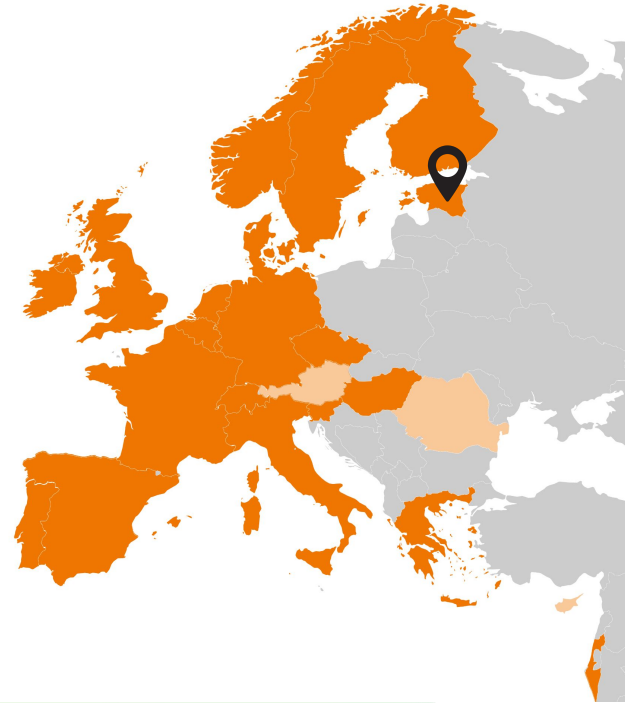# Open Data

(FAIR Data Management)

Diana Pilvar
Data management trainer

Who am I

Member countries of ELIXIR Europe

ELIXIR - European life-sciences infrastructure for biological information

ELIXIR Estonia training coordinator and data manager.

I give courses on Data Management, Licensing, Data management Plans, FAIR, Metadata, OpenRefine and Tableau

Member of Research Data Management Community

Participant in Data Stewards Interest Group

What is Open Data?

- **Open Science** - the movement to make scientific research and its dissemination accessible to all levels of society, amateur or professional

- **Open Data** is part of Open Science
  - Openly accessible, exploitable, editable and shareable by anyone for any purpose.
  - Licensed under an open license. CCo, CC-BY, CC BY-SA, Unlicense, Apache, MIT, GNU
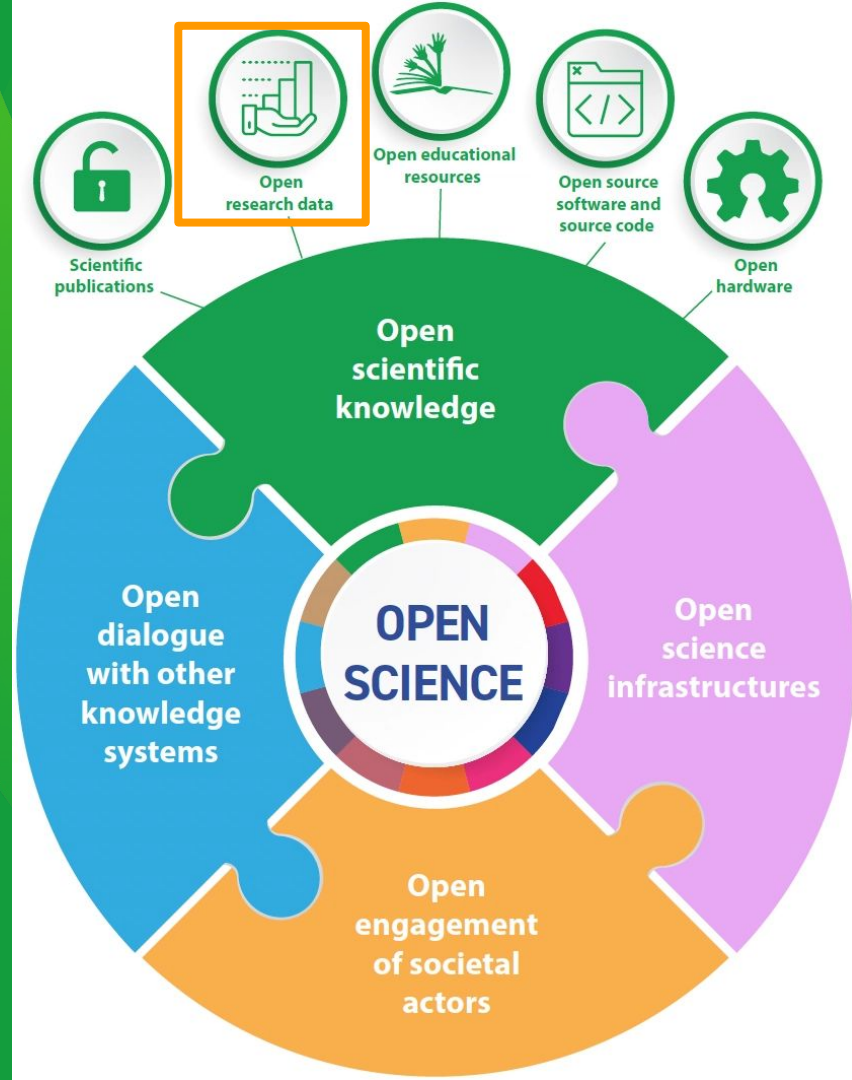
Why should I care about Open Data?

- Transparency
    - We have the right to know what is being done with government money
    - Analyse and visualise data, combine with other datasets
    - Improved efficiency and effectiveness
- Releasing social and commercial value
    - Improved or new products and services
    - Innovation
- Participation and engagement

https://okfn.org/en/library/why-open-data/
https://opendatahandbook.org/guide/en/why-open-data/

Open Data in practice

- [Open Data Handbook](https://okfn.org/en/library/how-to-open-data/)- legal, social and technical aspects of open data

- 3 key rules
  - Keep it simple
  - Engage early and engage often with potential (re)users
  - Address common fears and misunderstandings

- 4 steps
  - Choose your dataset(s)
  - Apply an open license
  - Make the data available
  - Make it discoverable

https://okfn.org/en/library/how-to-open-data/

Open Data in practice

- [Open Data Handbook](#)- legal, social and technical aspects of open data

- 3 key rules
  - Keep it simple
  - Engage early and engage ~~often~~ with potential (re)users
  - Address common fears and misun~~derstandings~~

- 4
  - ○
  - ○ Make the data available
  - ○ Make it discoverable

NOT ALWAYS POSSIBLE

https://okfn.org/en/library/how-to-open-data/

OPEN DATA ≠ FAIR

Not all data should be publicly available. Consider the following:

Regular Personal Data

- Name
- ID nr
- Address
- Phone nr
- Bank / Credit card nr
- Email address
- IP address
- Cookies
- Online identifiers
- Rare occupation

Special category Personal Data

- Biometric data
- Genetic data
- Health data, disability
- Race and ethnicity
- Political opinions
- Philosophical views
- Religious beliefs
- Trade union membership
- Sexual orientation
- Sex life

Sensitive data not listed in GDPR

- Confidential data (industry)
- Ecological data about rare and
- Endangered species

Open Seeds
Mentoring & Training Program

F indable   A ccessible   I nteroperable   R eusable

As open as possible,
as closed as necessary

https://www.go-fair.org/fair-principles/

FAIR in practice - Findable

The first step in (re)using data is to find them.

Metadata and data should be easy to find for both

humans and computers.

- Persistent Identifier (DOI, Handle, ORCID, ROR)

- Add enough metadata to the produced data to

  be findable - Keywords!

- (Meta)data is indexed in a searchable resource

FAIR in practice - Accessible

- Who can access the data and under which conditions?
  - Open access
  - Registered or authenticated
  - Controlled access
  - Access upon request
- How are the data backed up?
- How is the information stored? Storage = €£$
- Who is the owner of the data;
  - and can this put restrictions on the accessibility of the data while in collection phase or at the end of the project?

Metadata should be always findable, even if data is deleted!

FAIR in practice - Interoperable

It is recommended to keep one copy of the file in software independent file format.

- File formats (free and open format)
  - https://dans.knaw.nl/en/file-formats/
- Use domain specific vocabulary.
  - https://fairsharing.org/search?fairsharingRegistry=Standard&q=vocabulary
- Qualified references to other metadata
  - Don't just throw a list of references somewhere, write how they relate to one another

FAIR in practice - Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- README file (how data was created)
  - Template https://cornell.app.box.com/v/ReadmeTemplate
  - Good examples of READMEs from curated database https://dataverse.no/
- License
  - Creative Commons https://creativecommons.org/share-your-work/cclicenses/
  - https://choosealicense.com/

Open Seeds
Mentoring & Training Program

Data Management is becoming more complicated

- More data
- Complex technologies
- Data sharing
- Open Science

In a review of 1,082 retracted publications from the journal PubMed from 2013 to 2016, authors found that 32% of retractions were due to data management errors

Open Seeds
Mentoring & Training Program

Data Management has a life cycle

Data management is not just one-time thing. It needs continuous effort

Resources:
- Online guides: [The Research Data Management toolkit for Life Sciences](#)
- Online guides: [Data Management Expert Guide (DMEG)](#) (social sciences)
- Book: [Data Management in Large-Scale Education Research](#) - good for all fields

Making a plan negates a lot of risks

Data management Plan is a living document

completed throughout the project

If something changes during the project, it is written

into the DMP with explanation

How to write one?

- Look up resources from your local University (Library)
- My guide for general data management

https://doi.org/10.5281/zenodo.13801448

Summary

- Open Data and FAIR have similar goals but are not the same

- Open Data should also be FAIR!

- Sharing data can be complicated process

  - Writing data management plan helps

Open Seeds
Mentoring & Training Program