Data Management Planning

Speaker: Federico Bianchini
(Centre for Bioinformatics, University of Oslo)
Moderator: Espen Åberg (UiT)

elixir NORWAY

BioMedData

14 June 2022

FAIR Data Management in Life Sciences

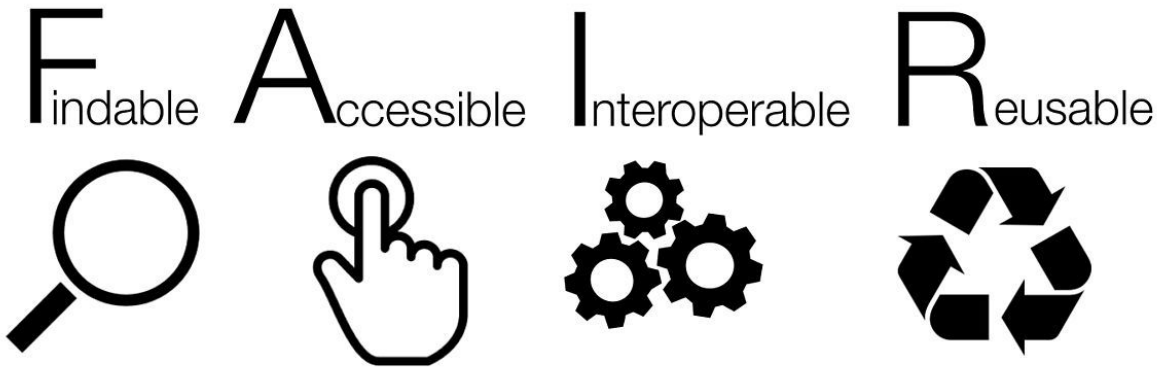
In this session, we will learn:

- What is a Data Management Plan (DMP)
- What information should be captured in a DMP
- When a DMP is to be submitted or updated
  - e.g. by requirement from funders
- How to use the Data Stewardship Wizard (DSW)

Session Take-Away

After completing this session, you will be:

- familiar with the concept of data management planning
- able to understand the crucial information to capture in a DMP
- able to understand how to use DSW to produce a DMP

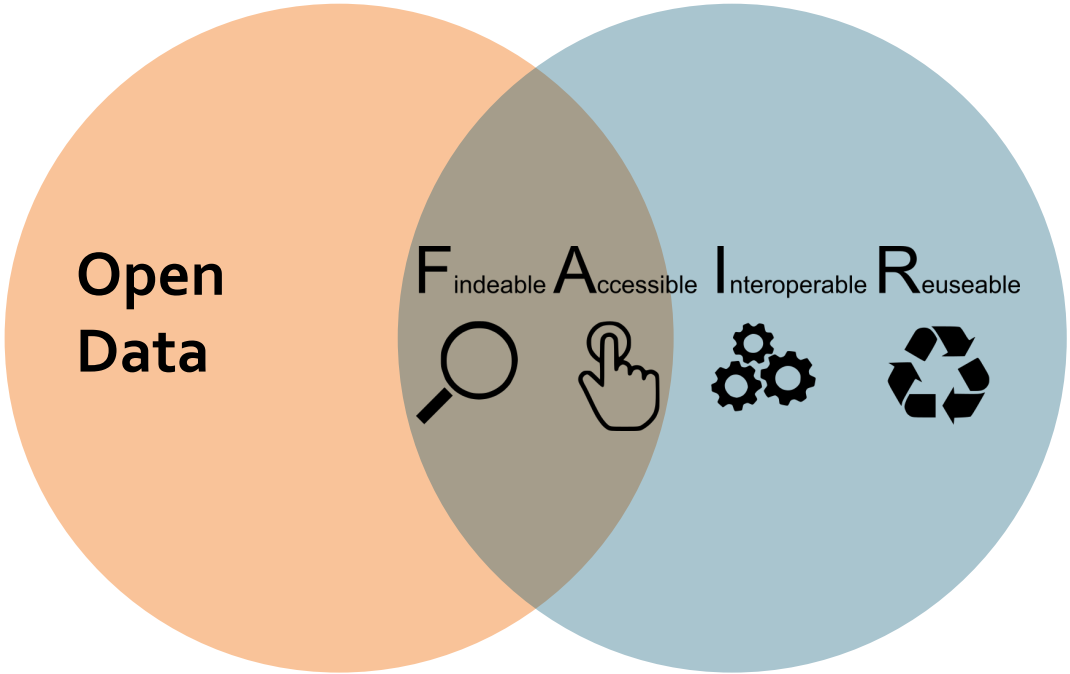


A quick introduction to the FAIR principles:

- Many concepts will become clearer as the workshop progresses,
- but it is useful to have a general idea at this early stage.

- General guiding principles
  - Not domain specific
  - Not refer to a specific technological implementation
  - Aimed at automatisation / machine-actionability

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



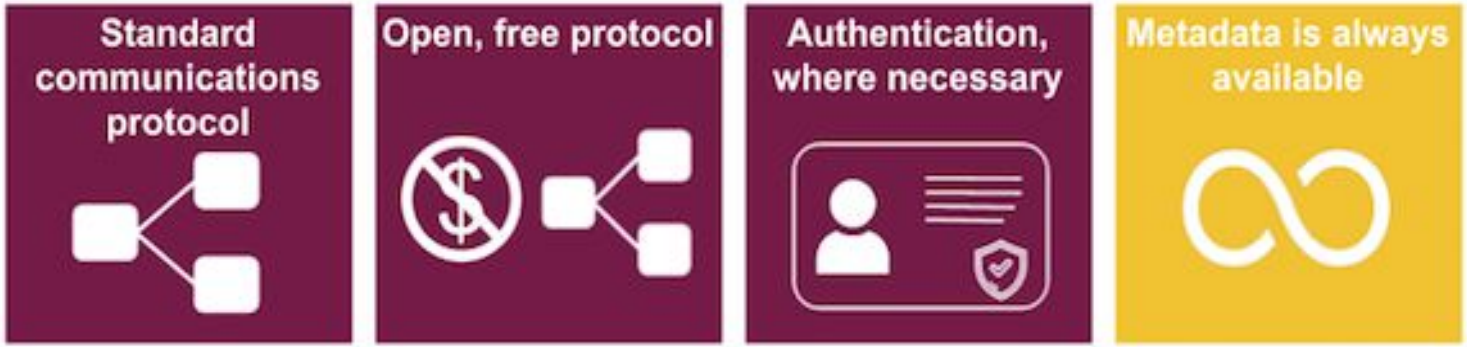
As open as possible and as close as necessary
Data that is not open should also be FAIR
A system for authentication and authorisation should be in place

FAIR: **Findable** Data

Metadata are crucial: the data are found through the metadata.

Deposition on a repository is one of the FAIR principles, such an (external) service has to be in place.

F**A**IR: **Accessible** Data



Technical implementations for accessing data, authentication and authorisation (FAIR not OPEN).

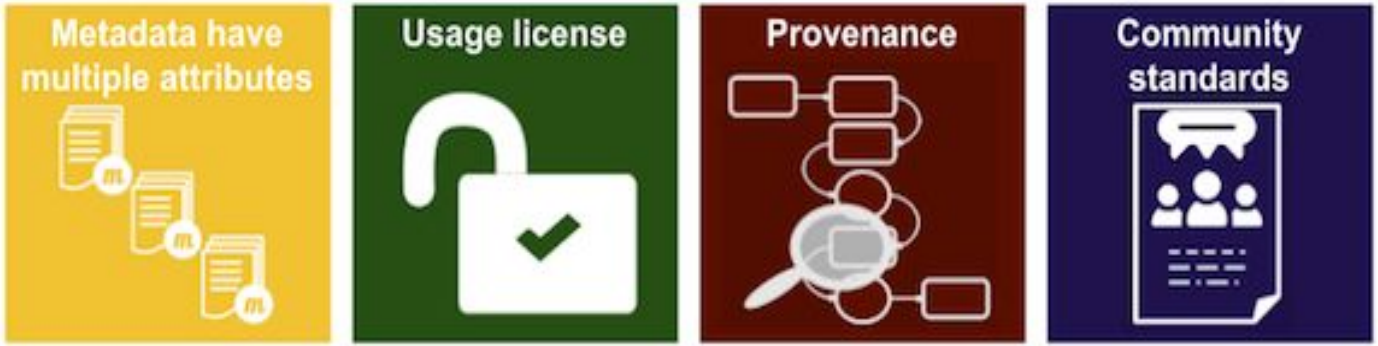
Metadata always available, also when the dataset no longer exist.

FAIR: **Interoperable** Data

Adoption of community-defined standards and definitions.

Consistent metadata annotation allows linking across datasets.

FAI**R**: **Reusable** Data



To ensure successful reusability, you need:

- Rich metadata
- License
- Data “history” (experimental technique, author, publication)
- Community standards (e.g. from domain repository of choice)

Back to Data Management Planning

- Why do I need a data management plan?
- Who will be looking at it?
- What information do I need to capture in it?
- Where can I get help?

Back to Data Management Planning

- Why do I need a data management plan?
- Who will be looking at it?
- What information do I need to capture in it?
- Where can I get help?

Advantages for the user (researcher, PI):

- DMP helps budgeting in advance necessary resources.
- DMP defines roles and responsibilities in the team.
- DMP helps identifying risks in data handling.
- DMP facilitates data sharing, reuse and preservation
- DMP monitors compliance to best practises.
- Funders require a DMP.

Government.no

National strategy on access to and sharing of research data

Plans/strategy | Date: 13/02/2018

Research data should be shared and reused more widely.

https://www.regjeringen.no/en/dokumenter/national-strategy-on-access-to-and-sharing-of-research-data/id2582412/

Government.no

National strategy on access to and sharing of research data

Plans/strategy | Date: 13/02/2018

Research data should be shared and reused more widely.

1. Research data must be as open as possible, as closed as necessary.
   1.1. Security, privacy protection, intellectual property rights apply
2. Research data should be managed and curated to take full advantage of their potential
   2.1. metadata
   2.2. license
3. Decisions concerning archiving and management of research data must be taken within the research community.
   3.1. Community standards, repositories

https://www.regjeringen.no/en/dokumenter/national-strategy-on-access-to-and-sharing-of-research-data/id2582412/

Government.no

National strategy on access to and sharing of research data

Plans/strategy | Date: 13/02/2018

Research data should be shared and reused more widely.

1. Research data must be as open as possible, as closed as necessary.
2. Research data should be managed and curated to take full advantage of their potential
3. Decisions concerning archiving and management of research data must be taken within the research community.

→ **Data Management Plan (DMP)**

https://www.regjeringen.no/en/dokumenter/national-strategy-on-access-to-and-sharing-of-research-data/id2582412/

Government.no

National strategy on access to and sharing of research data

Plans/strategy | Date: 13/02/2018

Research data should be shared and reused more widely.

Stipulations to draw up data management plans

“In its updated policy for open access to research data, the Research Council of Norway requires projects that generate data to provide data management plans at the contract phase. The individual research institutions will need to have good procedures and guidelines in place to meet these requirements.”

https://www.regjeringen.no/en/dokumenter/national-strategy-on-access-to-and-sharing-of-research-data/id2582412/

Government.no

National strategy on access to and sharing of research data

Plans/strategy | Date: 13/02/2018

Research data should be shared and reused more widely.

The Government expects:

The research institutions to develop procedures for (i) approving data management plans

The Government will:

Ask the Research Council of Norway to widely publicise its requirement regarding provision of a data management plan approved by the research institution at the time a contract is signed.

https://www.regjeringen.no/en/dokumenter/national-strategy-on-access-to-and-sharing-of-research-data/id2582412/

The Research Council of Norway

https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf

https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf

The Research Council of Norway

- Research data must be stored/archived in a safe and secure manner.
  - The data must be stored in secure archives, either in a trusted repository at the relevant institution or in national archives.
  - The research-performing institution is responsible for selecting relevant archiving solutions for the various research data in accordance with the FAIR Principles.
  - The projects should explicitly address how to manage research data that are considered to have long-term value.
  - The projects should explicitly address how research data that are not considered to have long-term value should be managed, or if relevant destroyed, after a certain period of time.

https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf

The Research Council of Norway

- Research data must be made accessible for reuse.
  - Research data must be made accessible to all relevant users, on equal terms, as long as there are no legal, ethical or security-related reasons to preclude this.
- Research data should be made accessible at an early stage.
  - The data used as the basis for scientific articles should be made accessible as soon as possible, and never later than at the time of publication.
  - Other data that may be of interest for other research should be made accessible within a reasonable amount of time, and never later than three years after the project has ended.

https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf

The Research Council of Norway

- Research data must be accompanied by standardised metadata.
  - The metadata must enable others to search for and use the data.
  - The metadata should follow international standards.
  - The metadata should describe the quality of the data.
- Research data must be provided with a license for access, reuse and redistribution.
  - The license should be internationally recognised.
  - The license should set as few restrictions as possible on the access, reuse and redistribution of the data.

https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf

The Research Council of Norway

- Research data should be made accessible at the lowest possible cost.
  - Metadata should be made accessible at no charge and be published in a way so that they may be harvested and used in research data searches.
  - Research data should preferably be made accessible at no charge.
  - The price of access to research data should never be higher than the actual costs of making them available.

https://www.forskningsradet.no/siteassets/publikasjoner/1254032622112.pdf

The Research Council of Norway

- The management of research data must be described in a data management plan.
  - The research-performing institution must approve that data management plans follow the institution's requirements and expectations. The plan should comply with the FAIR Principles.
  - As far as possible, data management plans should be public and made openly accessible by the research-performing institution to enable scientific groups to follow peer practice.

The Research Council of Norway

Alignment between RCN and Science Europe on RDM:

- Description & Re-use
- Documentation & Data quality
- Storage & Backups
- Legal & Ethical requirements
- Sharing & Preservation
- Adoption of the FAIR-principles

evaluation rubric for DMPs
not yet adopted by RCN

The Research Council of Norway

No DMP required

But you should consider:

- Ethical and Legal aspects
  - GDPR considerations, REK approval, IPR
- Storage and computing (budgeting)
- Budgeting for DM-personnel & time effort

Full DMP required and publicly available

- Documentation, formats, volume
- Quality assurance including Metadata standards
- Backup during project
- Sharing & Preservation
- Ethical and Legal aspects - see above
- Project responsibility
- Target repository for data deposition

DMP is constantly updated

- Identifiers to published datasets

Changes to the plan do not need to be reported to RCN

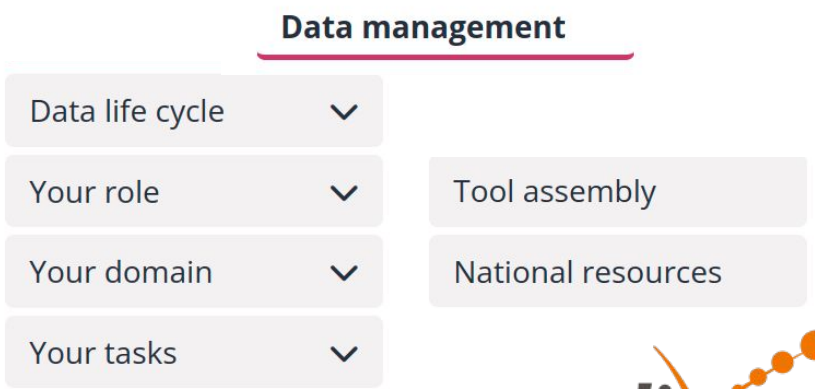
Final version of the DMP to be delivered with the final report of the project

- Data is deposited to trustworthy repository

Useful Resources

RDMkit

https://rdmkit.elixir-europe.org/



FAIRsharing.org
standards, databases, policies

https://fairsharing.org/

DSW
DATA STEWARDSHIP WIZARD

https://ds-wizard.org/
https://elixir-no.ds-wizard.org/

DSW

DATA STEWARDSHIP WIZARD

- A Tool for creating, customising and exporting DMPs in a collaborative way
- Provides guidance throughout the data life-cycle
- Intuitive for the researcher:
  - Clickable
  - No ambiguous open questions
- Questionnaire fully customisable (data stewards)
- FAIR metrics – easy to evaluate (funders)

https://ds-wizard.org/

Demo
General functions and features

elixir-no.ds-wizard.org

Learning Activity

- Data Planning Quiz (https://forms.gle/qykMAfo3FknUhAyLA)
- Group activity on day 3
  - review a DMP case study and discuss how to improve it

Thank you!

elixir-norway.org

@elixirnorway

contact@bioinfo.no