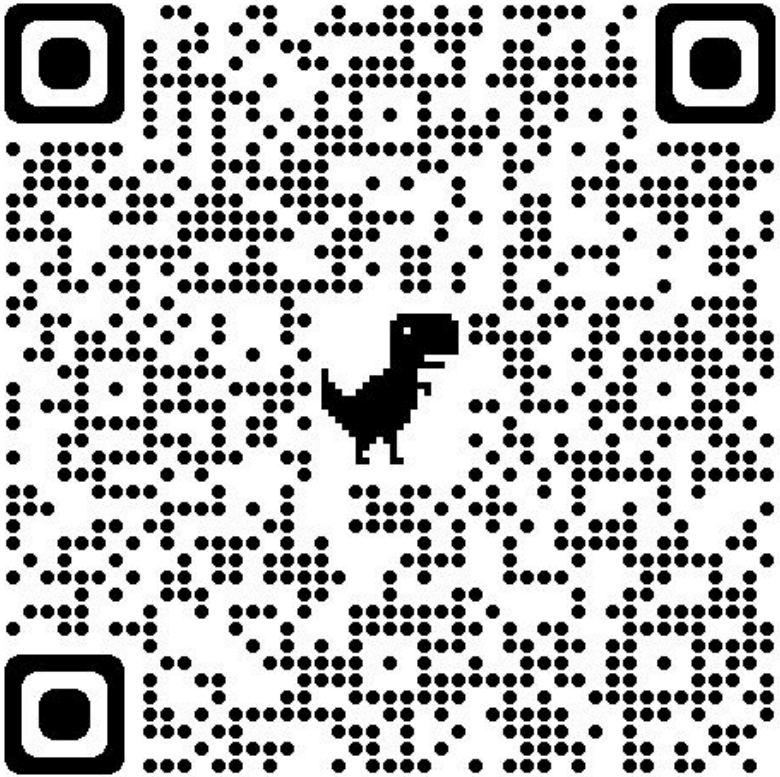
Data publishing

Irena Njezic
Institute of Virology, Vaccines and Sera “Torlak”



Q1. Why publish data?

Q2. Where to publish data?

Q3. What to deposit?

Clarify concepts and differences

Deposit: upload a digital object on a platform that allows to correctly describe the object through metadata and that implements long-term preservation

Give access: once the object is deposited, the authors can choose type of access that can be granted and assigns a licence to reuse the contents.

What is the difference between sharing, publishing & archiving?

SHARE: any way of sharing information, within a specific group for example

PUBLISH: citable artefact, discoverable

ARCHIVE: ensure long-term preservation

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

Q1. Why publish research data?

- Data sharing policy
- Visibility, credibility & usability
- Receive credit & track citations

Q2. Where to publish data

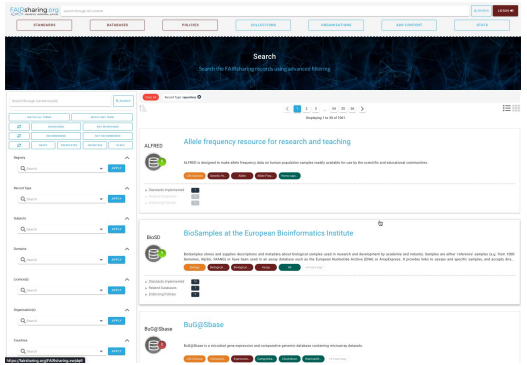
1. Data repositories
2. Data papers and data journals
3. Generalist repositories

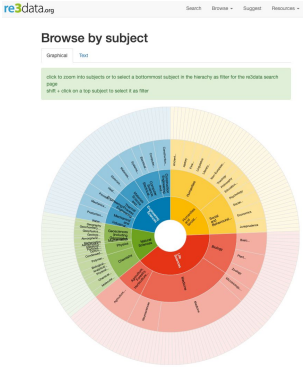
Data repositories

Data repositories are a centralized place to hold data, share data publicly, and organize data in a logical manner.

Where to find data repository

Funder specified?
Journal specified?
Domain specified?
Search registries - re3data, fair-sharing.org





Data papers and data journals

A Data Paper is a scholarly journal publication, which describes a dataset.

scientific data

How to find data journals

Walters, W. H. Data Journals: Incentivizing Data Access and Documentation within the Scholarly Communication System. 2020, *33* (1), 18. https://doi.org/10.1629/uksg.510.

https://www.gbif.org/data-papers

data

Journal	Publisher	Open Access (license)	APC estimate	Journal Impact Factor (2022)	Scopus CiteScore (2022)
Annals of Forest Science	Springer Nature	CC BY	EUR 1690	3.0	6.4
Arxius de Miscel·lània Zoològica	Nat Hist Museum of Barcelona	CC BY	0	-	1

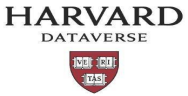
Generalist repositories

https://zenodo.org/records/7946938#.ZGe00nbMKUk

Curated
Data Processing Charges

20GB of free storage
Up to 5 GB files

Datasets can be shared privately between users
Personal accounts limit 10GB per dataset



Open-source web application
Free
Limit 2.5GB per file and 10GB per dataset

Project management tool that supports dataset upload
Free
5GB per file limit

Developed by CERN
Free
50GB limit per dataset

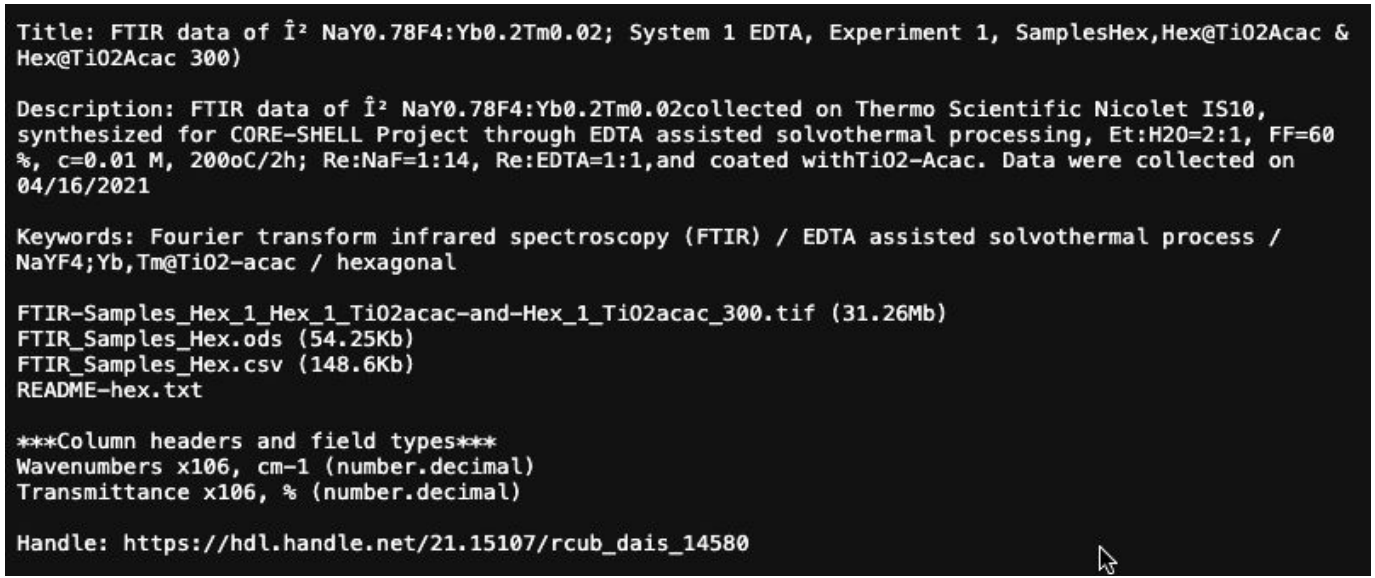
Q3. What to deposit

What to deposit? Everything needed to find, assess, understand & reuse data

DATA
Open file formats
Use relevant standards for interoperability

METADATA (data about data)
High 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



README files

A README is a plain text document that is stored alongside a data file.

Thank you