Create and a Handle Data Management Plan

Federico Bianchini & Nazeefa Fatima

Centre for Bioinformatics, University of Oslo

Wednesday 13 December 2023

elixir.no

https://elixir-europe.org/events/code-of-conduct

Time (CET)	Activity
9:10-10:00	Presentation
10:00-10:30	The Data Stewardship Wizard – Demo
10:30-10:45	Coffee break
10:45–11:45	Hands-on session
11:45-12:00	Evaluation and wrapping-up

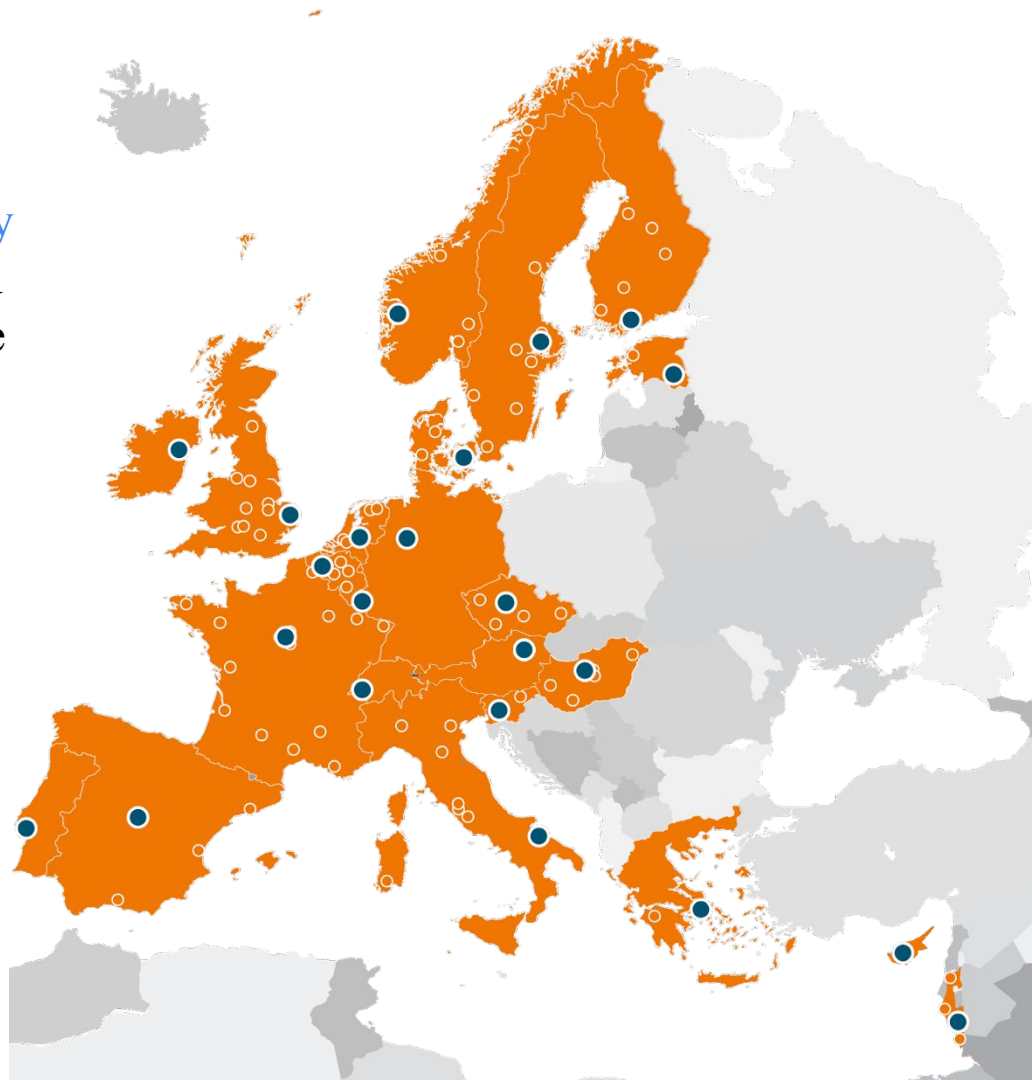
Introduction

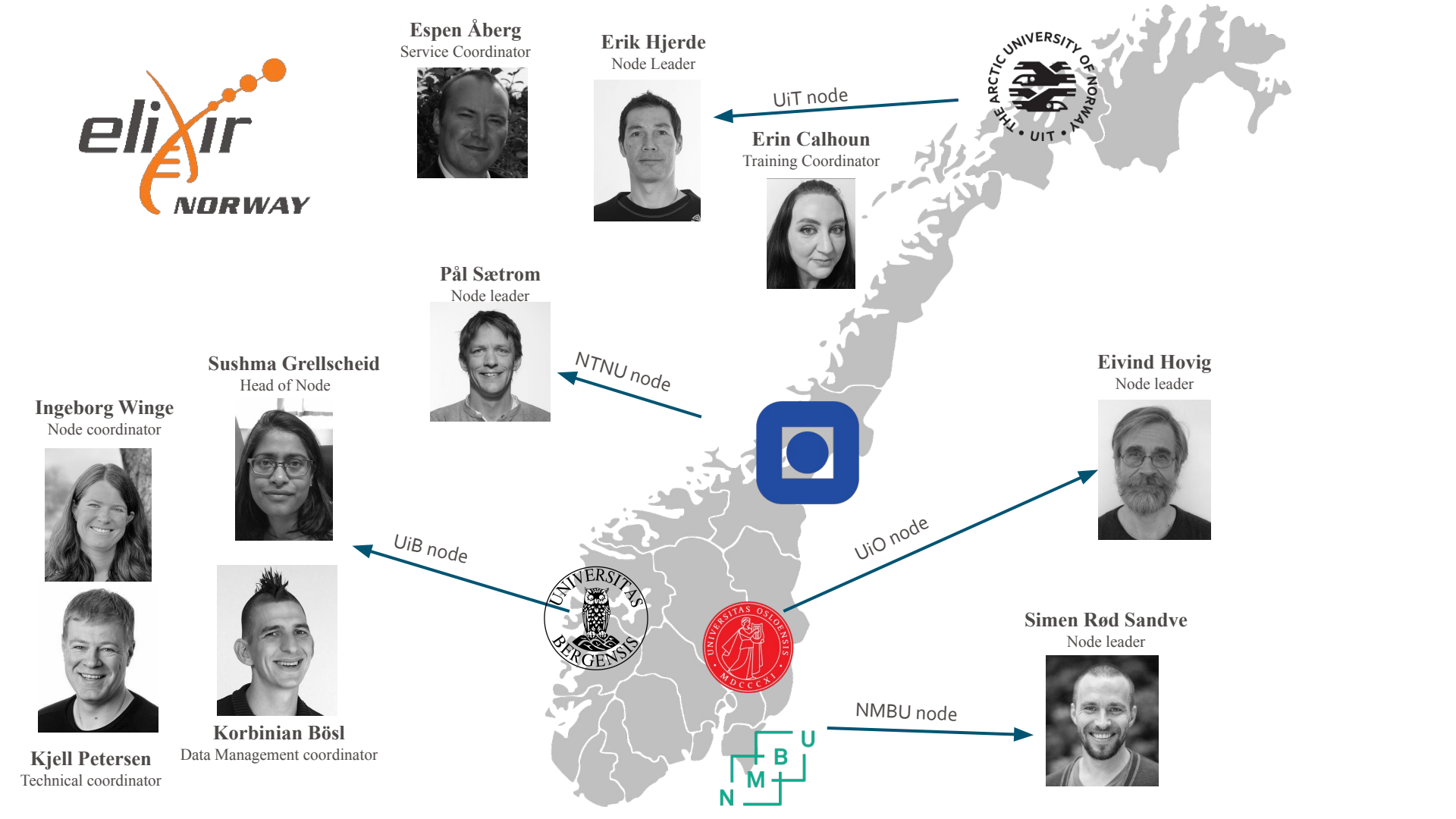
ELIXIR and ELIXIR Norway

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

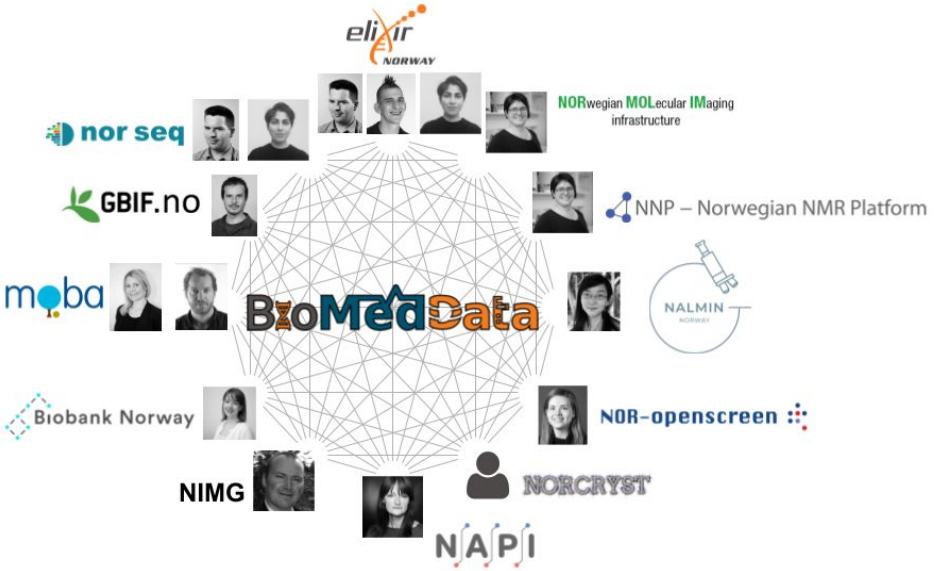



elixir NORWAY
Espen Åberg
Service Coordinator
Erik Hjerde
Node Leader
UiT node
THE ARCTIC UNIVERSITY OF NORWAY
UIT
Erin Calhoun
Training Coordinator
Pål Sætrom
Node leader
NTNU node
Sushma Grellscheid
Head of Node
Ingeborg Winge
Node coordinator
Eivind Hovig
Node leader
UiB node
UiO node
UNIVERSITAS BERGENSIS
UNIVERSITAS OSLOENSIS MDCCCXI
Simen Rød Sandve
Node leader
NMBU node
N M B U
Kjell Petersen
Technical coordinator
Korbinian Bösl
Data Management coordinator

Data Management

BioMedData

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



Joint Strategic Statement of Norwegian Life Science Research Infrastructures on Research Data Management

Image from https://rdmkit.elixir-europe.org/data_life_cycle

elixir

Introduction

Data Management plans and funders’ requirements

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"

https://www.forskningsradet.no/contentassets/e4cd6d2c23cf49d4989bb10c5eea087a/the-research-council-of-norways-policy-for-open-access-to-research-data.pdf

Forskningsrådet

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.

https://www.forskningsradet.no/contentassets/e4cd6d2c23cf49d4989bb10c5eea087a/the-research-council-of-norways-policy-for-open-access-to-research-data.pdf

The Research Council will work to:

- Develop guidelines for DMPs in line with international practice
- Ensure that Norwegian research data are in compliance with international standards and protocols for data and metadata.
- Make it possible to link citations and data reuse to researchers and research projects.

- Ensure that national infrastructures comply with international standards and certifications that make it possible to share data across national borders.
- Stipulate that data infrastructures cooperate with relevant national and international stakeholders and are accessible to the users
- Require all infrastructure projects funded by the Research Council to draw up a robust funding plan to ensure future operations and maintenance and the long-term storage, processing and accessibility of research output.

https://www.forskningsradet.no/siteassets/forskningspolitisk-radgivning/apen-forskning/nfr-policy-open-science-eng.pdf

Assessment of open science in grant applications

From 2023, open research will be incorporated into the assessment criteria for Researcher Projects and Collaborative and Knowledge-building Projects on open science practice.

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

- potential impact of the proposed research
- communication and exploitation

It is important that you make good plans for how you intend to reuse the research results that emerge in the project. In addition, it is important that research results are easy to verify, for example by making research data available.

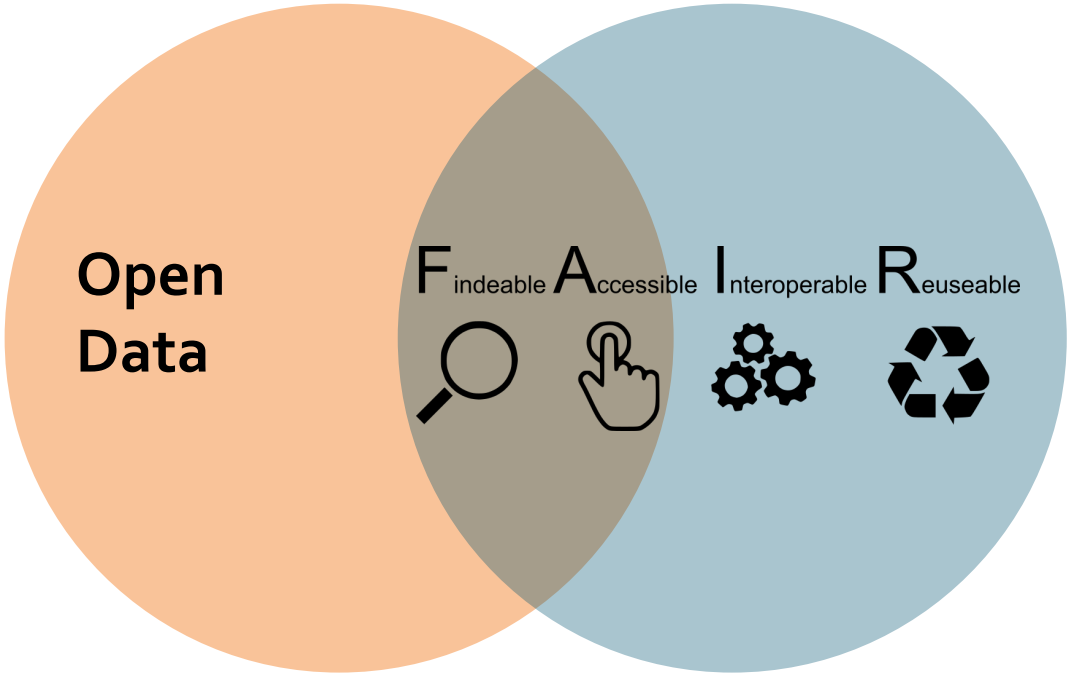
Describe to what extent and how you will adopt early and open sharing. For example, you can mention what type of early and open sharing is appropriate for your discipline and project, such as preprints or pre-registration and/or registration reports, and which publishing platforms you plan to use.

Open Science

https://www.forskningsradet.no/en/processing-grant-applications/processing-applications/assessment-open-science

Introduction

The FAIR data principles



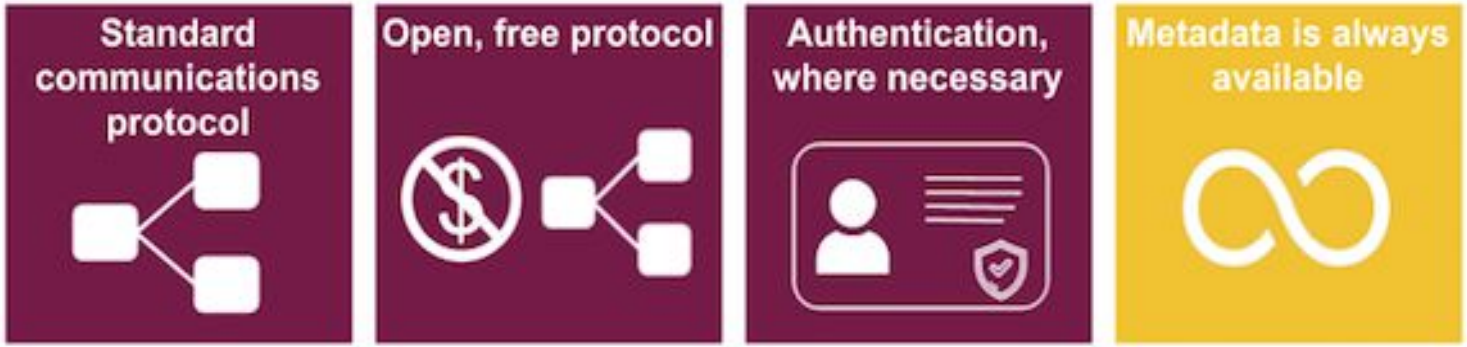
As open as possible and as close as necessary
Data that is not open should also be FAIR
A system for authentication and authorisation should be in place

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.

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



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

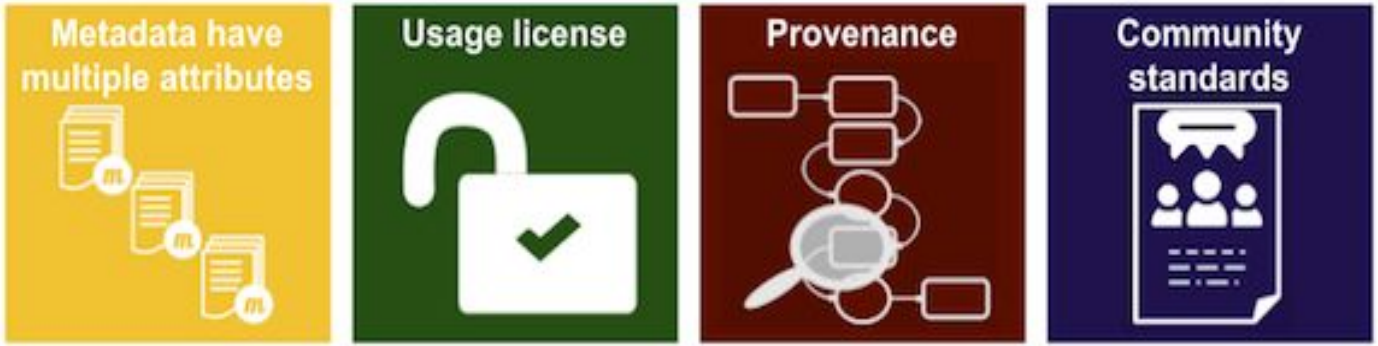
Metadata always available, also when the dataset no longer exist.

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)

Introduction

The Machine-Actionable DMP standard

The machine-actionable DMP standard (maDMP)

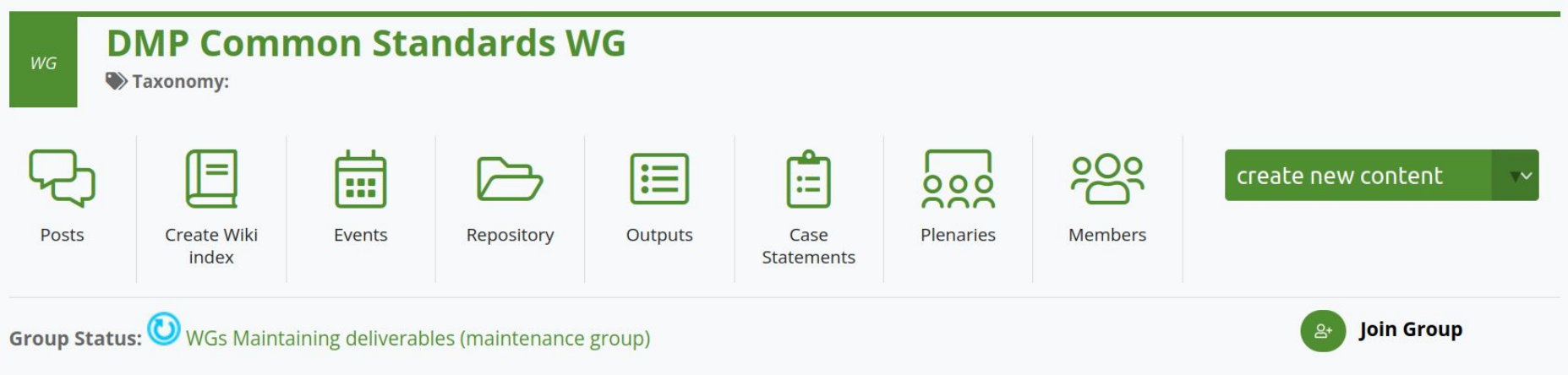
Machine-actionable:

- 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)

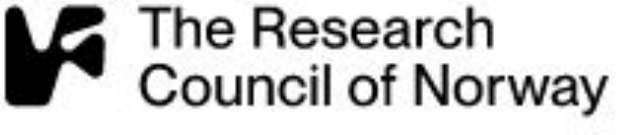


elixir CZECH REPUBLIC

DSW DATA STEWARDSHIP WIZARD

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: • 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.

DSW
DATA STEWARDSHIP WIZARD

Introduction

The Data Stewardship Wizard

The Data Stewardship Wizard

https://ds-wizard.org/

DSW
DATA STEWARDSHIP WIZARD

elixir NETHERLANDS

elixir CZECH REPUBLIC

The ELIXIR Norway DSW Story

Adoption in 2019

Norwegian funders require that research groups submit data management plans (DMPs) upon signing the contract for their research projects. Generating a DMP has been regarded by many researchers as a mere administrative burden, rather than a tool to revise their habits and support their projects.

Some developments funded through:

CENTRE FOR DIGITAL LIFE NORWAY

The Research Council of Norway

DSW
DATA STEWARDSHIP WIZARD

Data Stewardship Wizard

elixir-no.ds-wizard.org

Questionnaire → **Template** → **DMP in various formats**

pre-filled suggestions based on domain

.docx, .tex, .html, .json

Machine-actionable DMP standard, recommended by NFR

Adapted for Norwegian users

Login

Feide

LS LOGIN

Full compliance:

SCIENCE EUROPE

European Commission

https://github.com/ds-wizard

DSW jargon

Knowledge Models are “templates” determining the structure of Questionnaires. When creating a new Questionnaire, you select the Knowledge Model the most suitable for your research.

A **document template** describes how the replies from a questionnaire are composed into a document. This allows to produce various kinds of documents from a single questionnaire by using different templates.

KMs are typically customised by data stewards. While document templates can be customised, they usually reflect more strict rules (e.g. Science Europe requirements, or the maDMP standard)

Introduction

Knowledge bases and signposting

RDMkit

Research **D**ata **M**anagement **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

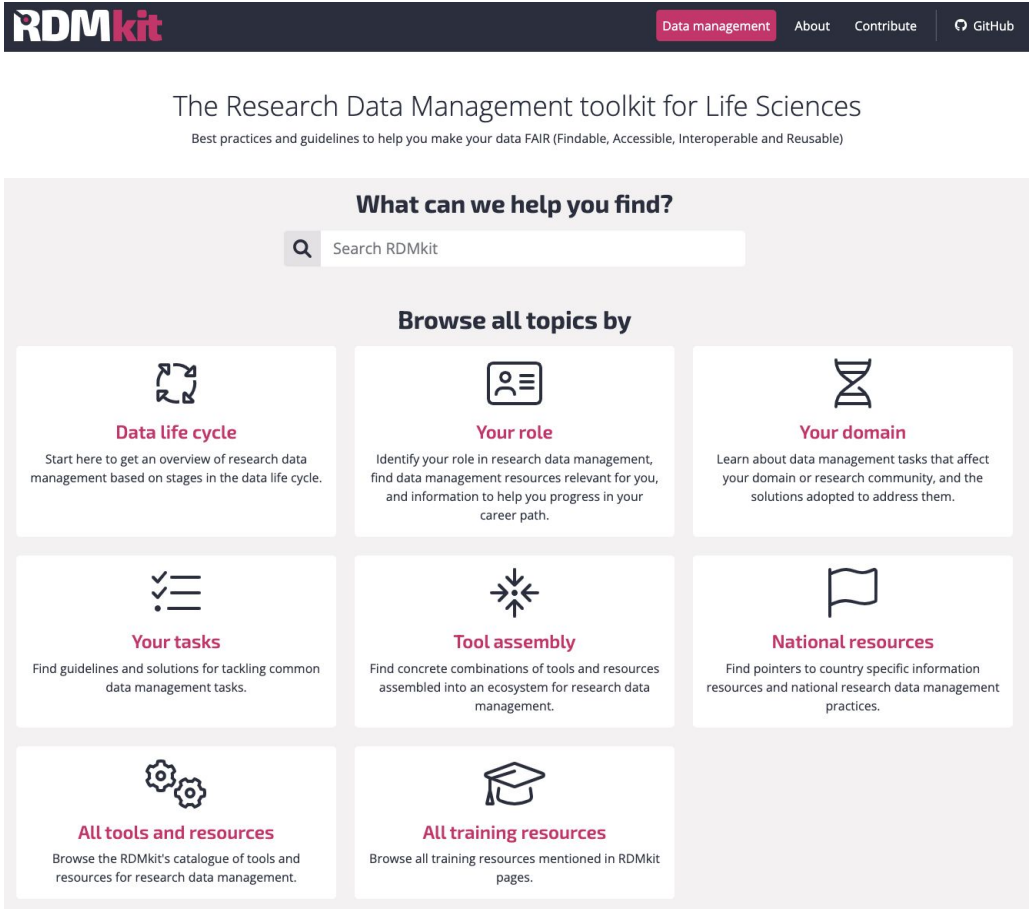
pool the expertise of the community for the community

Built by life scientists and data stewards

A sustainable, open, ongoing community effort

https://rdmkit.elixir-europe.org

RDMkit: Website Walkthrough



RDMkit: National resources in Norway

- Your tasks
- Tool assembly
- National resources
  - Belgium
  - Germany
  - Estonia
  - Spain
  - Finland
  - France
  - Italy
  - Norway
  - Portugal
  - Sweden
  - United Kingdom

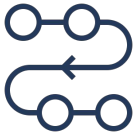
- Funder policies on research data
- Institutional policies on research data
- Support services
- Domain-specific infrastructures/resources
- Ethical committees and general authorities
- Relevant ethical guidelines
- Laws and regulations relevant to life sciences research data
- Tools and resources
-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

GitHub

Simple, sustainable platform

Guidelines, process documents and data are made available under a CC-BY license.

Software is made available under an MIT license.

FAIRsharing.org
standards, databases, policies

search through all content

SEARCH LOGIN

STANDARDS DATABASES POLICIES COLLECTIONS ORGANISATIONS ADD CONTENT STATS

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:

elixir CONVERGE

DSW FAIRsharing integration

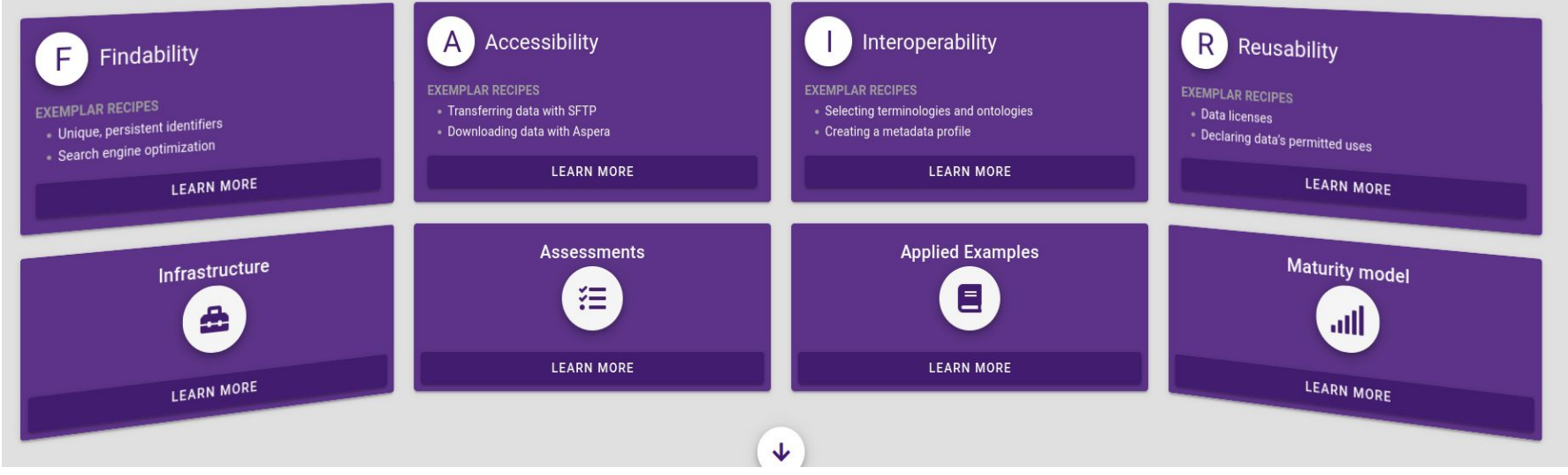
FAIRCOOKBOOK

The FAIR Cookbook

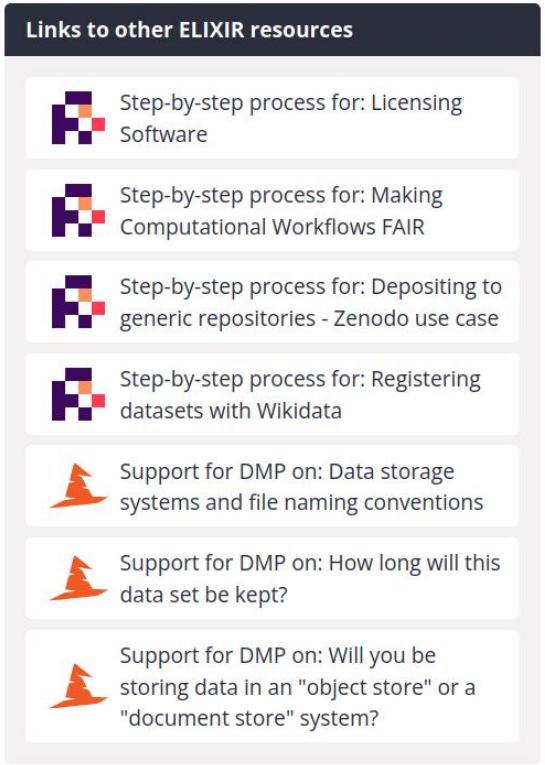
for FAIR doers

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/

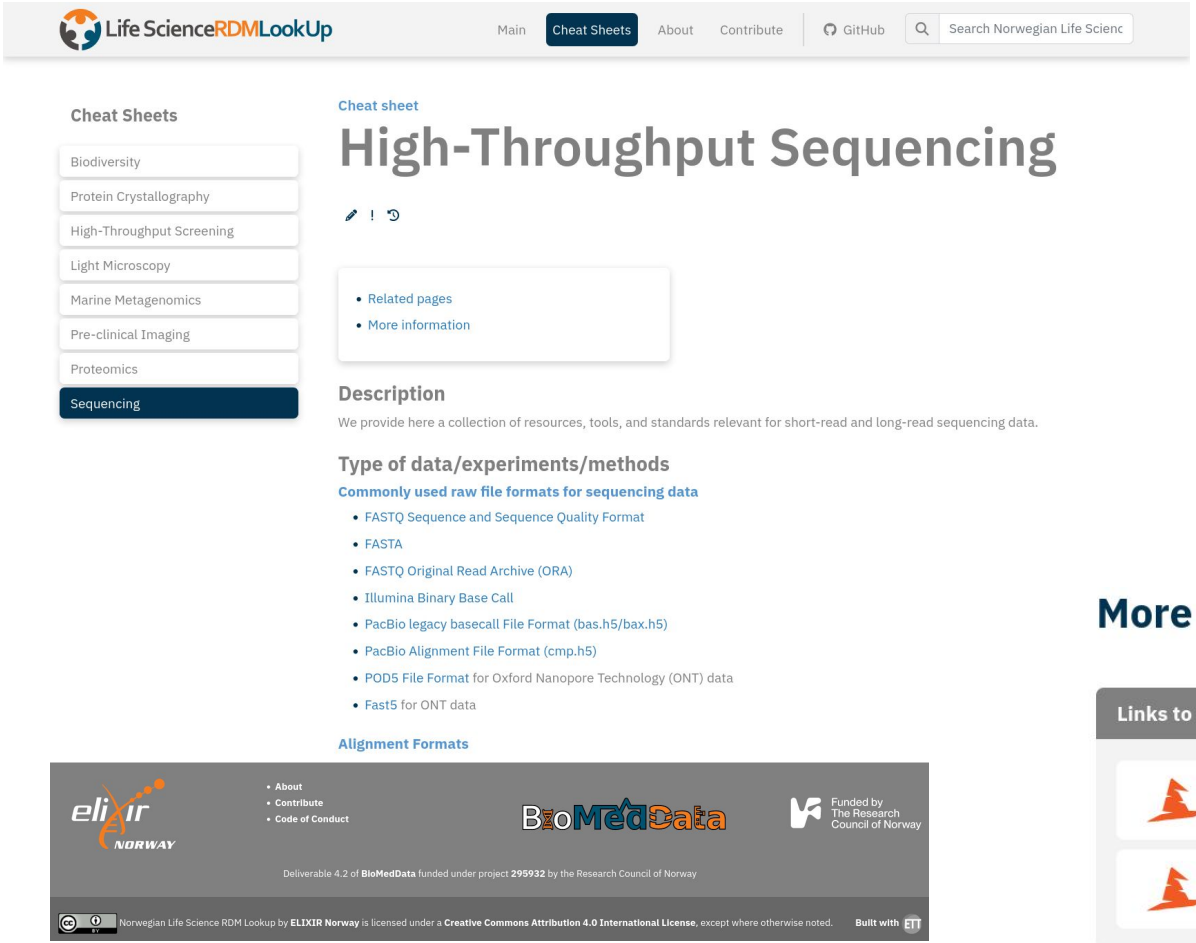


RDMkit: Integrations



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



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

DSW

DATA STEWARDSHIP WIZARD

Live demo

Coffee Break

10-15 minutes

III. Creating and collecting data 32

- Are you running the project in a collaboration between different groups or institutes?
- Will you be collecting physical samples?
- How will you do file naming and file organization?
- What existing data formats/types will you be using?
- What existing encodings/terminologies/vocabularies/ontologies will you be using?
- Will you be using new types of data?
- How will you be collecting and keeping your metadata?
- Will you be acquiring data using measurement equipment?
- Do you have any non-equipment data capture?
- Is there a data integration tool that can handle and combine all the data types you are dealing with in your project?
- Will you collect any data connected to a person, "personal data"?
- Is the data collection subject to ethical legislation?
- Data use restrictions
- How are the rights of the collected data arranged?
- Will you monitor data integrity once it has been collected?
- Do you need guidance on what data formats/types to use?
- Do you need guidance on what encodings/terminologies/vocabularies/ontologies to use?

IV. Processing data

- Will you be using a shared working space to work with your data?
- Data storage systems and file naming conventions
- Workflow development
- How will you make sure to know what exactly has been run?
- How will you validate the integrity of the results?
- Do you need to do compute capacity planning?
- Is the risk of information loss, leaks and vandalism acceptably low?
- Do you have a contingency plan?

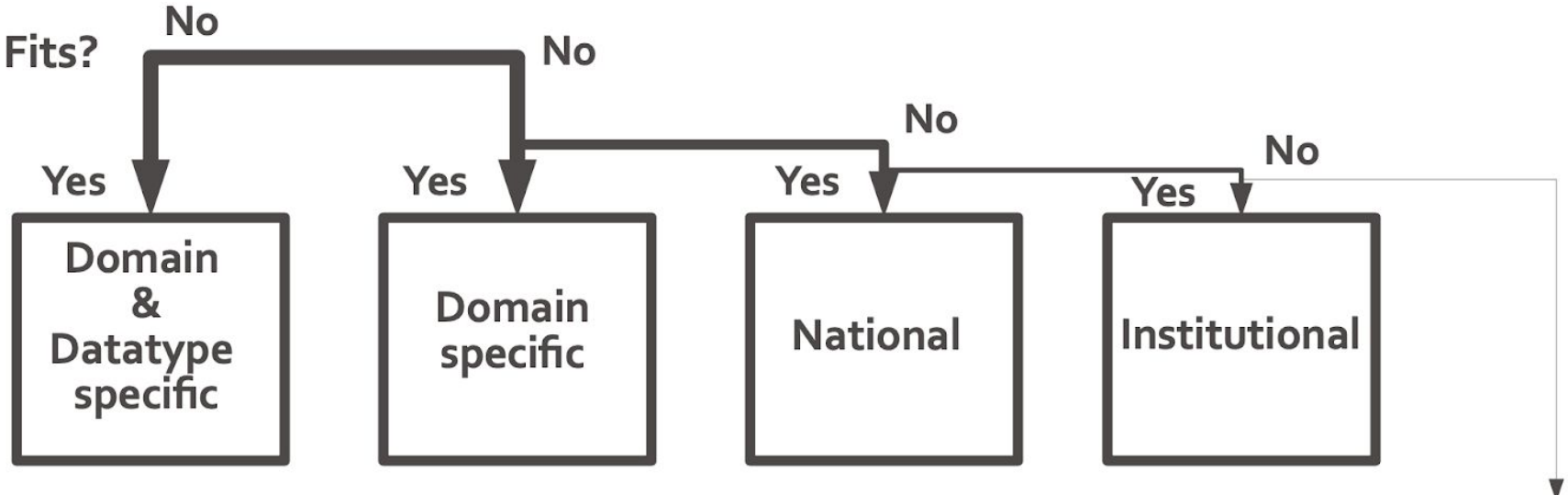
VI. Preserving data 22

- Specify a list of data sets you will be producing
- Will you be archiving data (using so-called 'cold storage') for long term preservation already during your project?
- Will you be archiving your data in 'cold storage' after the project finishes?
- Will any of the repositories you use charge you for their services?
- Are there any other recurring fees to keep data or documents available?
- Did you budget for the time and effort it will take to prepare the data for publication?
- Will you be making sure that blocks of data deposited in different repositories can be recognized as belonging to the s…
- Specify a list of software packages you will be publishing
- Will reference data be created?

VII. Giving access to data 4

- Will you be working with the philosophy 'as open as possible' for your data?
- Can all of your data become completely open over time?
- Will you use temporary restrictions on the reuse of the data (embargo)?
- Will metadata be available openly?

Where should I deposit my data?



ENA
European Nucleotide Archive

NCBI SRA

BioStudies.

NIRD RESEARCH DATA ARCHIVE

DataverseNO
Dataverse Network Norway

NTNU Open Research Data
NTNU

NMBU Open Research Data
Norwegian University of Life Sciences

UiB Open Research Data

UiT Open Research Data

zenodo

EUDAT

figshare

Wrapping-up

Further comments and questions

Evaluation form

https://forms.gle/URXJfkbPB49vEBLm8

ELIXIR Norway group photo, Bergen 2021

elixir NORWAY

https://elixir.no/

Horizon 2020 Research Infrastructure Program

European Commission

With funding from The Research Council of Norway

BioMedData

CENTRE FOR DIGITAL LIFE NORWAY

Thank you!

elixir-norway.org

@elixirnorway

support@elixir.no