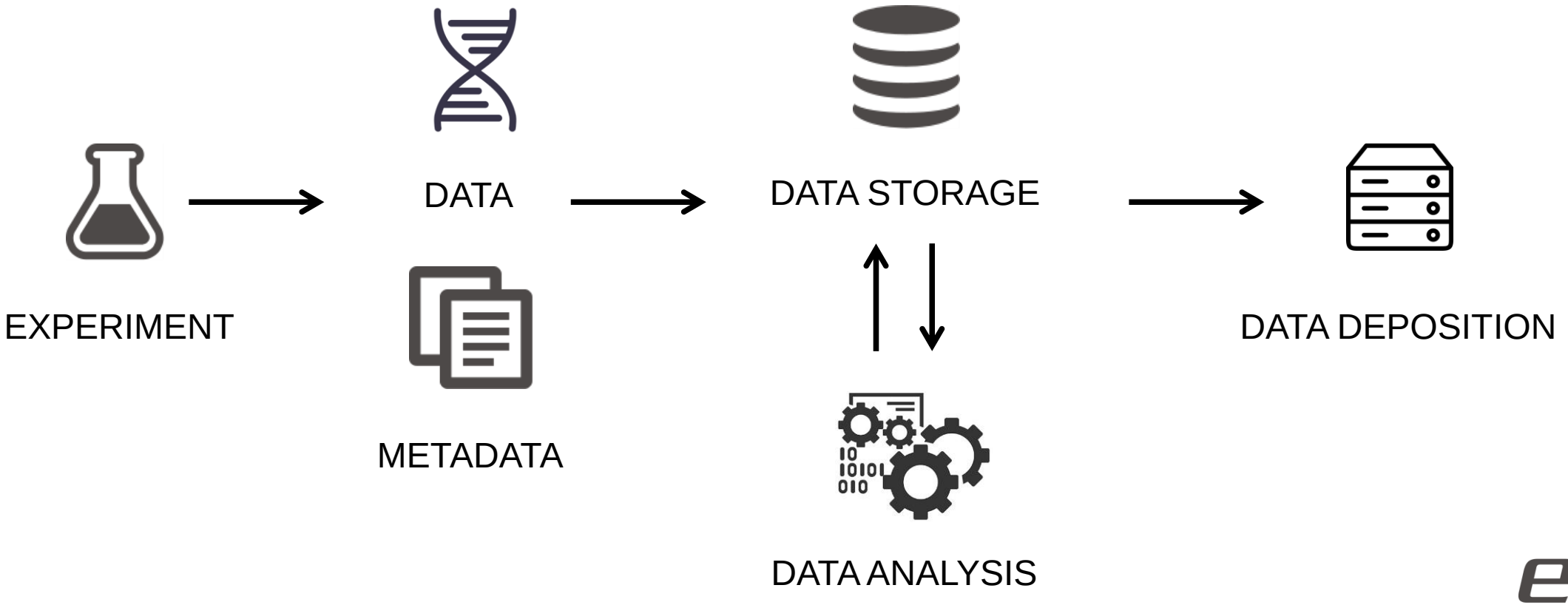
National storage infrastructures

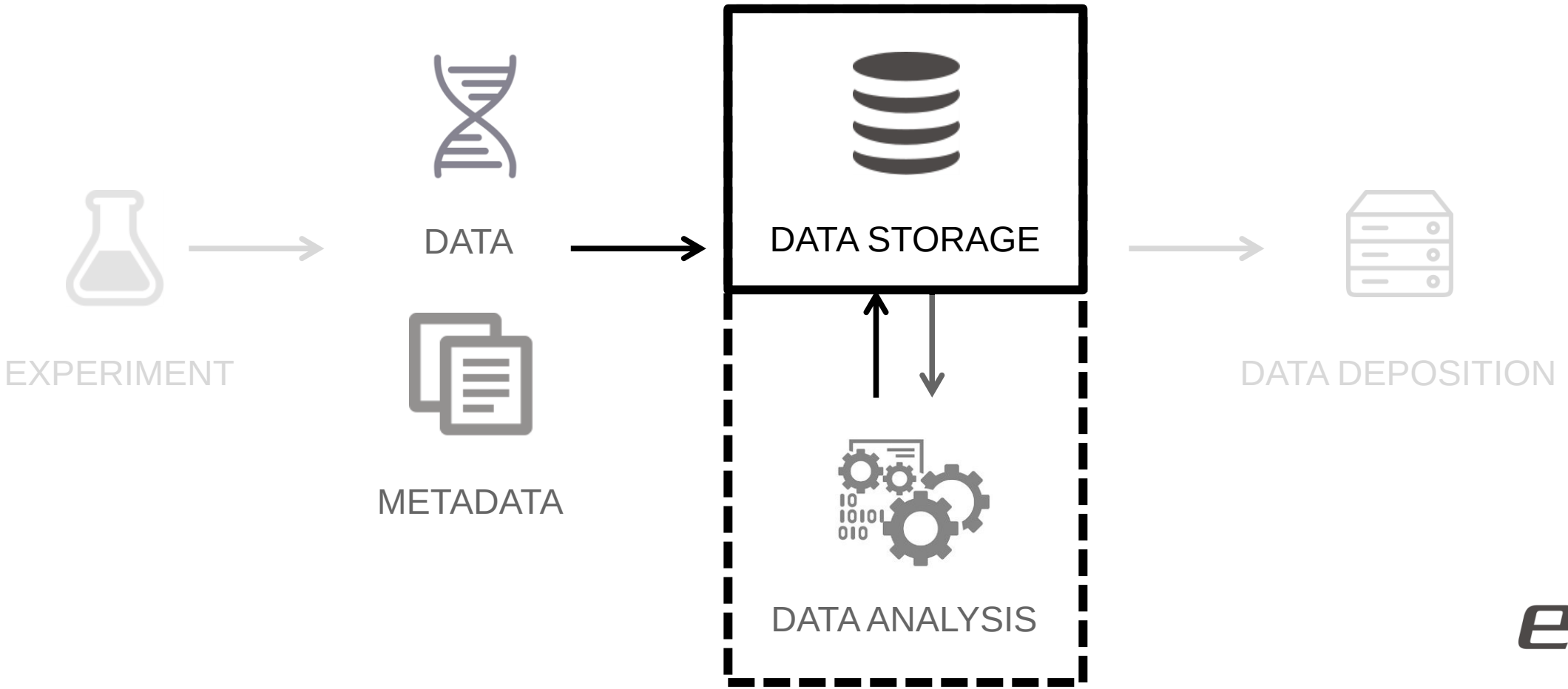
NeLS/StoreBioInfo, NIRD, TSD

Erik Hjerde
Arturo Vera-Ponce de Leon
ELIXIR Norway

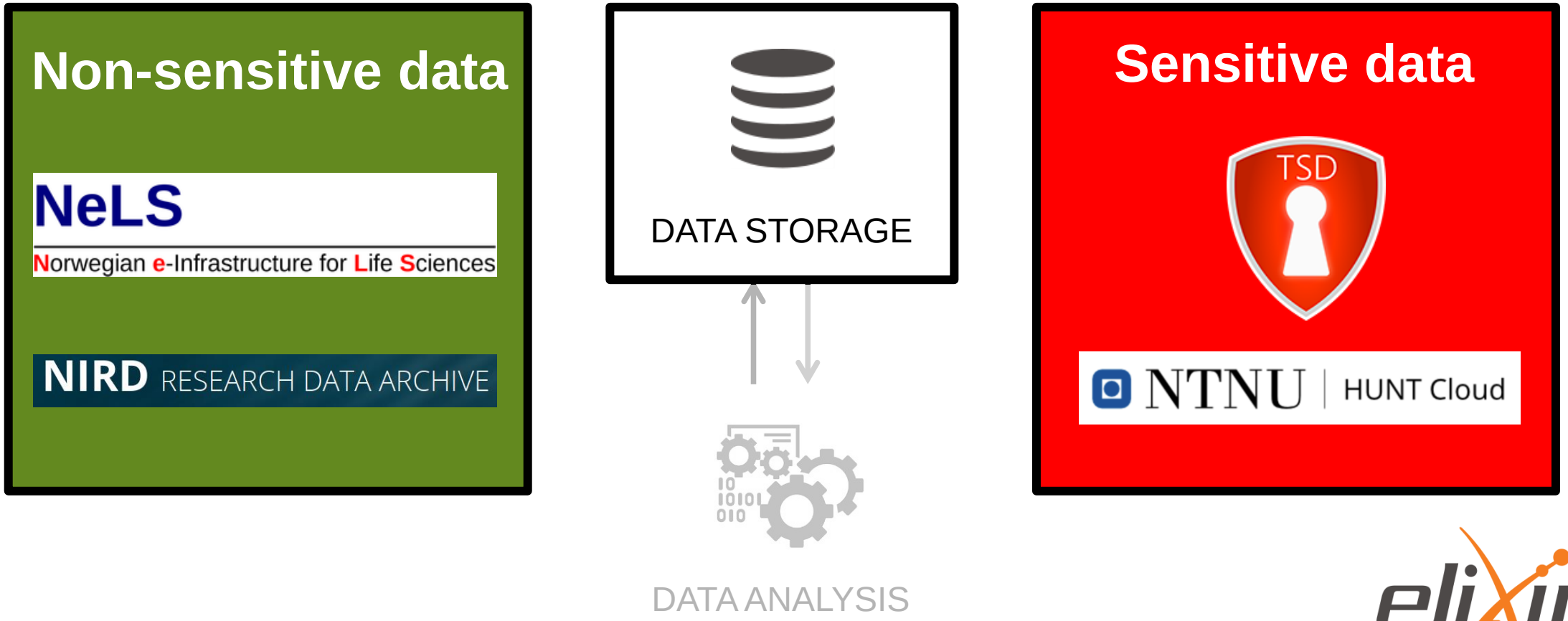
Data storage – from the researcher perspective



Data storage – from the researcher perspective



National storage infrastructures. Where to put my data???



elixir NORWAY

Data storage – General considerations

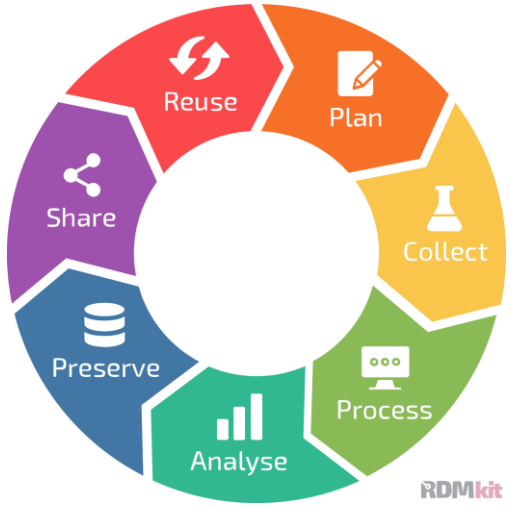
The ELIXIR Research Data Management kit

- Data life cycle +
- Your role +
- Your domain +
- Your problem -
  - Compliance monitoring
  - Data analysis
  - Data management plan
  - Data organisation
  - Data protection
  - Data publication
  - Data quality
  - Data storage
  - Data transfer
  - Identifiers
  - Licensing
  - Documentation and metadata
  - Sensitive data
- All tools and resources

- What features do you need in a storage solution when collecting data?
- How do you estimate computational resources for data processing and analysis?
- Where should you store the data after the end of the project?
- Relevant tools and resources

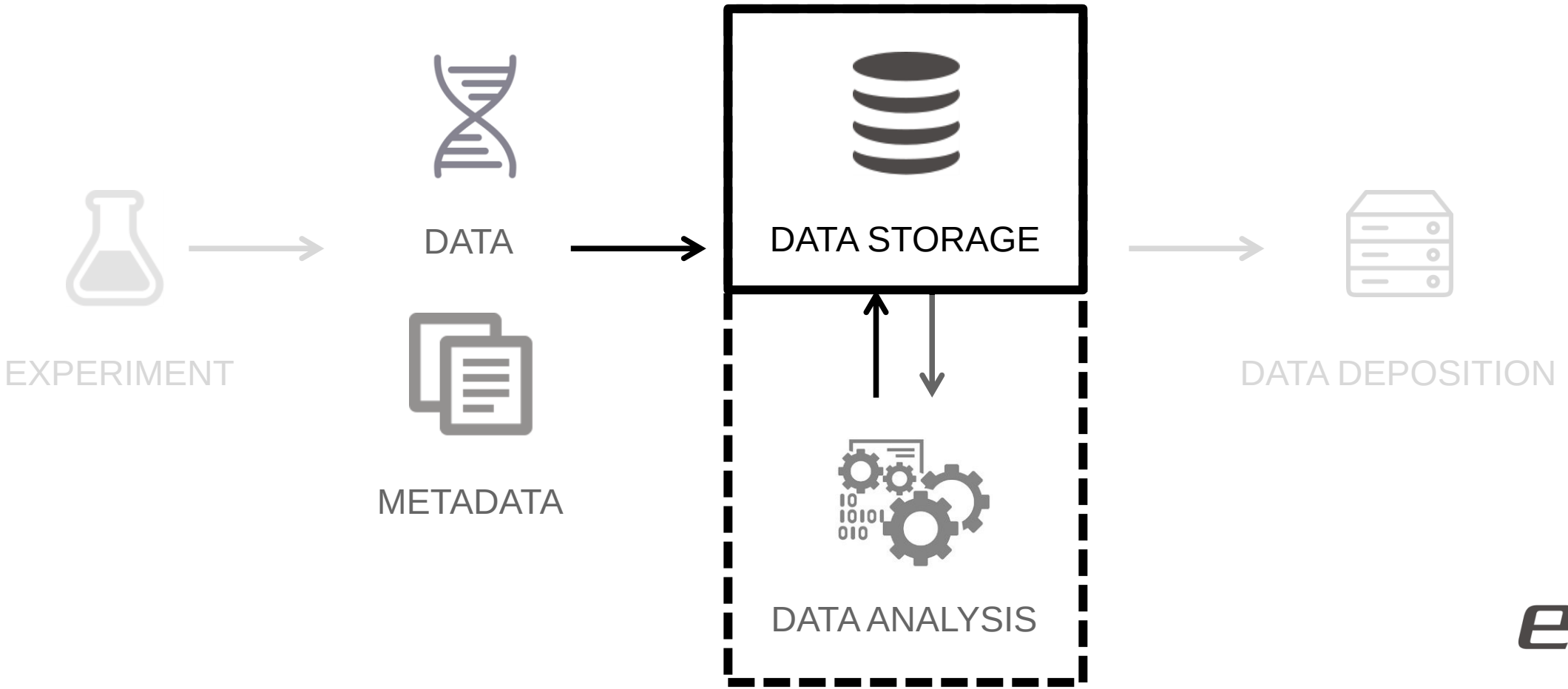
Relevant tools and resources

Tool or resource	Description	Tags	Registry
CERNBox	CERNBox cloud data storage, sharing and synchronization	storage	
CS3	Cloud Storage Services for Synchronization and Sharing (CS3)	storage	
DATAVERSE	Open source research data respository software.	storage, researcher, data manager, IT support	



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

Data storage – from the researcher perspective



Elixir Norway

https://elixir.no/

Scan me!

Data storage in NeLS and StoreBioInfo (SBI)

Norwegian e-infrastructure for Life Sciences - developed and operated by Elixir Norway

NeLS in general

Available to all Norwegian researchers

Enable collaborative projects

Multiple storage layers

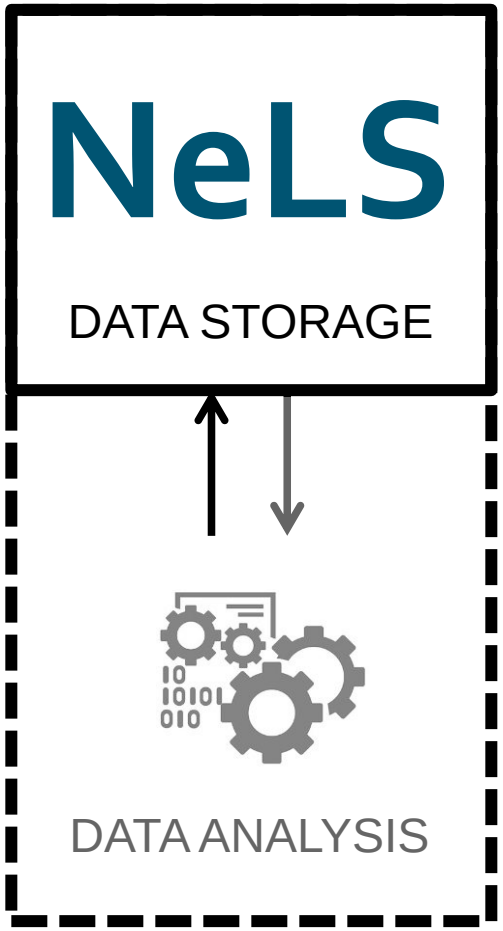
Integrated with compute

Multiple permission roles - shared data

Free of charge*

Support for data deposition

User support: Elixir Norway helpdesk



Type of data

Non-sensitive data

Support multiple data types

SEEK integration for metadata

elixir NORWAY

NeLS

Norwegian e-Infrastructure for Life Sciences

SERVICES

ELIXIR Norway offers services and infrastructure for life science, including software tools and workflows, databases and storage. In addition ELIXIR Norway provides support through its national helpdesk and offers trainings. Moreover, ELIXIR Norway provides services to the international research community.

E-infrastructure

- NeLS
- Galaxy
- TSD@UiO
- Other sensitive data platforms

The Norwegian e-infrastructure for life sciences (NeLS),
is a unique national portal for sharing, analysis and **midterm** storage of **non-sensitive** molecular life science data.

Free up to 10Tb

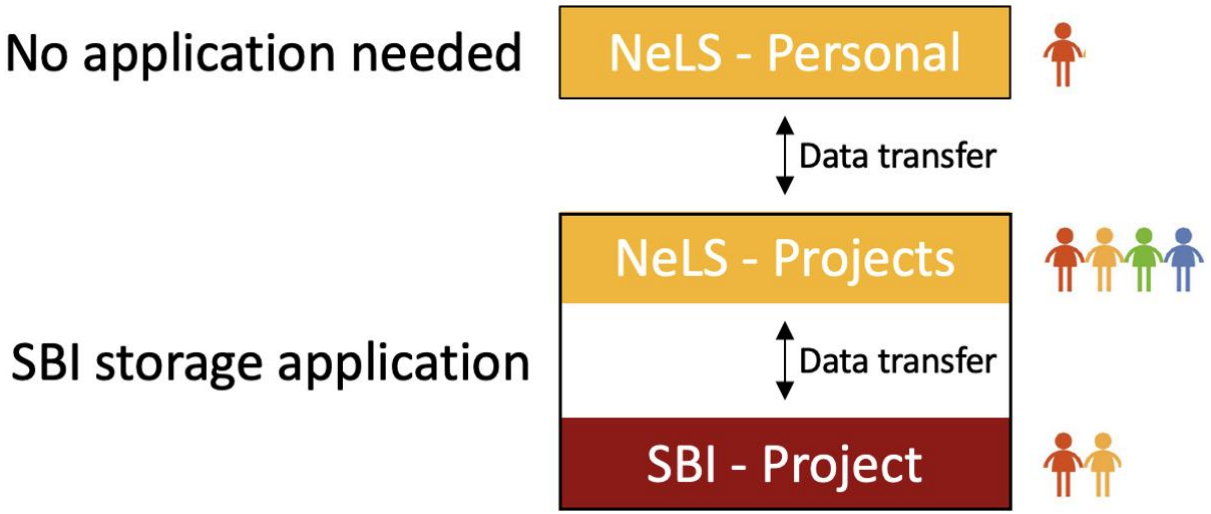
elixir NORWAY

Data storage in NeLS/SBI – How to get access

Storage application needed for projects where data is shared by many users

Access via FEIDE user or NeLS idp can be made for non-FEIDE users

Free storage < 10 TB



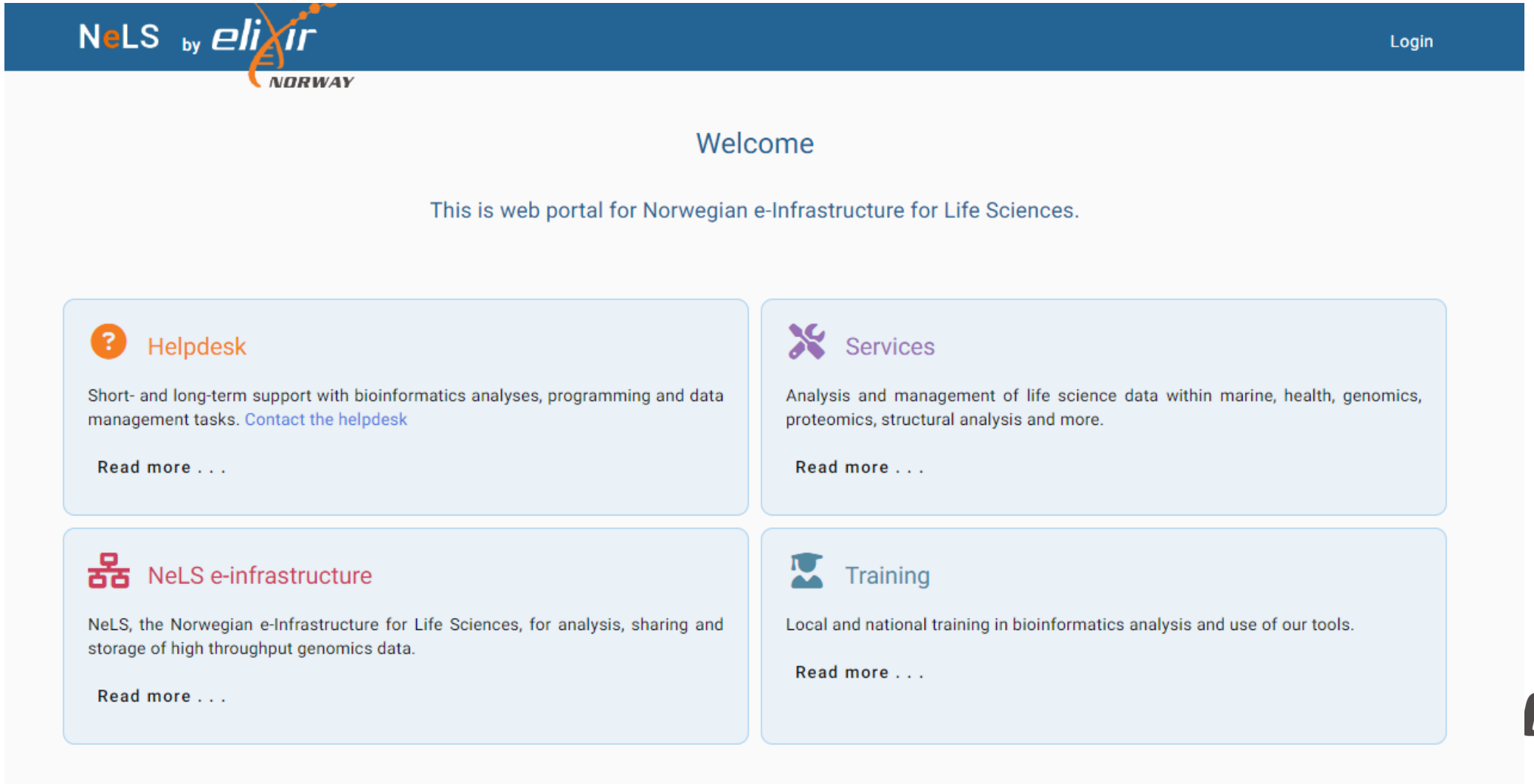
Apply for storage: contact@bioinfo.no

Wiki for usage: https://nels-docs.readthedocs.io/en/latest/

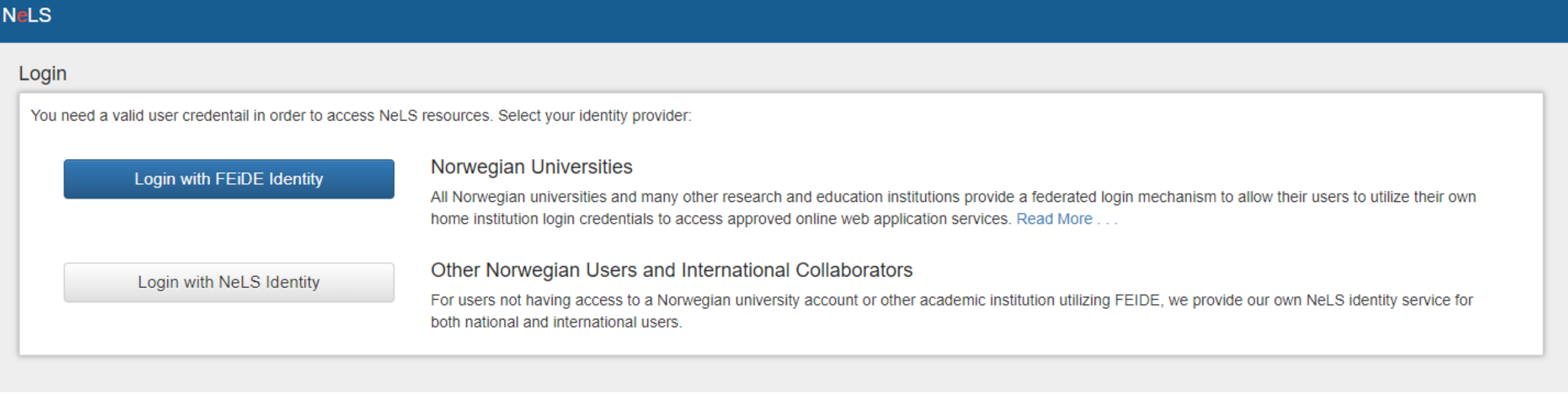
elixir NORWAY

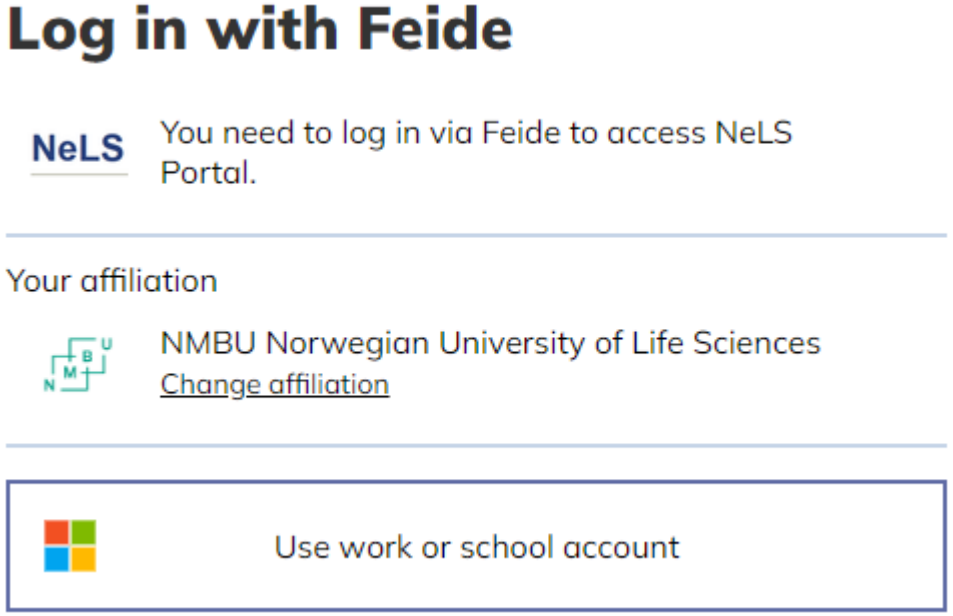
Login to NeLS using the WebPortal

https://nels.bioinfo.no/nels-web/#/



Scan me!



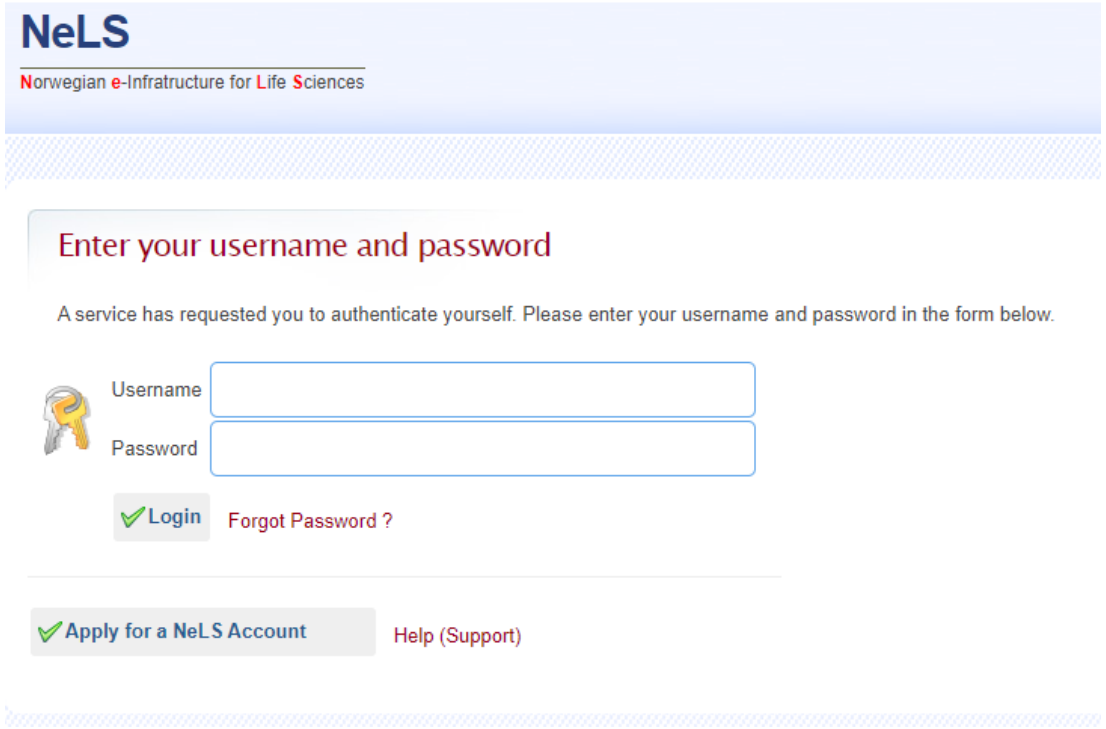


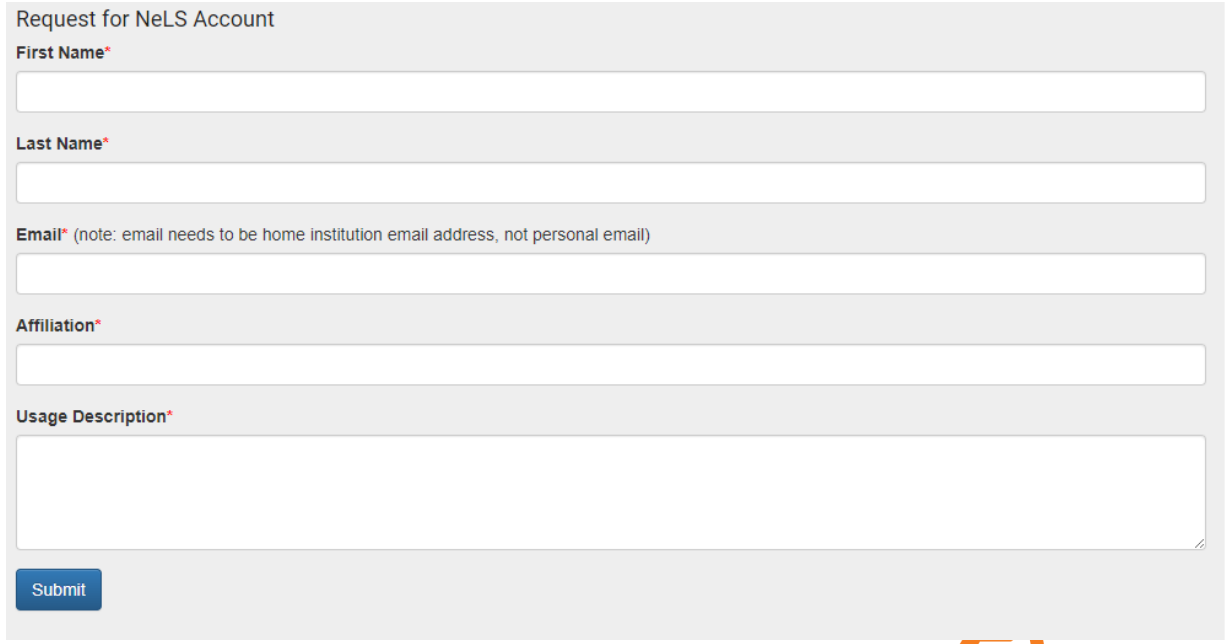
All users with FEIDE login credentials can access the NeLS Portal

https://nels.bioinfo.no/nels-web/#/

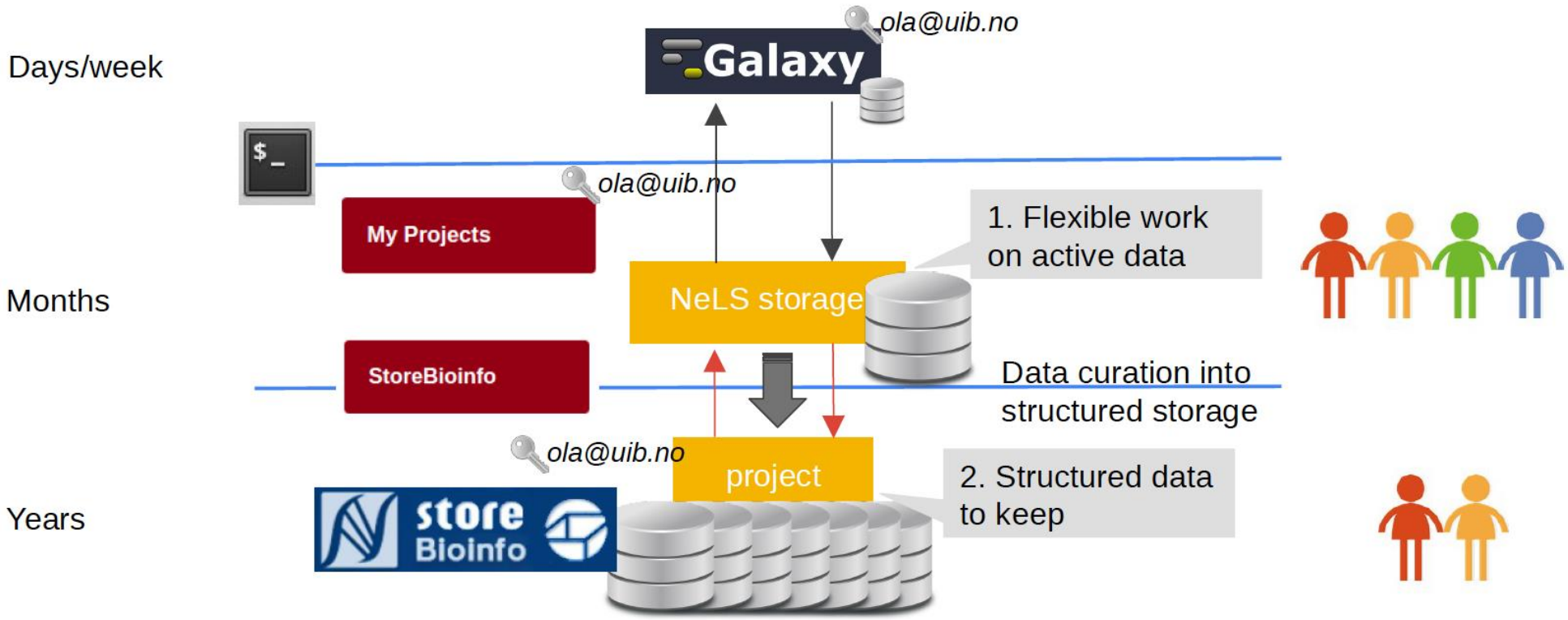
I don't have FEIDE login credentials (Non-Norwegian collaborators)

- It is also possible to login to NeLS via a NeLS idp. You can apply for a NeLS idp





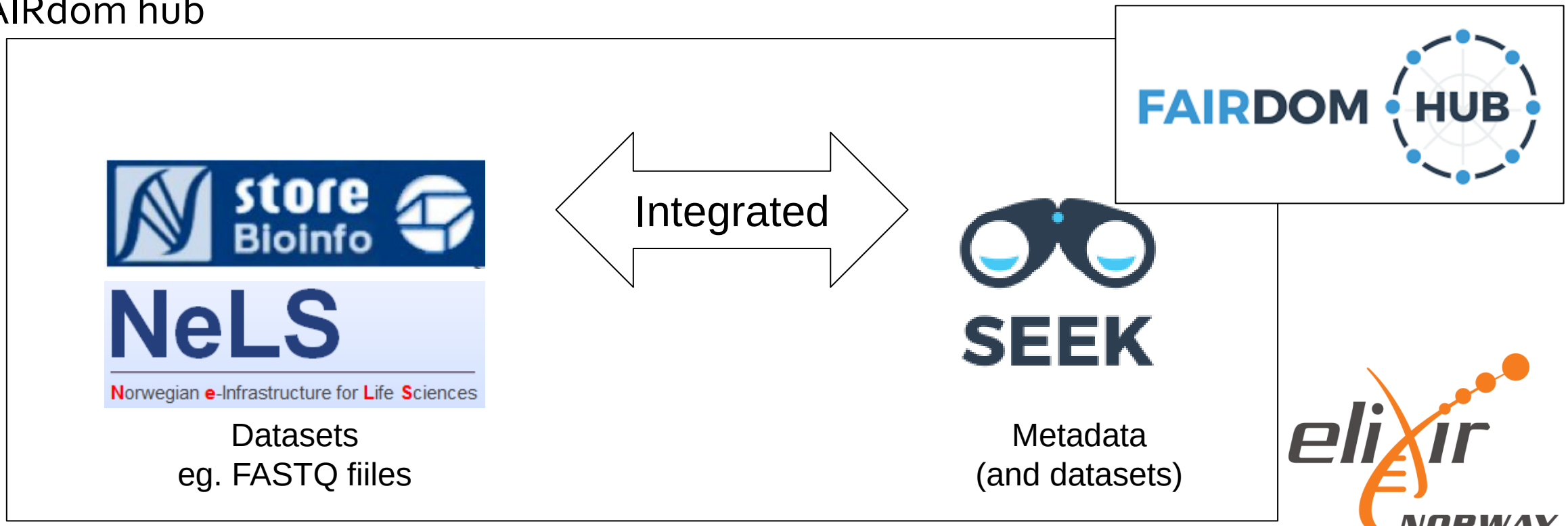
NeLS architecture



SEEK - sharing heterogeneous scientific research datasets, models or simulations, processes and research outcomes

The SEEK platform is a web-based tool for organising and storing data, and for exploring and annotating data

Norwegian users can link datasets stored in NeLS to a SEEK project using FAIRdom hub



Data storage in NIRD

National e-Infrastructure for Research Data - owned and operated by UNINETT Sigma2

NIRD in general

Available to all users with account

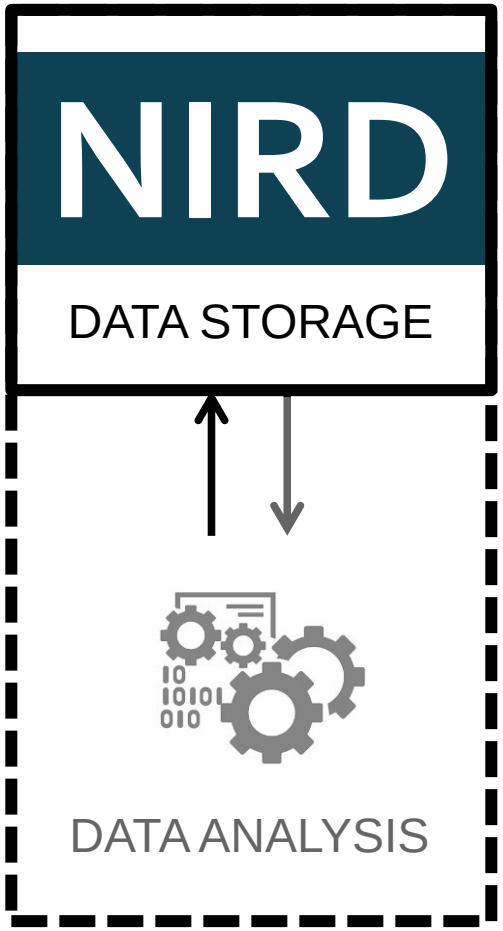
Enable collaborative projects

Integrated with compute

Multiple permission roles - shared data

Free of charge*

User support: Metacenter support



Type of data

Non-sensitive data

Support multiple data types

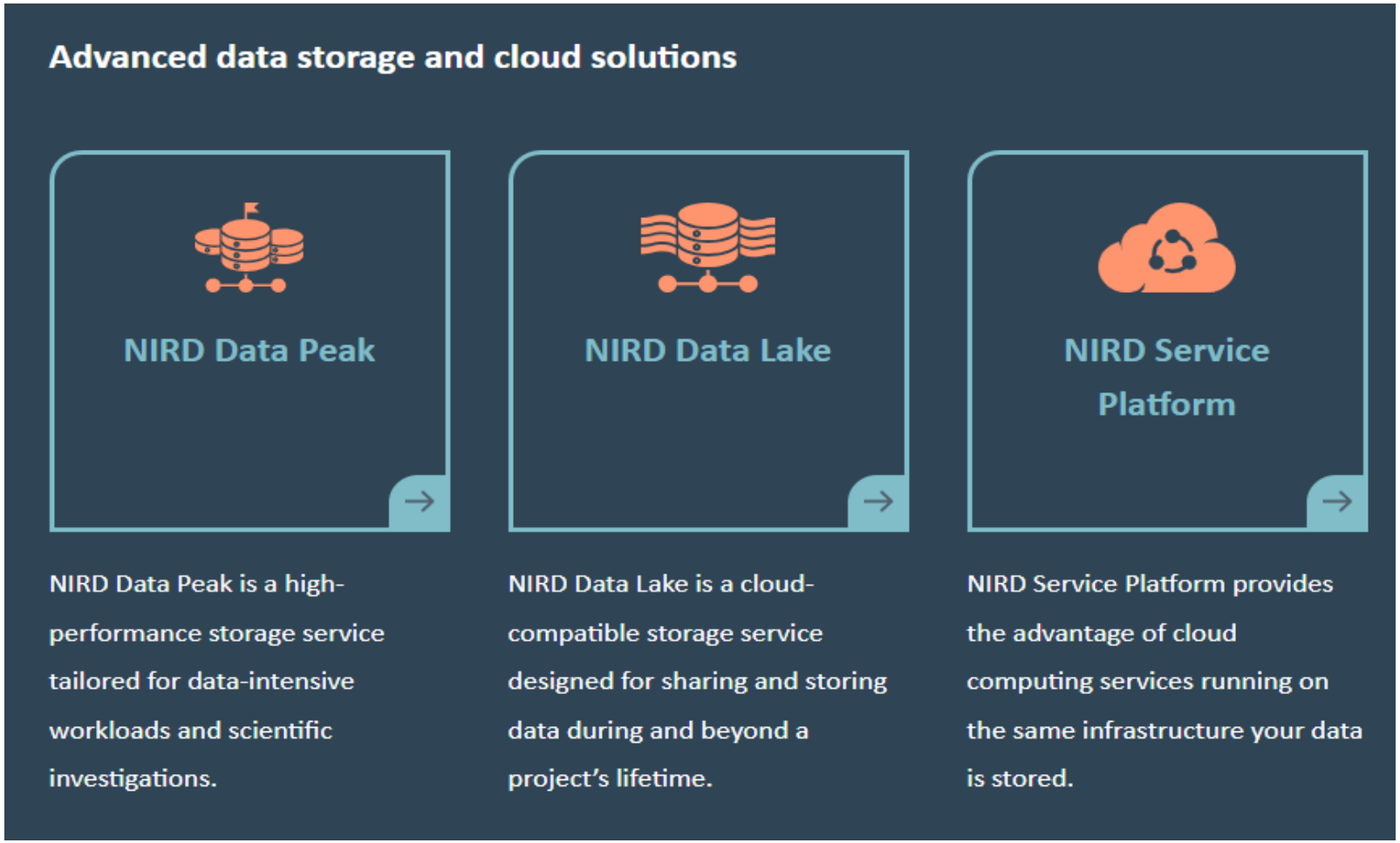
All type of scientific data*

Scan me!

elixir NORWAY

Data storage in NIRD

National e-Infrastructure for Research Data - owned and operated by UNINETT Sigma2



intensive I/O operations

Long term-storage

NIRDToolKit

Data storage in NIRD – How to get access

Apply for a UNINETT SIGMA2 user

Apply for a storage quota - Free storage < 10 TB

Four price categories

The model introduces four different categories of contributors.

- Category A
  Large projects with funding from the Norwegian Research Council of 15 MNOK or more or storage needs above 10 TiB, paying for operational expenses
- Category B
  Non-commercial projects needing Dedicated Resources, paying for capital and operational expenses
- Category C
  Commercial research and industry which will pay the full cost price
- Category 0
  Non-contributing projects which are smaller projects not in any of the above categories

https://www.sigma2.no/apply-e-infrastructure-resources

The process explained

1. **Submit user application**
2. **Signature by project management**
3. **Review by Sigma2**
4. **Set a password and log in**

Scan me!

Access data storage in NIRD

The NIRD storage can be accessed via the command line using SSH

Data import/export via command line tools such as SCP or SFTP or other file transfer tools (e.g. WinSCP)

```
(base) auve@L004538:~$ NIRD
Last login: Wed Nov  6 20:22:33 2024 from 158.36.102.145
Welcome to nird.sigma2.no!

Documentation:  https://documentation.sigma2.no/
Support email:  support@metacenter.no
Request resources:
                https://www.sigma2.no/content/apply-e-infrastructure-resources/
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Latest news from:       https://opslog.sigma2.no

  o 2024-11-25: Reduced capacity due to NTNU cooling upgrade week 48
  o 2024-11-05: Slowness on Saga
  o 2024-10-30: Saga is partially down and very slow.
  o 2024-10-28: NIRD datalake full
  o 2024-10-21: Issues with login to Fram
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Module command is not yet stable. Please report any issue encountered.
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
(miniforge3) [auve@login0-nird-lmd: ~]$
```

elixir NORWAY

All data storage in NIRD can be accessed in all the Sigma2 HPC (e.g. SAGA)

NIRD / NIRD mounts on clusters

NIRD mounts on clusters

The NIRD project storage areas, namely NIRD Data Peak (TS) and NIRD Data Lake (DL) are mounted on the login nodes of Betzy, Fram, and Saga. One can directly access the NIRD project area from the login nodes of the aforementioned compute clusters.

The path to Data Peak (TS) project areas is

`/nird/datapeak/NSxxxxK`

while the path to Data Lake (DL) project is

`/nird/datalake/NSxxxxK`

where `NSxxxxK` is the ID of the project.

```
(BASICS)[auve@login-4: ~]$ ls /nird/datapeak/
acl@                                NS10016K  NS10082K  NS11026K  NS11105K  NS5001K  NS8044K  NS8302K  NS9020K  NS9059K  NS9188K  NS9419K  NS9561K  NS9623K  NS9748K  NS9819K  NS9879K
files-ns9012k                       NS10022K  NS10086K  NS11027K  NS12002K  NS5011K  NS8045K  NS8402K  NS9021K  NS9061K  NS9191K  NS9461K  NS9564K  NS9630K  NS9754K  NS9825K  NS9880K
fram                                NS10023K  NS10088K  NS11028K  NS2345K   NS5023K  NS8052K  NS8403K  NS9029K  NS9063K  NS9206K  NS9478K  NS9576K  NS9635K  NS9765K  NS9828K  NS9883K
migrator                            NS10024K  NS10091K  NS11030K  NS2506K   NS6000K  NS8060K  NS8999K  NS9033K  NS9064K  NS9207K  NS9481K  NS9588K  NS9640K  NS9769K  NS9829K  NS9889K
migrator-betzy                      NS1002K   NS10100K  NS11032K  NS2615K   NS8001K  NS8061K  NS9000K  NS9034K  NS9065K  NS9252K  NS9486K  NS9593K  NS9641K  NS9772K  NS9832K  NS9897K
mmbackup.unsupported.PROJECT1       NS1003K   NS1012K   NS11043K  NS2649K   NS8002K  NS8065K  NS9001K  NS9035K  NS9066K  NS9279K  NS9493K  NS9595K  NS9644K  NS9777K  NS9835K  NS9898K
mmbackup.unsupported.protect1_tsminst1  NS10044K  NS1014K   NS11048K  NS2806K   NS8006K  NS8075K  NS9003K  NS9036K  NS9067K  NS9295K  NS9518K  NS9599K  NS9648K  NS9781K  NS9841K  NS9970K
NS0000K                             NS10047K  NS1015K   NS11051K  NS2875K   NS8014K  NS8085K  NS9005K  NS9039K  NS9073K  NS9305K  NS9525K  NS9600K  NS9672K  NS9787K  NS9853K  NS9985K
NS10003K                            NS1004K   NS1018K   NS11054K  NS2980K   NS8015K  NS8087K  NS9008K  NS9042K  NS9081K  NS9312K  NS9529K  NS9603K  NS9689K  NS9794K  NS9864K  NS9988K
NS10005K                            NS10054K  NS11002K  NS11082K  NS2993K   NS8021K  NS8095K  NS9010K  NS9043K  NS9083K  NS9320K  NS9530K  NS9607K  NS9691K  NS9806K  NS9868K  NS9989K
NS10006K                            NS10055K  NS11007K  NS11083K  NS4605K   NS8028K  NS8112K  NS9013K  NS9048K  NS9114K  NS9370K  NS9531K  NS9608K  NS9696K  NS9809K  NS9869K  NS9990K
NS1000K                             NS10064K  NS11009K  NS11084K  NS4654K   NS8029K  NS8121K  NS9014K  NS9052K  NS9121K  NS9378K  NS9540K  NS9610K  NS9699K  NS9813K  NS9873K  NS9992K
NS10013K                            NS10068K  NS11016K  NS11085K  NS4659K   NS8035K  NS8122K  NS9015K  NS9054K  NS9122K  NS9403K  NS9559K  NS9611K  NS9715K  NS9817K  NS9874K  NS9997K
NS10014K                            NS10069K  NS11022K  NS11096K  NS4704K   NS8039K  NS8128K  NS9019K  NS9057K  NS9181K  NS9408K  NS9560K  NS9616K  NS9745K  NS9818K  NS9876K  NS9999K
(BASICS)[auve@login-4: ~]$
```

https://documentation.sigma2.no/files_storage/nird/mounts_lmd.html

elixir NORWAY

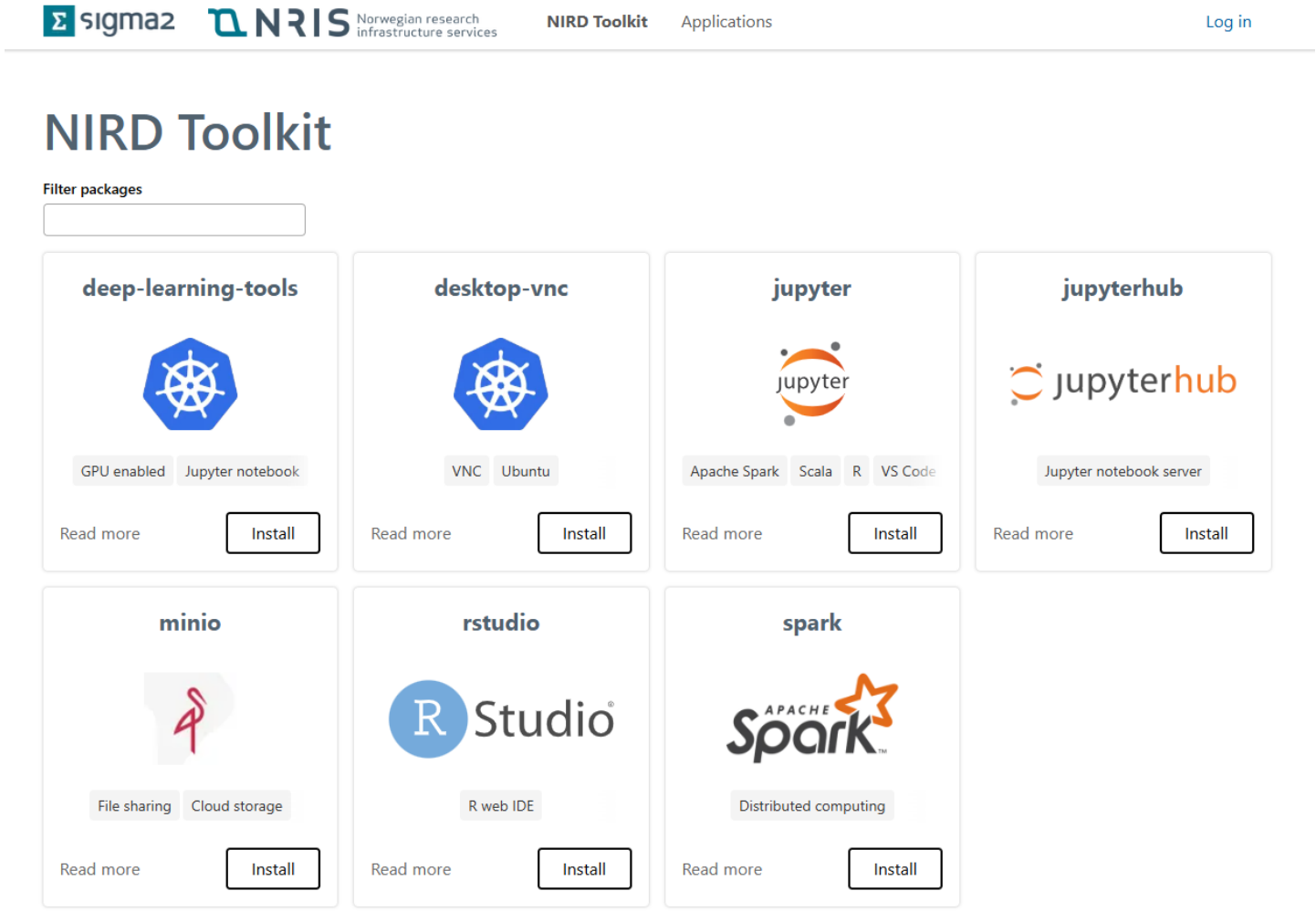
NIRD ToolKit

What is the NIRD Service Platform

NIRD Service Platform is a platform to run cloud services, such as web services, domain and community-specific portals, and tools for data visualization, pre/post-processing, data discovery and data sharing.

The service platform offers dedicated resources for data sharing and computing with large datasets. Here you can effortlessly process and share data with your preferred tools, such as

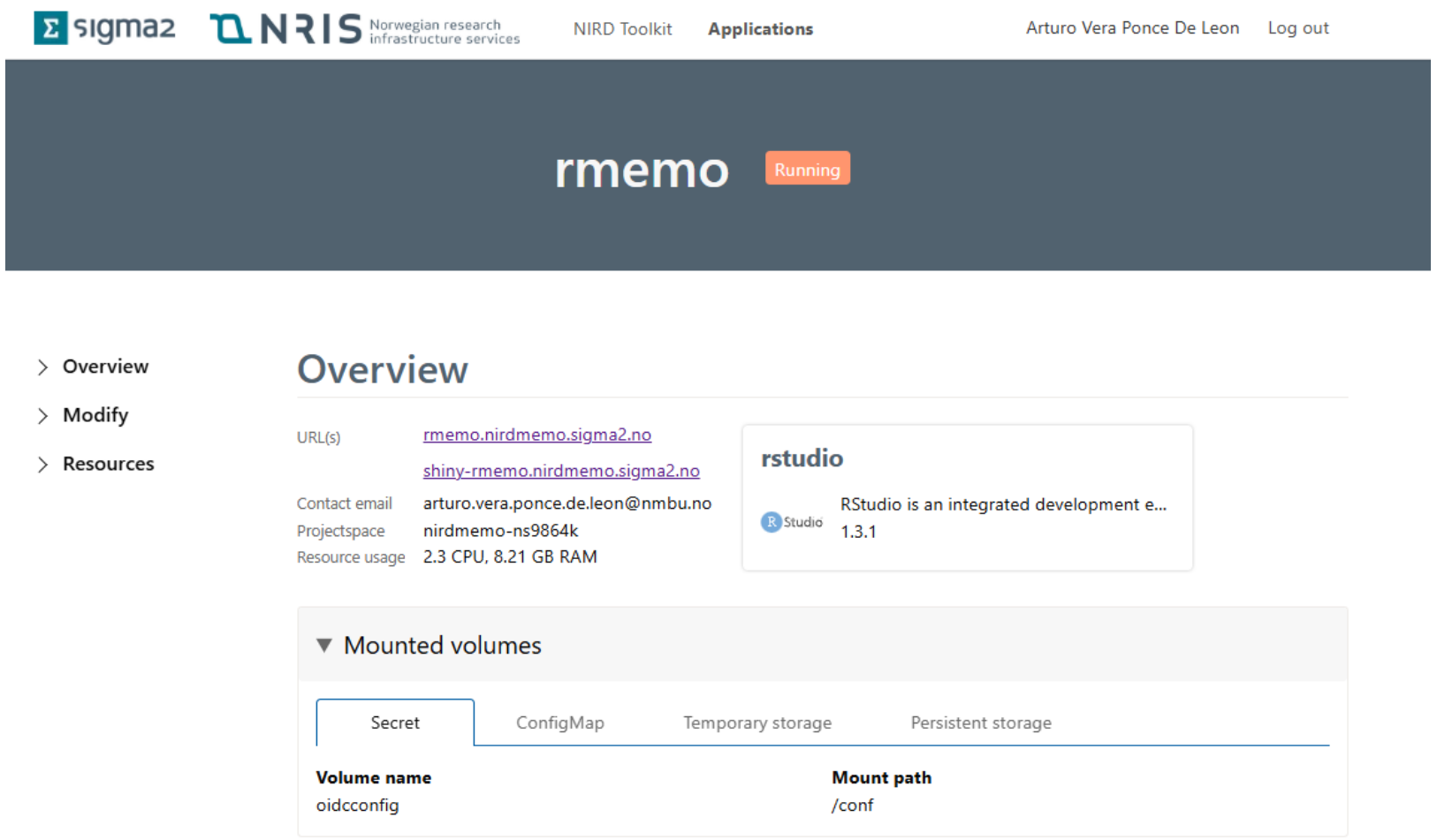
- Jupyter Notebook
- Jupyter Hub
- Spark
- R-Studio



https://apps.sigma2.no/

elixir NORWAY

NIRD ToolKit: Run Rstudio to analize data directly on NIRD



elixir NORWAY

Sensitive data storage

Non-sensitive data

NeLS

Norwegian e-Infrastructure for Life Sciences

NIRD RESEARCH DATA ARCHIVE

DATA STORAGE

Sensitive data

TSD

NTNU | HUNT Cloud

elixir NORWAY

Data storage in TSD

National service for sensitive data - developed and operated by UiO

TSD in general

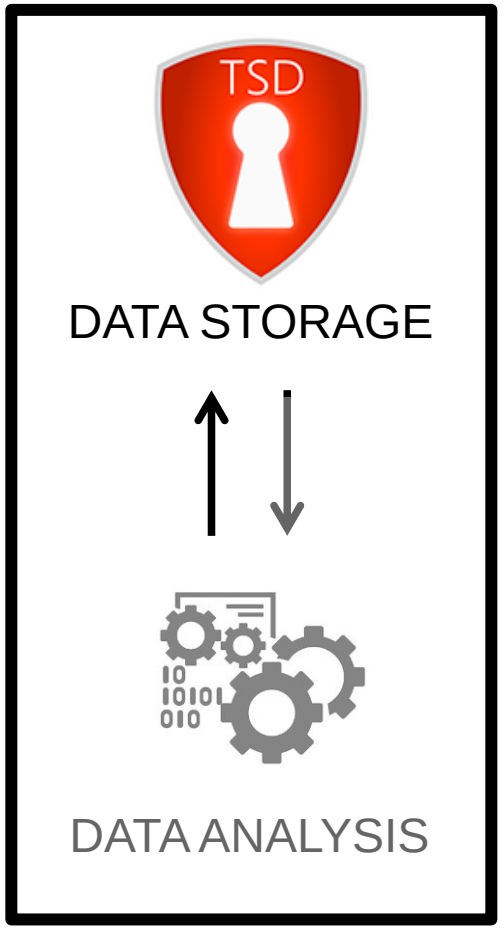
Available to all users with account

Enable collaborative projects

Integrated with compute

Multiple permission roles - shared data

User support: Tech and admin support



Type of data

Sensitive data

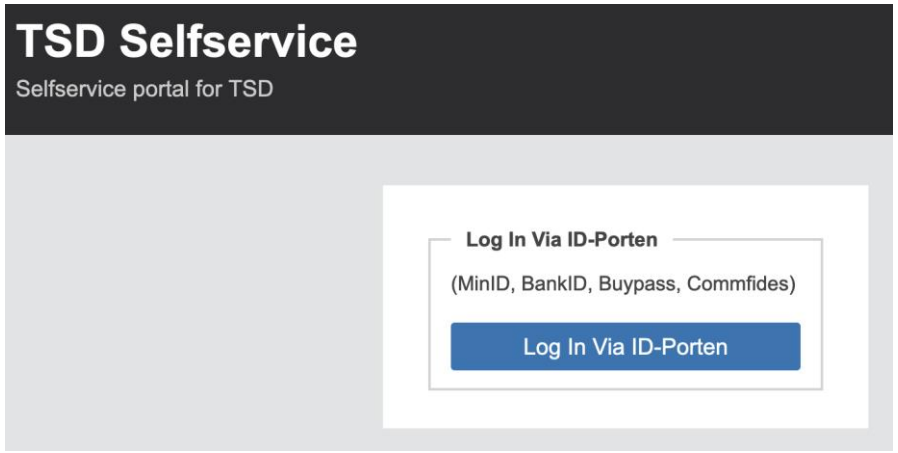
Support multiple data types

Data storage in TSD – How to get access

Apply for a project – need to document ethical approval (e.g. from REC)

Generate a TSD user – require 2-factor authentication

Payment models for storage



Apply for a TSD project through ELIXIR.NO: ELIXIR Norway has a quota in TSD that, and users can get subsidised quota through ELIXIR in TSD

Generate user account: https://selfservice.tsd.usit.no/

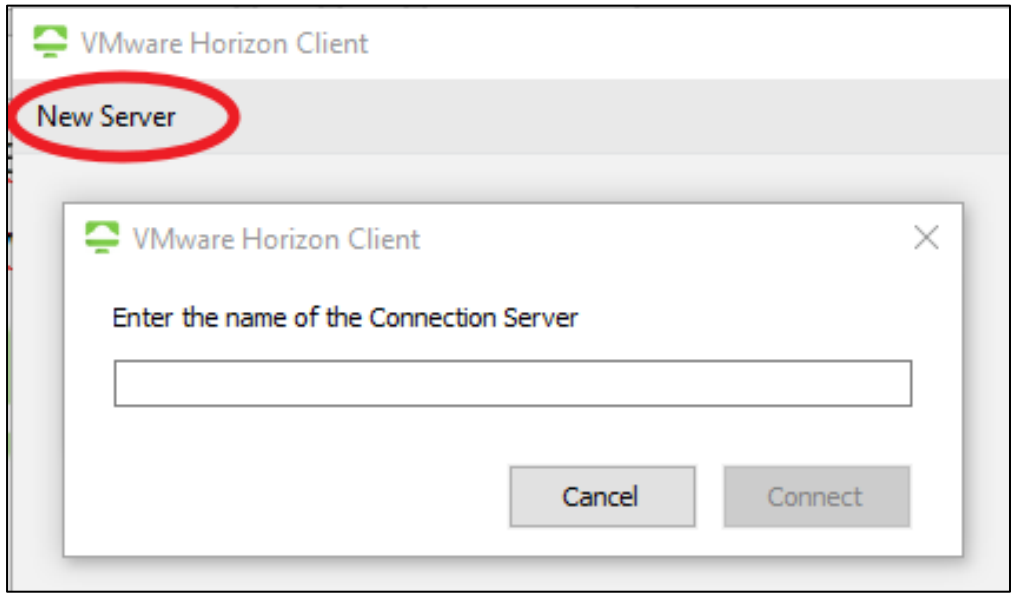
Apply for project: https://www.uio.no/tjenester/it/forskning/sensitiv/hjelp/start/registrer.html

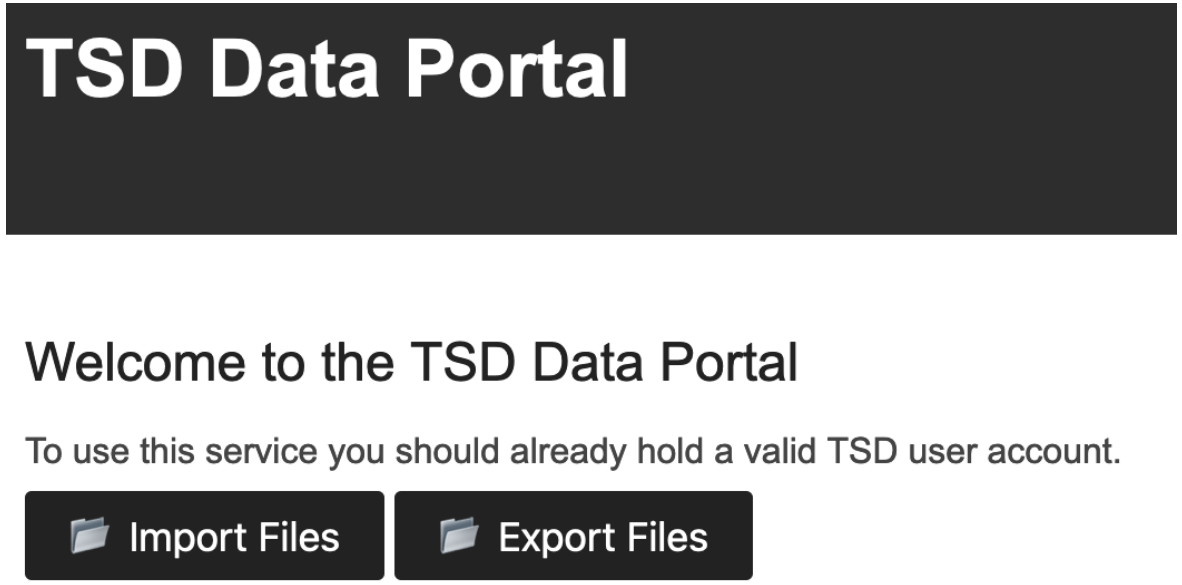
Wiki for usage: https://www.uio.no/tjenester/it/forskning/sensitiv/hjelp/start/index.html

Access to data storage in TSD

The TSD storage can be accessed via VMware Horizon + 2-factor authentication

Import/export data via the web file upload service





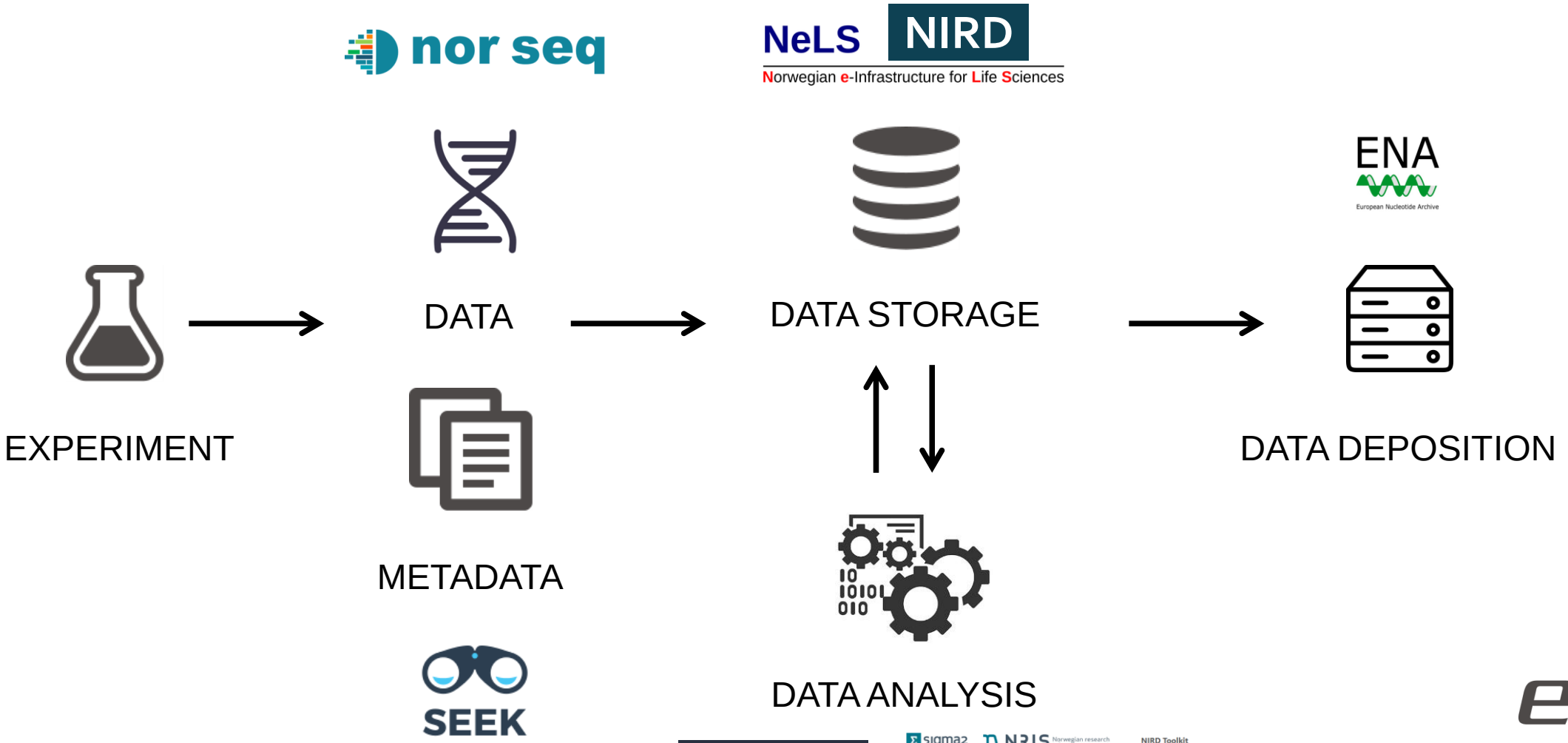
Where should I store and share my life science data?

	Sensitive data	Direct transfer from data producer	Web portal access*	Command line access	Shared projects	Integration for metadata	Support for data deposition	Free of charge
NeLS/SBI	✗	✓	✓	✓	✓	✓	✓	✓
NIRD	✗	✓	✗	✓	✓	✗	✗	✓
TSD	✓	✓	✗	✓	✓	✗	✗	✗

*Direct access to the data and project management through the web browser

elixir NORWAY

Example - Data flow/handle using ELIXIR Norway



elixir NORWAY