



Data Management Planning

Instructors: Federico Bianchini, Nazeefa Fatima
Centre for Bioinformatics, University of Oslo

1

DIVERSITY



3

FOSTER
TRUST AND
COLLABORATION



2

ACT ETHICALLY
AND WITH
INTEGRITY



4

RESPECT
AND SHOW
COURTESY TO
EACH OTHER





Agenda

Time	Activity
09:00	Welcome and Introduction
09:15	Data Management Planning
09:45	The Data Stewardship Wizard - Login Help
10:00	The Data Stewardship Wizard – Demo
10:30	BREAK
10:45	Exercise: DMP evaluation – breakout rooms
11:15	Exercise: DMP evaluation – discussion
11:45	Wrapping up: <u>Short quiz</u> + workshop evaluation

Introduction

ELIXIR and ELIXIR Norway

ELIXIR – what do we do

We build life science informatics **capacity** and **infrastructure** in Europe, connect and develop a **network of experts** and provide hundreds of high quality **services** and **resources** available to all



Databases



Training



Software tools



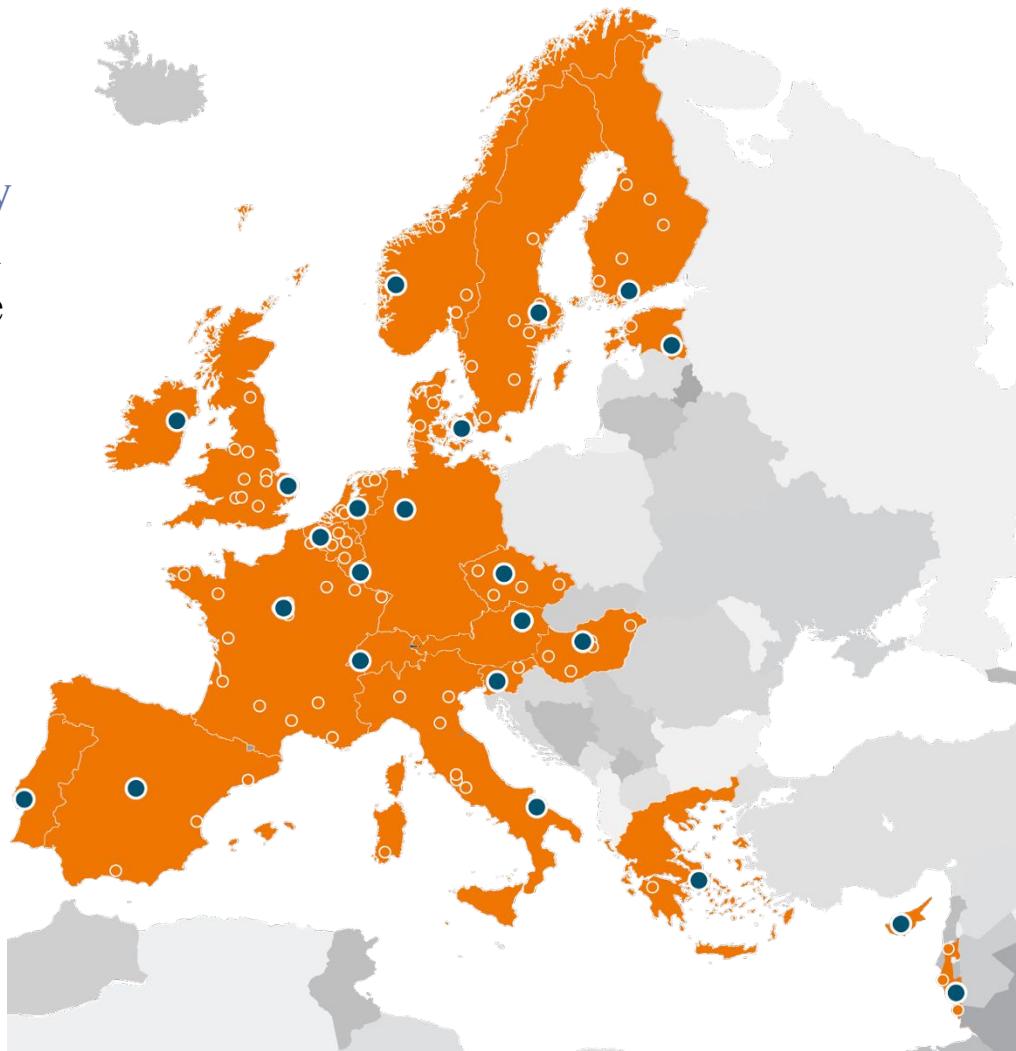
Data standards



Compute resources



Scientific & technical experts





Espen Åberg
Service Coordinator



Erik Hjerde
Node Leader



UiT node

Erin Calhoun
Training Coordinator



Pål Sætrom
Node leader



NTNU node

Sushma Grellscheid
Head of Node



Ingeborg Winge
Node coordinator



Korbinian Bösl
Data Management coordinator

Kjell Petersen
Technical coordinator

Eivind Hovig
Node leader



UiO node



Simen Rød Sandve
Node leader

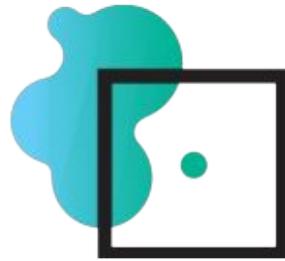


Our services to Norwegian users

- Support - through our nationally coordinated Helpdesk:
 - Analysis
 - Scripting
 - Workflow development
 - Data management
- Training:
 - Using our tools and workflows
 - Data management
- e-infrastructure: analysis, sharing and storage of data
 - Non-sensitive data:
 - NeLS - the Norwegian e-infrastructure for Life Sciences
 - Sensitive data:
 - TSD
 - Containerised workflows on SAFE and HUNT Cloud

support@elixir.no





Centre for Digital Life Norway (DLN 2.0)

facilitates transdisciplinary research, innovation, and education in
Life Science



Transdisciplinary collaborations



Education and career development



Responsible research and innovation



Innovation and commercialisation



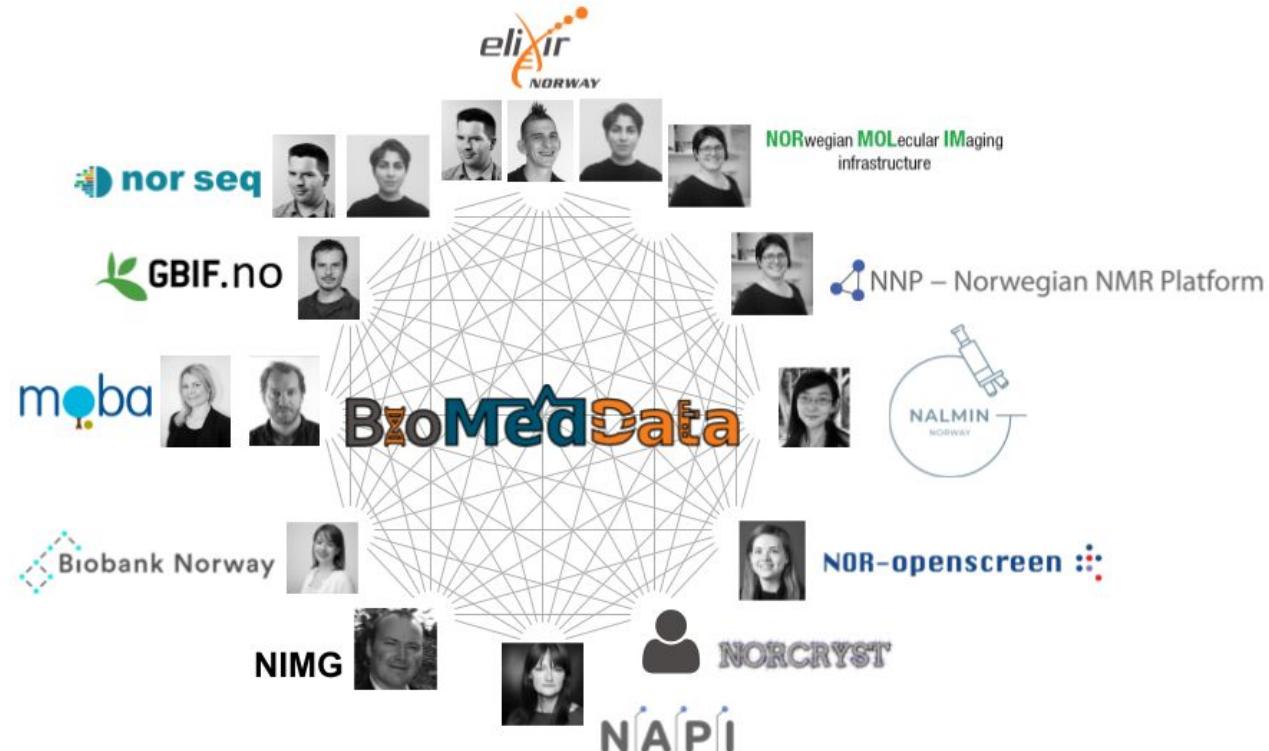
Data management

www.digitallifenorway.org



BioMedData

A national research infrastructure to promote FAIR management of life science data along the data life-cycle



Data Life Cycle (image) from https://rdmkit.elixir-europe.org/data_life_cycle

Introduction

Data Management plans and funders' requirements

Research data...

-  ...must be stored/archived in a safe and secure manner.
-  ...must be made accessible for reuse.
-  ...should be made accessible at an early stage [latest at publication]
-  ...must be accompanied by standardised metadata.
-  ...must be provided with a license for access, reuse and redistribution.
-  ...should preferably be made accessible at no charge.
-  ... must be described in a data management plan.

Research Council of Norway's Policy for Open Access to Research Data

Originally 2014, updated Dec/2017



New:

“... requiring that R&D-performing institutions or companies should assess whether projects receiving funding from the Research Council must develop a **data management plan**.”

and

“The **FAIR Guiding Principles** for scientific data management and stewardship are included as a main principle in the Research Council’s policy”

Assessment of open science in grant applications

Published 17 Nov 2022 |

The referees are to assess open science practice through two subsections of the criterion 'Impact':

- [potential impact of the proposed research](#)
- [communication and exploitation](#)



Core requirements to data management plan



6 Core Requirements:

1. Data description & Collection/re-use
2. Documentation & Data quality
3. Storage & Backup
4. Legal & Ethical requirements
5. Data sharing & Preservation
6. Responsibilities & Resources

The Research Council recommends using a service for DMPs that allows the project to generate a machine-actionable DMP, for example according to the [RDA Common Standard](#).

Until further notice, the project must upload the data management plan in the format of pdf, .doc(x) or similar. NFR is working on developing systems to facilitate machine-actionable DMP.

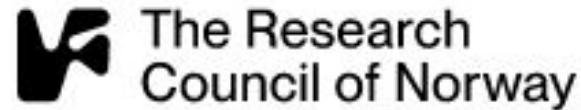
Assigning your data management plan a persistent identifier is strongly recommended.

Tools and service providers to create a good data management plan

- Data Stewardship Wizard (DSW), ELIXIR Norway ↗
- Norwegian Centre for Research Data (NSD) ↗
- Digital Curation Centre ↗
- easyDMP ↗
- Argos (openaire.eu) ↗

Data management resources

- Social sciences and humanities ↗
- Life sciences ↗



Datahåndteringsplan (DMP)

En datahåndteringsplan er et dokument som beskriver hvordan data i et forskningsprosjekt skal håndteres, helt fra oppstart til etter avslutning av prosjektet.



<https://www.forskningsradet.no/en/research-policy-strategy/open-science/research-data/>

F

indable

A

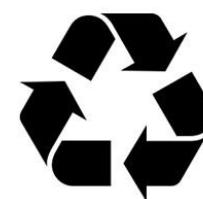
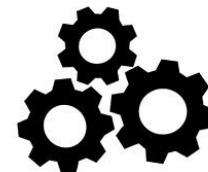
ccessible

I

nteroperable

R

eusable

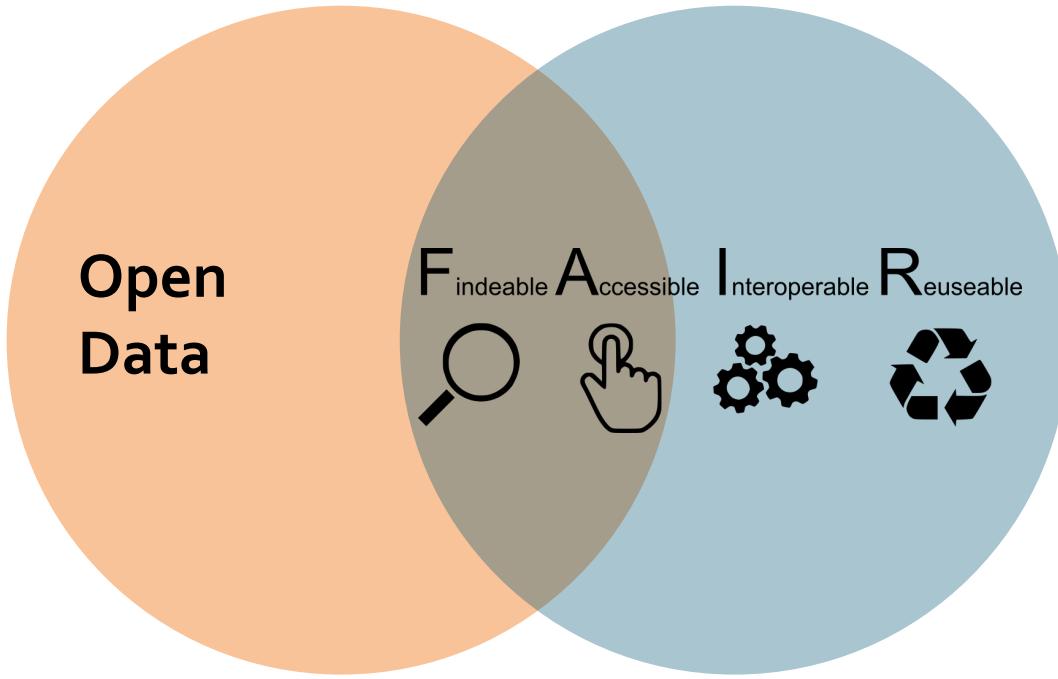


A quick introduction to the FAIR principles:

- 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 automation / machine-actionability

Introduction

The FAIR data 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

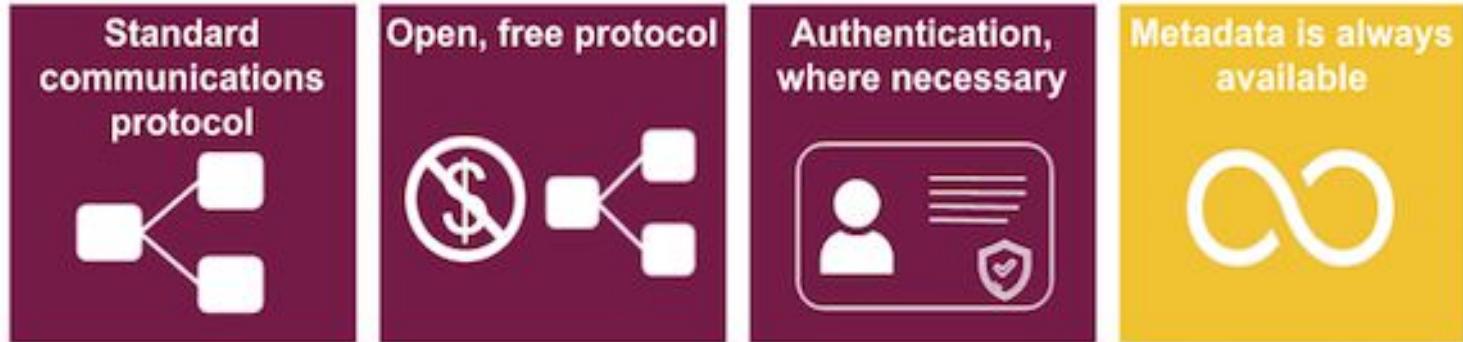
FAIR: Findable Data



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.

FAIR: Accessible Data



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

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

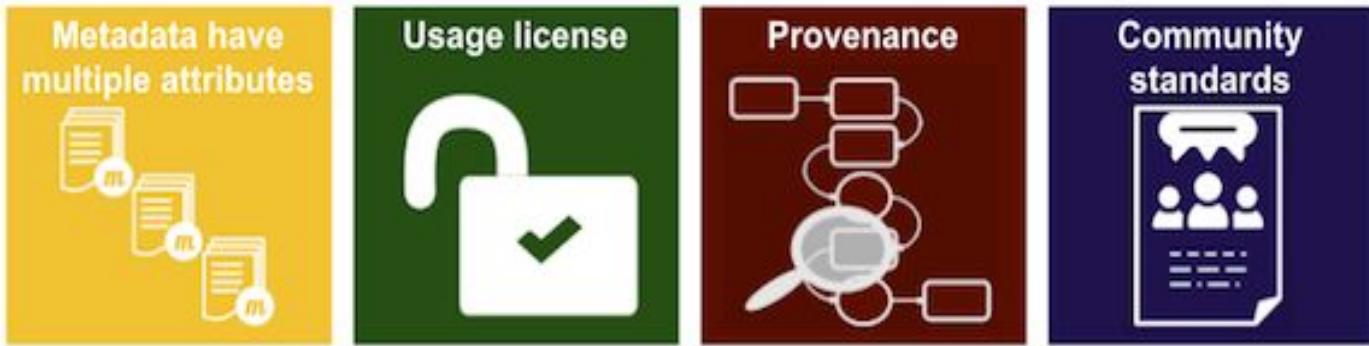
FAIR: Interoperable Data



Adoption of community-defined standards and definitions.

Consistent metadata annotation allows linking across datasets.

FAIR: 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)

Introduction

The Machine-Actionable DMP standard

The machine-actionable DMP standard (maDMP)

Machine-actionable:

- Information structured in a consistent way so that machines can be programmed against the structure.
- Formats such as JSON, XML, RDF

Standard:

- Contains “minimum information”, but can also be customised.
- All tools supporting this format become interoperable

Research Data Alliance (RDA) working group (WG)

WG DMP Common Standards WG

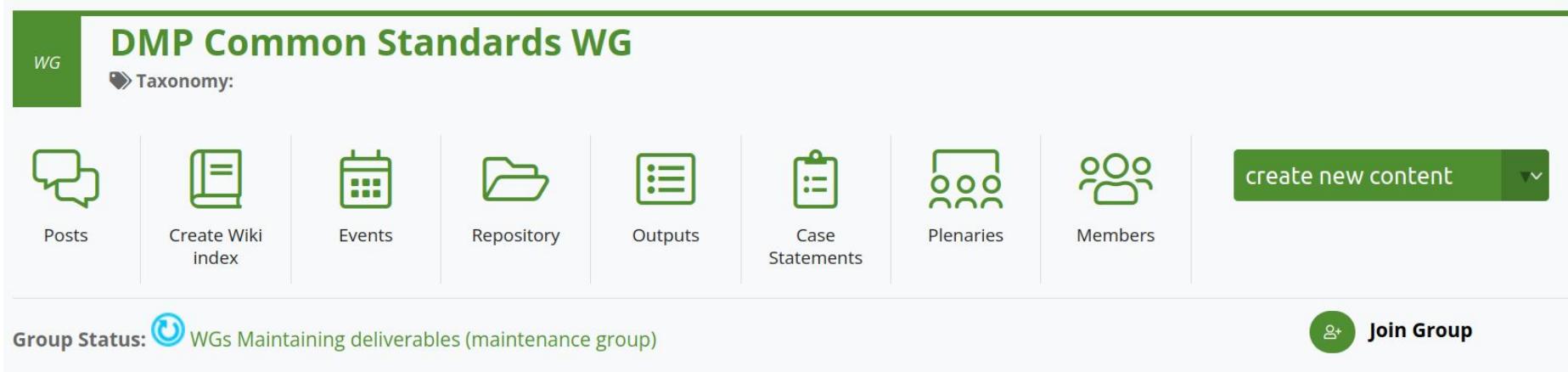
_taxonomy:

create new content ▾

Posts Create Wiki index Events Repository Outputs Case Statements Plenaries Members

Group Status:  WGs Maintaining deliverables (maintenance group)

Join Group



<https://www.rd-alliance.org/groups/dmp-common-standards-wg>

Who recommends/requires a maDMP 1



What a data management plan should include

The Research Council recommends using a service for data management plans that allows the project to generate a machine-actionable data management plan, for example according to the [RDA Common Standard](#). Until further notice, the project must upload the data management plan in the format of pdf, .doc(x) or similar. We are working on developing our systems to facilitate machine-actionable data management plans. We also recommend assigning your data management plan a persistent identifier, such as a DOI. Several services for data management plans offer this.

<https://www.forskningsradet.no/en/research-policy-strategy/open-science/research-data/>

Who recommends/requires a maDMP 2

December 12, 2022

Report

Open Access

UB-BOTT-samarbeid om datahåndteringsplaner: kartlegging og anbefalinger

 Gabrielsen, Ane;  Kvale, Live;  Ostrop, Jenny;  Sarre, Aili

Project member(s)

 Heggland, Ingrid;  Bertheussen, Lene;  Flatby, Helene;  Bochynska, Agata;  Conzett, Philipp;  Longva, Leif

<https://doi.org/10.5281/zenodo.7428542>

Who recommends/requires a maDMP 2

December 12, 2022

Report

Open Access

UB-BOTT-samarbeid om
datahåndteringsplaner: kartlegging og
anbefalinger

*Prosjektgruppen slutter seg også til anbefalingen fra Forskningsrådet om å
bruke verktøy som kan generere maskinhåndterbare datahåndteringsplaner
(f.eks. etter RDA Common Standard) og publisering av DMPer.*

<https://doi.org/10.5281/zenodo.7428542>

JSON schema

A JSON schema is a specification of the structure of a JSON file.
It allows e.g. to restrict the types that can enter the various fields of the JSON file.
Schemas can thus be used to validate the content of a file.

Schema for the maDMP is available at:

<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/examples/JSON/JSON-schema/1.1/maDMP-schema-1.1.json>

JSON schema of maDMP

Properties in 'contact'

Name	Description	Data Type	Cardinality	Example Value
contact_id	Identifier for a contact person	Nested Data Structure	1	
mbox	E-mail address	String	1	cc@example.com
name	Name of the contact person	String	1	Charlie Chaplin

Properties in 'contact_id'

Name	Description	Data Type	Cardinality	Example Value
identifier		String	1	
type	Identifier type Allowed Values: <ul style="list-style-type: none">• orcid• isni• openid• other	Term from Controlled Vocabulary	1	orcid

JSON schema of maDMP

Properties in 'contact'

Name	Description	Data Type	Cardinality	Example Value
contact_id	Identifier for a contact person	Nested Data Structure	1	
mbox	E-mail address	String	1	cc@example.com
name	Name of the contact person	String	1	Charlie Chaplin

- This specific format is already implemented in DSW.
- Filling out a DMP produces a document compliant with maDMP
- This will be demonstrated in the next session.

Introduction

Knowledge bases and signposting



Research Data Management Kit

A user-oriented guide to the FAIR RDM practices in life sciences



increase
self-sufficiency



support researchers
to know and utilise
RDM services



build capacity and
skills in every
research institute



pool the expertise
of the community
for the community

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



RDMkit: Website Walkthrough



Data management

About

Contribute

GitHub

The Research Data Management toolkit for Life Sciences

Best practices and guidelines to help you make your data FAIR (Findable, Accessible, Interoperable and Reusable)

What can we help you find?

Search RDMkit

Browse all topics by

- Data life cycle**

Start here to get an overview of research data management based on stages in the data life cycle.
- Your role**

Identify your role in research data management, find data management resources relevant for you, and information to help you progress in your career path.
- Your domain**

Learn about data management tasks that affect your domain or research community, and the solutions adopted to address them.
- Your tasks**

Find guidelines and solutions for tackling common data management tasks.
- Tool assembly**

Find concrete combinations of tools and resources assembled into an ecosystem for research data management.
- National resources**

Find pointers to country specific information resources and national research data management practices.
- All tools and resources**

Browse the RDMkit's catalogue of tools and resources for research data management.
- All training resources**

Browse all training resources mentioned in RDMkit pages.





RDMkit: National resources in Norway



Your tasks ▼

Tool assembly ▼

National resources ^

Belgium

Germany

Estonia

Spain

Finland

France

Italy

Norway

Portugal

Sweden

United Kingdom

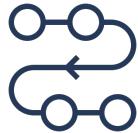
- Funder policies on research data
- Institutional policies on research data
- Support services
- Domain-specific infrastructures/resources
- Ethical committees and general authorities
- Relevant ethical guidelines
- Laws and regulations relevant to life sciences research data
- Tools and resources
- Related pages
- More information

Introduction

This page provides an overview of the data management resources in Norway. The target audience is the Norwegian scientific community in the life sciences and collaborators. The [Data Stewardship Wizard instance](#) from [ELIXIR-Norway](#) provides an interactive way to navigate this recommendations and resources.

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

RDMkit: A sustainable, open, expanding community effort



Contribution and editorial processes



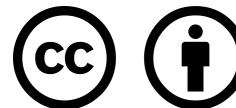
Contentathons and focus groups



Low barrier for contributions



Simple, sustainable platform



Guidelines, process documents and data are made available under a [CC-BY](#) license.



Software is made available under an [MIT](#) license.





Cheat Sheets

Biodiversity

Protein Crystallography

High-Throughput Screening

Light Microscopy

Marine Metagenomics

Pre-clinical Imaging

Proteomics

Sequencing

Cheat sheet

High-Throughput Sequencing



- Related pages
- More information

Description

We provide here a collection of resources, tools, and standards relevant for short-read and long-read sequencing data.

Type of data/experiments/methods

Commonly used raw file formats for sequencing data

- FASTQ Sequence and Sequence Quality Format
- FASTA
- FASTQ Original Read Archive (ORA)
- Illumina Binary Base Call
- PacBio legacy basecall File Format (bas.h5/bax.h5)
- PacBio Alignment File Format (cmp.h5)
- PODS File Format for Oxford Nanopore Technology (ONT) data
- Fast5 for ONT data

Alignment Formats



- About
- Contribute
- Code of Conduct



Funded by
The Research
Council of Norway

Deliverable 4.2 of BioMedData funded under project 295932 by the Research Council of Norway



Norwegian Life Science RDM LookUp by ELIXIR Norway is licensed under a Creative Commons Attribution 4.0 International License, except where otherwise noted.

Built with 

More information

Links to other ELIXIR resources



Support for DMP on: Non-quantitative next generation sequencing of non-human data



Support for DMP on: Non-quantitative next generation sequencing of human data

<https://elixir.no/rdm-lookup>

A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.

Guides consumers to discover, select and use these resources with confidence.

Helps producers to make their resources more visible, more widely adopted and cited.

Provides humans and tools with access to trustworthy content to enable data management tasks.

Example top page for the European Nucleotide Archive:

Recommendation

GENERAL INFORMATION



ENA
European Nucleotide Archive

European Nucleotide Archive (ENA)

doi 10.25504/FAIRsharing.dj8nt8



DSW FAIRsharing integration

✓ VII.1.a.5.b.1.a.1 What repository will this data be stored in?

[Horizon 2020 DMP](#)[Horizon Europe DMP](#)[Science Europe DMP](#) a. A domain-specific repository[Findability](#)

✓ VII.1.a.5.b.1.a.1.a.1 What repository?

[Horizon 2020 DMP](#)[Horizon Europe DMP](#)[Science Europe DMP](#) Desirable: Before Finishing the Project

enal

European Nucleotide Archive

BioSamples at the European Bioinformatics Institute

GenBank Nucleotide Sequence Database

RegulonDB

The European Genome-phenome Archive


FAIRsharing.org
standards, databases, policies doi [10.25504/FAIRsharing.dj8nt8](https://doi.org/10.25504/FAIRsharing.dj8nt8) 

FAIRCOOKBOOK

The FAIR Cookbook for FAIR doers

An online, **open** and **live** resource for the Life Sciences with recipes that help you to make and keep data Findable, Accessible, Interoperable and Reusable; in one word FAIR.

<https://faircookbook.elixir-europe.org/>

F Findability

EXEMPLAR RECIPES

- Unique, persistent identifiers
- Search engine optimization

[LEARN MORE](#)

A Accessibility

EXEMPLAR RECIPES

- Transferring data with SFTP
- Downloading data with Aspera

[LEARN MORE](#)

I Interoperability

EXEMPLAR RECIPES

- Selecting terminologies and ontologies
- Creating a metadata profile

[LEARN MORE](#)

R Reusability

EXEMPLAR RECIPES

- Data licenses
- Declaring data's permitted uses

[LEARN MORE](#)

Infrastructure



[LEARN MORE](#)

Assessments



[LEARN MORE](#)

Applied Examples



[LEARN MORE](#)

Maturity model



[LEARN MORE](#)

Links to other ELIXIR resources



Step-by-step process for: Licensing Software



Step-by-step process for: Making Computational Workflows FAIR



Step-by-step process for: Depositing to generic repositories - Zenodo use case



Step-by-step process for: Registering datasets with Wikidata



Support for DMP on: Data storage systems and file naming conventions



Support for DMP on: How long will this data set be kept?



Support for DMP on: Will you be storing data in an "object store" or a "document store" system?

RDMkit: Integrations

Your tasks

Data storage



<https://rdmkit.elixir-europe.org/storage>

- Integration with DSW for the relevant portions of a standard DMP
- Integration with FAIR Cookbook for associated technical solutions
- Integration with registries such as FAIRsharing for standards and databases and TeSS for training

DATAVERSE

Open source research data repository software.



Different instances available

Plant Phenomics

Plant sciences

Machine actionability



Standards/Databases



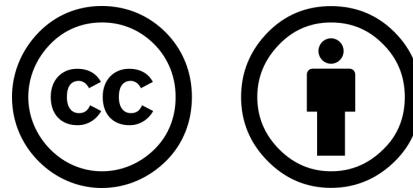
Training

Thank you!

 Elixir-norway.org

 @elixirnorway

 support@elixir.no



Except where otherwise noted, this work is licensed under a
Creative Commons Attribution 4.0 International License

<https://creativecommons.org/licenses/by/4.0/>