

Research Data Management and Open/FAIR Data

6th International PhD Summer School

13 Jul 2022 | KTU | Lithuania

eva.hnatkova@techlib.cz



European legislation on open data

The Directive on open data and the re-use of public sector information, also known as the 'Open Data Directive' ([Directive \(EU\) 2019/1024](#)) entered into force on 16 July 2019.

Article 10

Research data

1. Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'), following the principle of 'open by default' and compatible with the FAIR principles. In that context, concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests, shall be taken into account in accordance with the principle of 'as open as possible, as closed as necessary'. Those open access policies shall be addressed to research performing organisations and research funding organisations.
2. Without prejudice to point (c) of Article 1(2), research data shall be re-usable for commercial or non-commercial purposes in accordance with Chapters III and IV, insofar as they are publicly funded and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository. In that context, legitimate commercial interests, knowledge transfer activities and pre-existing intellectual property rights shall be taken into account.

What is RDM?

Research Data Management is a broad term that includes all practices and actions to ensure that research data are:

- secure,
- sustainable,
- and easy to find, understand and (re)use

not only during a research project, but also in the longer term.

RDM is the **process** of organization, storage, preservation, and sharing of data collected and used in a research project.

Taking proper care of data is an essential part of **good research practice**

Why manage research data?



- The drive is broken and I didn't make a backup!
- A former student/collaborator worked on the data and he/she disappeared
- I can't find the password to access the data
- I don't have the software anymore
- I had troubles in understanding my code written years ago
- I cannot reproduce experiments done years ago



Why manage research data?

- Make your research easier
- Stop yourself drowning in irrelevant stuff
- Save data for later
- Avoid accusations of fraud or bad science
- Share your data for re-use
- Get credit for it
- Meet funder/institution requirements

Well-managed data opens up opportunities for re-use, sharing and makes for better science!

Research Data Lifecycle



It describes the different stages research data go through

- before,
- during,
- and after a research project

Source: [Elixir RDMkit](#)



Data Management Plan

- A practical tool for researchers working on individual or collaborative projects.
- Most funders and institutions now require that projects have a DMP, as this may facilitate workflow and FAIR-ness

Should be a **'living document'** which is updated along the research process to reflect what actually happened with the data.

DMP example

Evaluating the sustainability and nutritional benefits of alternative production systems and foods - focusing on vertical farming systems

A Data Management Plan created using dmponline

Creator: Ciara Byford

Affiliation: Cranfield University

Funder: Biotechnology and Biological Sciences Research Council (BBSRC)

Template: BBSRC Template

ORCID iD: 0000-0001-7429-5815

Project abstract:

Literature review looking at emerging trends, developments and innovations in the field of urban farming and vertical farming. Literature review of the social, environmental and economic risks and opportunities posed by vertical farming systems.

Last modified: 20-04-2021

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Evaluating the sustainability and nutritional benefits of alternative production systems and foods - focusing on vertical farming systems

Data areas and data types

Outline the volume, type and content of data that will be generated e.g. experimental measurements, models, records and images

Literature review is to be conducted for research paper, dataset to be created is a reference list. A reference list will be compiled of publicly available research papers, online newspapers, online market reports, business websites and online books for use in literature review paper. Reference list consists of approximately 116 references.

Standards and metadata

Outline the standards and methodologies that will be adopted for data collection and management, and why these have been selected

A broad horizon scanning methodology has been developed to gather, organise and rank insights on the subject of vertical farming, these insights are then to be used to identify key drivers of development and innovation in the field of vertical farming. All horizon scans conducted use publicly available research papers, online newspapers, online market reports, business websites and online books.

Metadata will be provided and added to a repository (CORD). Information gathered from publicly available literature will be organised into a reference list and all work relating to the project will be securely stored on Cranfield University secure cloud storage system (OneDrive).

File will be saved in folder named 'Vertical Farming', file name will consist of date in following format: day, month, year, file name 'vertical farming literature review' file will be saved in pdf and .doc format on both harddrive and Cranfield University secure onedrive cloud system.

Relationship to other data

State the relationship to other data available in public repositories

Dataset in the form of reference list will be compiled. A reference list consisting of publicly available research papers, online newspapers, online market reports, business websites and online books will be made. Approximately 116 references will be included.

Secondary Use

Outline the further intended and/or foreseeable research uses for the completed dataset(s)

Horizon scans will be made using publicly research papers, online newspapers, online market reports, business websites and online books. Insights will be developed based on this information and this will help identify key drivers of development and innovation in the field of vertical farming.

The paper and reference list will be made publicly available for transparency.

Methods for data sharing

Outline the planned mechanisms for making these data available, e.g. through deposition in existing public databases or on request, including access mechanisms where appropriate

Data will be retained securely in Cranfield's institutional data repository, CORD, which uses the figshare platform and preserves data for at least 10 years after project end, with datasets assigned a DOI for long-term accessibility, in accordance with Cranfield's Management of Research Data Policy. A CC BY license will be used.

Proprietary data

Outline any restrictions on data sharing due to the need to protect proprietary or patentable data

Entire literature review and reference list will be shared, no barriers to sharing. It is foreseen that information gathered in reference list of publicly available research papers, online newspapers, online market reports, business websites and online books could be accessed by other users looking for further information on the subject of vertical farming.

[DMPonline: Public DMPs](#)



Data Management Plan (DMP)

- DMP is designed to help you plan for your research and (sometimes hidden) costs well in advance
- Plans are partially domain specific and institution specific

First step of DMP

Find out if Institution/Lab has Research Data Management (**RDM**) defined



Source: [Elixir RDMkit](#)



"In proposal" DMP declaration



Adhere to FAIR principles and specify:

- How/where you will get data
- What type of data (type, format, size)
- Where you will store them (with long/short backup strategy)
- Who will have access to them
- If they are sensitive
- How they will be analyzed

Define:

- What data will be published and where
- Under what licence
- Who will own them

Source: [Elixir RDMkit](#)



Funder template:

- Horizon Europe DMP Template (EU)

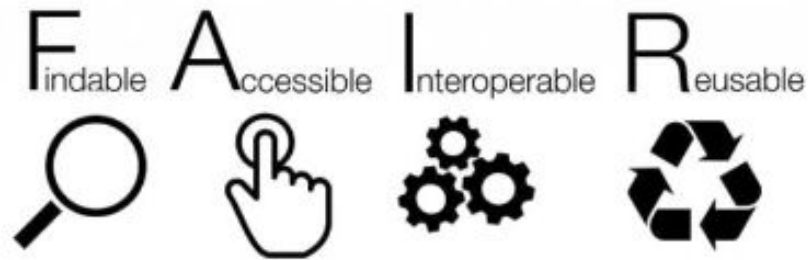
Online tools for creation of DMPs:

- [DS Wizard](#) (EU level), [DPM Online](#) (UK), [Argos](#) (OpenAIRE)

	 DSW	About	Features	Resources	Get Started		
DSW Compared to Other DMP Tools							
	 DSW	DMPonline Free	DMPonline Subscription	DMPTool	Argos	DataWiz	easyDMP
Template Selection	✔	✔	✔	✔	✔	Limited	✔
Detailed Export Settings	✔	✔	✔	✔	✔	Limited	Limited
Template Creation	✔	✘	✔	Limited	✔	✘	Limited
Template Styling	✔	✘	✔	Limited	Unknown	Limited	Limited

Data Stewardship Wizard

Open vs FAIR Data



OPEN
DATA

FAIR principles

- **Findable** - rich metadata, keywords, persistent identifiers (e.g. DOI)
- **Accessible** - data in trusted repositories
- **Interoperable** - standards
- **Re-usable** - licenses



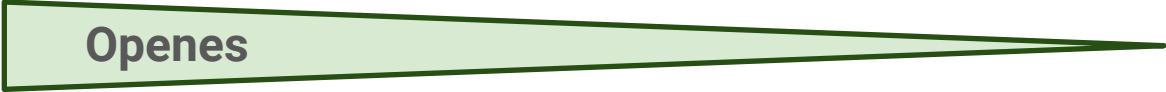
[Go FAIR](#) | [Hansen & al \(2018\)](#)



FAIR Data – open by default

FAIR Data: **“As open as possible, as closed as necessary”**

Opens



Open	Restricted / Controlled	Closed
Can be freely used, modified & shared by anyone for any purpose	Limits on who can access & use data, how, and/or for what purpose: • only subset of data • only certain (types of) users • only certain types of use •	Under embargo Unable to share



www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically

Received: 10 December 2015

Accepted: 12 February 2016

Published: 15 March 2016

<https://www.nature.com/articles/sdata201618.pdf>

FAIR principles

They **contain** guidelines for **good data management practice** that aim at making data FAIR:

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**e-usable



Go
FAIR

Each letter refers to a list of principles with a total of 15 principles altogether



FAIR principles



FAIR\$FAIR
Fostering Fair Data Practices in Europe

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

F_{indable}

- F1. (meta)data are assigned a globally **unique and persistent identifier**
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are **registered or indexed in a searchable resource**

A_{ccessible}

- A1. (meta)data are retrievable by their identifier using a **standardized communications protocol**
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. **metadata are accessible**, even when the data are no longer available

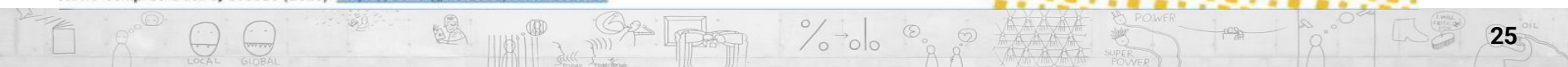
I_{nteroperable}

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (meta)data use **vocabularies that follow FAIR principles**
- I3. (meta)data include qualified references to other (meta)data

R_{eusable}

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible **data usage license**
 - R1.2. (meta)data are associated with **detailed provenance**
 - R1.3. (meta)data meet **domain-relevant community standards**

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>



The implementation of the principles should be adapted to each research area => each community will make the principles work in their respective contexts

Understanding FAIR principles

- **both humans and machines** are intended as digesters of data.
- **FAIR principles apply to both data and its metadata** – i.e. records about data sets.
- FAIR principles are **not only about open data**, you can work in a FAIR *manner* with data that is not intended for public availability



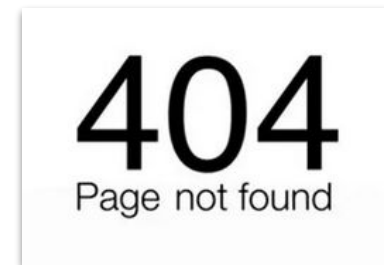
Findable

Findable #1

(Meta)data are assigned globally unique and **persistent identifiers**

One of the big problems with data concerns

- the ability to **cite** your own and other **people's data**,
- and keep pointing to its **exact location** in cyberspace, because the location might change.



=> this can be solved by using **persistent identifiers**

They work like a **big index or registry** where you assign a **unique key** (the identifier) to each data set. If someone tries to follow the identifier (often referred to as “resolving a persistent identifier”), the resolver will point to the correct web address (URL). If the URL changes – i.e. if data is moved – the one who made the key is responsible for providing the new location to the resolver. In this way, you do not end up in blind alleys of “page not found”. That is why it is called persistent.

Findable #1

(Meta)data are assigned globally unique and **persistent identifiers**

- The most common identifier (for data sets) is DOI

What is DOI?

- DOI stands for **Digital Object Identifier**
- A DOI is a **standardized system** for identifying an electronic or digital object
- A DOI looks like this: **DOI: 10.1080/1557533090327963**
(prefix/suffix)



Findable #1

Data availability

Underlying data

Zenodo: Raw data for manuscript "Simple and robust method for determination of laser fluence thresholds for material modifications: an extension of Liu's approach to imperfect beams". <http://doi.org/10.5281>

[/zenodo.4421003](http://doi.org/10.5281/zenodo.4421003)³⁰



This project contains the following underlying data:

- PSFlab_raw_1030nm-f50mm-T25.txt (PSFlab raw data of $P_T = 25\%$ profile represented in Figure 2).
- PSFlab_raw_1030nm-f50mm-T40.txt (PSFlab raw data of $P_T = 40\%$ profile represented in Figure 2).

References

30. Garcia-Lechuga M, Grojo D: Raw data for manuscript "Simple and robust method for determination of laser fluence thresholds for material modifications: an extension of Liu's approach to imperfect beams" [Data set]. *Zenodo*. 2021. <http://www.doi.org/10.5281/zenodo.4421003>

What is metadata?

- **Metadata** is defined as the data providing information about one or more aspects of the data, in other words '**data about data**'
- it is used to summarize **basic information** about data which can **make tracking and working with specific data easier**

Metadata may be **generic** and **specific** for each discipline

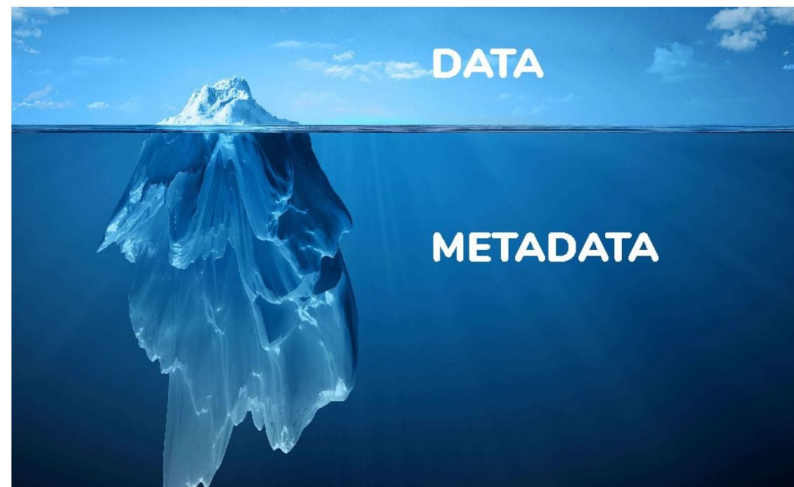


Findable #2

Data are described with **rich metadata**

Metadata includes (e.g.):

- title, author (ORCID iD)
- key words
- creation/last modification date
- funder
- type, size, file names
- unique & persistent identifiers



Findable #2

Data are described with **rich metadata**

When humans and machines look for data

- metadata are often the **first point of contact**, as they are usually **indexed in search engines** etc.
- It is often the metadata that **determine whether the data set they describe is perceived as relevant or not** for a given usage scenario.
- If you asked a human being, the same query words they would use to find a data set should be available in the metadata.
- This is metadata **about the context and/or prerequisites** for the data set, quality issues etc., as well as a number of discipline-specific data, e.g. sample size, equipment etc.

Findable #3

Metadata clearly and explicitly include the identifier of the data they describe

Metadata and data are two separate things

- They **should be treated as such, they can have a life of their own.**
- E.g. metadata that are harvested for big indexes that do not hold or index the contents of the data files.
- Just like article metadata that are fetched for indexing in commercial and non-commercial search engines, without them looking at the content of the article itself.
- If the metadata do not include reference to the data they describe, it is doubtful - especially for a machine - that the data described by the metadata will ever be found.

Findable #4

(Meta)data are **registered or indexed in a searchable resource**

- **Repositories** are websites – most often – and represent a common way for building structured indexes of metadata and data sets that are uploaded to the repository.
- The indexes often adhere to a specific way of describing the data using a common standard.
- This will allow both repository and other search engines to harvest and index these registries, often aggregating them to larger indexes that eventually can be cross-searched

Accessible

Accessible #1

(Meta)data are retrievable by their identifier using a **standardised communication protocol**

Once the data or metadata have been found, the user must know how to access, possibly including authentication and authorization (standardised mechanisms)

- This principle states that access should be provided through a standardised protocol. Most often, these are protocols we know from the internet – e.g. http(s) or FTP.
- **This is usually the case, when data are deposited in a trusted repository**
- However, there might be cases where you will need additional mechanisms such as contract information or similar before someone can access your data.
- This is perfectly in line with the FAIR principles, if you clearly account for this in the metadata.
- This may be in the form of contact options that are broadly accepted and easy to use. Examples of this are telephone numbers and email addresses

Accessible #1.1

The protocol is open, free and universally implementable

A protocol is the technical term for the standard of how data are transported via a network, e.g. the internet.

- The Internet protocol suite is the conceptual model and set of communications protocols used in the Internet and similar computer networks.
- TCP/IP (Transmission Control Protocol and the Internet Protocol) is an example of an open, free and universally implementable protocol that is used as protocol for most of the internet.
- This means that **anyone can use it without having to pay usage fees**. If you choose a protocol with restrictions on usage, you might prevent other people from accessing to your (meta)data, thus making it hard – or even impossible – to use the data you have published. **If you use a trusted repository for data publication, the repository will make sure that you are in line with this principle.**

Accessible #1.2

The protocol allows for authentication and authorisation when required

One of the most widely believed myths on FAIR data is that FAIR data must be available as open data, meaning that they should be free for everyone to download. That is not the case.

- This principle states that if you place data behind some sort of digital wall, being a paywall or a **simple approval system for access**, the **system must allow for some type of authentication and authorisation**. This holds for both humans and machines.
- **Authentication** is all about telling a system **who you** - or your system - **are**.
- **Authorisation** is the **process where the system is evaluating, if you are allowed to access a given resource**.

There are many good reasons to place data behind authentication and authorisation mechanisms.

- **Manual processes** - who should have access to the data, the repository or storage system must be able to contact you and seek approval.
- **Other systems** will let users or systems register themselves and then provide access to the data, while maintaining a logbook of who has been given access to the data.

Accessible #2

Metadata are accessible even when the data are no longer available

There are plenty of good reasons why data disappear from repositories and similar places

- They may be withdrawn **due to cost** of having them online, if no one is left to maintain access to the data, or for other reasons.
- However, **if the metadata about the data set disappear as well, it will leave humans and machines in an “unfulfilled” state**, when they try to retrieve the data, e.g. when resolving a persistent identifier, or following a link to the entry of a repository.
- Therefore, make sure to leave metadata telling that “**yes, the data were here, but they are no longer available**”.
- **Leaving metadata may also offer information on the context, the authors and the institution, where the data were created**, for those looking for further details.

Interoperable

Interoperable #1

(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

Humans, and especially machines, can have a hard time interpreting data (heterogeneity: a well-known problem)

- Words are ambiguous, and the multitude of spoken and written languages add further to the complexity.
- Problems range from **not being able to interpret the value of a cell due to missing information** on the metrics used,
- to more complex situations where you will have to **look for many different terms describing the same object**

Interoperable #1

FAIR principles address this issue

by recommending the **use of shared data standards** for representing data, and the **use of vocabularies** and **ontologies** to represent values and mark-up data.

=> **Vocabularies** and **ontologies** are often defined within the research communities

Interoperable #2

(Meta)data use **vocabularies that follow the FAIR principles**

A vocabulary is only good, if it is accessible and allows for the right interpretation of the data.

This principle highlights the importance of using vocabularies that are

- common to the community and
- well documented, and
- can be referred to using persistent identifiers.

Usually, you will find this type of vocabulary, taxonomy etc. within your research discipline or maybe in other disciplines, where these are developed.

Interoperable #3

(Meta)data include qualified references to other (meta)data

It is important to be able to trace the connections between your data set and data sets that are related to it.

- by linking to other data sets that are not included in your work.
- through connections that **show how your data set is derived from a previous version or e.g. is processed data based on some raw data.**
- it is important to maintain these connections by referencing between data sets. If your data set relies on other's data - or your own - this is also an appropriate method to ensure that proper credit is given to the people who created the data that your data is based upon.

Reusable

Reusable #1

(Meta)data are richly described with a plurality of accurate and relevant attributes

Labelling your data with relevant attributes - most often in the form of metadata

- This can be in the form of purpose and **processing statements, equipment used, software versions** etc.
- Imagine finding your own data. Now think of the contextual information that would benefit you in determining whether the data is relevant to your specific needs - and whether you would be able **to understand how the data were created**.
- **Speaking of machines**, your best choice would be to use e.g. **controlled vocabularies, persistent identifiers** or similar to make the contextual description unambiguous

Reusable #1.1

(Meta)data are released with a clear and accessible **data usage license**

Licensing data and metadata is an important aspect of the FAIR principles

- both if you retain some or all rights, or if you set your **data free as completely open data**.
- In a FAIR context, a license is a **standardised machine-readable statement** that **tells the end user exactly how he/she can use the data, and under which conditions**.
- You can apply many different licenses to data. One of the **most common is the Creative Commons**
- you explicitly **state how and if you are to be cited** when your data are reused
- Typically, you **choose a license when depositing data in a repository**.
- **You do not have to apply the same license to all your data,**
- **The worst thing you can do is not to apply any license** at all. Because then your data is most likely protected by copyright as default, even though the data are publicly accessible.

Reusable #1.2

(Meta)data are associated with **detailed provenance**

Much of the value in data is the ability for a machine or a human to judge the origin of the data.

- This includes the ability to know **how the data was created, by whom it was created, and with which type of equipment?**
- Also, has the data been **processed, or is it raw data?** If it was processed, how was the **workflow?**
- This is quite **similar to a section on method in a paper or article**, and **you can refer to this type of documentation** from your data set. However, keep in mind that this might not be readable for a machine.
- **Remember to include provenance of who you are, and how you would like the data set to be cited/credited if used elsewhere.**
- The easiest way to get started, is to try and think of yourself as a re-user of your own data. What details would you need to evaluate and trust a given data set?

Reusable #1.3

(Meta)data meet domain-relevant community standards

Working with data sets from a variety of resources is much simpler, if everybody agrees to a certain standard way of organising and describing the data.

- That is why **many disciplines have created metadata standards** for describing data, and created **lists of recommended file formats** etc.
- However, **not everything can be standardized**, of course, and many research disciplines are breaking new ground where there are currently no standards – and then you will turn to more generic standards, or begin inventing new ones.
- Notice that **standard in this sense is not a quality measurement** indicating a level of high or low quality of the (meta)data. It should always be judged by the people who are re-using the data in their specific contexts.

Which repository to use?

The general steps for finding a data repository are:

1. use a **disciplinary** repository if there is one;
2. alternatively, use the **institutional** repository, if you have one where the data will also be available for the long term;
3. use the **catch-all** repository [Zenodo](https://zenodo.org), maintained by CERN;
4. or search the global [re3data.org portal](https://re3data.org) for a fitting repository - this provides several filtering options.

Preparing and depositing a dataset

Making sure the files are in a persistent format, that you have all the necessary metadata, and that you have documented the dataset in a readme file.

Data Guidelines - example

1. [Background](#)
 - 1.1 [Open Data Policy](#)
 - 1.2 [FAIR Data Principles](#)
2. [Share Your Data in 4 Steps](#)
 - 2.1 [Prepare Your Data for Sharing](#)
 - 2.2 [Select a Repository](#)
 - 2.3 [Add a Data Availability Statement to Your Article](#)
 - 2.4 [Link Your Datasets to Your Article](#)

Disciplinary repositories



This list is part of the [Open Access Directory](#).

- This is a list of OA disciplinary repositories (also called central or subject repositories). Unless otherwise noted, they accept relevant deposits regardless of the author's institutional affiliation.
- This list includes both preprint and postprint repositories.
- When possible, please indicate the institution hosting or maintaining the repository. If a repository only accepts submissions in certain languages, please indicate the languages. If a repository is open in some respects but not others, please include it with an annotation rather than exclude it.
- For a bibliography of works on disciplinary repositories, see the section on [Disciplinary Archives](#) in the OAD [Bibliography of open access](#).
- Related lists in OAD: [Data repositories](#).
- Also see the [Wikipedia list of preprint repositories](#), organized by field or discipline.
- For news about disciplinary repositories, including some newly launched repositories not yet listed here, follow the [oa.repositories.disciplinary](#) tag of the [Open Access Tracking Project](#).
- Wikidata identifier for this page: [Q56229148](#).
- Alphabetical by field.

Contents [\[hide\]](#)

- 1 Agriculture
- 2 Anthropology
- 3 Architecture and Civil Engineering
- 4 Arts
- 5 Art history
- 6 Astrophysics
- 7 Biology / Life sciences
- 8 Business
- 9 Classics
- 10 Cognitive science
- 11 Computer science
- 12 Cryptology
- 13 Development
- 14 Digital preservation

http://oad.simmons.edu/oadwiki/Disciplinary_repositories



Registry of Research Data Repositories



Re3data offers

- Researchers
 - Funding
 - Organization
 - libraries and
 - publishers
- an overview of
existing
international
repositories for
research data

<https://www.re3data.org/>



- Subjects ⊕
- Content Types ⊕
- Countries ⊕
- AID systems ⊕
- API ⊕
- Certificates ⊕
- Data access ⊕
- Data access restrictions ⊕
- Database access ⊕
- Database access restrictions ⊕
- Database licenses ⊕
- Data licenses ⊕
- Data upload ⊕
- Data upload restrictions ⊕
- Enhanced publication ⊕
- Institution responsibility type ⊕
- Institution type ⊕
- Keywords ⊕
- Metadata standards ⊕
- PID systems ⊕

Re3data - search & filters



VYSOKÁ ŠKOLA
CHEMICKO-TECHNOLOGICKÁ
V PRAZE

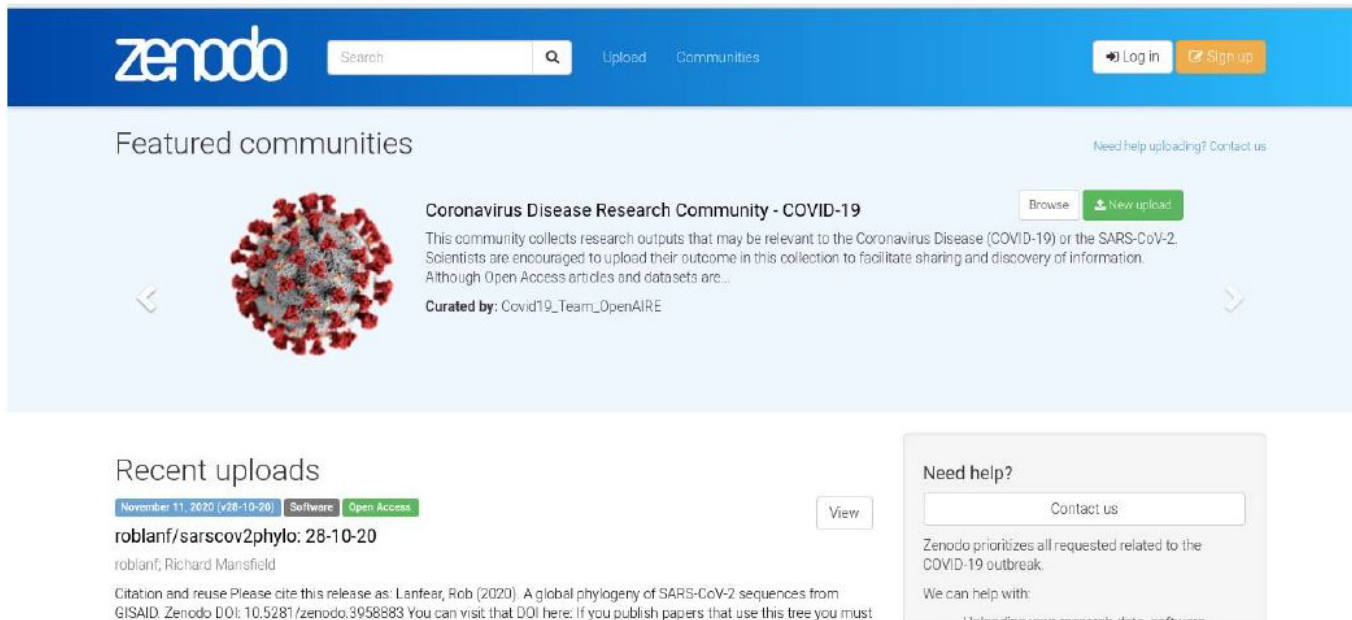


<https://www.re3data.org/search>



Public repository

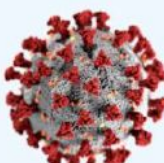
EU level (e.g. Zenodo)



The screenshot shows the Zenodo website interface. At the top is a blue header with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. On the right of the header are 'Log in' and 'Sign up' buttons. Below the header, the 'Featured communities' section is displayed. It features a large image of a coronavirus particle and the title 'Coronavirus Disease Research Community - COVID-19'. A description states that the community collects research outputs related to COVID-19 or SARS-CoV-2. A 'Browse' button and a green 'New upload' button are present. Below this, the 'Recent uploads' section shows a release titled 'roblanf/sarscov2phylo: 28-10-20' by roblanf and Richard Mansfield, dated November 11, 2020. It includes a 'View' button and a citation/reuse section. On the right side of the interface, there is a 'Need help?' section with a 'Contact us' button and text stating that Zenodo prioritizes requests related to the COVID-19 outbreak.

zenodo Search Upload Communities Log in Sign up

Featured communities Need help uploading? Contact us

 **Coronavirus Disease Research Community - COVID-19** Browse New upload

This community collects research outputs that may be relevant to the Coronavirus Disease (COVID-19) or the SARS-CoV-2. Scientists are encouraged to upload their outcome in this collection to facilitate sharing and discovery of information. Although Open Access articles and datasets are...

Curated by: Covid19_Team_OpenAIRE

Recent uploads

November 11, 2020 (v28-10-20) Software Open Access View

roblanf/sarscov2phylo: 28-10-20

roblanf; Richard Mansfield

Citation and reuse Please cite this release as: Lantear, Rob (2020). A global phylogeny of SARS-CoV-2 sequences from GISAID. Zenodo DOI: 10.5281/zenodo.3958883 You can visit that DOI here: If you publish papers that use this tree you must

Need help? Contact us

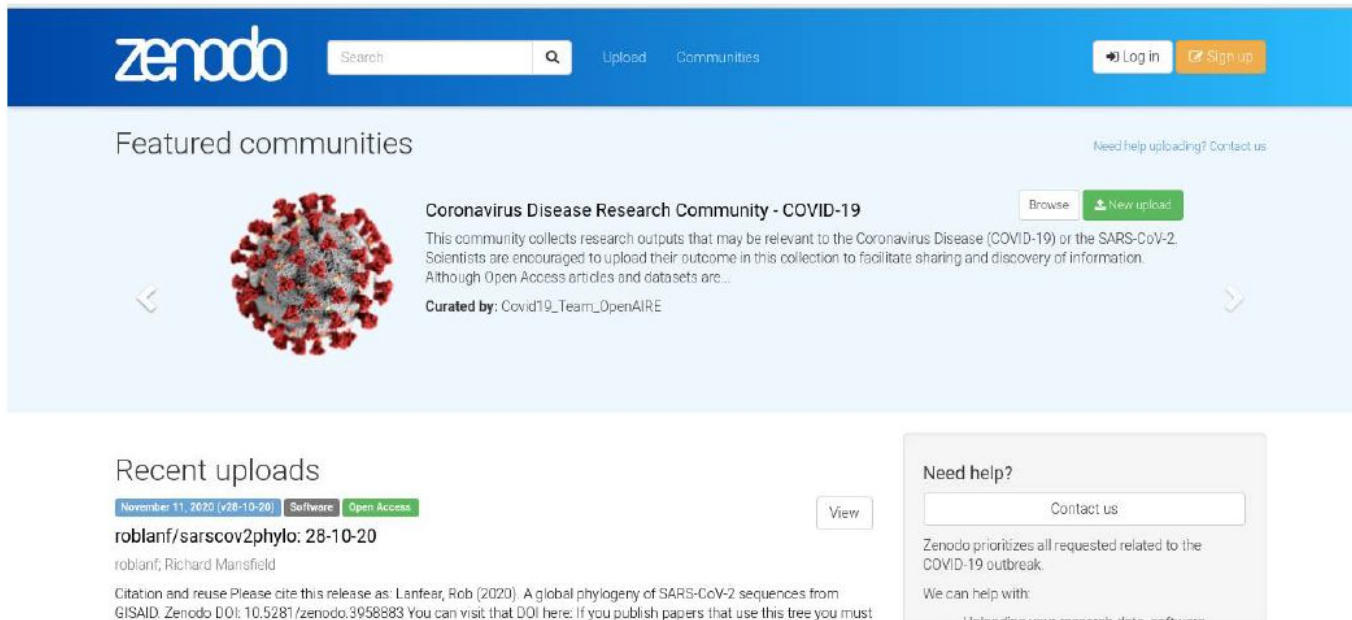
Zenodo prioritizes all requested related to the COVID-19 outbreak.

We can help with:

<https://zenodo.org/>

Public repository

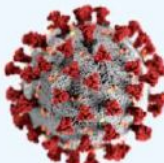
EU level (e.g. Zenodo)



The screenshot shows the Zenodo website interface. At the top is a blue header with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. On the right of the header are 'Log in' and 'Sign up' buttons. Below the header, the 'Featured communities' section is displayed. It features a large image of a coronavirus particle and the title 'Coronavirus Disease Research Community - COVID-19'. A description states that the community collects research outputs related to COVID-19 or SARS-CoV-2. A 'Browse' button and a green 'New upload' button are present. Below this, the 'Recent uploads' section shows a release titled 'roblanf/sarscov2phylo: 28-10-20' by roblanf, dated November 11, 2020. It includes a 'View' button and a citation/reuse section. On the right side of the interface, there is a 'Need help?' section with a 'Contact us' button and text stating that Zenodo prioritizes requests related to the COVID-19 outbreak.

zenodo Search Upload Communities Log in Sign up

Featured communities Need help uploading? Contact us

 **Coronavirus Disease Research Community - COVID-19** Browse New upload

This community collects research outputs that may be relevant to the Coronavirus Disease (COVID-19) or the SARS-CoV-2. Scientists are encouraged to upload their outcome in this collection to facilitate sharing and discovery of information. Although Open Access articles and datasets are...

Curated by: Covid19_Team_OpenAIRE

Recent uploads

November 11, 2020 (v28-10-20) Software Open Access View

roblanf/sarscov2phylo: 28-10-20

roblanf; Richard Mansfield

Citation and reuse Please cite this release as: Lantear, Rob (2020). A global phylogeny of SARS-CoV-2 sequences from GISAID. Zenodo DOI: 10.5281/zenodo.3958883 You can visit that DOI here: If you publish papers that use this tree you must

Need help? Contact us

Zenodo prioritizes all requested related to the COVID-19 outbreak.

We can help with:

<https://zenodo.org/>



{ Findable
Accessible



{ Interoperable
Reusable

What is Zenodo?

Zenodo is a **general-purpose open-access repository** developed under the European OpenAIRE program (2013) and **operated by CERN**.

It allows researchers to deposit (**up to 50GB**)

- research papers,
- data sets,
- research software,
- reports,
- and any other research related digital artifacts.



<https://zenodo.org/>

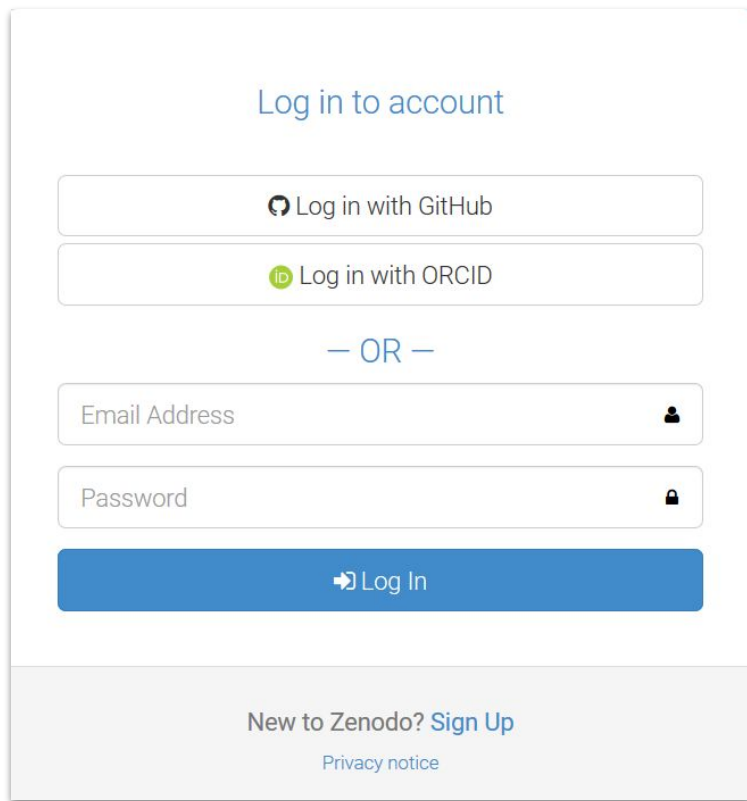
It provides a **persistent digital object identifier (DOI)** to make the work easier to cite and supports **various license type**

Why to use Zenodo?

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** — Easily update your dataset with our versioning feature.
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo.
- **Usage statistics** — All uploads display standards compliant usage statistics

<https://zenodo.org/>

Zenodo - Login-in



The screenshot shows the Zenodo login interface. At the top, it says "Log in to account". Below this are two buttons: "Log in with GitHub" (with a GitHub icon) and "Log in with ORCID" (with an ORCID icon). In the center, there is a separator "— OR —". Below the separator are two input fields: "Email Address" (with a user icon) and "Password" (with a lock icon). At the bottom of the form is a blue button labeled "Log In" with a right-pointing arrow icon. Below the form, in a light gray box, it says "New to Zenodo? [Sign Up](#)" and "[Privacy notice](#)".

Zenodo has integrated **authentication mechanism:**

- Login via ORCID or Github credentials

More information: [FAIRcookbook: Deposit to Zenodo Archive](#)

Zenodo - Data upload

Files can be dragged and dropped, with the following **limitations:**

- **5 GB** max each
- **50 GB** total / dataset

More information: [FAIRcookbook: Deposit to Zenodo Archive](#)

Zenodo - Data upload

Files 

Choose files Start upload

Filename (2 files)	Size	Progress	Delete
20201012_Data a datové repozitáře_NTK_VSCHT.pdf	198 kB		
2020_What is ERA and EHEA_by Eva.pdf	290 kB		

Note: File addition, removal or modification are not allowed after you have published your upload. This is because a Digital Object Identifier (DOI) is registered with [DataCite](#) for each upload.

(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.



More information: [FAIRcookbook: Deposit to Zenodo Archive](#)

Zenodo - Data upload

Upload type required ▾

 Publication ☒	 Poster ☐	 Presentation ☐	 Dataset ☐	 Image ☐	 Video/Audio ☐	 Software ☐	 Lesson ☐	 Physical object ☐	 Other ☐
--	--	--	---	---	--	--	--	---	---

Publication type ▾
Publication type.

More information: [FAIRcookbook: Deposit to Zenodo Archive](#)

Zenodo - providing metadata



VYSOKÁ ŠKOLA
CHEMICKO-TECHNOLOGICKÁ
V PRAZE

Basic information required ▾

Digital Object Identifier

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 *

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

Title *

Required.

Authors *

ORCID (e.g.: 0000-0002-1825-0097)

Optional.

[+ Add another author](#)

Description *

More information: [FAIRcookbook: Deposit to Zenodo Archive](#)

Zenodo - providing metadata

 **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](https://iso639.org) for more information.

 **Keywords**

[+ Add another keyword](#)

 **Additional notes**

Optional.

More information: [FAIRcookbook: Deposit to Zenodo Archive](#)

Zenodo - Access and License

License required 

Access right *

☒  Open Access

☐  Embargoed Access

☐  Restricted Access

☐  Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

License *

Start typing a license name...

- Creative Commons Attribution 4.0 International
- Creative Commons Attribution 1.0 Generic
- Creative Commons Attribution 2.0 Generic
- Creative Commons Attribution 3.0 Unported
- Creative Commons Attribution 3.0 Austria
- Creative Commons Attribution 3.0 United States

Access right *

☐  Open Access

☒  Embargoed Access

☐  Restricted Access

☐  Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

Embargo date



License *



September 2025

Sun	Mon	Tue	Wed	Thu	Fri	Sat
31	01	02	03	04	05	06
07	08	09	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	01	02	03	04

More information: [FAIRcookbook: Deposit to Zenodo Archive](#)

Zenodo - Miscellaneous information

Funding

recommended

Zenodo is integrated into reporting lines for research funded by the European Commission via [OpenAIRE](#). Specify grants which have funded your research, and we will let your funding agency know!

Grants

European Commission (EU)

Start typing a grant number, name or abbreviation...

Optional. OpenAIRE-supported projects only. For other funding acknowledgements, please use the *Additional Notes* field.
Note: a human Zenodo curator will need to validate your upload - you may experience a delay before it is available in OpenAIRE.
[+ Add another grant](#)

Related/alternate identifiers

recommended

Contributors

optional

References

optional

Journal

optional

Conference

optional

Book/Report/Chapter

optional

Thesis

optional

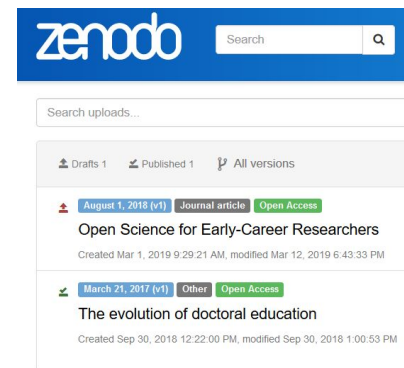
Subjects

optional

Delete

Save

Publish



More information: [FAIRcookbook: Deposit to Zenodo Archive](#)

