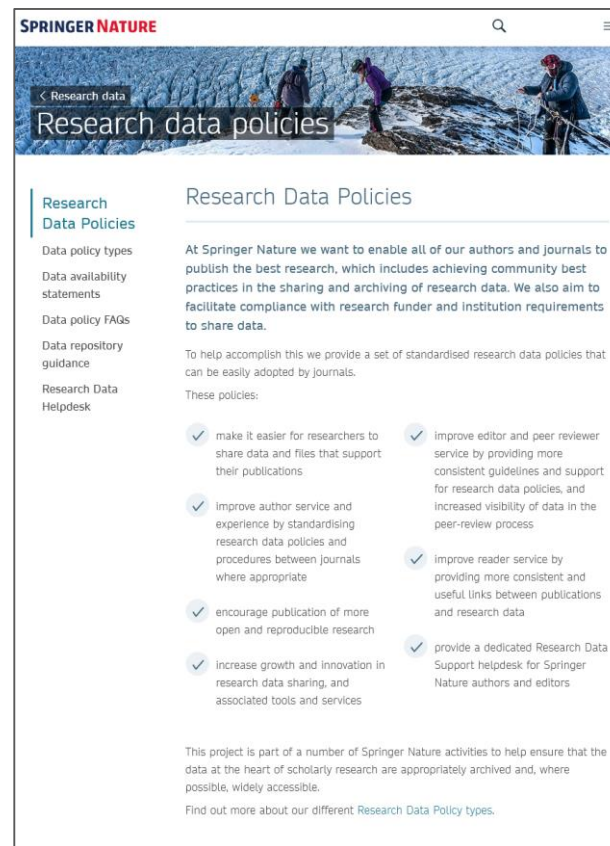
This project is part of a number of Springer Nature activities to help ensure that the data at the heart of scholarly research are appropriately archived and, where possible, widely accessible.

Find out more about our different Research Data Policy types.

# 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.



The screenshot shows the Springer Nature website's 'Research Data Policies' page. The header includes the Springer Nature logo and a search icon. Below the header is a banner image of three people in winter gear on a snowy mountain peak, with a 'Research data' breadcrumb and the title 'Research data policies'. The main content area is titled 'Research Data Policies' and contains an introductory paragraph, a list of policy goals, and a concluding statement. A sidebar on the left lists navigation options: 'Research Data Policies', 'Data policy types', 'Data availability statements', 'Data policy FAQs', 'Data repository guidance', 'Research Data Helpdesk', and 'Data policy types'.

**SPRINGER NATURE** Q ☰

[Research data](#)  
**Research data policies**

## Research Data Policies

At Springer Nature we want to enable all of our authors and journals to publish the best research, which includes achieving community best practices in the sharing and archiving of research data. We also aim to facilitate compliance with research funder and institution requirements to share data.

To help accomplish this we provide a set of standardised research data policies that can be easily adopted by journals.

These policies:

- ✓ make it easier for researchers to share data and files that support their publications
- ✓ improve editor and peer reviewer service by providing more consistent guidelines and support for research data policies, and increased visibility of data in the peer-review process
- ✓ Improve author service and experience by standardising research data policies and procedures between Journals where appropriate
- ✓ improve reader service by providing more consistent and useful links between publications and research data
- ✓ encourage publication of more open and reproducible research
- ✓ provide a dedicated Research Data Support helpdesk for Springer Nature authors and editors
- ✓ Increase growth and innovation in research data sharing, and associated tools and services

This project is part of a number of Springer Nature activities to help ensure that the data at the heart of scholarly research are appropriately archived and, where possible, widely accessible.

[Find out more about our different Research Data Policy types.](#)

**Research Data Policies**

- Data policy types
- Data availability statements
- Data policy FAQs
- Data repository guidance
- Research Data Helpdesk

# 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

Microbiological Data for Tomato-Treatments & Sanitization Trials (USDA-NIFA funded research)

(Texas A and M University)

[Texas Data Repository](#) > [Texas A&M University Database Repository](#) > [Microbiological Data for Tomato-Treatments & Sanitization Trials \(USDA-NIFA funded research\)](#) >

## Microbiological Data: Pathogen-Inoculated and Non-Inoculated Tomatoes

Version 1.0



Bylin, Theresa. 2021. "Microbiological Data: Pathogen-Inoculated and Non-Inoculated Tomatoes". <https://doi.org/10.18731/SDR/5161>. Texas Data Repository, V1. UNF:P22qP0SdmdhAzmyW0IqQ-- [fileURL]

[Cite Dataset](#) > [Learn about Data Citation Standards](#)

Access Dataset >

Contact Owner

Share

Dataset Metrics

147 Views

0 Downloads

0 Citations

Description

Data file contains raw and log transformed microbiological data for tomato skin samples prepared and inoculated with bacterial pathogens or left uninoculated. Data are presented in log CFU/gram<sup>2</sup> in a column based subtitle for importing into RMP or other statistical analysis software. Data were generated by research experiments supported by USDA-NIFA.

Subject

Agricultural Sciences

Keyword

Food Safety

License/Data Use Agreement



CC BY-NC-SA

Files

Metadata

Terms

Versions

1 File



Tomato Microbiological Data Manuscript tab  
Tabular Data - 35.2 KB  
Published Date: 23 2021  
3 Downloads  
1.7 Varieties, 261 Observations, UNF:81m126\_q24w--

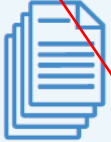


# Institutional data repository

Texas Data Repository > Texas A&M University Dataverse Repository > Forensic Science Study Dataverse >

## Forensic Science Information Seeking 2017 Data

Version 2.2



Sare, Laura; Bankston, Sarah, 2019, "Forensic Science Information Seeking 2017 Data", <https://doi.org/10.18738/T8/2BAQE0>, Texas Data Repository, V2

[Cite Dataset](#) [Learn about Data Citation Standards.](#)

**Access Dataset** ▾

Contact Owner Share

Dataset Metrics ?

448 Views ?

14 Downloads ?

0 Citations ?

**Description** ?

Response data from survey posted to the American Academy of Forensic Sciences in 2017 in fulfillment of NIJ Grant. (2019-06-29)

**Subject** ?

Social Sciences; Other

**Keyword** ?


DOI and citation

Usage statistics

Version control


# Institutional data repository


**License/Data Use Agreement**

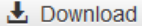
 **PUBLIC DOMAIN** CC0 1.0




**Terms of use and access**

Files Metadata Terms Versions

Search this dataset... 

Filter by  
File Type: All Access: All File Tag: All 

**1 to 2 of 2 Files**  Download

<input type="checkbox"/>		<a href="#">American Academy of Forensic Sciences Survey Data.csv</a> Comma Separated Values - 137.4 KB Published Jun 29, 2019 13 Downloads MD5: e8a...c34  Survey Data in CSV format <b>Data</b>	
--------------------------	---	--	---



# General purpose repository

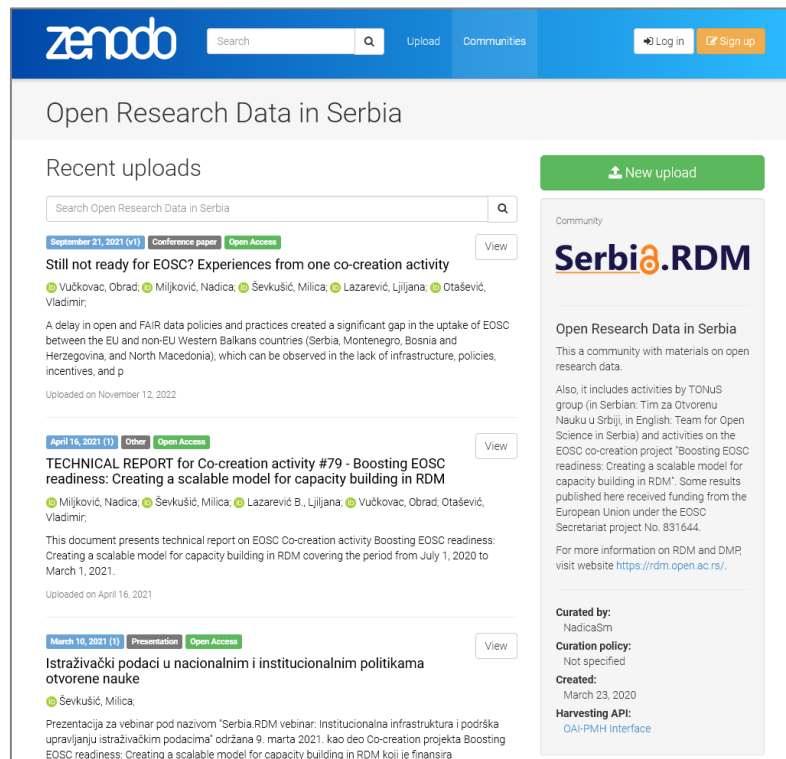
## PROS:

- 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



The screenshot displays the Zenodo interface for the 'Open Research Data in Serbia' community. At the top, the Zenodo logo and navigation links (Search, Upload, Communities, Log in, Sign up) are visible. The main heading is 'Open Research Data in Serbia'. Below this, there is a 'Recent uploads' section with a search bar and a 'New upload' button. The first upload is dated 'September 21, 2021 (v1)', categorized as a 'Conference paper' with 'Open Access' status. The title is 'Still not ready for EOSC? Experiences from one co-creation activity'. The authors listed are Vučković, Obrad; Mijković, Nadica; Ševkušić, Milica; Lazarević, Ljiljana; Otašević, Vladimir. A short paragraph of text follows, and the upload date is noted as 'November 12, 2022'. The second upload is dated 'April 16, 2021 (1)', categorized as 'Other' with 'Open Access' status. The title is 'TECHNICAL REPORT for Co-creation activity #79 - Boosting EOSC readiness: Creating a scalable model for capacity building in RDM'. The authors are Mijković, Nadica; Ševkušić, Milica; Lazarević B., Ljiljana; Vučković, Obrad; Otašević, Vladimir. A short paragraph of text follows, and the upload date is noted as 'April 16, 2021'. The third upload is dated 'March 10, 2021 (1)', categorized as 'Presentation' with 'Open Access' status. The title is 'Istraživački podaci u nacionalnim i institucionalnim politikama otvorene nauke'. The author is Ševkušić, Milica. A short paragraph of text follows. On the right side, there is a community profile for 'Serbi8.RDM' with a description of the community's focus on Open Research Data in Serbia and a list of metadata including 'Curated by: NadicaSm', 'Duration policy: Not specified', 'Created: March 23, 2020', and 'Harvesting API: OAI-PMH interface'.

# 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

The screenshot displays the Zenodo interface for a dataset titled "Open Research Data in Serbia". The page is highlighted with a red border. Key information includes:

- Publication date:** January 23, 2023
- DOI:** 10.5281/zenodo.7559361
- Keyword(s):** COVID-19, coronavirus, scientometrics, bibliometrics
- Related identifiers:** Cites <https://pages.semanticscholar.org/coronavirus-research> (Dataset) and <https://github.com/diwis/PaperRanking> (Software)
- Supplement to:** [www.biorxiv.org/content/10.1101/2020.04.11.037093v2](http://www.biorxiv.org/content/10.1101/2020.04.11.037093v2) (Preprint)
- Communities:** Coronavirus Disease Research Community - COVID-19, Zenodo
- License (for files):** Creative Commons Attribution 4.0 International

On the right side of the page, there is a community profile for "Serbi8.RDM" (Open Research Data in Serbia) with a description of the community's mission and funding sources.

# 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

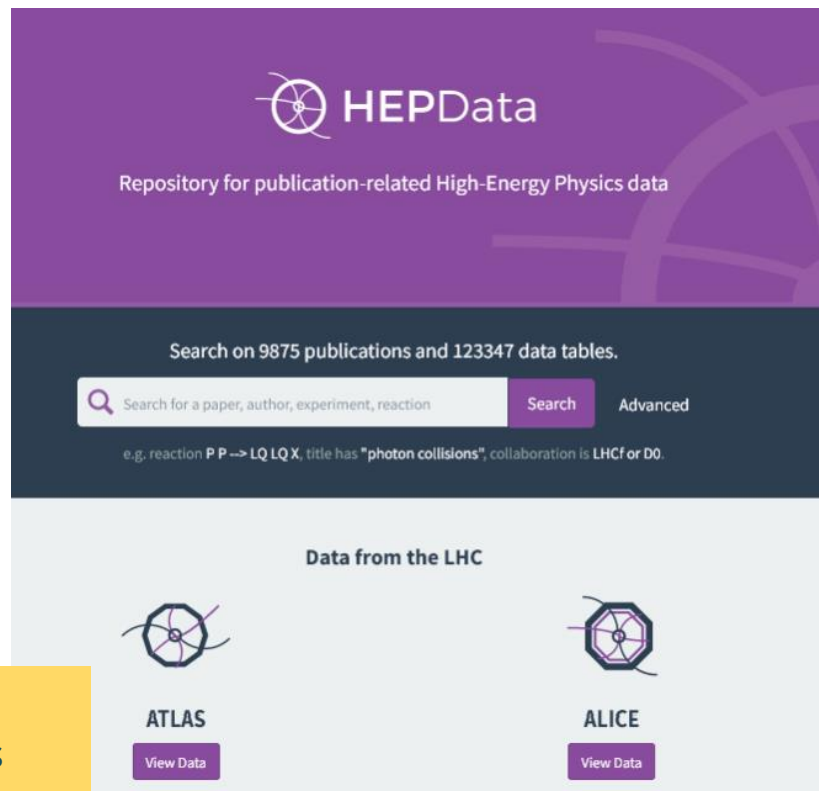
## PROS:

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

## CONS:

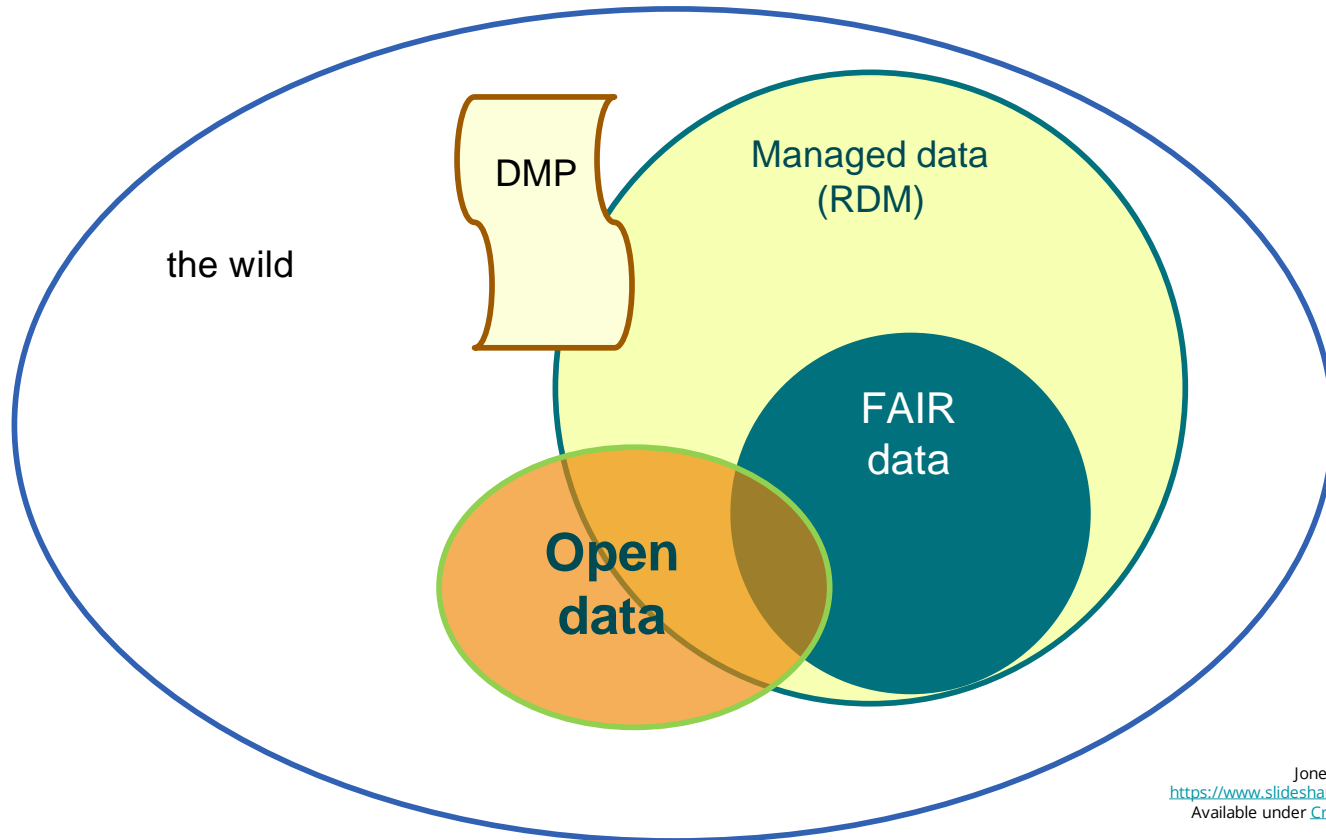
- may include costs
- require high standards in (meta)data quality

Point to the  
community services



The screenshot shows the HEPData website interface. At the top, there is a purple header with the HEPData logo and the text "Repository for publication-related High-Energy Physics data". Below the header, a dark blue section contains a search bar with the text "Search on 9875 publications and 123347 data tables." and a search button. Below the search bar, there is a search input field with the placeholder text "Search for a paper, author, experiment, reaction" and a search button. Below the search input field, there is a small text snippet: "e.g. reaction P P → LQ LQ X, title has 'photon collisions', collaboration is LHCf or D0." Below the search section, there is a light gray section titled "Data from the LHC". This section contains two columns of data. The first column features the ATLAS logo, the text "ATLAS", and a "View Data" button. The second column features the ALICE logo, the text "ALICE", and a "View Data" button.

# Summary





Institut za nuklearne  
nauke Vinča

# Questions?

Obrad Vučkovic

University of Belgrade

Vinča Institute of Nuclear Sciences - Library

ORCID: [0000-0001-5616-2680](https://orcid.org/0000-0001-5616-2680)



Except as otherwise noted, this presentation is licensed under the Creative Commons Attribution 4.0 International Licence.  
To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.