



# Data Preservation

**Speakers:** Thu-Hien To (NMBU), and Korbinian Bösl (University of Bergen, UiB)

**Moderator:** Even Birkeland (UiB)

# Data storage & Repository



## Learning Objectives

After this session, you will be able to understand:

- What to consider when preserving your data
- How to select a relevant repository for archiving your data
- Having an overview of existing deposition database for life sciences

# Data preservation

**Ensure safety, integrity and accessibility of data for as long as necessary**

- Guarantee that your research can be verified and reproduced for several years after the end of the project
- Allow the reuse of your data

# Before starting the project

- Anticipate your needs:
  - Estimate the volume of raw data
  - What kind of access/transfer speed and access frequency
  - Where the data will come from
  - How long the raw data, analysis workflow need to be stored
- Check the storage solutions that you can have from institutional/national service:
  - Backup regulations
  - Tiered storage (“Cold” and “Hot” Storage)

# During the project

- Control access of the data:
  - Have the raw data read only – maybe have a separate copy
  - Be aware of who has which permission to the data
- Keep track of the changes, include information about provenance of data
- Ensure that data is well organized:
  - Check if your institute provides a (meta)data management system
  - Ensure that a consistent naming scheme, folder structures and versioning systems.
  - Make sure to generate well described and self-explanatory documentation (i.e., README file) and (rich) metadata together with the data.
- Update file format and software so that they do not become outdated
- Include sufficient licensing information

# At the end of the project

- Archiving your data
  - Unique data or that cannot be easily re-generated (e.g. raw data, analysis workflow)
  - Data that will probably being reused in the future
  - Data of great value for society (scientifically, historically or culturally)
  - Check with funder, publisher and institution policies
  - Check with legal or ethical requirements
- **Deposit your data into data repositories**
  - ➔ Increased visibility, added values → more citations
  - ➔ FAIRification - Funding agencies requirement
  - ➔ Safe money on storage

# FAIR Data Repository

<b>Findable</b> 	Persistent Identifiers (PIDs) 	Rich metadata 	Indexed data repositories 	PIDs in metadata 
<b>Accessible</b> 	Standard communications protocol 	Open, free protocol 	Authentication, where necessary 	Metadata is always available 
<b>Interoperable</b> 	Vocabularies 	Vocabularies are FAIR 	Linked metadata 	
<b>Reusable</b> 	Metadata have multiple attributes 	Usage license 	Provenance 	Community standards 

# Looking for relevant data repositories

- [fairsharing.org](https://fairsharing.org) : 1863 registered database  
Community-developed platform (including standards, repositories and policies across all disciplines)
- [re3data.org](https://re3data.org) : 2834 registered database  
Registry of Research Data Repositories (overview of existing international repositories of research data from all academic disciplines).



Registry

Search  APPLY

Record Type

Search  APPLY

Subjects

Search  APPLY

Domains

Search  APPLY

Licence(s)

Search  APPLY

Organisation(s)

Search  APPLY

Countries

Search  APPLY

Species

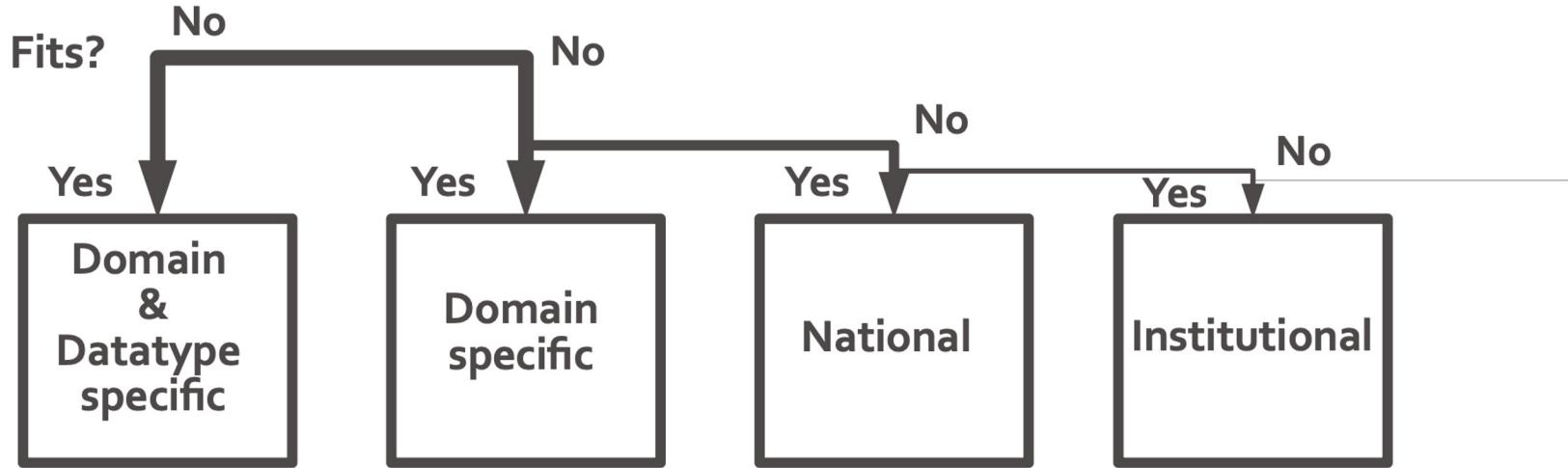
Search  APPLY

User defined tags

Search  APPLY

- Subjects 
- Content Types 
- Countries 
- AID systems 
- API 
- Certificates 
- Data access 
- Data access restrictions 
- Database access 
- Database access restrictions 
- Database licenses 
- Data licenses 
- Data upload 
- Data upload restrictions 
- Enhanced publication 
- Institution responsibility type 
- Institution type 
- Keywords 
- Metadata standards 
- PID systems 
- Provider types 
- Quality management 
- Repository languages 
- Software 
- Syndications 
- Repository types 
- Versioning 

# Repositories specification

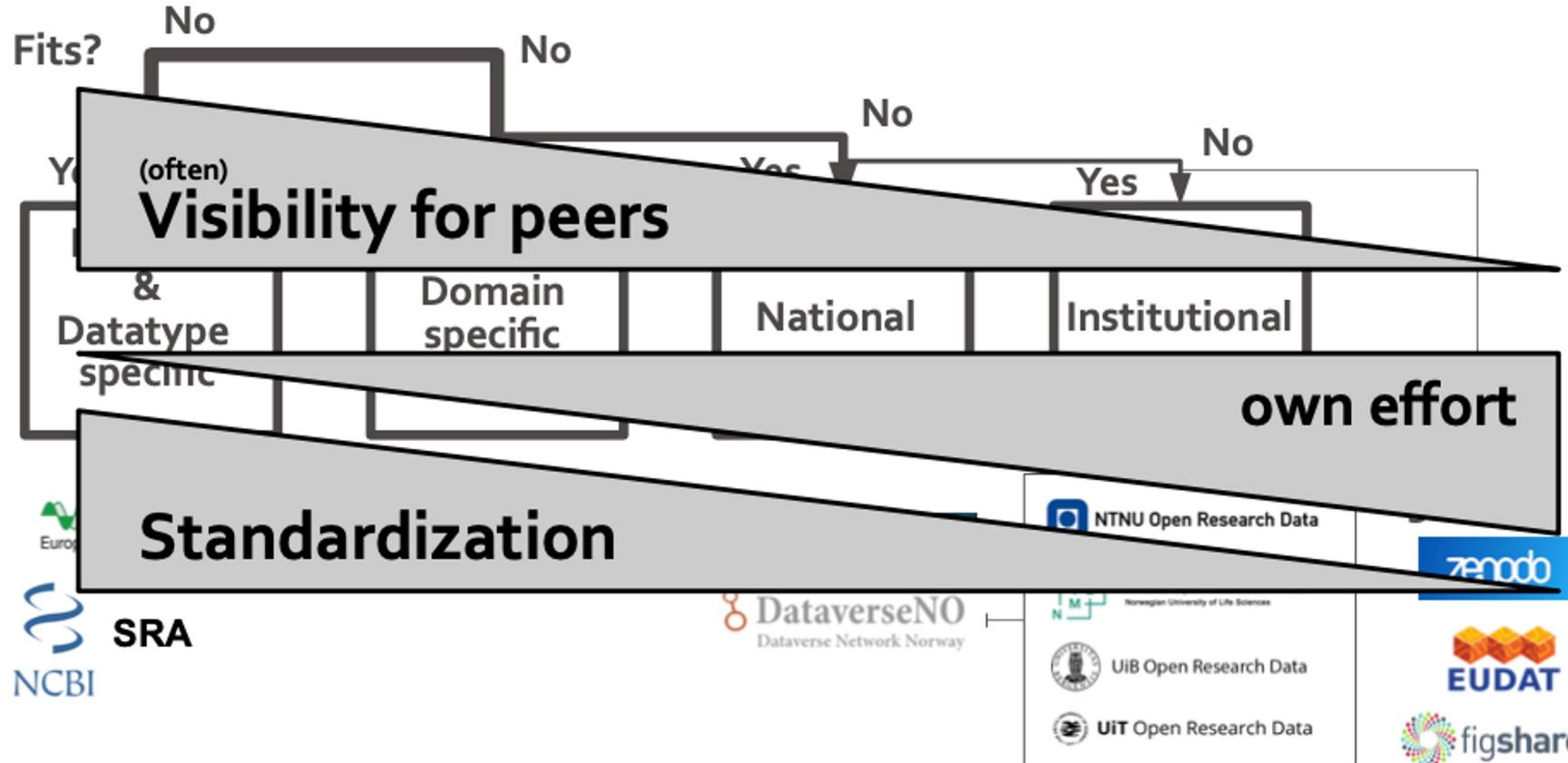


NIRD RESEARCH DATA ARCHIVE

DataverseNO  
Dataverse Network Norway



# Repositories specification



# ELIXIR Deposition Databases



## BioStudies.



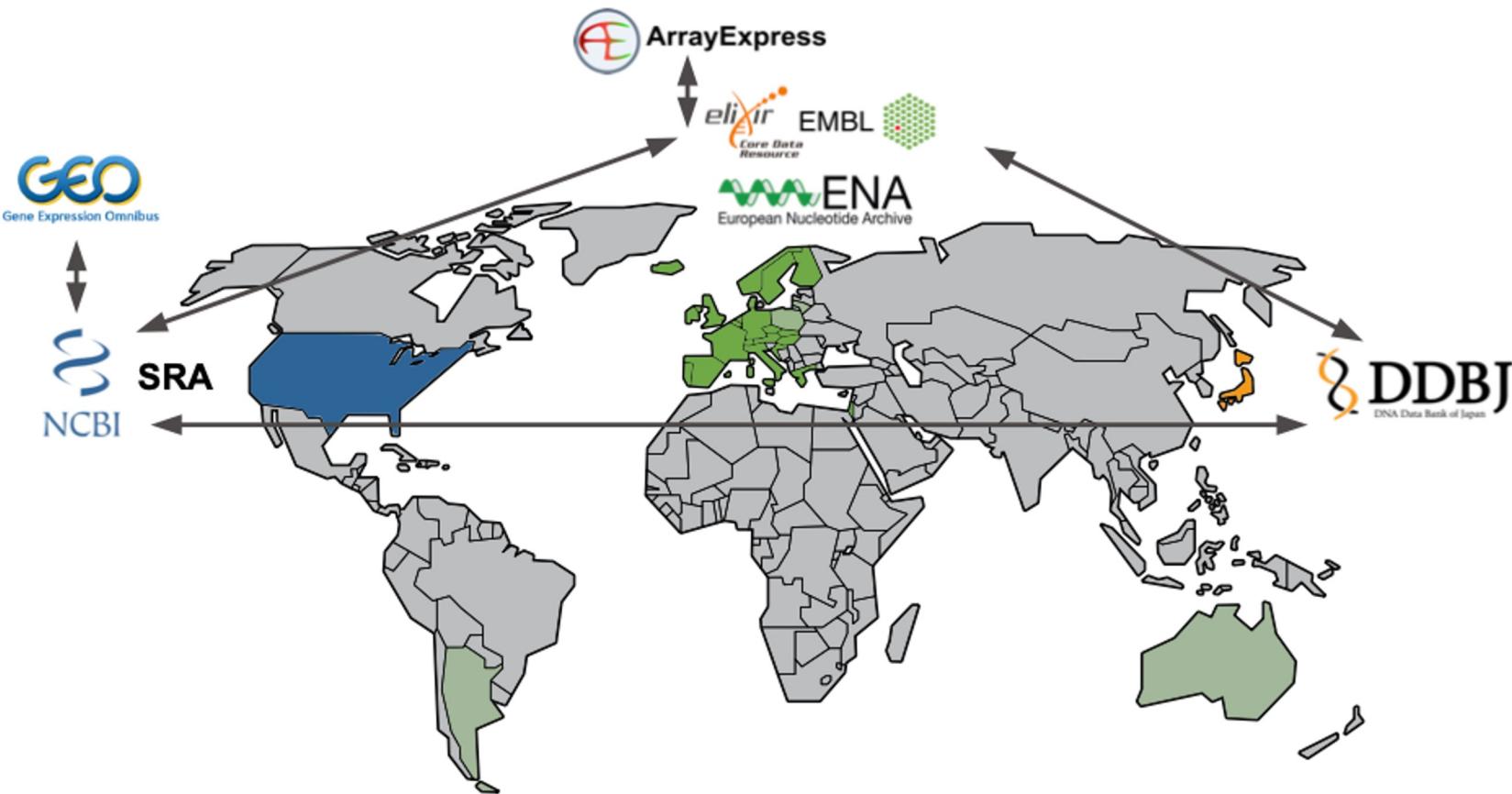
Data submission wizard which helps to find the right archive  
for your data through a series of questions  
<https://www.ebi.ac.uk/submission/>

- Recognized & recommended
- Free of charge
- Quick upload
- Not operated by single university/government
- Guided submission process
- Generous embargo regulation (2yrs)
- API access (upload from NeLS planned)

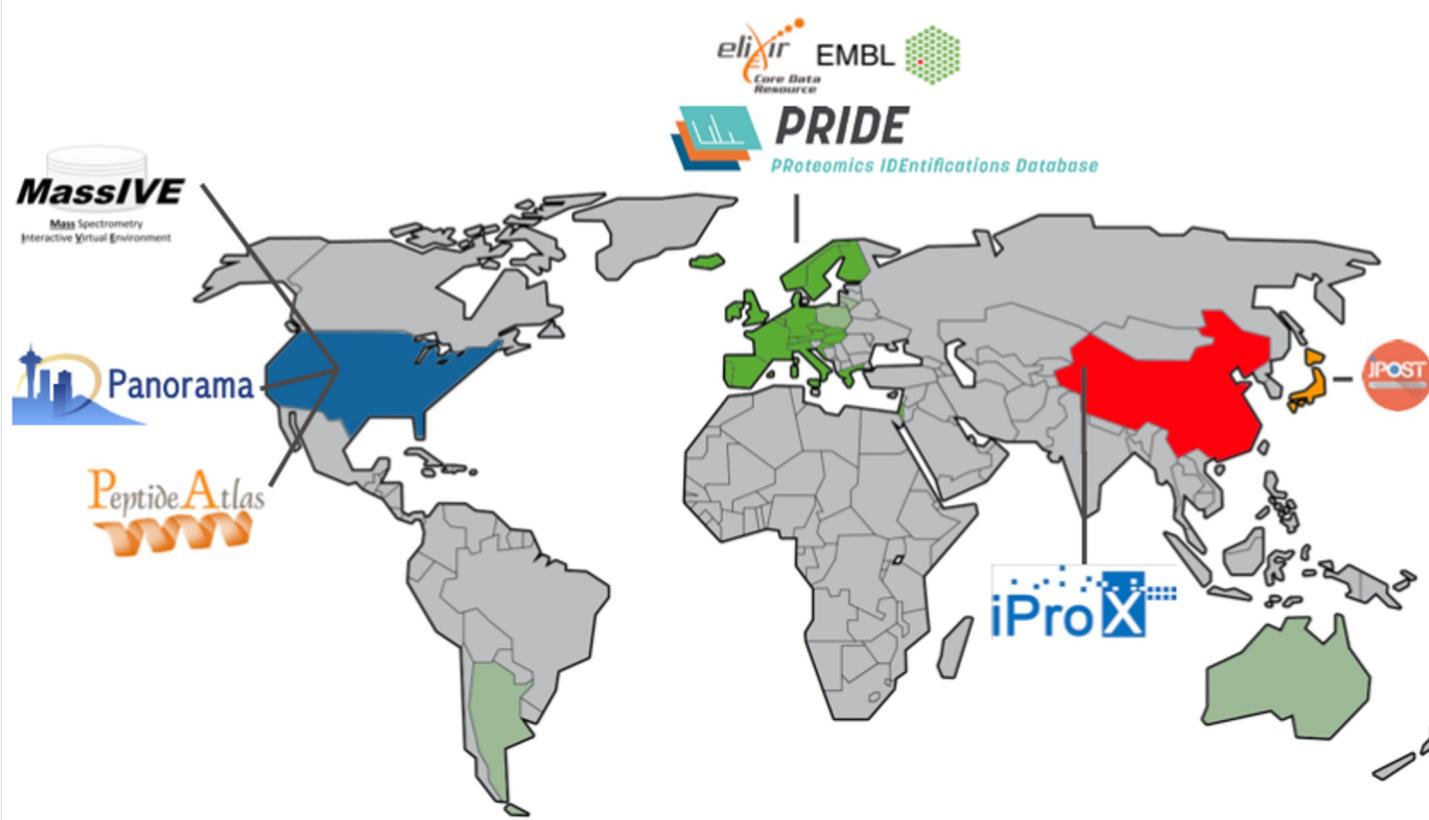
# ELIXIR Deposition Database



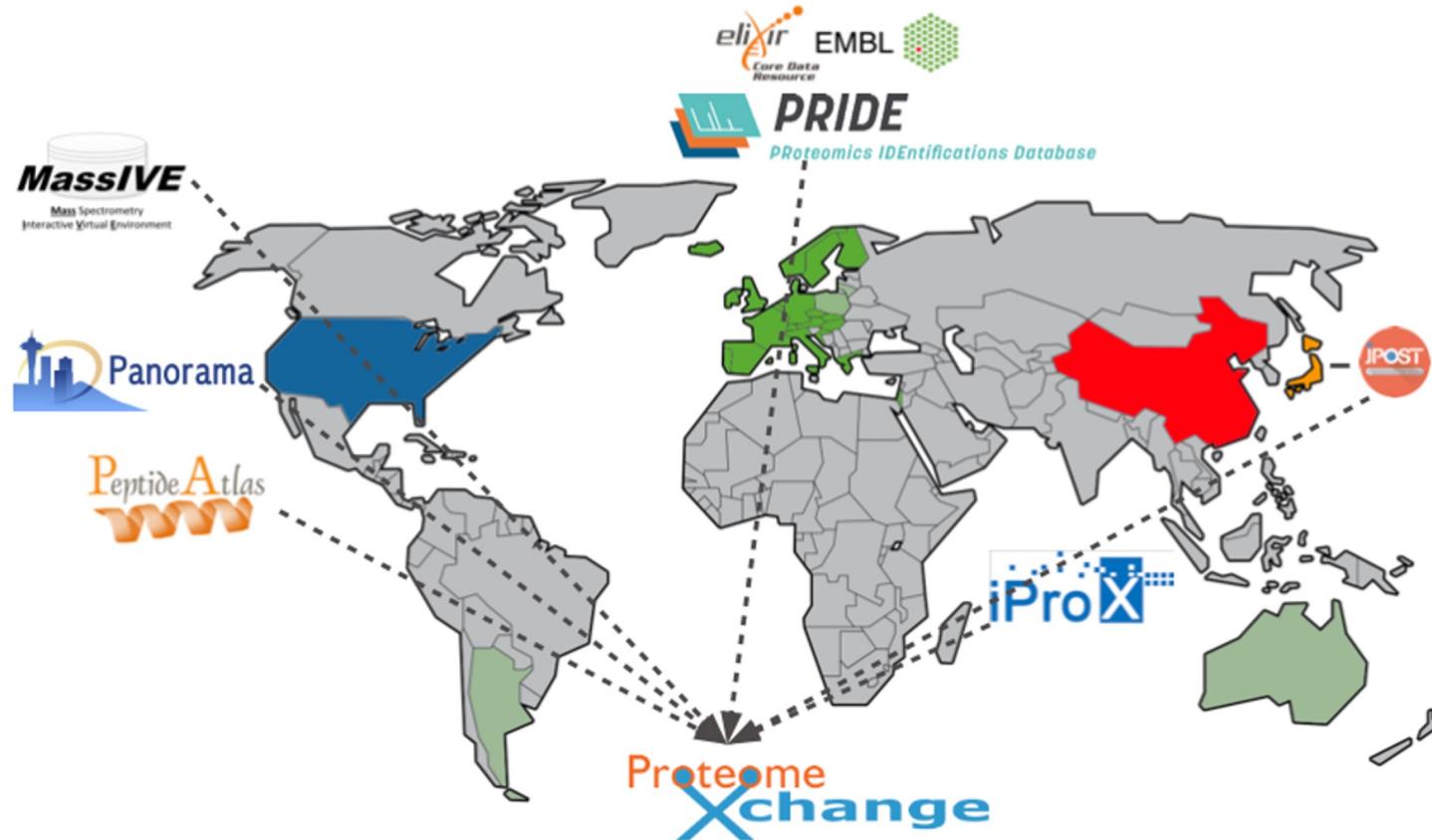
# Sequencing data



# Proteomic data



# Proteomic data



# Neuroscience data



**EBRAINS**

- a FAIR place to share your neuroscience data

All types of neuroscience data are welcome



Histology



2-Photon



Ephys

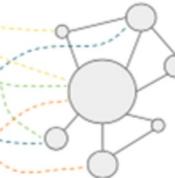


Models

Data Curation and support



Integration of data into the EBRAINS Knowledge Graph Search



Visit our website:

Or contact us directly at:

Anne Marte Kvello

Data and metadata management and storage

Data curation and integration

Data security

Collaboration and sharing of data

Copyright, licensing and citation

Long-term storage of data

# National, institutional archives

- NIRD

- National, free depositing repository
- Domain agnostic
- Dublin-core metadata standard
- DOI accessible
- Supports machine readable metadata harvesting



- NSD (Sikt from 2022)

- Domain agnostic
- DOI accessible
- Implemented data access committees
- Main data type: Questionnaires



- DataverseNO

- National/Institutional, free depositing repositories
- Domain agnostic
- Manually curated by libraries
- DOI accessible
- 8Gb max/upload - contact support >50Gb



# Generic archives



Less standard metadata → Re-usability ↓

Domain agnostic → Findability ↓

Guided submission process

Commercial factors: Size limitations, ...



---

<https://zenodo.org/>

<https://b2share.eudat.eu/>

<https://figshare.com>

# Archiving codes/scripts



Good integration - easy to implement

Archival of each release

beta version for gitlab

<https://zenodo.org/>

<https://guides.github.com/activities/citable-code/>

<https://gitlab.com/sbeniamine/gitlab2zenodo>

# Archiving codes/scripts

1. Add doi-button to GitHub README (copy & paste from Zenodo)

Markdown

```
[![DOI] (https://zenodo.org/badge/DOI/10.5281/zenodo.5772661.svg)] (https://)
```

2. Add & fill out GitHub CITATION.cff template  
>> import in Zenodo  
>> supported by Zotero

Example CITATION.cff file:

```
cff-version: 1.2.0
message: "If you use this software, please cite it as below."
authors:
- family-names: "Lisa"
  given-names: "Mona"
  orcid: "https://orcid.org/0000-0000-0000-0000"
- family-names: "Bot"
  given-names: "Hew"
  orcid: "https://orcid.org/0000-0000-0000-0000"
title: "My Research Software"
version: 2.0.4
doi: 10.5281/zenodo.1234
date-released: 2017-12-18
url: "https://github.com/github/linguist"
```

<https://docs.github.com/en/github/creating-cloning-and-archiving-repositories/creating-a-repository-on-github/about-citation-files>



# Legal requirements

## Learning Objectives

After this session, you will be able to:

- Identify potential sources of legal implications
- Identify if the data you work with is personal
- Having an overview of relevant support services and authorities

## Learning Activity

### Quiz

# Disclaimer

This is not a legal advice

# Categories of legal requirements

## Contractual obligations

- Funder
- Employer



## Requirements originating from

- Laws
- Regulations
- Ethical guidelines



# Categories of legal requirements

## Contractual obligations

- Funder
- Employer



## Requirements originating from

- Laws
- Regulations
- Ethical guidelines



# The 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”

# RCNs guidelines for storage, dissemination and sharing of research data



## Research data...



## Research data...



...must be stored/archived in a safe and secure manner.



## Research data...



...must be stored/archived in a safe and secure manner.



...must be made accessible for reuse.



## 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]



## 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.



## 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.



## 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.



## 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.



## Example: Fripro

You must **prepare a data management plan** for any research data handled in the project. The **data must be made available** in accordance with the FAIR principles (Findable, Accessible, Interoperable and Reusable). The Project Owner is responsible for selecting which archiving solution(s) to use for storing research data generated during the project.

# FAIR-principles

## Data Management Plan

Storage

Archival

Ethics

Deposition



<https://www.uib.no/en/foremployees/142184/university-bergen-policy-open-science>

<https://innsida.ntnu.no/wiki/-/wiki/English/NTNU+Open+Data>

[https://www.nmbu.no/en/research/for\\_researchers/researchdata/node/34680](https://www.nmbu.no/en/research/for_researchers/researchdata/node/34680)

<https://www.uio.no/english/for-employees/support/research/research-data-management/policies-and-guidelines/index.html>

[https://intranett.uit.no/Content/532111/cache=20170109133727/Principles%20and%20guidelines%20for%20research%20management%20at%20UiT\\_010917.pdf](https://intranett.uit.no/Content/532111/cache=20170109133727/Principles%20and%20guidelines%20for%20research%20management%20at%20UiT_010917.pdf)



# Categories of legal requirements

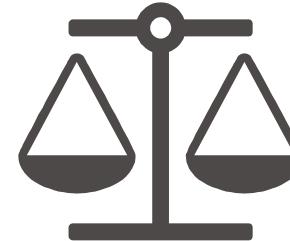
## Contractual obligations

- Funder
- Employer



## Requirements originating from

- Laws
- Regulations
- Ethical guidelines



# Important Relevant Legislation

Research Ethics Act

Health Research Act

Health Registry Act

Biotechnology Act

Archive Act

Patents Act

Copyrights Act

...



- Health Register Act: <https://lovdata.no/dokument/NL/lov/2014-06-20-43>
- Biotechnology Act: <https://lovdata.no/dokument/NL/lov/2003-12-05-100>
- Archives Act: <https://lovdata.no/dokument/NL/lov/1992-12-04-126>
- Patents Act: <https://lovdata.no/dokument/LTI/lov/2019-06-21-49>
- Copyright Act: <https://lovdata.no/dokument/NL/lov/2018-06-15-40>

# Important Relevant Legislation

Personal Data Act

Regulations on the processing of personal data

General Data Protection Regulation (GDPR)



**The GDPR is an EU Regulation that applies directly to Norway, as a member of the EEA.**



# Research Ethics Act



Withhold-, mislead about-, or selectively/secretly dispose of undesired results.



Conceal of scientific efforts and / or scientific achievements.  
Improper allocation of authorship etc.



Destruction of research data / material to prevent investigations of misconduct.

Research Ethics Act: <https://lovdata.no/dokument/NL/lov/2017-04-28-23>



**General guidelines on  
research ethics**

**Guidelines for Research Ethics  
in Science and Technology**

**Guidelines for clinical  
trial of drugs**

**Guidelines for the use of  
genetic studies of humans**

**Guidelines for the inclusion  
of women**

**Guidelines for research ethics and  
scientific assessment of qualitative  
research projects**

Read more at: [forskningsetikk.no](http://forskningsetikk.no)



**The Vancouver Recommendations**

**Declaration of Helsinki**

**Oviedo Convention**

Otherwise known as: **The Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine**

Read more at: [forskningsetikk.no](http://forskningsetikk.no)



# Health Research Act

- Prior approval for health research
- Consent from participants
- Data access rights for participants
- Biobank Regulations
- Maximum data storage time for non-archived data  
(Default: 5 years after end of project – exemptions: approval)
- Regional Committees for Medical and Health Research Ethics (REK)





# GDPR: General Data Protection Regulation

-  Protection of Personal Data
-  Privacy by Design
-  Data Protection Impact Assessment (DPIA)
-  Processing of Personal Data
-  Technical and organisational measure to secure data
-  Records of Processing Activities
-  Access rights, Right to be forgotten, Right on Information
-  Fines: Up to 20 million Euros, ... up to 4 % of their total global turnover of the preceding fiscal year



# Personal Data Act & Personal Data Regulations

## National Implementation of GDPR



Consent from participants (13 years and older)



Exceptions for archival, public interest, and scientific reasons



Authorities: Privacy Ombudsman, Privacy Committees,  
Data Inspectorates (datatilsynet)

More information at:

- Regulations on the processing of personal data: <https://lovdata.no/dokument/SF/forskrift/2018-06-15-876>
- Transitional rules on the processing of personal data: <https://lovdata.no/dokument/SF/forskrift/2018-06-15-877>
- Personal Data Act: <https://lovdata.no/dokument/NL/lov/2018-06-15-38>

# Let's find out: What is...?

Personal Data

Sensitive Personal Data

Genetics Data

Pseudonymisation

Anonymisation

# What defines personal data?

-  Any information that relates to an identified or identifiable living individual
-  What if data can be de-identified, encrypted, or pseudonymised?
-  **Still, it is considered personal data!**
-  Truly anonymised → anonymisation must be irreversible.

Further reading: [https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en)

Illustrative Icons from [Font Awesome](#) licensed under [Creative Commons Attribution 4.0 International License](#)

# What is personal data? Examples:

-  A name and surname
-  A home address
-  An email address such as  
name.surname@company.com;
-  Location data
-  An Internet Protocol (IP) address
-  A cookie
-  Data held by a hospital or doctor, which could be a symbol that uniquely identifies a person.

Further reading: [https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en)

Illustrative Icons from [Font Awesome](#) licensed under [Creative Commons Attribution 4.0 International License](#)

# What personal data is considered special?

- Personal data revealing racial or ethnic origin, political opinions, religious, or philosophical beliefs
- Trade-union membership
- Genetic data, biometric data processed solely to identify a human being
- Health related data

Further reading - what personal data is considered sensitive?: [https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en)

# What defines genetic data?



Genetic data is defined as: “personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person.” - GDPR article 4(13)

- ⚠ Always both: Personal identifier and sensitive information!

Further reading - what personal data is considered sensitive?: [https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en)

# Data Psuedonymisation



“The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person” - GDPR article 4(13)

- De-identified and not back traceable for the researcher without the identifier
- ⚠ Pseudonymised data is still personal data!

# Anonymous Data

→ Anonymous data cannot in any way be used to identify individuals in a data material, either:

- ◆ directly by name or personal identification number or
- ◆ indirectly by additional information

⚠ Not possible for many data types, such as genetic data

# What is...?



Personal Data



Sensitive Personal Data



Genetics Data



Pseudonymisation



Anonymisation

# What does GDPR say about...?

Data subjects, processing, controller, and processors

Storage and processing of personal (sensitive) data: How & When?

- Fairness and Transparency
- Data Minimisation
- Data Accuracy
- Storage Limitation
- Integrity and Confidentiality
- Accountability



# What does GDPR say about:

## Data subjects, processing, controller, and processors

→ **Data subject:**

“the natural person information relates to.”- GDPR article 4(1)

→ **Data Processing:**

“any operation or set of operations which is performed on personal data or on sets of personal data.” - GDPR article 4(2)

→ **Data Controller:** determines the purposes and means of the processing of personal data

→ **Data Processor:** processes personal data on behalf of the controller

# Conditions to store/process personal (sensitive) data

- Lawful and transparent manner ('**lawfulness, fairness and transparency**')
- Specific purposes ('**purpose limitation**')
- Only the personal data that is necessary to fulfil that purpose ('**data minimisation**')
- Stored for no longer than necessary ('**storage limitation**')
- **Data Accuracy**
- Technical and organisational safeguards that ensure the security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technology ('**integrity and confidentiality**')
- **Accountability** → Immediate reporting of incidents

# Criteria to store/process personal (sensitive) data



Explicit consent from participant

Health Research Act



- The data is processed for **archiving, scientific or historical research** purposes or statistical purposes on the basis of EU or national law.
- “Tasks carried out in the public interest” - Article 6, GDPR
- “It is not always possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection” - GDPR, Recital 33

## Further readings:

- [https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/under-what-conditions-can-my-company-organisation-process-sensitive-data\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/under-what-conditions-can-my-company-organisation-process-sensitive-data_en)
- [https://lovdata.no/dokument/NL/lov/2008-06-20-44/KAPITTEL\\_4#KAPITTEL\\_4](https://lovdata.no/dokument/NL/lov/2008-06-20-44/KAPITTEL_4#KAPITTEL_4)



# Criteria to store/process personal (sensitive) data



Explicit consent from participant

Health Research Act



If you are using consent as your legal basis:

**Make sure your consent form (& REK\* approval)  
allows controlled access deposition (e.g. to EGA\*\*)  
...before you start!**

\* REK: Regional Committees for Medical and Health Research Ethics

\*\* EGA: European Genome Archive

Further readings:

- [https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/under-what-conditions-can-my-company-organisation-process-sensitive-data\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/under-what-conditions-can-my-company-organisation-process-sensitive-data_en)
- [https://lovdata.no/dokument/NL/lov/2008-06-20-44/KAPITTEL\\_4#KAPITTEL\\_4](https://lovdata.no/dokument/NL/lov/2008-06-20-44/KAPITTEL_4#KAPITTEL_4)

# Storage Limitation

- Personal data shall not be stored for longer than necessary to fulfil the purpose
- Once the purpose has been achieved, the data shall in principle be deleted or made anonymous
- REK usually sets requirements for storage beyond the project period for reasons of verifiability - **you may need to include this in your application!**
- Processing personal data for the purpose of verifiability is legitimate
- Personal data may be stored for longer periods so far as the personal data is processed solely for scientific, historical, and/or statistical research purposes

## Further readings:

- [https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/under-what-conditions-can-my-company-organisation-process-sensitive-data\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/under-what-conditions-can-my-company-organisation-process-sensitive-data_en)
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e1888-1-1>

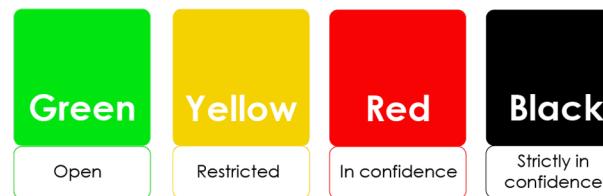
# Integrity and Confidentiality

→ Personal data must be processed in a manner that ensures:

- ◆ appropriate security of the personal data,
- ◆ protect personal data against unauthorised access,
- ◆ unlawful processing,
- ◆ accidental loss, distribution, amendment, or damage



Follow institutional guidelines



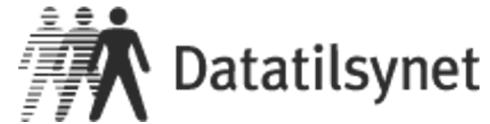
→ Pseudonymisation

- ◆ Dedicated data analysis platforms

Image Source: <https://ui.no/english/for-employees/support/research/funding/units/hf/imv/data-ethics/colors.html>

More about TSD: <https://www.ui.no/english/services/it/research/sensitive-data/>

# Responsible Authorities



# Contact Information

## Data Protection Officers

UiB Janecke.Veim@uib.no

NTNU: thomas.helgesen@ntnu.no

UiT personvernombud@uit.no

NMBU personvernombud@nmbu.no

UiO personvernombudet@uio.no

## REK

rek-vest@uib.no

rek-midt@mh.ntnu.no

rek-nord@asp.uit.no

[rek-sorost@medisin.uio.no](mailto:rek-sorost@medisin.uio.no)

Datatilsynet - [datatilsynet.no](http://datatilsynet.no)

SIKT - sikt.no



[rekportalen.no/#home/REK](http://rekportalen.no/#home/REK)



[sikt.no](http://sikt.no)



Your domain



Plant sciences

Marine metagenomics

Human data

Biomolecular simulation data



## Data Management Life Cycle

licensed under a [Creative Commons Attribution 4.0 International License](#)

Link to RDMkit: <https://rdmkit.elixir-europe.org/>

[https://rdmkit.elixir-europe.org/human\\_data.html](https://rdmkit.elixir-europe.org/human_data.html)



## Useful Resources

- Data Preservation <https://rdmkit.elixir-europe.org/preserving>
- <https://fairsharing.org>
- <https://www.re3data.org>
- [ELIXIR Deposition Databases](#)
- [EMBL-EBI submission wizard](#)
- [NIRD research data archive](#)
- [DataverseNO](#)
- [NSD](#)
- [RDMkit overview of national regulations and laws](#)
- [Biomedical data Ethical, legal and social implications - online course EBI](#)



# Thank you!



elixir-norway.org



@elixirnorway



contact@bioinfo.no



Except where otherwise noted, this work is licensed under:

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

Content adapted from Nazeefa Fatima