



Archiving, preserving, sharing



Korbinian Bösl
Data management coordinator
ELIXIR Norway & Centre for Digital Life Norway
06 June 2024



Data management

Data life cycle ▼

Your role ▼

Your domain ▼

Your tasks ^

Compliance monitoring

Data analysis

Data management plan

Data organisation

Data protection

Data publication

Data quality

Data storage

Data transfer

Documentation and metadata

Existing data

Identifiers

Licensing

Machine actionability

Sensitive data

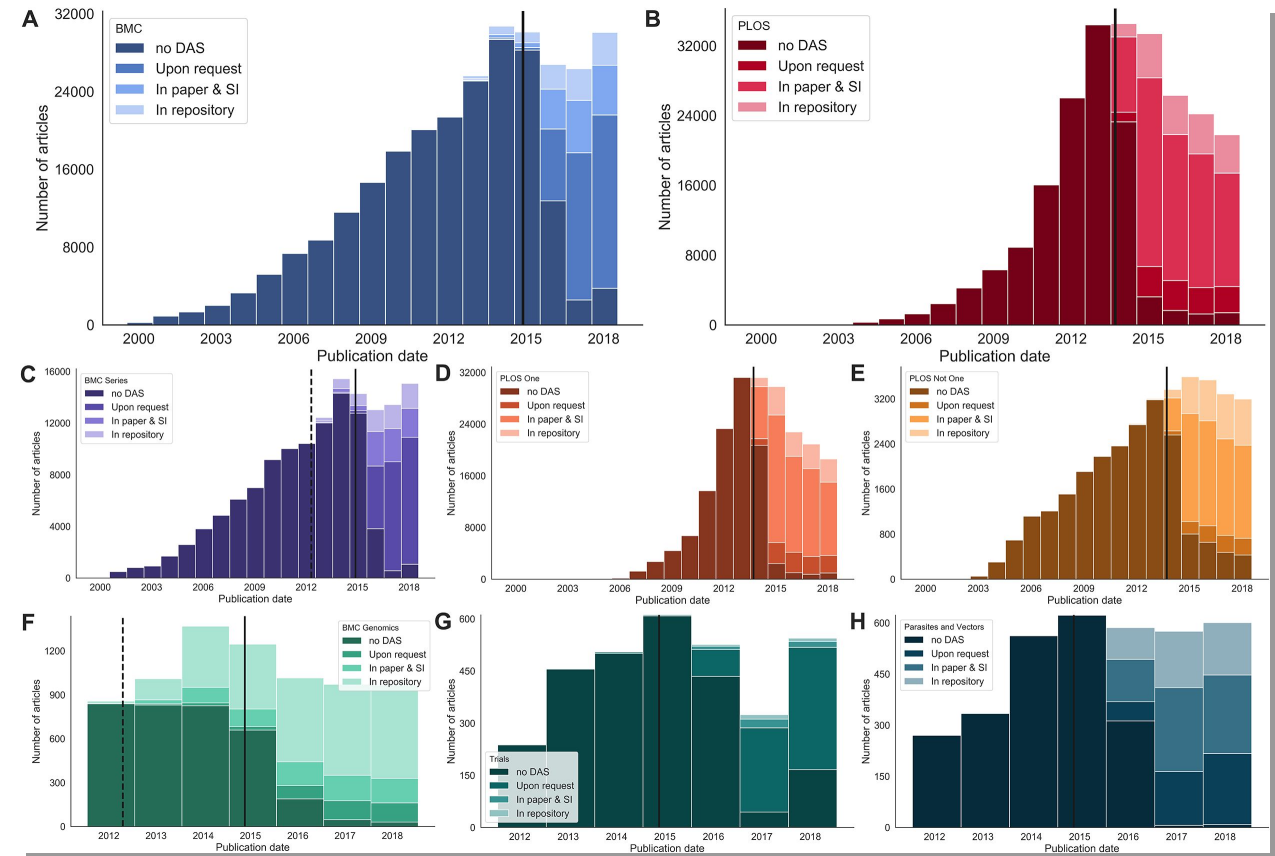
- Can you really deposit your data in a public repository?
- Which repository should you use to publish your data?
- How do you prepare your data for publication in data repositories?
- More information
- Relevant tools and resources



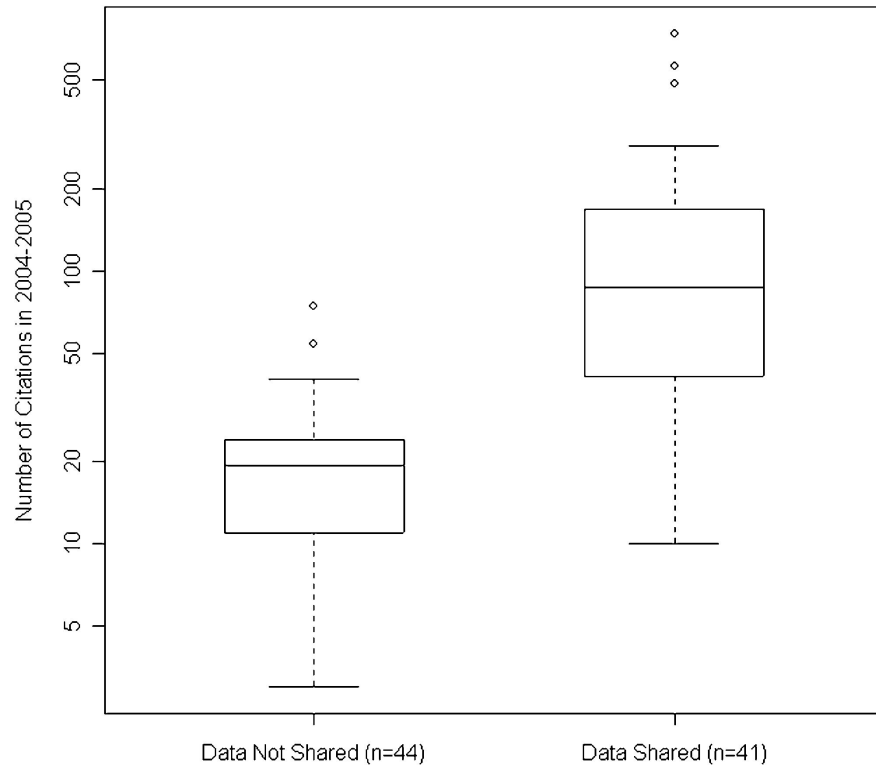
<https://rdmkit.elixir-europe.org/>

Why should I deposit my data?

- Increased Visibility (SEO), added value → 25% more citations
- Funding agencies requirement; institutions, and journal publishers policies
- Save money on storage



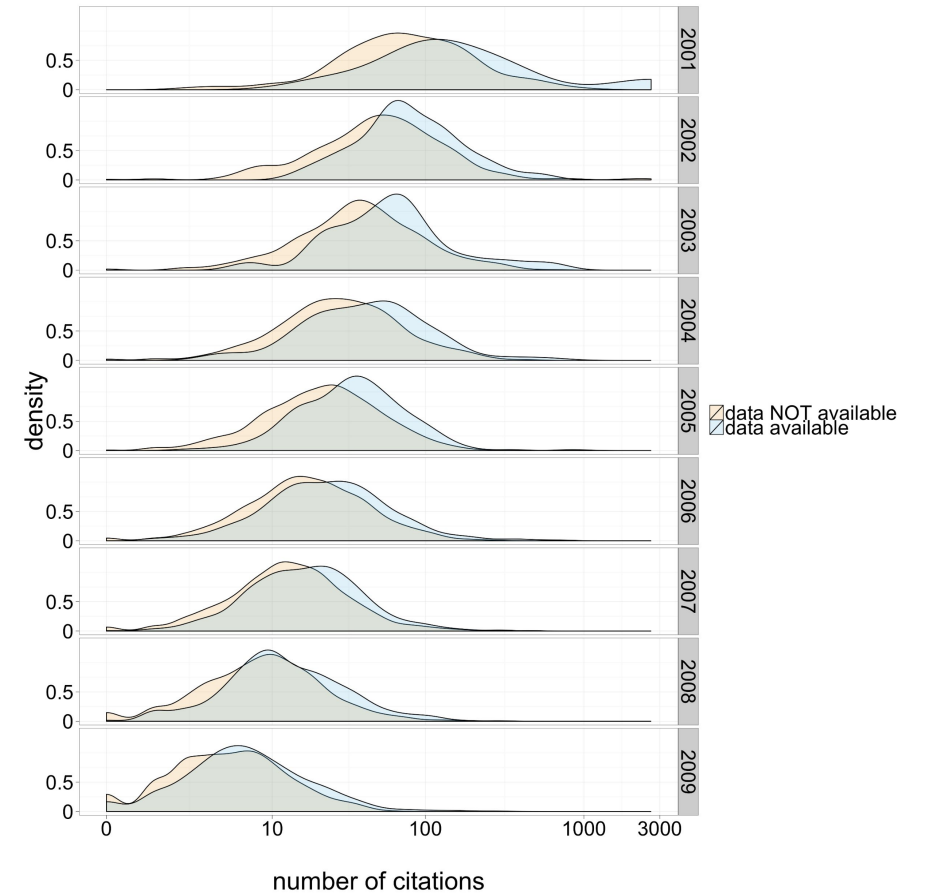
Articles with shared data have higher citation rates



Distribution of 2004–2005 citation counts of 85 trials by data availability.

“The 41 clinical trial publications which publicly shared their microarray data received more citations”

Heather A. Piwowar, Roger S. Day, and Douglas B. Fridsma. 2007. PLOS ONE, [doi:10.1371/journal.pone.0000308](https://doi.org/10.1371/journal.pone.0000308). Figure licensed CC BY.



Citation density for papers with and without publicly available microarray data, by year of study publication.

Heather A. Piwowar, Roger S. Day, and Douglas B. Fridsma. 2007. PLOS ONE, [doi:10.1371/journal.pone.0000308](https://doi.org/10.1371/journal.pone.0000308). Figure licensed CC BY.

The Research Council Policy for Open Science

European Open Science Cloud (EOSC):

- European network of **FAIR** data and services

Since 2009: Open Access requirement

“...research data must be stored in a safe and secure manner for a minimum of 10 years.”

Early deposition & sharing

Data is deposited to trustworthy repository

Adopted Plan S

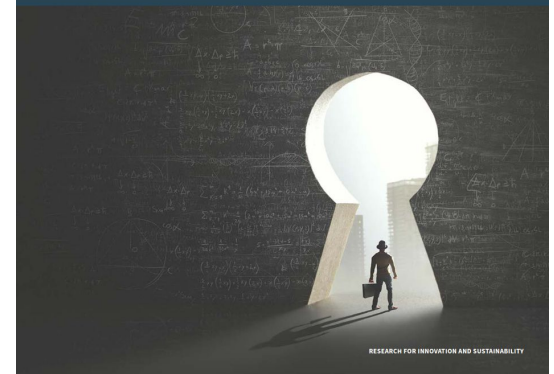


The Research
Council of Norway



The Research Council Policy
for Open Science

In effect from 2020



Plan S

Research may be made openly available in one of the three following ways:

1. Publication via open access journals/platforms
2. Making peer-reviewed research works immediately available in **open repositories**;
3. Publication in journals participating in collective agreements (e.g. “publish and read” agreements) aimed at transitioning from a subscription-based model to open access



Plan S

Making full & immediate
Open Access a reality

The University of Oslo's Open Access Policy

The following applies to peer-reviewed scientific articles:

1. Mandatory institutional archiving of all peer-reviewed articles in UiO's institutional repository

All members of staff employed by UiO after 4 July 2013 are obliged to deposit a post-print version of scientific articles produced in connection with the employment relationship into the institutional repository.

UiO also encourages all members of staff whose employment commenced prior to 4 July 2013 to follow this practice.

The "post-print" version means the final manuscript version following peer-review. When the publisher permits institutional archiving of the publisher's pdf, then this is the preferred choice. If publication takes place in a journal that does not allow institutional archiving, and the employee, after inquiring, is not granted such permission by the publisher, the person concerned is relieved of this requirement. If one or more co-authors use their right of refusal, the same applies.

2. Making scientific articles deposited into the institutional repository openly available

All members of staff employed by UiO after 4 July 2013 shall undertake to do their best to ensure that scientific articles deposited into the institutional repository can be made openly available as soon as possible.(1)

UiO also encourages all members of staff whose employment commenced prior to 4 July 2013 to comply with this practice of making scientific articles openly available.

3. Choosing where to publish

When choosing where to publish, UiO recommends that all employees select journals that allow the article to be openly available. These may either be Open Access journals or those that permit articles to be deposited and made openly available in an institutional repository.

(1) See Section 2.3.6. in UiO's IPR policy dated 6 December 2011.



UNIVERSITY
OF OSLO

Policy for Open Science at NTNU

The policy is approved by the NTNU Rector 2 October 2020, and is effective from this date. NTNU University Library manages the document.

6. Principles of open science

As a general rule, results of research and teaching carried out at NTNU must be made publicly available where this is practically and legally possible. In many cases, however, open access to research data or learning resources may not be appropriate. This may be because the research is of a sensitive nature, de-identification of personal information is not possible, or the research or teaching includes the use of material protected by copyright.

Researchers and teaching staff may also be bound by commercial agreements that make open access impossible. It is still important to follow good scientific practice for documentation of research results and data management, and more limited access to research results may be relevant in certain cases, such as replication studies.

As far as possible, NTNU will strive to ensure that the University or its employees retain copyright protection for their work in connection with entry into publishing contracts and licences for use, and that the author's moral rights are safeguarded.



NTNU

Kunnskap for en bedre verden

Employee Pages

Employment conditions ▾

Work support ▾

Service functions ▾

Competence ▾

Policy and strategy ▾

UiB > Employee Pages >

OPEN SCIENCE

The University of Bergen Policy for Open Science

Openness, transparency, and knowledge exchange are core values for the University of Bergen. Technological changes and increased digitalization have created new opportunities for research, education, innovation, and artistic research.

*adopted by the University Board
24.09.2020*

Open access to research publications and artistic research

Open access to research data

Open Innovation

Citizen Science

Open educational resources

UiB Policy for Open Science



<https://www.ub.uio.no/english/writing-publishing/open-access/documents/open-access-policy.html>

<https://www.ntnu.edu/policy-for-open-science>

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

Where should I deposit my data?

Preferably somewhere where most things are taken care of?



Where should I deposit my data?

1959 databases for molecular biology (2024)

Each journal publisher has its own policies and recommendations

fairsharing.org : 2123 registered database

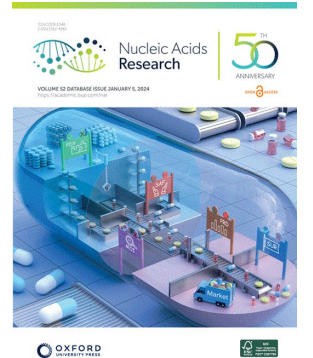
Community-developed platform (including standards, repositories and policies across all disciplines)

re3data.org : 3217 registered database

Registry of Research Data Repositories (overview of existing international repositories of research data from all academic disciplines).



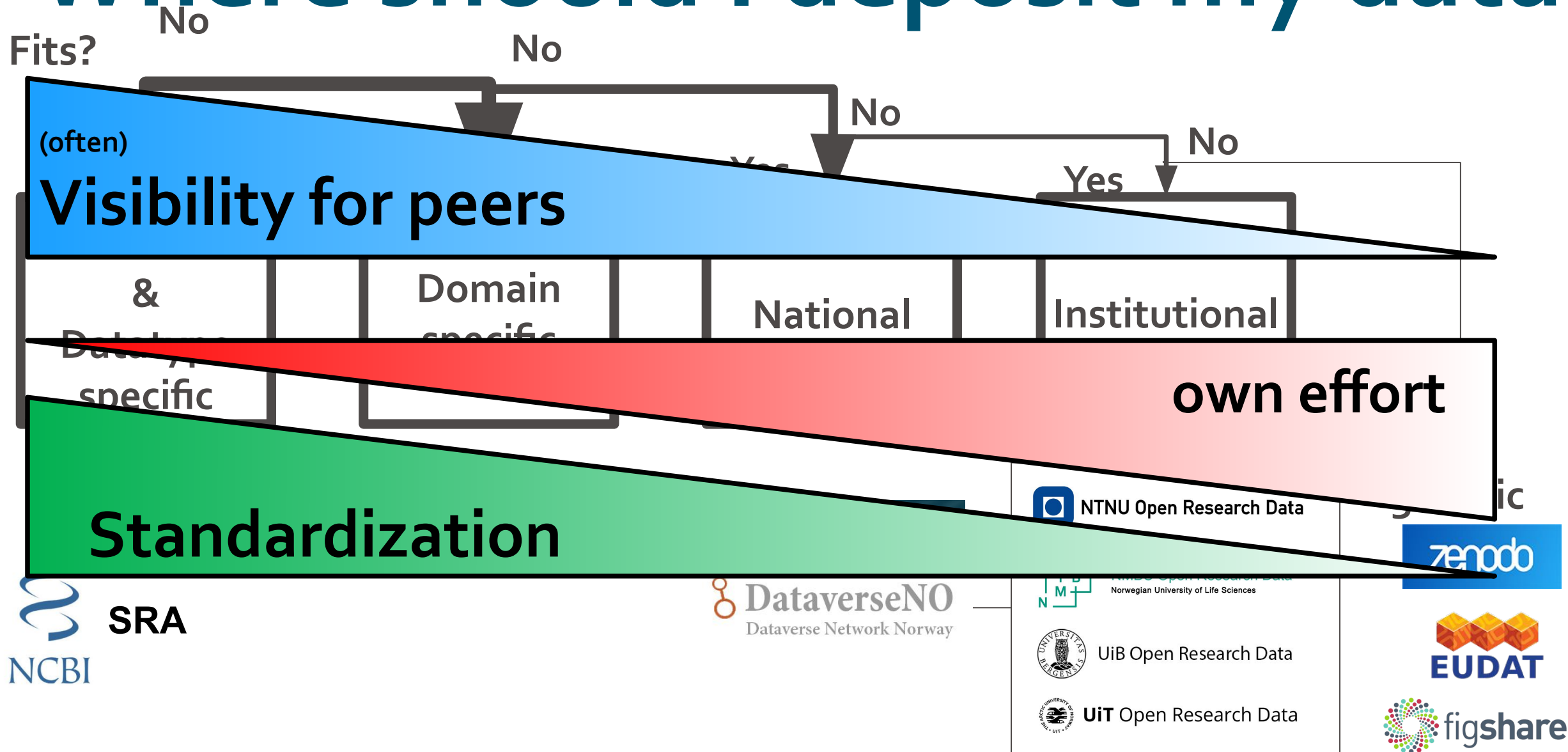
Nucleic Acids
Research



FAIRsharing

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Where should I deposit my data?



Some domain-specific archives

(and connected considerations)

Multiple Repositories – similar data type



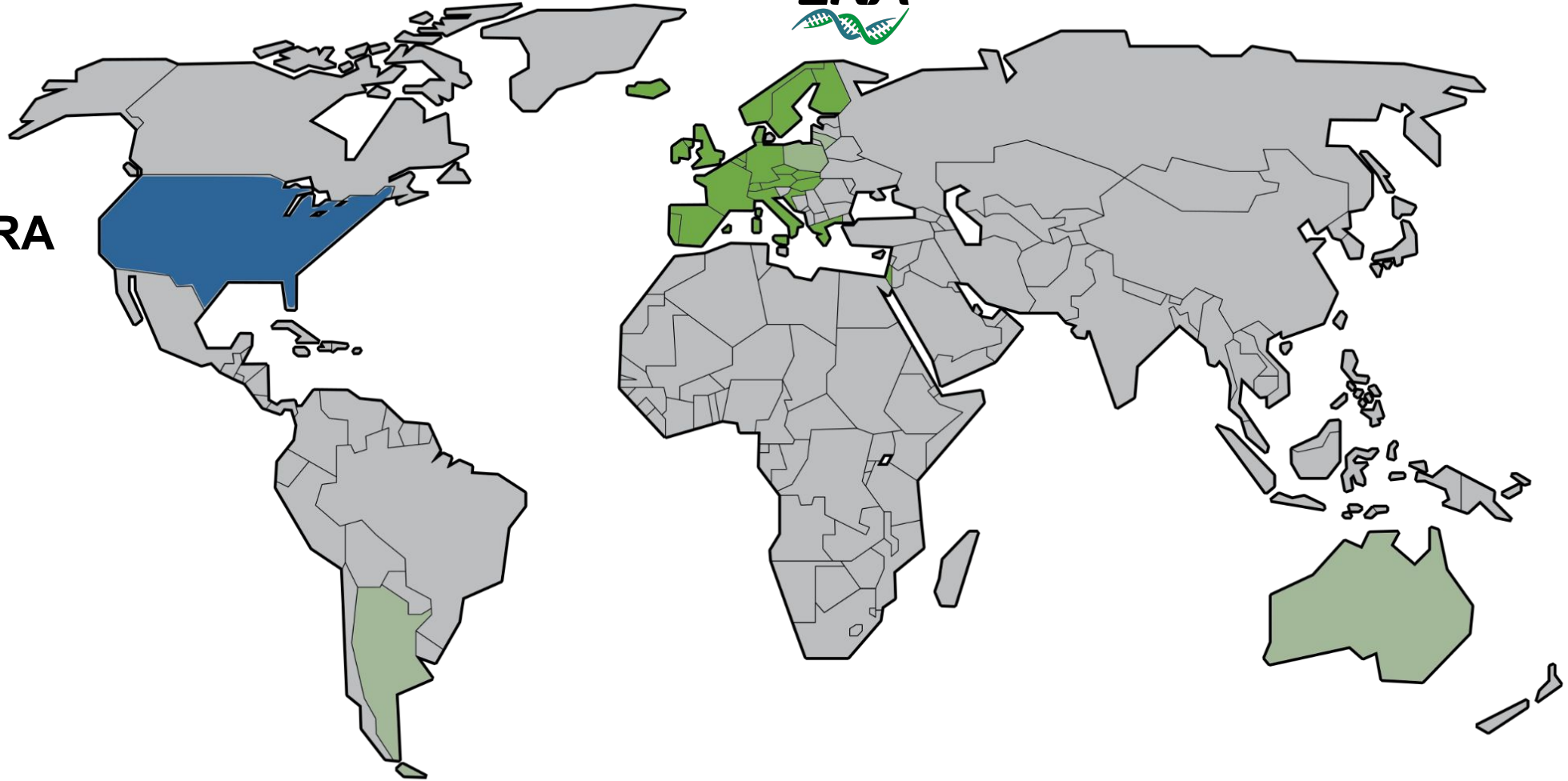


SRA





SRA



EMBL



2017



EMBL



ENA



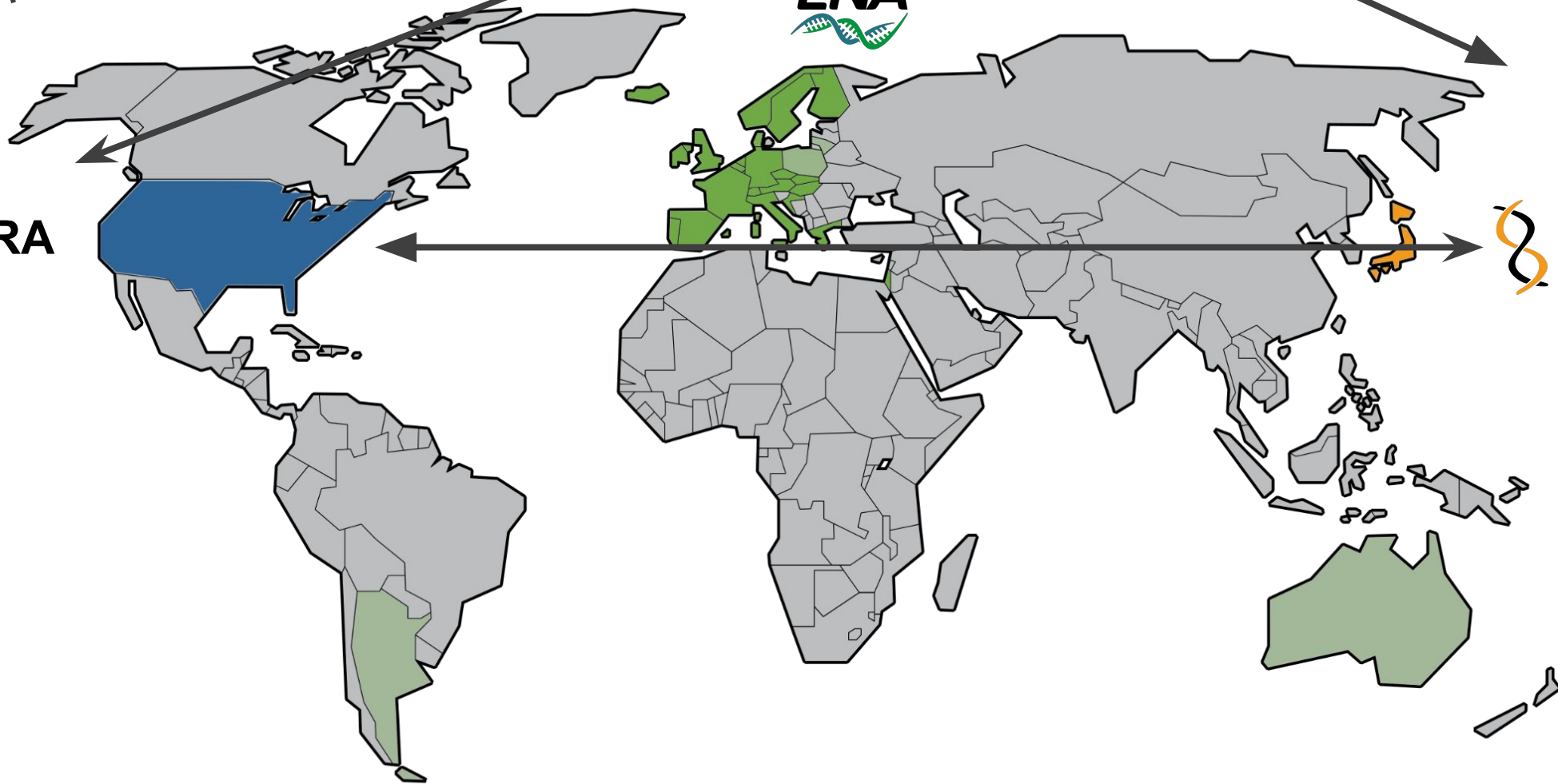
NCBI

SRA



DDBJ

DNA Data Bank of Japan





PRIDE

Proteomics IDentifications Database



MassIVE

Mass Spectrometry
Interactive Virtual Environment



Panorama



Elixir Data resource services

22 resources of fundamental importance to the wider life-science community and the long-term preservation of biological data

[ELIXIR data resource services](#)

elixir
Core Data
Resource

elixir
Deposition
Databases

12 deposition database for experimental biomolecular data

<https://elixir-europe.org/platforms/data/core-data-resources>

<https://elixir-europe.org/platforms/data/elixir-deposition-databases>

Elixir Deposition Databases



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)

<https://elixir-europe.org/platforms/data/core-data-resources>
<https://elixir-europe.org/platforms/data/elixir-deposition-databases>

Elixir Deposition Databases



Data submission

Use this data submission wizard to find the right archive for your data in a few simple steps.

1 What **type of data** do you have?

DNA/RNA sequence

Expression data

Protein data

Structures

Systems

Chemical biology

Ontologies

Images

Multi-omics or other cross-domain study

Other biological research data

EMBL-EBI



2 Are your data from human research subjects and do the data **require controlled access**?

Yes

No

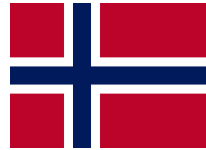
You can submit your data to the following database:



European Genome-Phenome Archive

<https://www.ebi.ac.uk/submission/>

National repositories



NIRD archive

NIRD RESEARCH DATA ARCHIVE



National, free depositing repository

Domain agnostic

Dublin-core metadata standard

DOI accessible

Supports machine readable metadata harvesting

<https://archive.norstore.no>

DataverseNO



National/Institutional, free depositing repositories

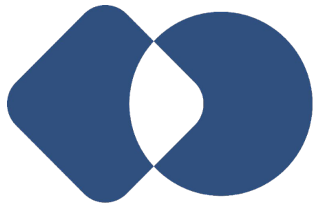
Domain agnostic

Manually curated by libraries

DOI accessible

8Gb max/upload - contact support >50Gb

Sikt archive



Sikt
Kunnskapssektorens
tjenesteleverandør



Domain agnostic (focus social science)

DOI accessible

Implemented data access committees

Main data type: Questionnaires, ...

Reorganisation: from 2022 NSD has merged with Uninett and Unit to form [Sikt](https://sikt.no/en/archiving-research-data)
– Norwegian Agency for Shared Services in Education and Research

<https://sikt.no/en/archiving-research-data>

Other types of data

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 code/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>

What about sensitive data?



Federated
European
Genome-phenome
Archive
Norway

ega-archive.org
ega.elixir.no

Central metadata accessibility

Secure storage

Implemented data access committees

[Norwegian EGA](#)

(for data that can not leave the country)



What about sensitive data?



Federated
European
Genome-phenome
Archive
Norwegian

**Ensure your consent form & REC
approval allow controlled access
archiving!**

Central metadata accessibility

Implemented data access committees

ega-archive.org
ega.elixir.no

[Norwegian EGA](#)

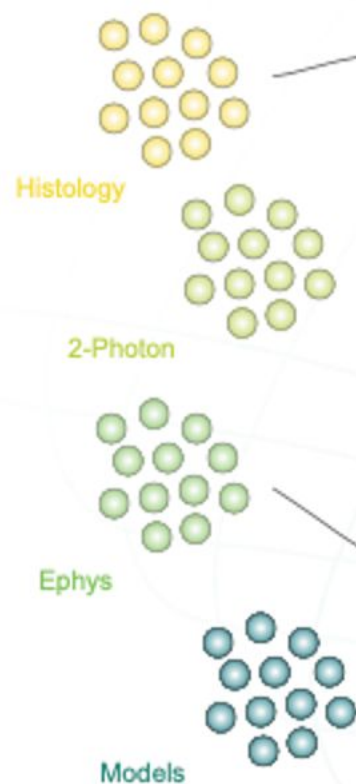
(for data that can not leave the country)





EBRAINS - a FAIR place to share your neuroscience data

All types of neuroscience data are welcome



Data Curation and support



Integration of data into the EBRAINS Knowledge Graph Search



Visit our website:

<https://ebrains.eu/services/data-knowledge/share-data/>

Or contact us directly at:

curation-support@ebrains.eu

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

What to consider before submitting to a long-term repository?

Outlined in the DMP:

- Can your data be submitted to an open repository?
- Who is the **long-term** contact person for the submitted dataset?

Familiarize yourself with the repository standards

- What data, metadata or other information does the repository require me to submit?
- Are my data compatible with the repository format standards?

When should I know which repository I should

submit to? Preferably already before the project starts

