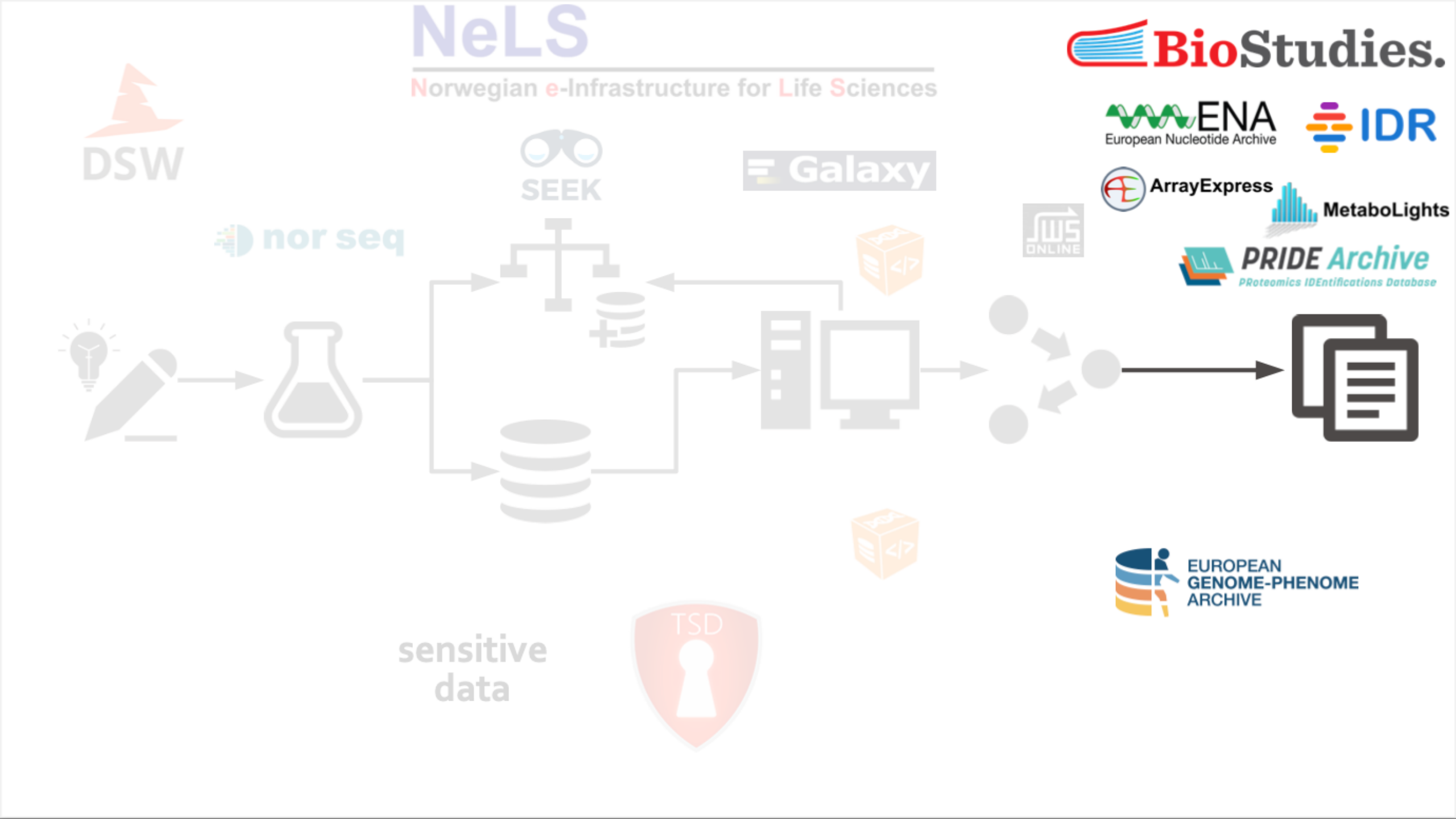
Archiving, preserving, sharing

Illimar Rekand,
Illimar.rekand@uib.no
ELIXIR Norway, UiB

CENTRE FOR DIGITAL LIFE NORWAY

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

Illustrative Icons from Font Awesome licensed under Creative Commons Attribution 4.0 International License


NeLS
Norwegian e-Infrastructure for Life Sciences
DSW
SEEK
Galaxy
nor seq
JWS ONLINE
BioStudies.
ENA
European Nucleotide Archive
IDR
ArrayExpress
MetaboLights
PRIDE Archive
PRoteomics IDEntifications Database
EUROPEAN
GENOME-PHENOME
ARCHIVE
sensitive
data
TSD

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

RDMkit

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
- Save money on storage

http://doi.org/10.1371/journal.pone.0230416

Where should I deposit my data?

Preferably somewhere where most things are taken care of?

FAIR illustrations icons from ARDC 2018 licensed under Creative Commons Attribution 4.0 International License

Where should I deposit my data?

- 1641 databases for molecular biology
- Each journal publisher has its own policies and recommendations
- fairsharing.org : 1863 registered database

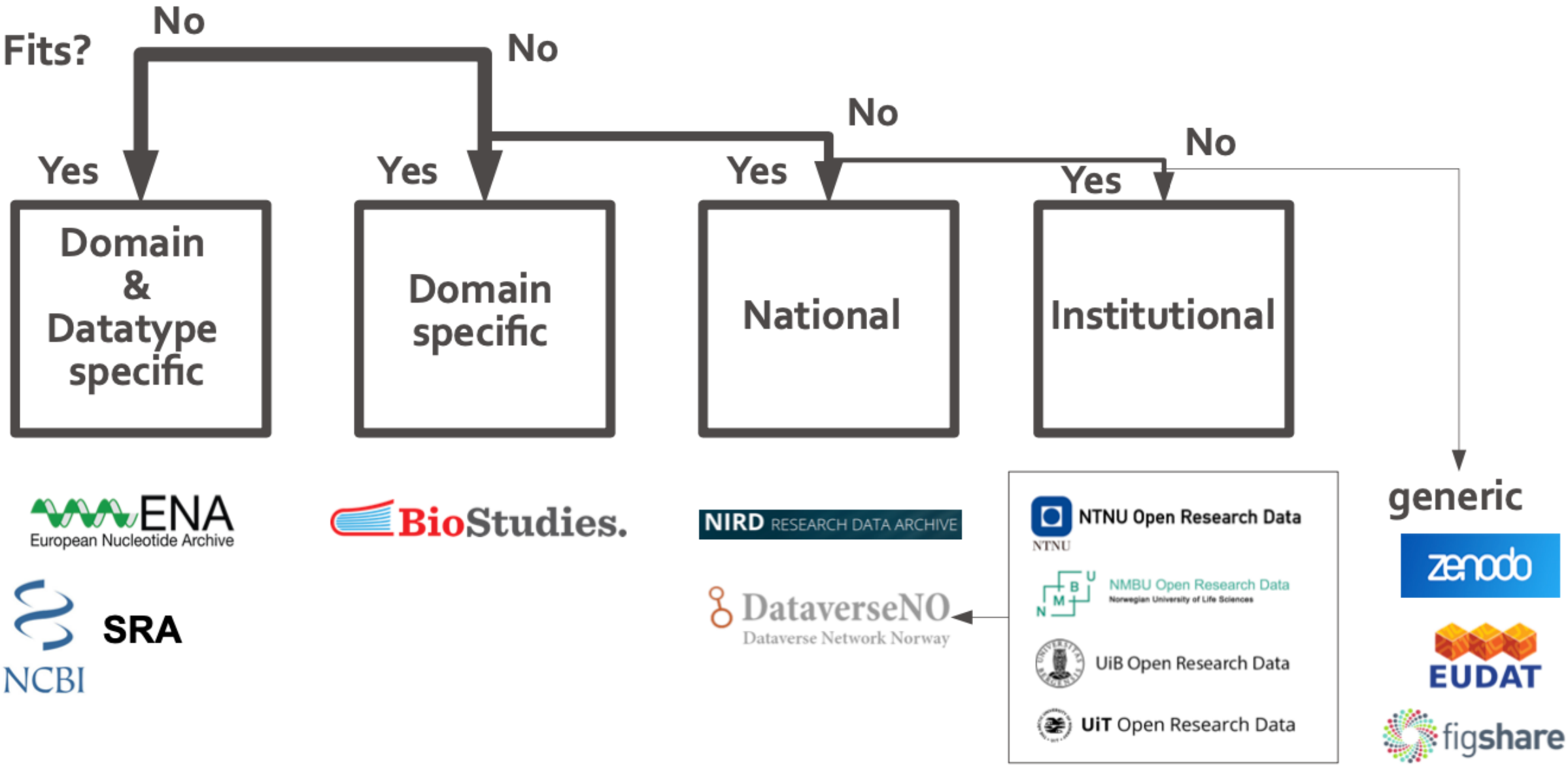
Community-developed platform (including standards, repositories and policies across all disciplines)

- re3data.org : 2834 registered database

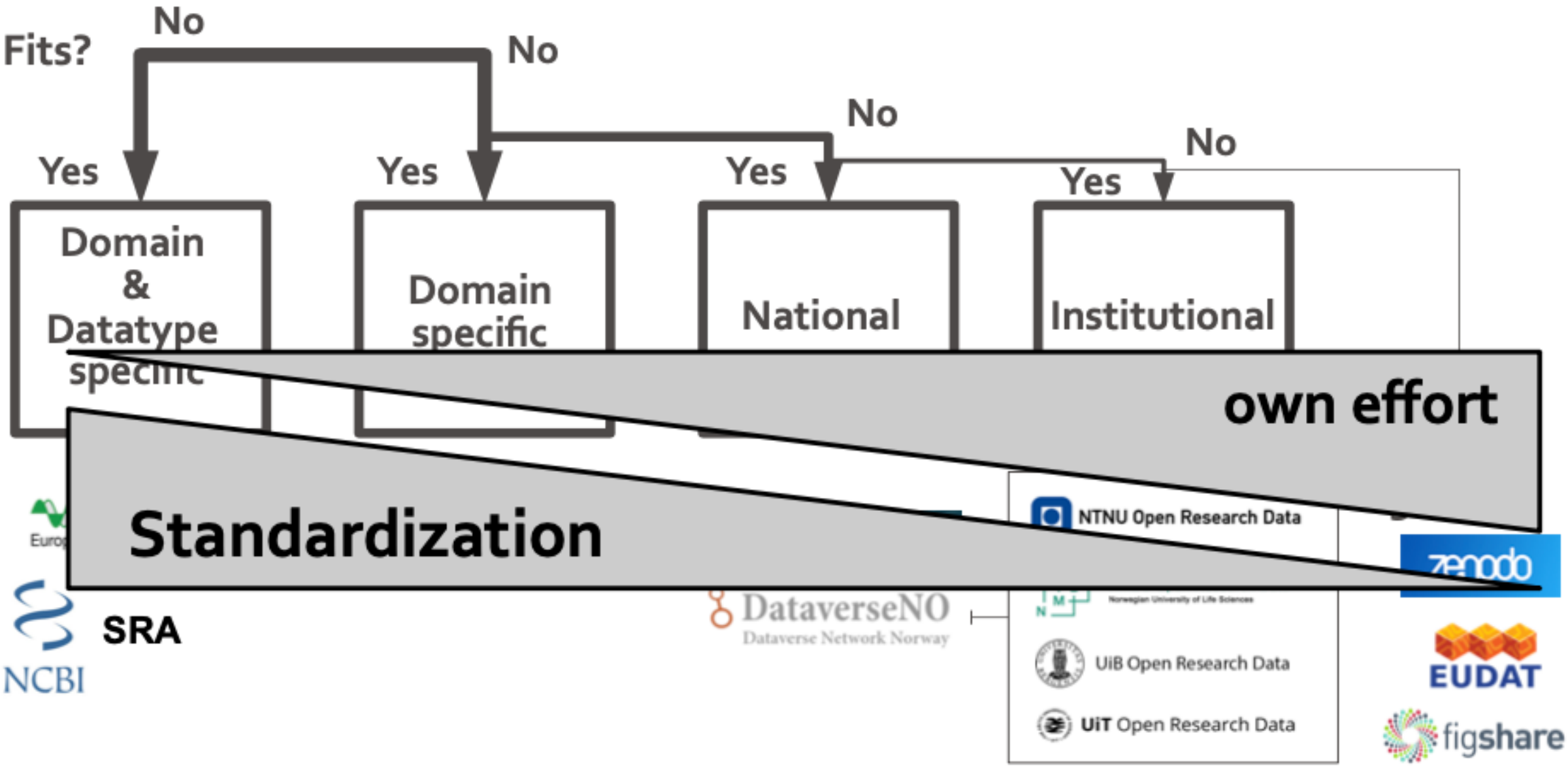
Registry of Research Data Repositories (overview of existing international repositories of research data from all academic disciplines).

Rigden D.J. & Fernández X.M., The 2021 Nucleic Acids Research database issue and the online molecular biology database collection. Nucleic Acids Research 49, D1-D9 (2021)

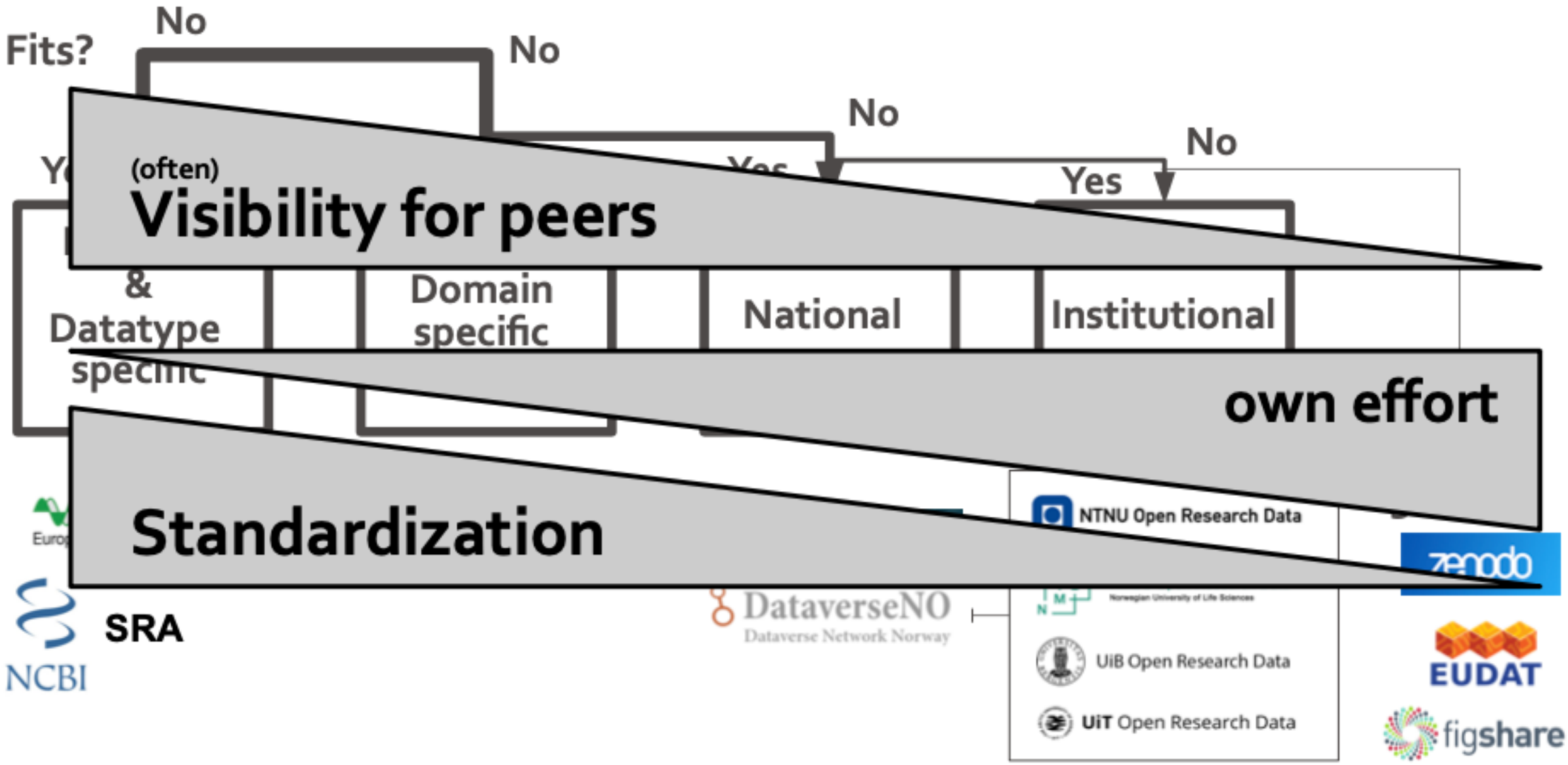
Where should I deposit my data?



Where should I deposit my data?

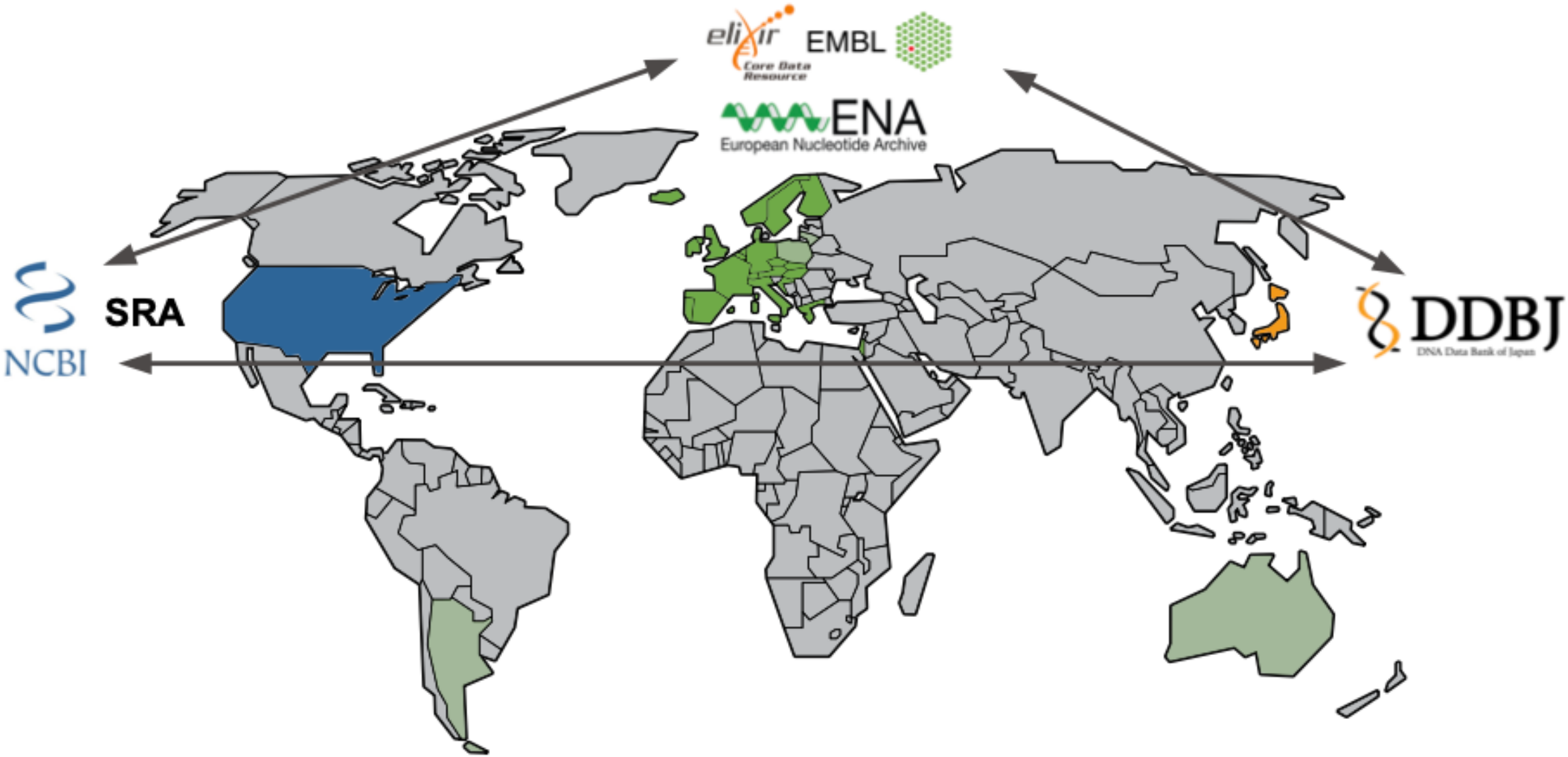


Where should I deposit my data?

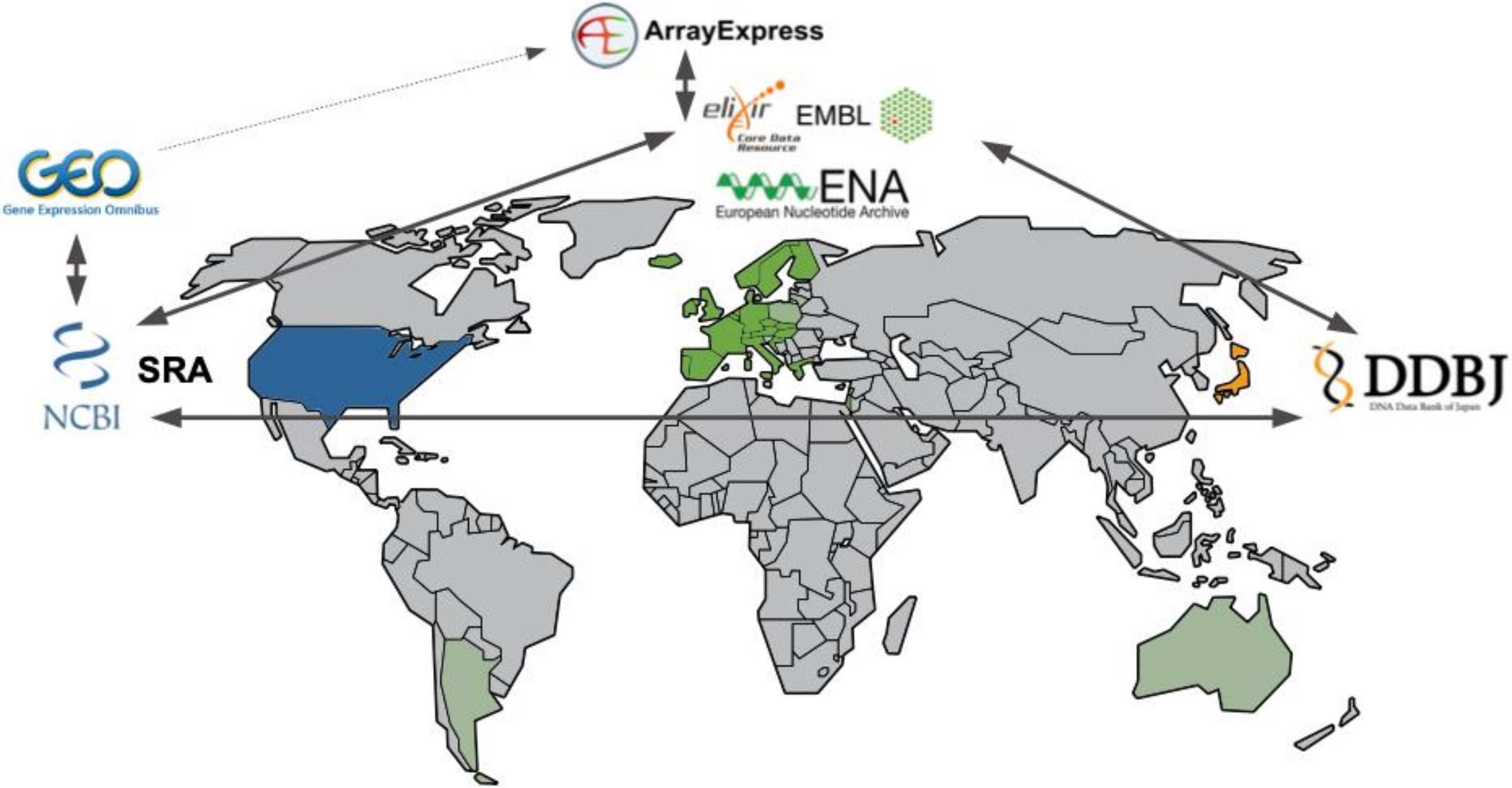


Some domain specific archives

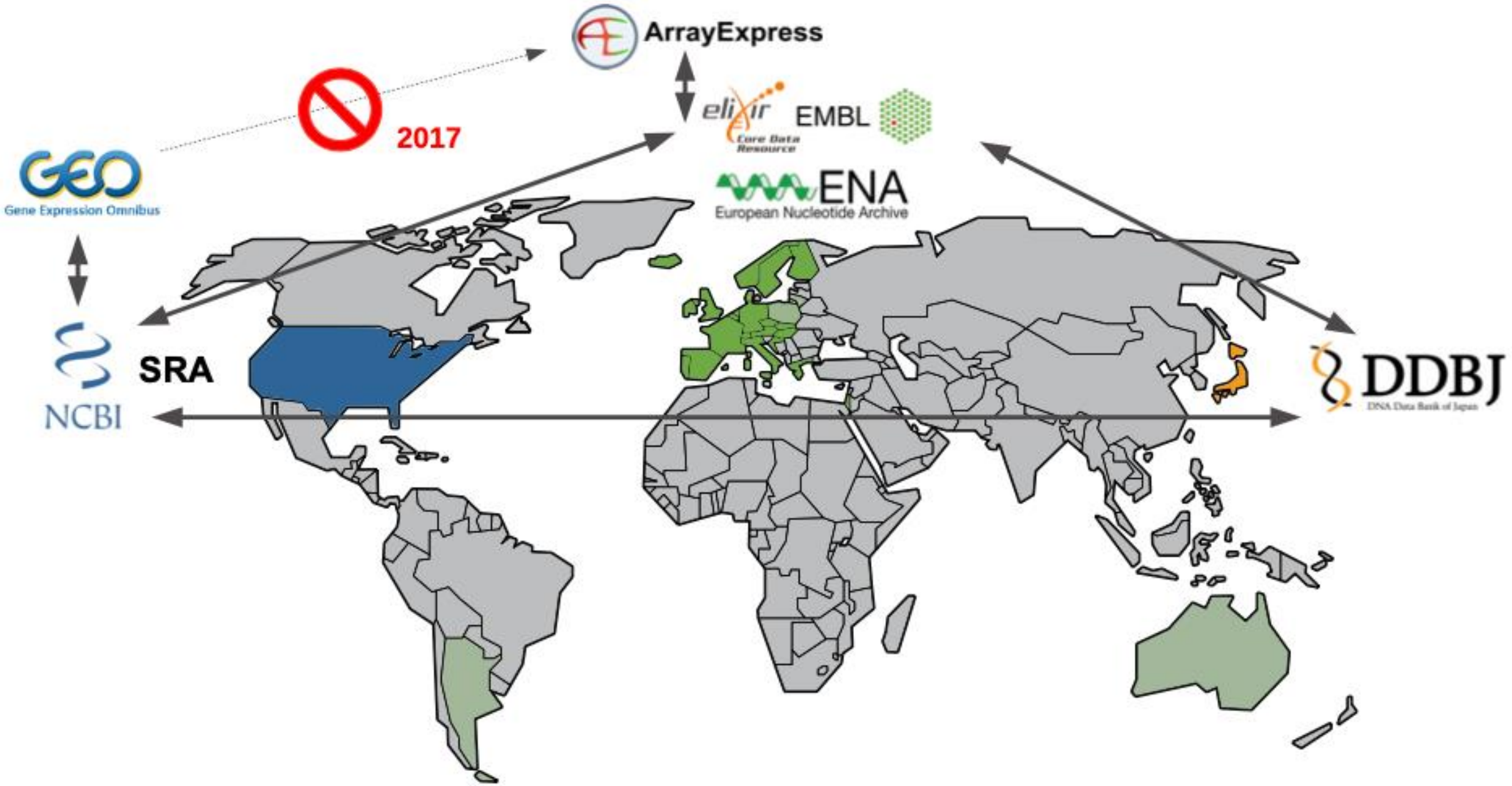
Sequencing data: Multiple Repositories – similar datatypes



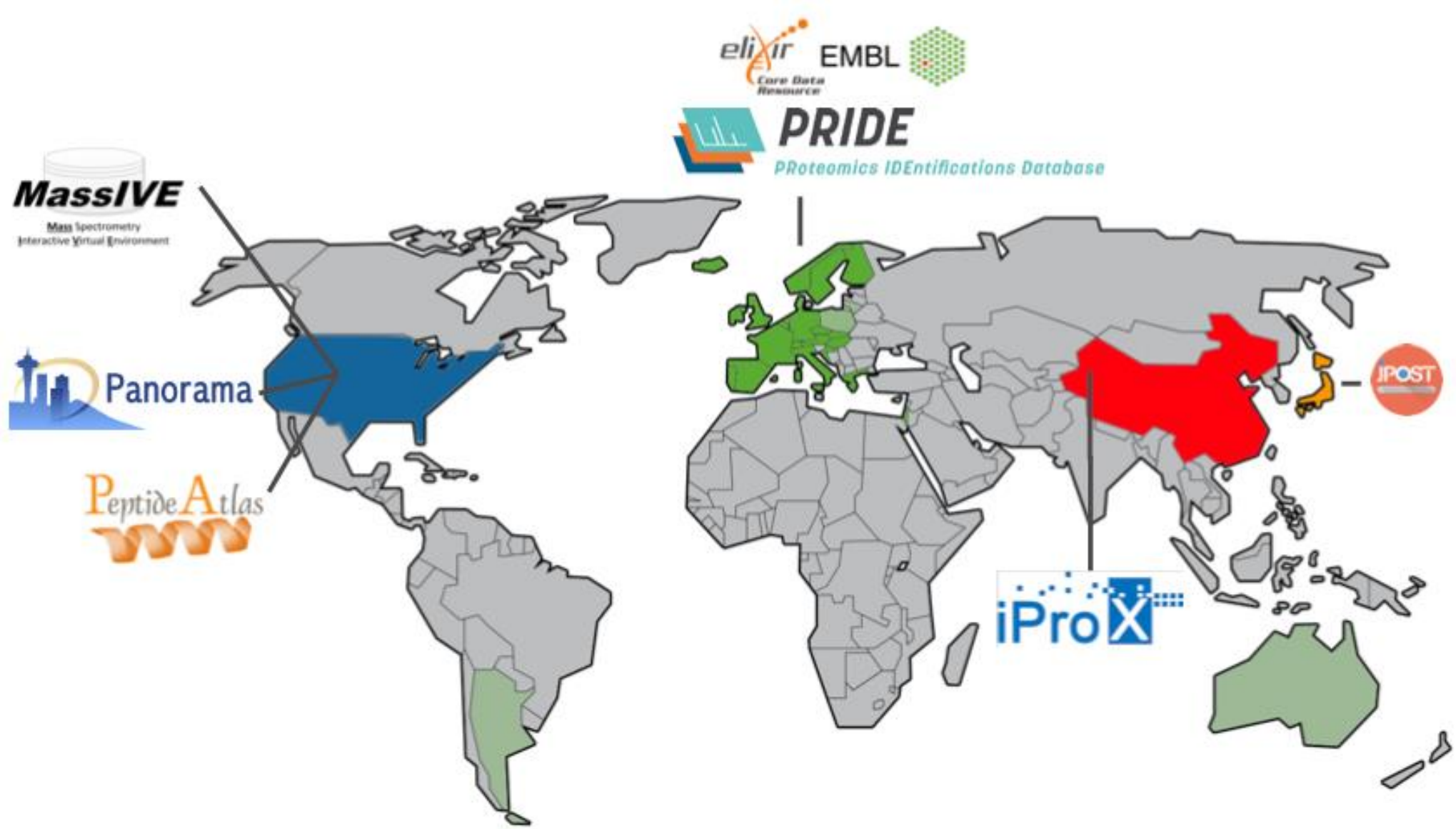
Sequencing data: Multiple Repositories – similar datatypes



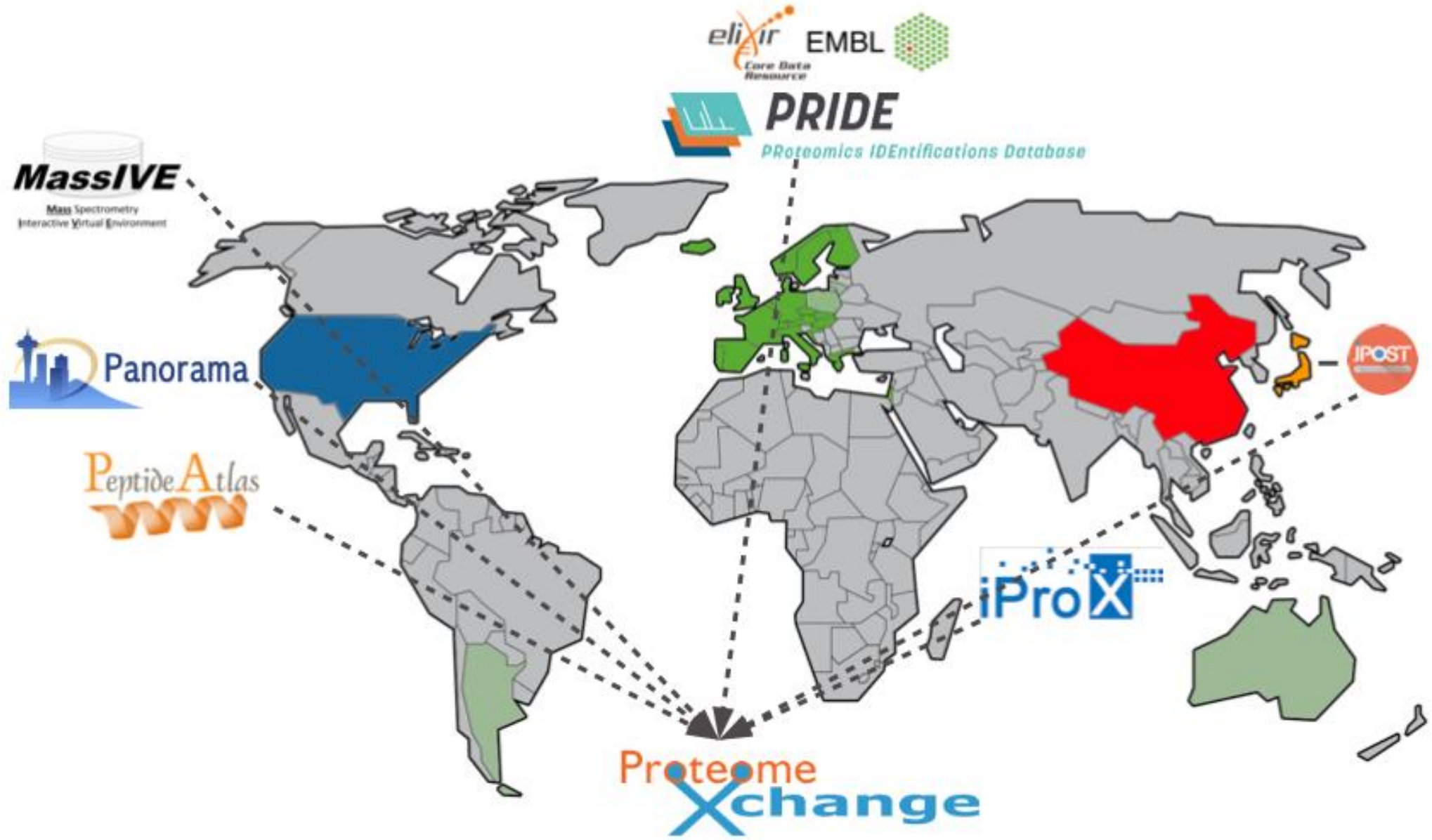
Sequencing data: Multiple Repositories – similar datatypes



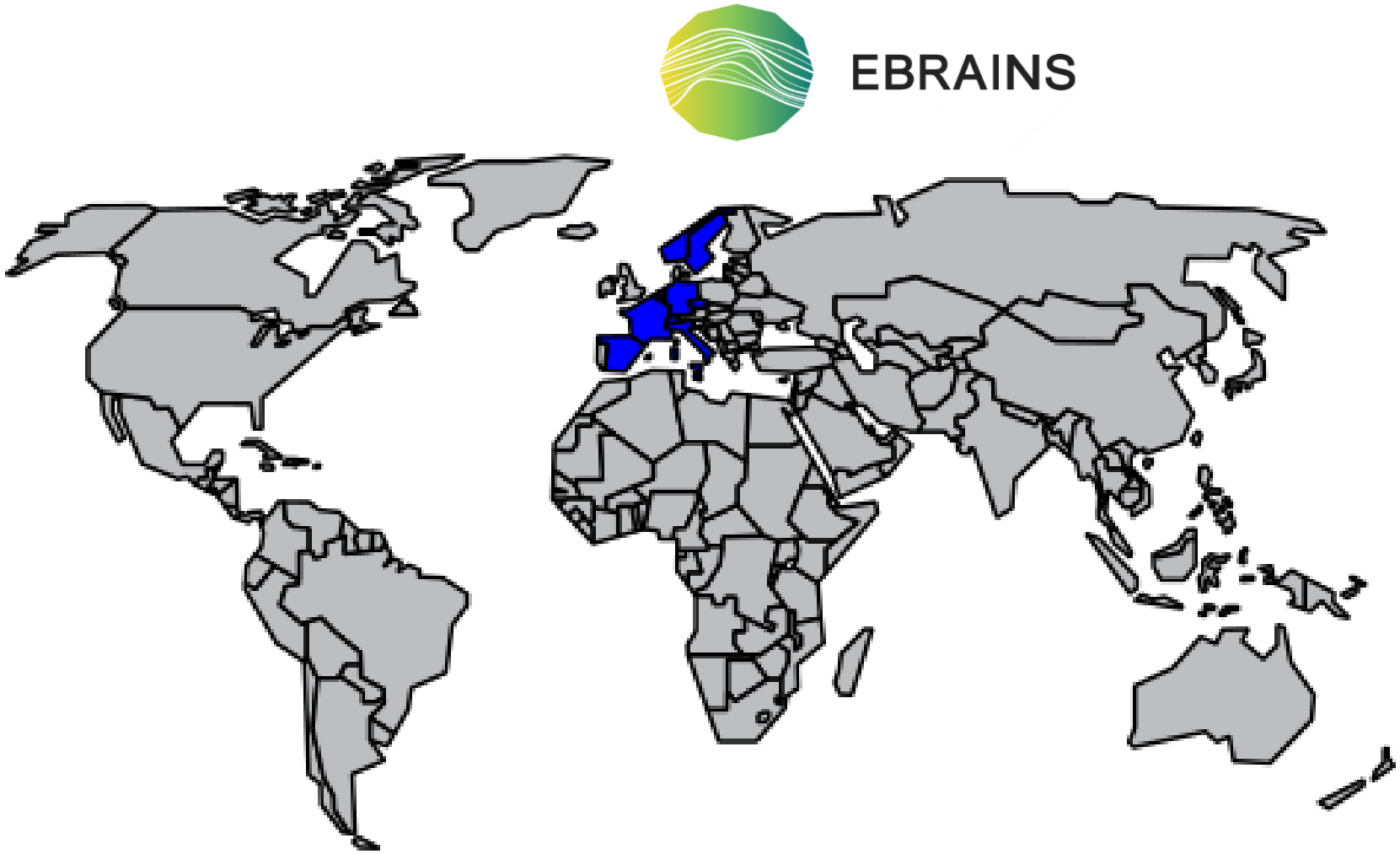
Proteomics data: Multiple Repositories – similar datatypes



Proteomics data: Multiple Repositories – similar datatypes



Neuroscience data: Multiple organisations association



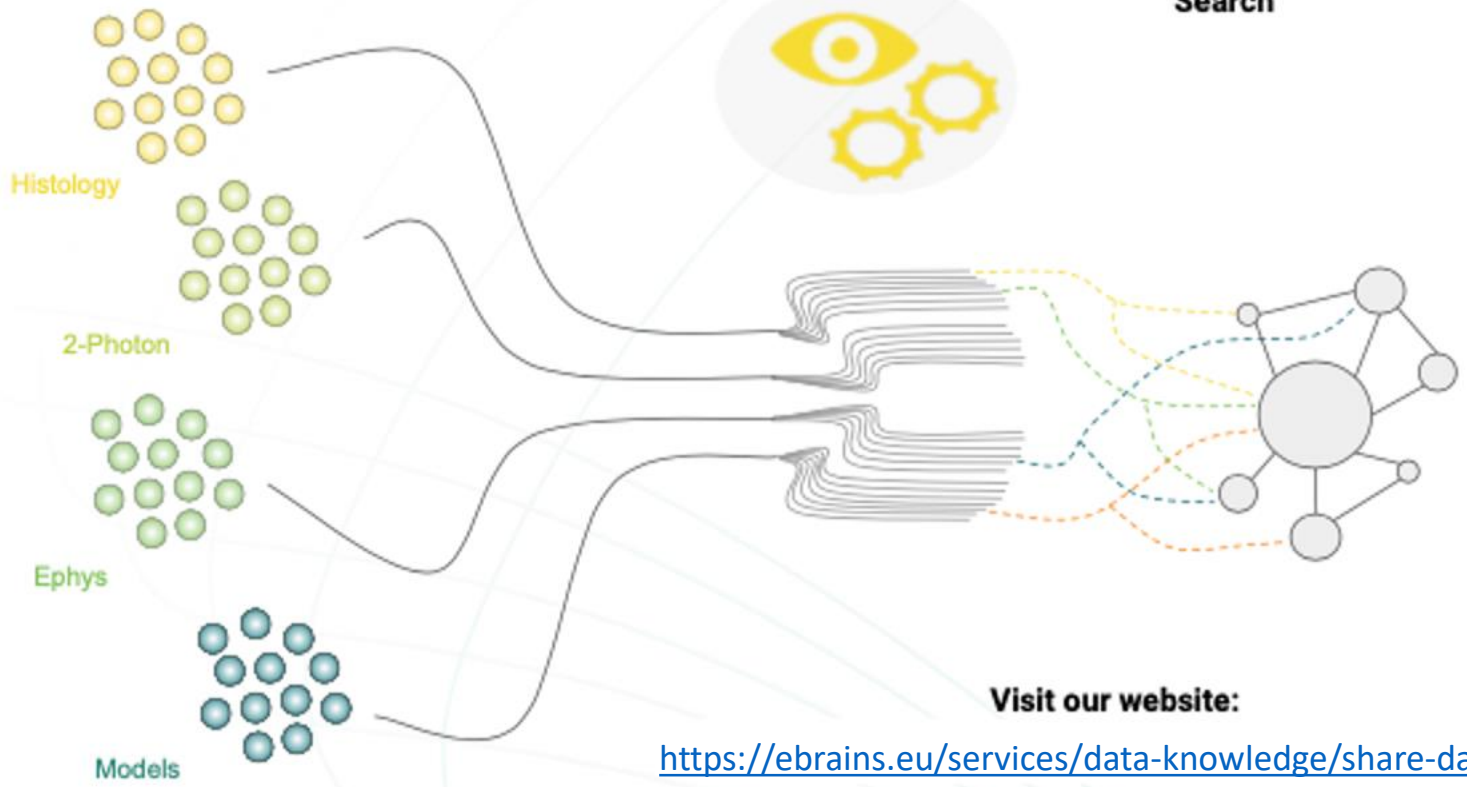
A shared digital brain research infrastructure for the EU

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



- 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

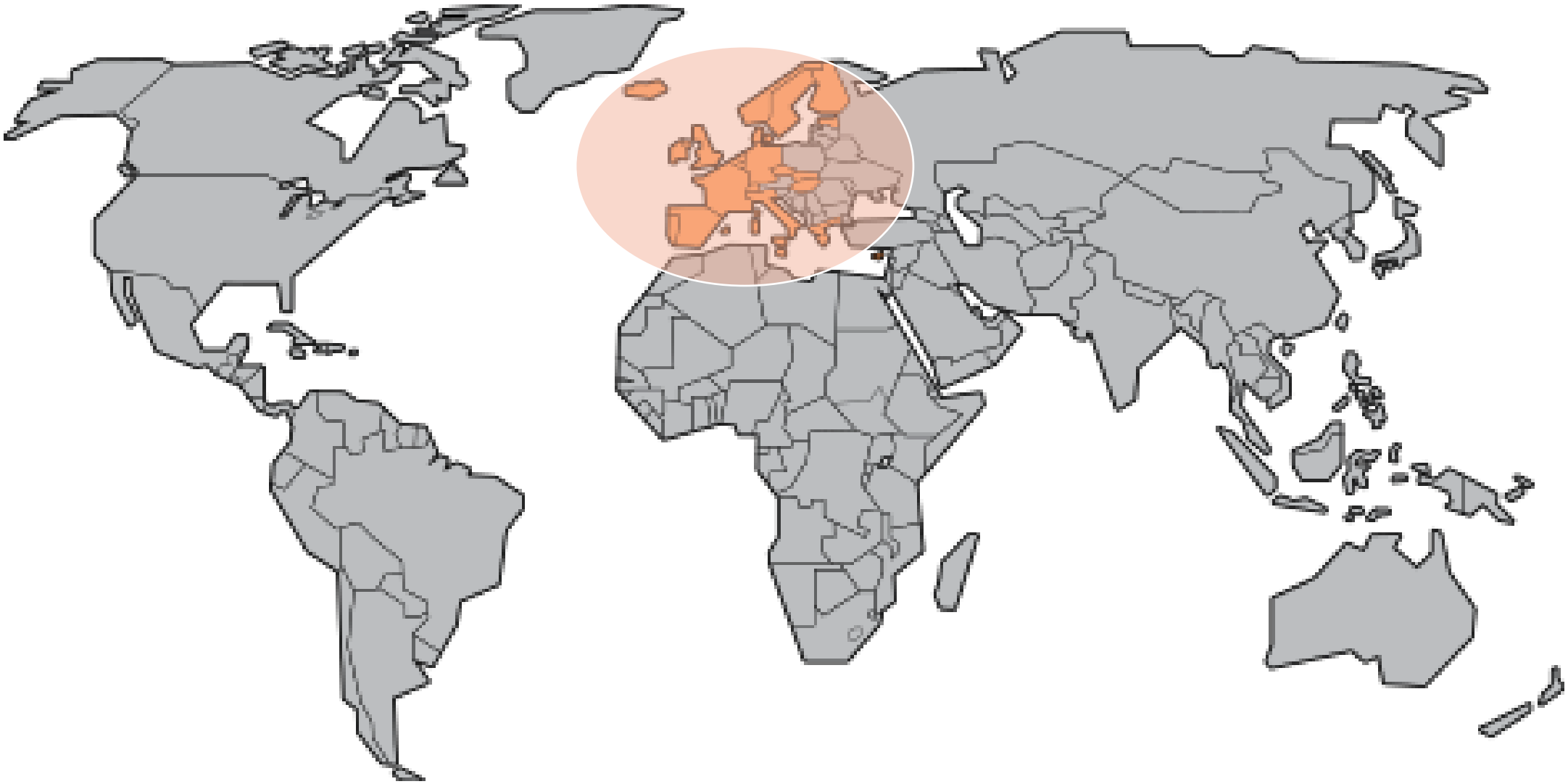
Visit our website:

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

Or contact us directly at:
curation-support@ebrains.eu

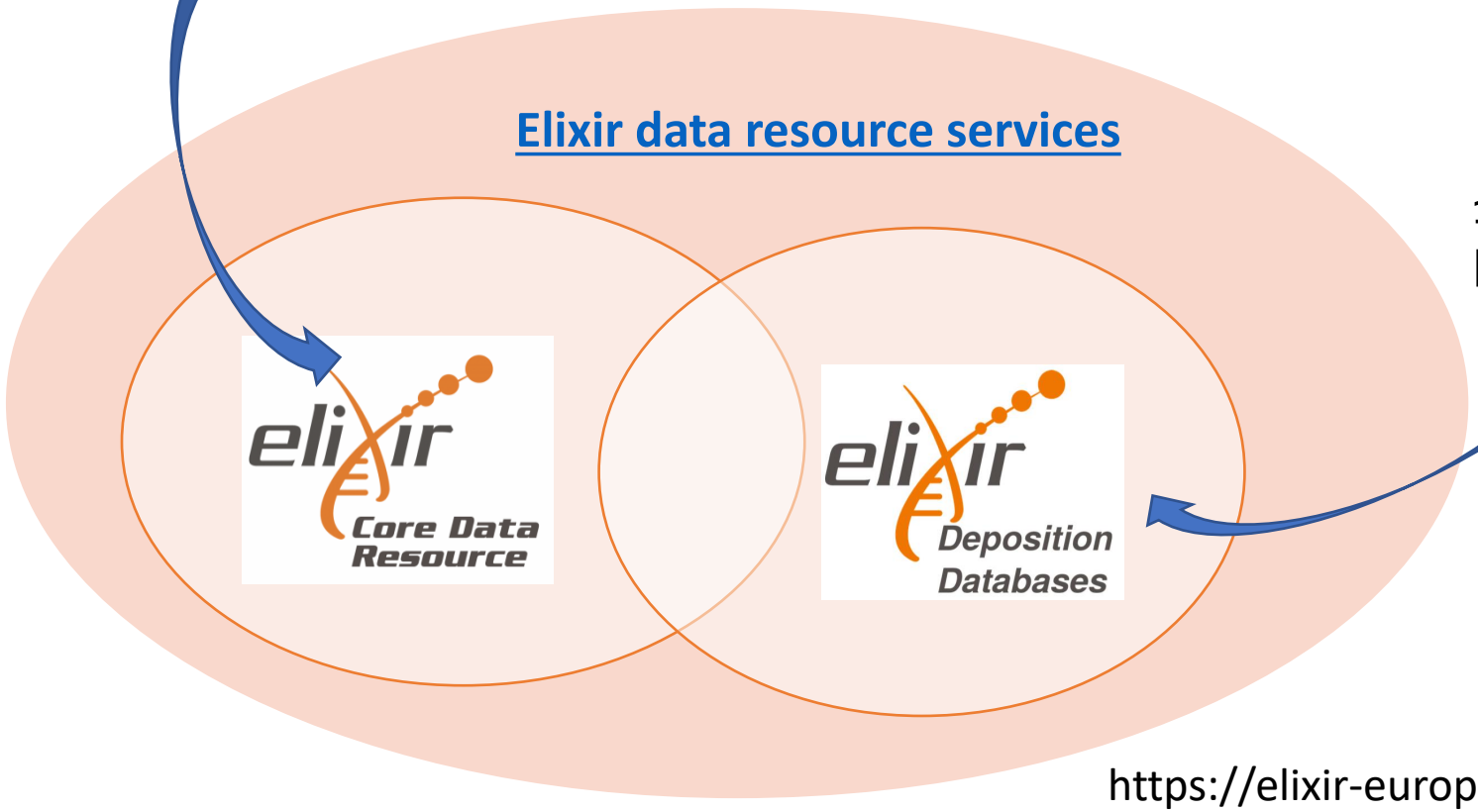
CC-BY 4.0 Anne Marte Kvello

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



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

Variation Archive

IntAct

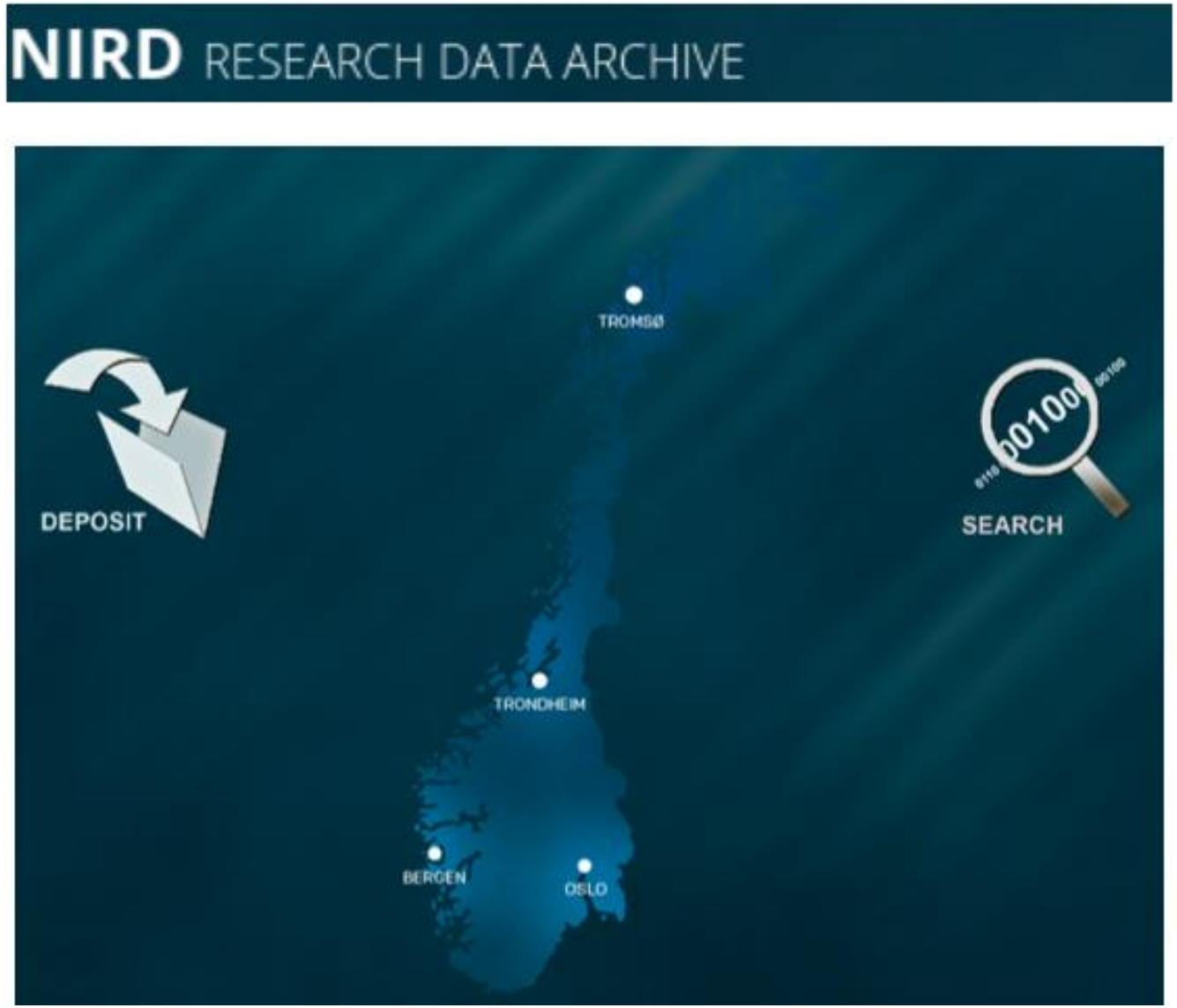
MetaboLights

PDBe Protein Data Bank in Europe

PRIDE PRoteomics IDEntifications Database

National repositories

NIRD 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

https://dataverse.no

NSD archive

NSD NORSK SENTER FOR FORSKNINGSDATA

CORE TRUST SEAL

Sikt
Kunnskapssektorens tjenesteleverandør

Domain agnostic

DOI accessible

Implemented data access committees

Main data type: Questionnaires, …

Reorganisation: from 2022 NSD 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.orgs/
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?

EUROPEAN GENOME-PHENOME ARCHIVE

ega-archive.org

Central metadata accessibility

Secure storage

Implemented data access committees

Norwegian EGA
(for data that can not leave the country)

Tryggve

elixir

neic

What about sensitive data?

Central metadata accessibility

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

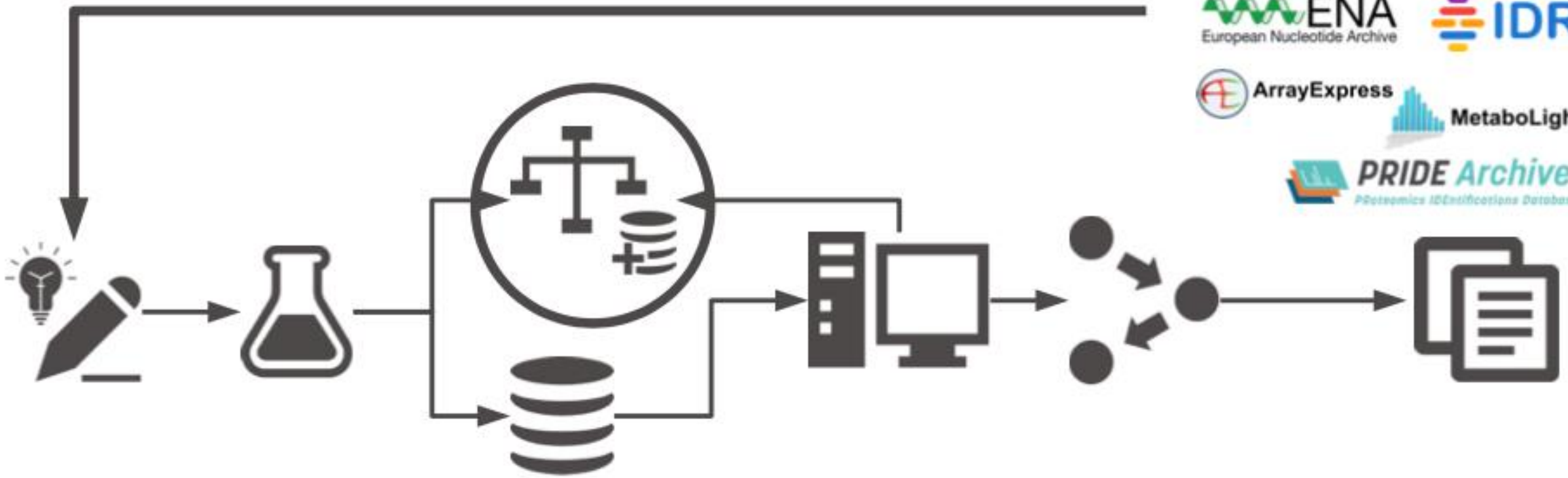
Secure storage

Implemented data access committees

ega-archive.org

[Norwegian EGA]

(for data that can not leave the country)



BioStudies.

ENA
European Nucleotide Archive

IDR

ArrayExpress

MetaboLights

PRIDE Archive
PRoteomics IDEntifications Database

EUROPEAN
GENOME-PHENOME
ARCHIVE

Illimar Rekand,
Illimar.rekand@uib.no
ELIXIR Norway, UiB