



# Archiving, preserving, sharing

Thu-Hien TO  
ELIXIR NMBU



This work is adapted from the work of Korbinian Bösl,  
and is licensed under <https://creativecommons.org/licenses/by/4.0/>



# NeLS

Norwegian e-Infrastructure for Life Sciences

 **nor seq**

  
**SEEK**

 **Galaxy**

 **BioStudies.**

 **ENA**  
European Nucleotide Archive

 **IDR**

 **ArrayExpress**

 **MetaboLights**

 **PRIDE Archive**  
Proteomics IDentifications Database

 **JWS ONLINE**

 **EUROPEAN  
GENOME-PHENOME  
ARCHIVE**



sensitive  
data



## 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



Link to RDMkit: <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
- Safe money on storage

# Where should I deposit my data?

Preferably somewhere where most things are taken care of?



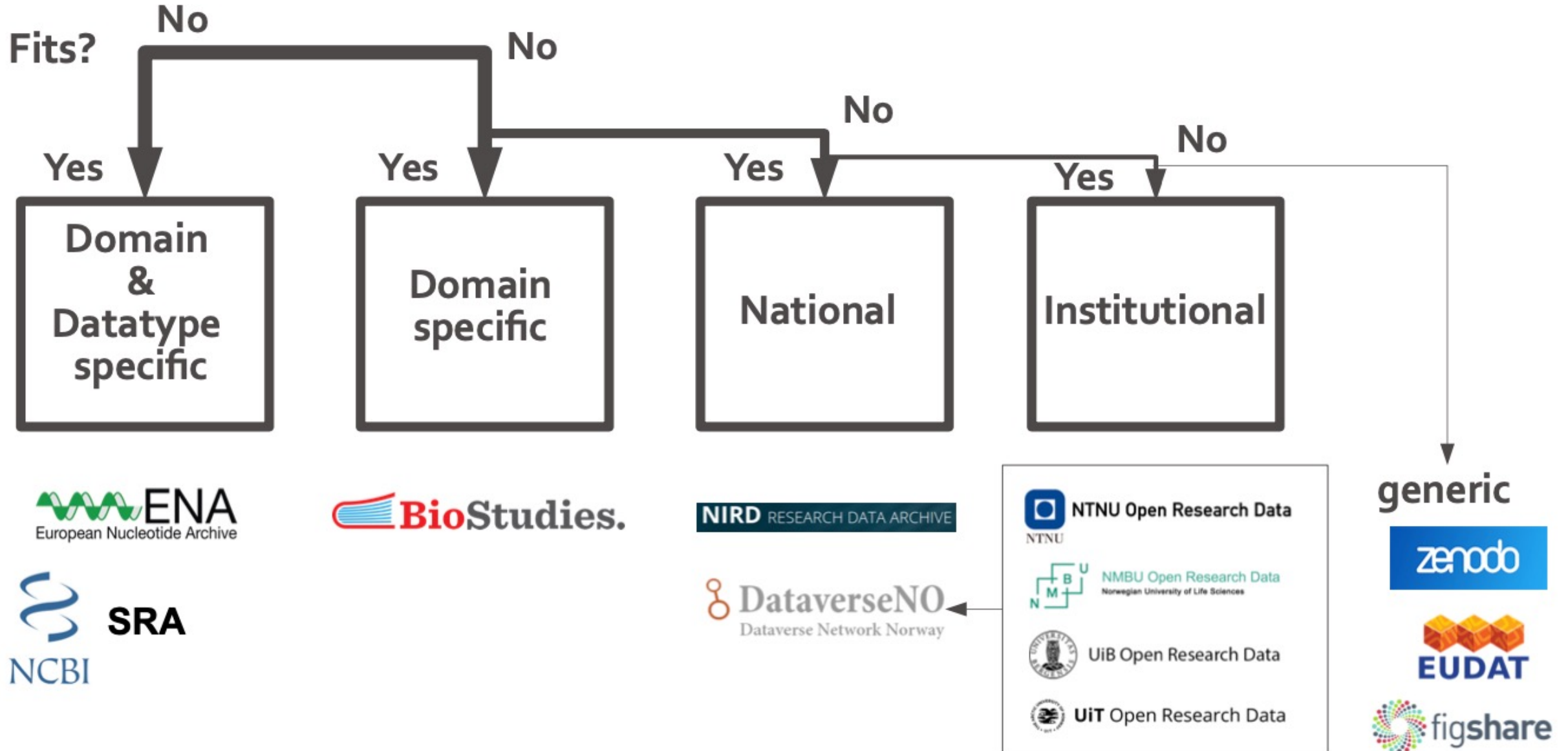
# Where should I deposit my data?

- 1641 databases for molecular biology
- Each journal publisher has its own policies and recommendations
- [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).

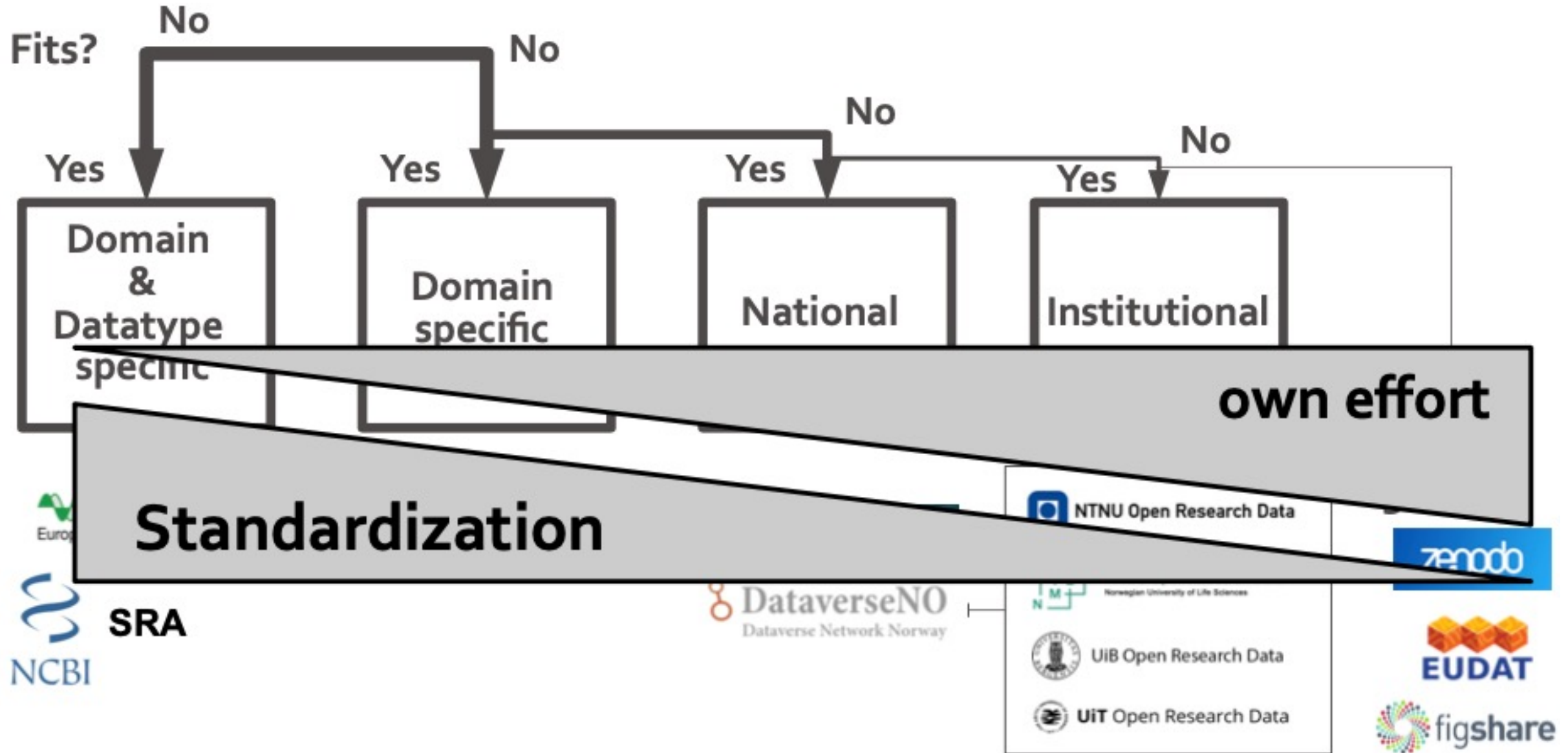




# Where should I deposit my data?

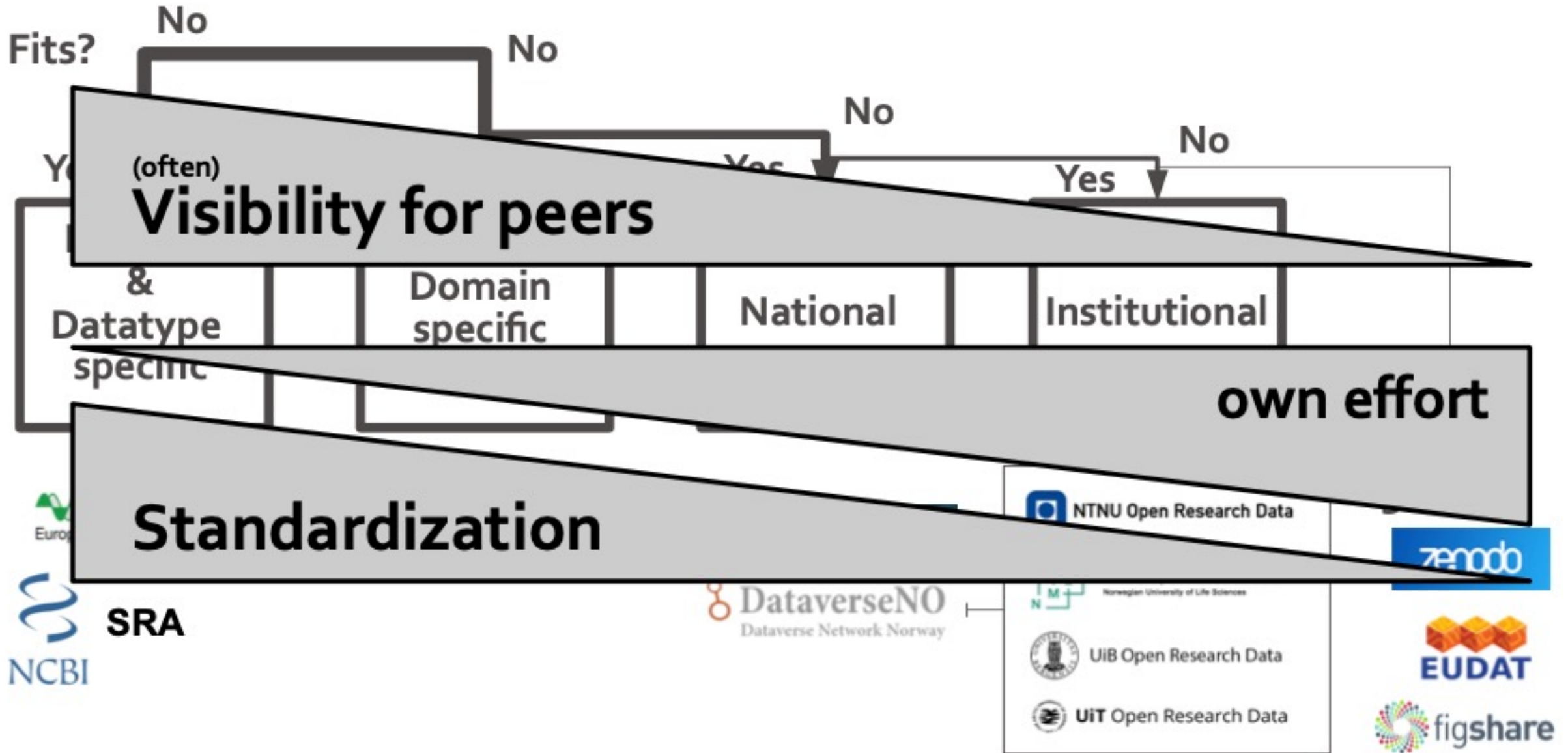


# Where should I deposit my data?





# Where should I deposit my data?

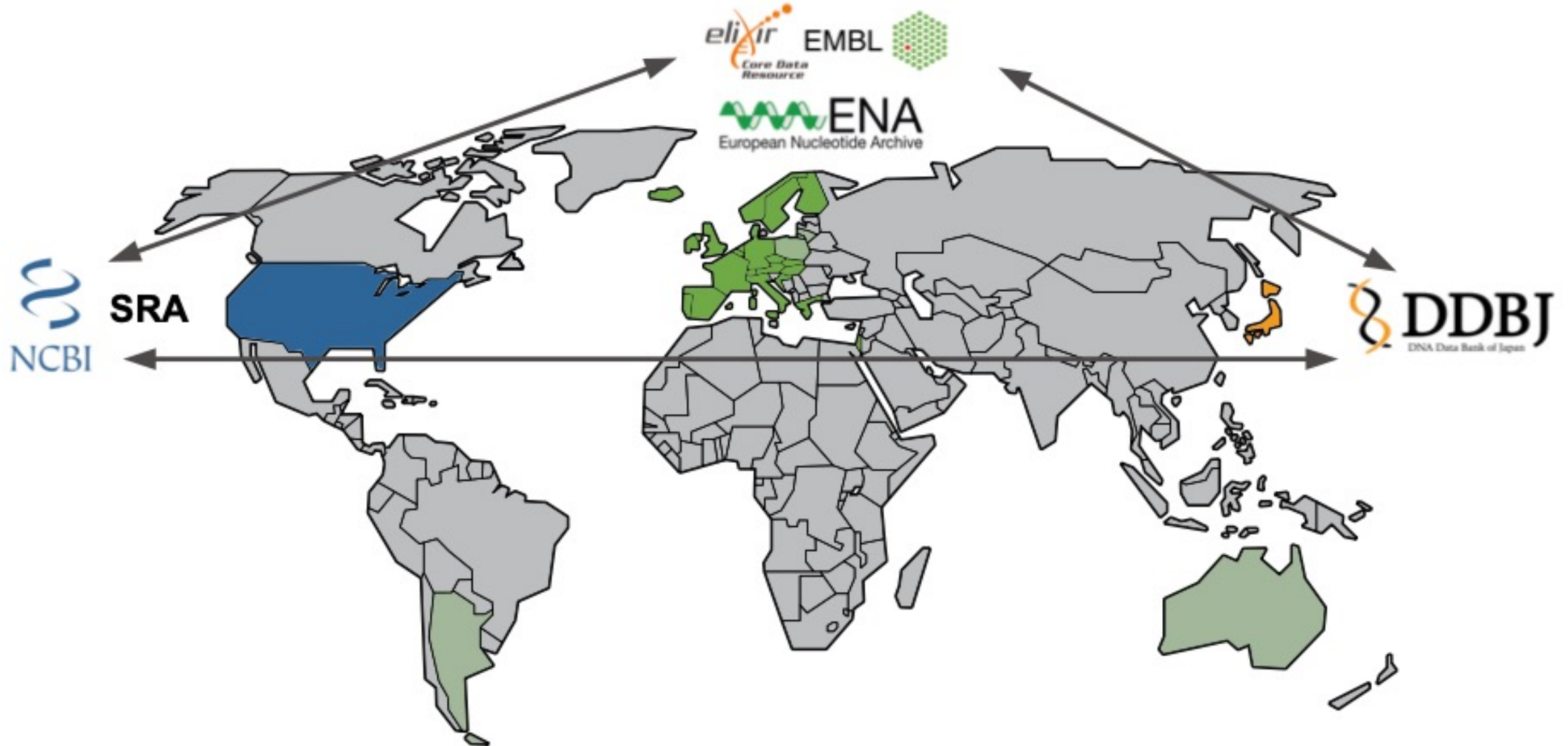




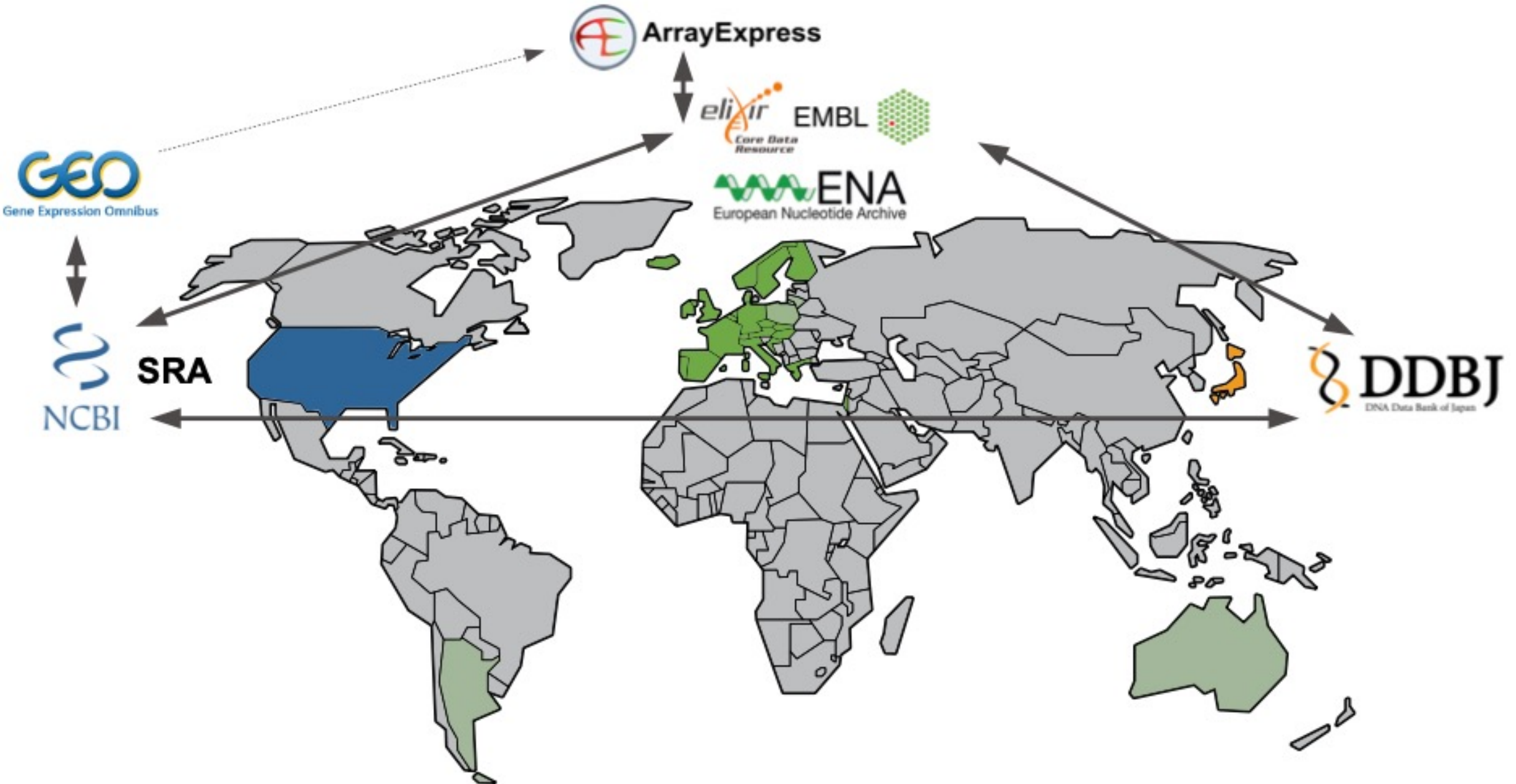
# Some domain specific archives

---

# Sequencing data: Multiple Repositories – similar datatypes

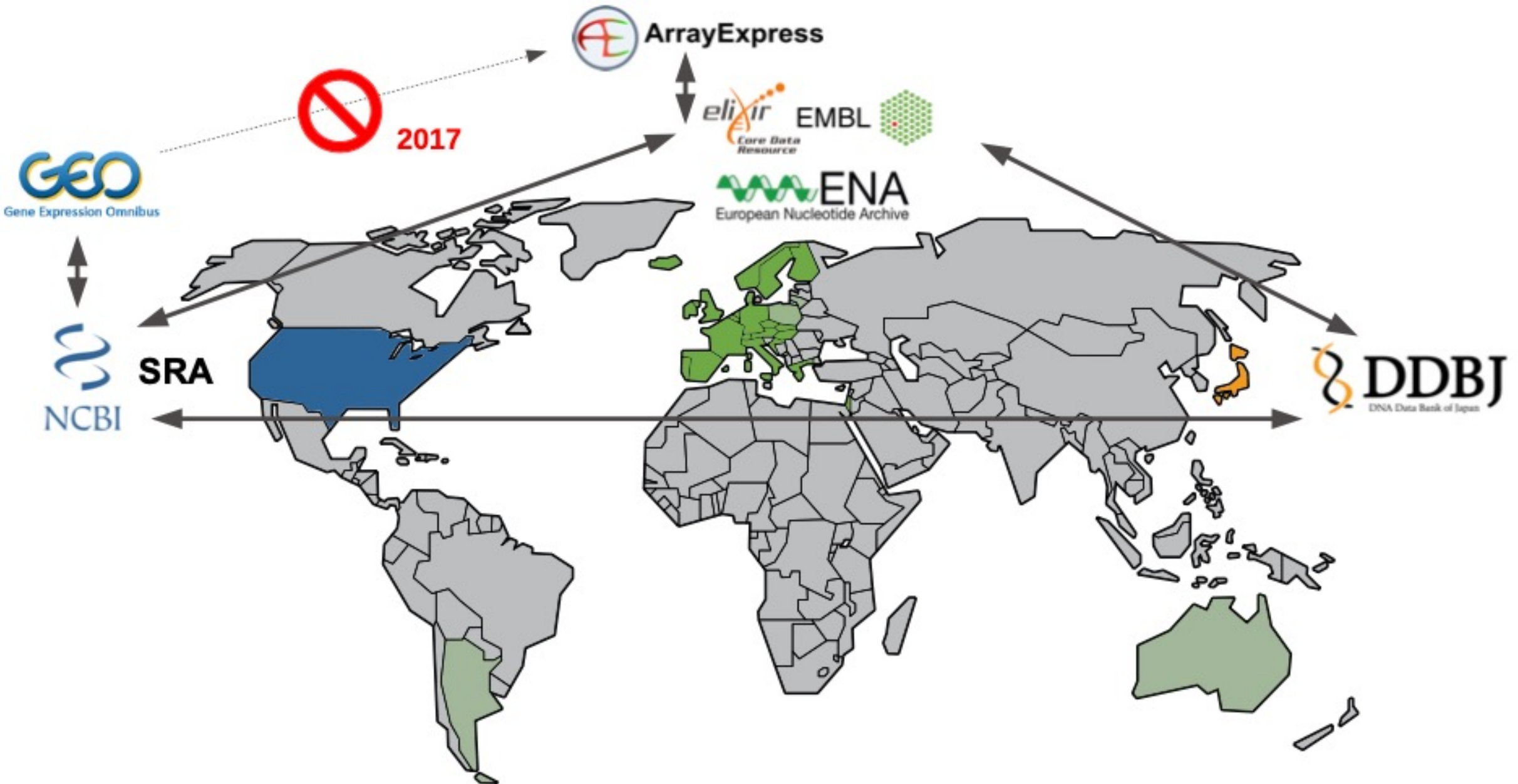


# Sequencing data: Multiple Repositories – similar datatypes

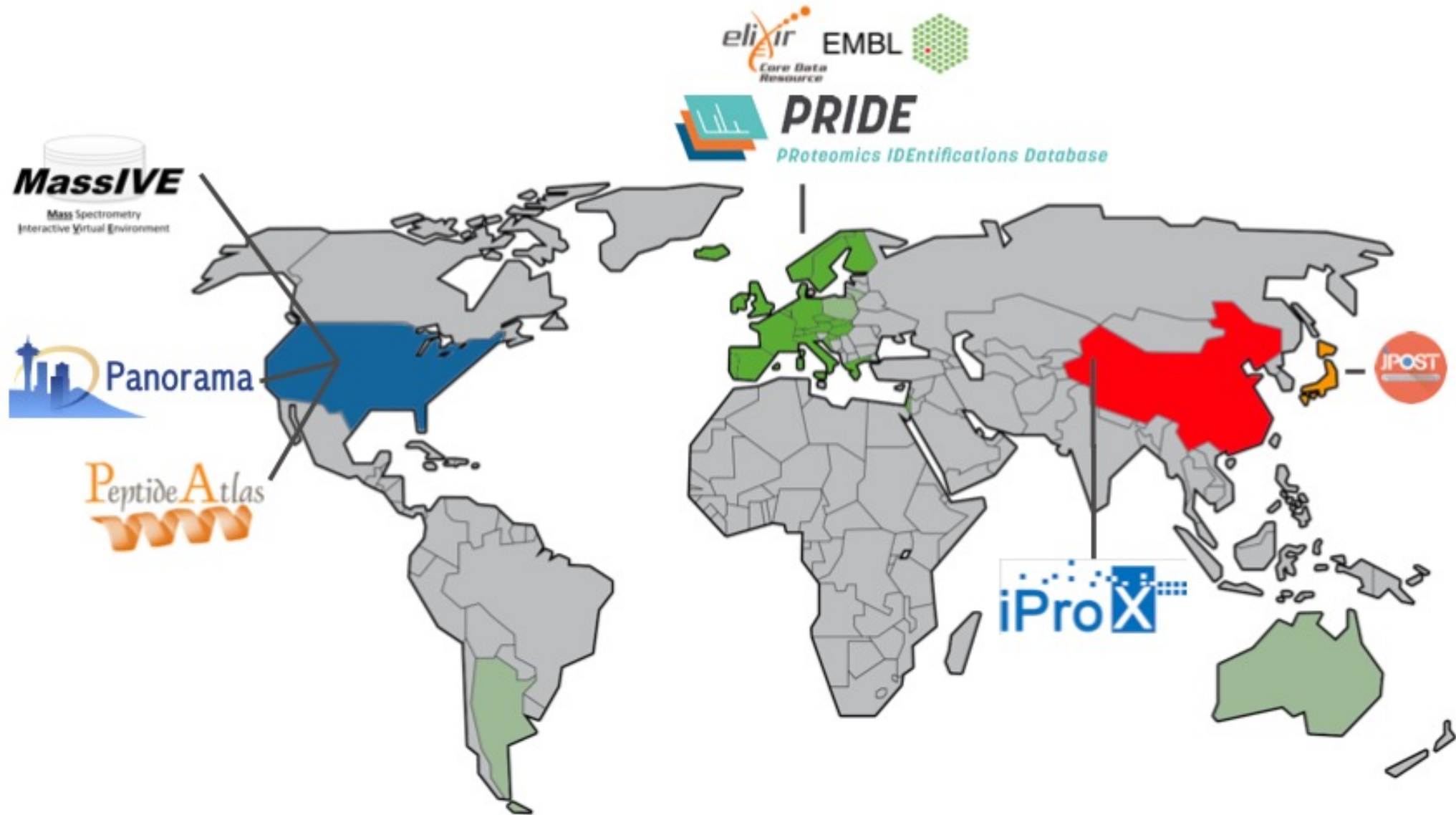




# Sequencing data: Multiple Repositories – similar datatypes

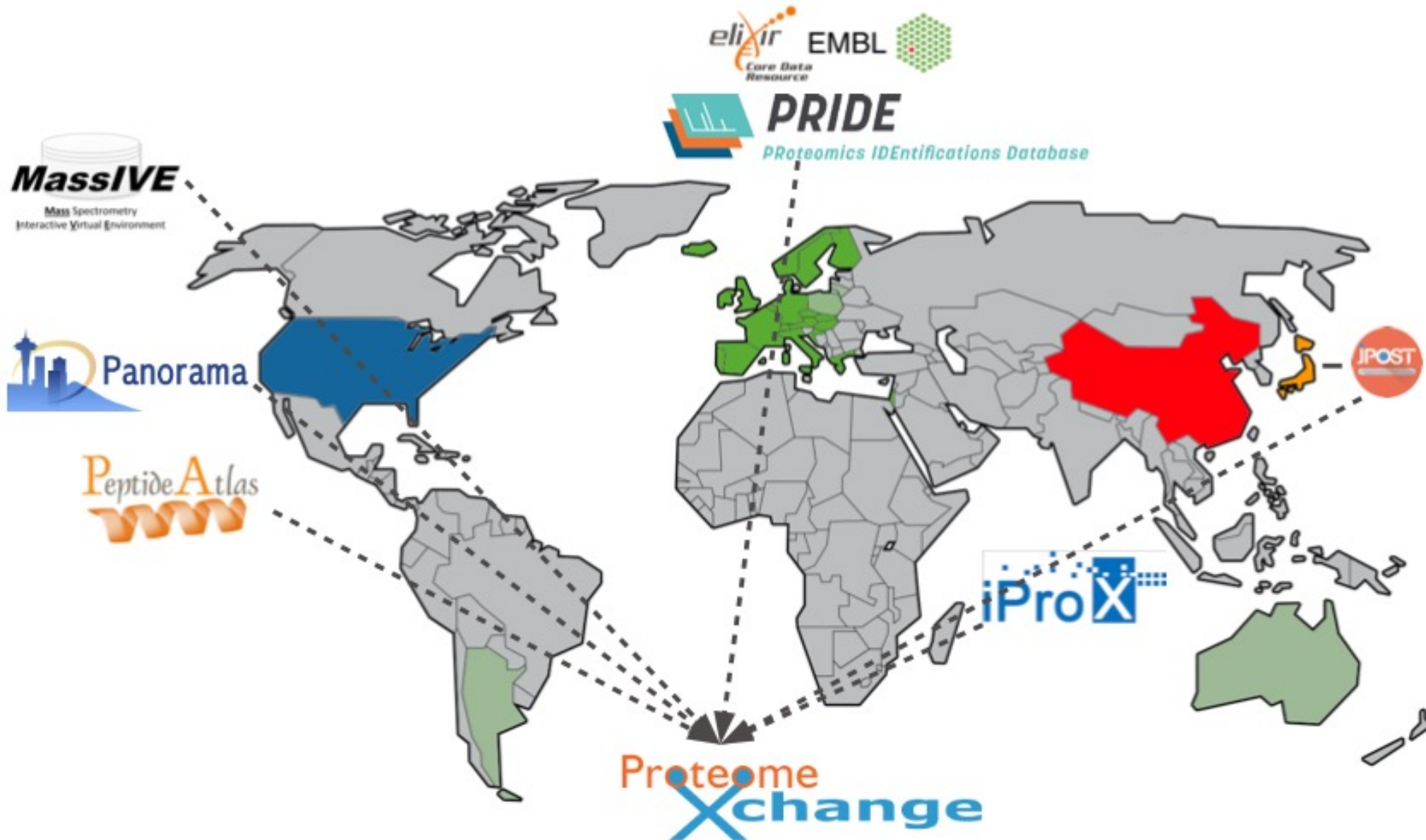


# Proteomics data: Multiple Repositories – similar datatypes

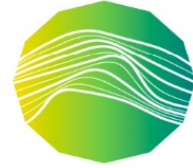




# Proteomics data: Multiple Repositories – similar datatypes



# Neuroscience data: Multiple organisations association



EBRAINS



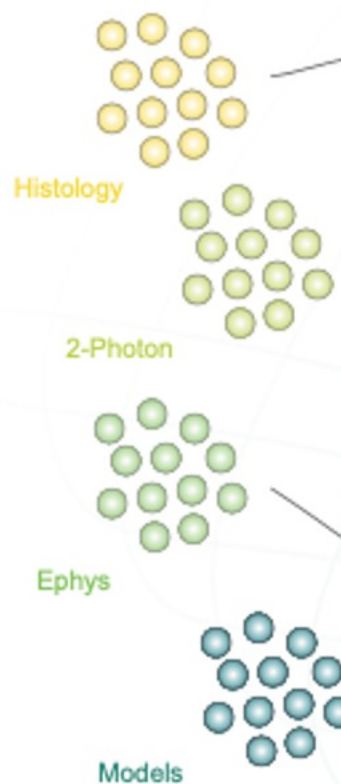
A shared digital brain research infrastructure for the EU



# 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](mailto:curation-support@ebrains.eu)

Data and metadata management and storage

Data curation and integration

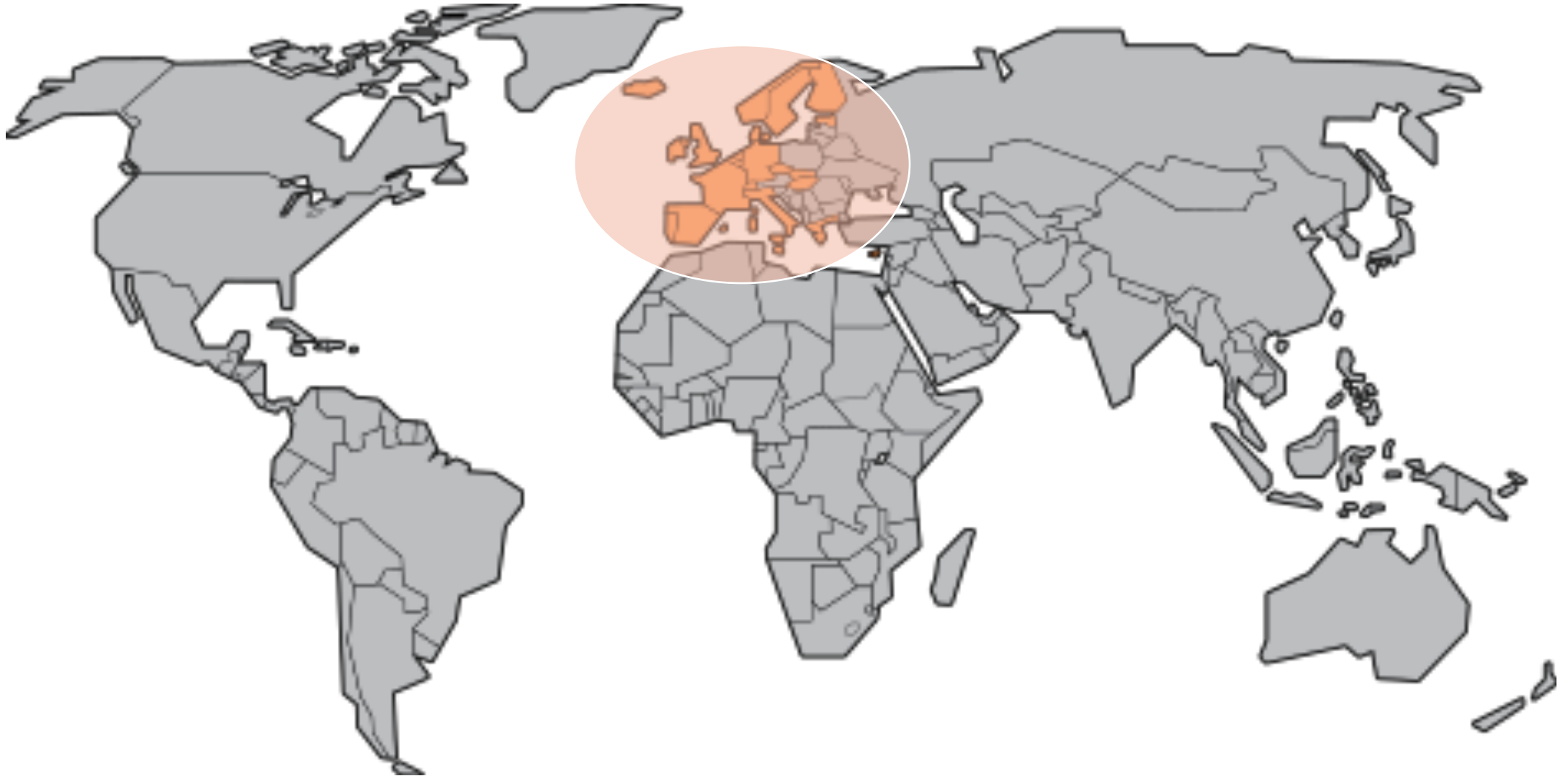
Data security

Collaboration and sharing of data

Copyright, licensing and citation

Long-term storage of data

**Elixir Data resource services:** delivered by different Elixir nodes



# 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



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 planned)

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

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



# Elixir Deposition Databases



# Elixir Deposition Databases

EMBL-EBI



Arra

**Data submission wizard** which helps to find the right archive for your data through a series of questions

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



EUROPEAN  
GENOME-PHENOMI  
ARCHIVE

European Variation Archive

IntAct



MetaboLights

PDBe  
Protein Data Bank in Europe



**PRIDE**

PROteomics IDentifications Database



# 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**

# NSD archive



NSD NORSK SENTER FOR FORSKNINGSDATA



Domain agnostic

DOI accessible

Implemented data access committees

Main data type: Questionnaires, ...

Reorganisation: from 2022 NSD has merged with Uninett and Unit to form [Sikt](#)

– Norwegian Agency for Shared Services in Education and Research

<https://arkivering.nsd.no>





## 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?

Central metadata accessibility

Secure storage

Implemented data access committees

[Norwegian EGA](#)

(for data that can not leave the country)



[ega-archive.org](http://ega-archive.org)



# What about sensitive data?

Central metadata accessibility



EUROPEAN  
GENOMIC  
ARCHIVE

[ega-archive.org](http://ega-archive.org)

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

Secure storage  
Implemented data access committees

[Norwegian EGA](#)

(for data that can not leave the country)



