Introduction to Data Management
— What, why and how?

Gunn Inger Lyse
NSD – Norwegian Centre for Research Data
www.nsd.no
2018

For NMBU Talent program 2016-2019,
Meeting 5: Holmsbu Bad- og Fjordhotell at 7-8 June 2018
With focus on Teaching and Data Management
(Internt saksnummer NSD: 201800211)

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD – Norwegian Centre for Research Data

NSD has made data available for researchers for almost 50 years.
We work continually to improve our services.

- Access to data for researchers

- Archiving/long-term perspective
  (requires curation to keep data alive)

- Open access balanced against personal data protection

- Overview (DBH and other services)

NSD
NORSK SENTER FOR FORSKNINGSDATA

NSD – Norwegian Centre for Research Data

NSD has made data available for researchers for almost 50 years.
We work continually to improve our services.

OPEN - IN USE

Data management plan
Tool to generate a data management plan. Compliant with the requirements from Horizon 2020 and Research Council of Norway.

Status: in production. Open for everyone.

OPEN - IN USE

Archiving portal
Service to deposit digital data for re-use. FAIR-compliant, long-term perspective and supports different access levels.

Status: in production. Open for everyone.

BETA VERSION

Search portal
Interface to find research data.

Status: beta version in production. Open for everyone.

OPEN - IN USE

Register data (RAIRD)
Tool to facilitate access to register data - with built-in protection for person data.

Status: testing with pilot users.

ONGOING

Training
Training program (digital and non-digital) aimed at researchers, students, administration and high school.

Status: ongoing and under development.

NEW VERSION

Personal data protection portal: Institution
A service providing an overview over projects with personal data.

Status: will replace existing solution in June 2018.

NEW VERSION

Personal data protection portal: researchers/students
A new service to register the use of personal data in research projects (GDPR).

Status: will replace existing solution in June 2018.

DEVELOPMENT

Institution portal
Wil give the institutions an overview over all activities related to data management plans, archiving, re-use of data and personal data.

Status: prototype summer 2018.

ALPHA VERSION

My page
Portal to all activities related to NSD, for researchers, institutions, etc.

Outline

WHAT – Basic concepts of data management
Funder requirements

WHY – Why is good data management good for you?

HOW – To establish good data management routines? How to share data?
Data management tools at NSD.

NSD
NORSK SENTER FOR FORSKNINGSDATA

Example 1: AVOID DUPLICATION OF EFFORT

An example from research on sea ice by

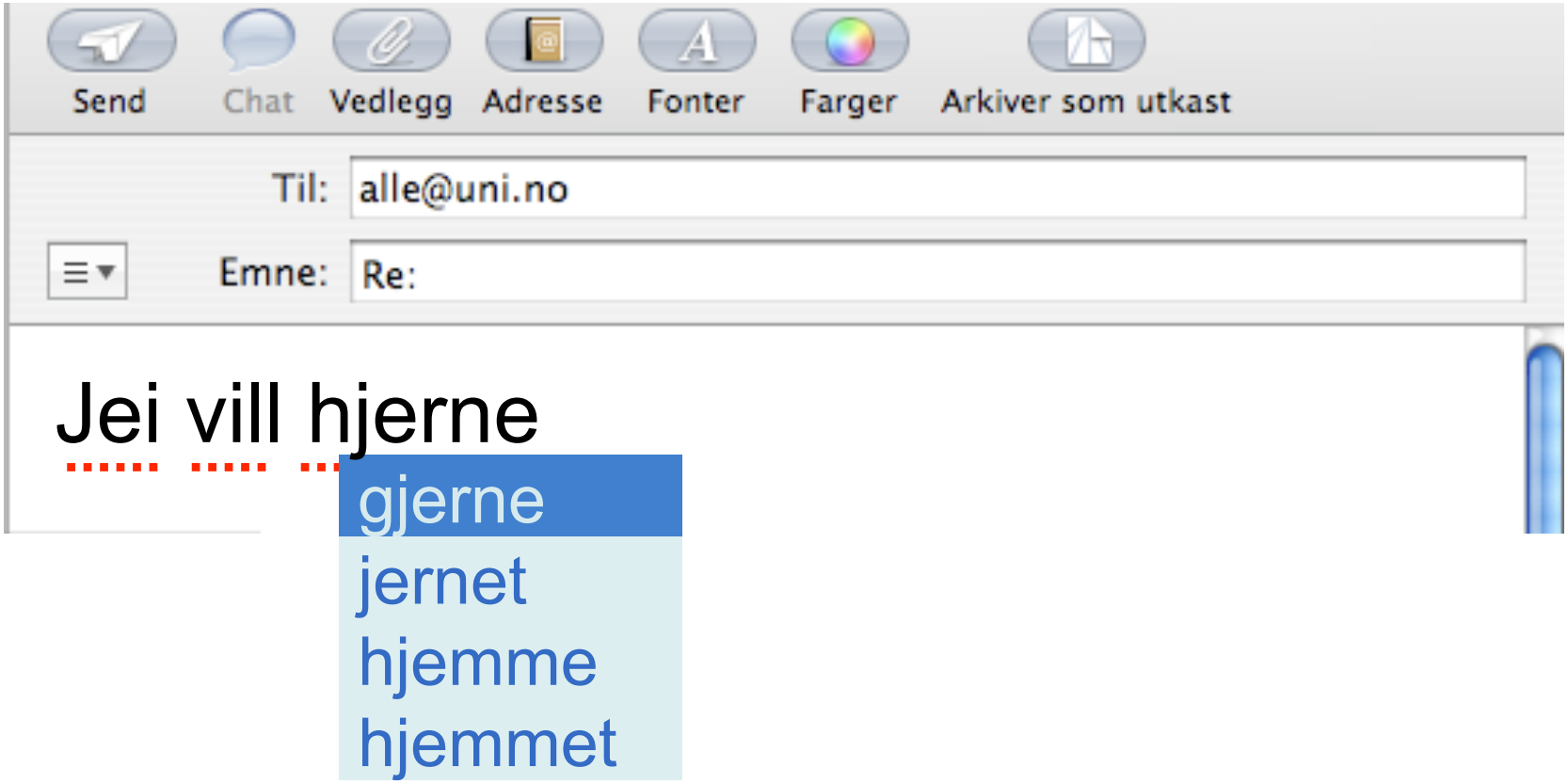
Marta Zygmuntowska,
University of Bergen

By NASA. Public Domain, https://commons.wikimedia.org/w/index.php?curid=15837631

NORSK SENTER FOR FORSKNINGSDATA

(Example from a presentation by Beate Krøvel Humberset, University of Bergen Library)

Example 2: SHARING DATA FOR RE-USE (1/3)

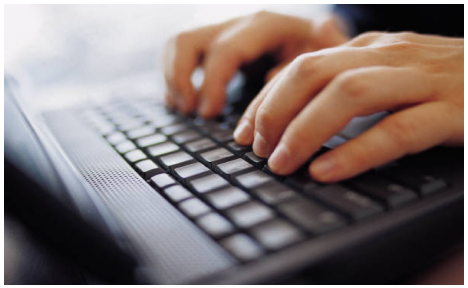


NORSK SENTER FOR FORSKNINGSDATA

Example 2: SHARING DATA FOR RE-USE (2/3)

Analysis
- Classification
  - sports, politics, …
  - bokmål/nynorsk
- Automatic analysis of words
  - gender, part of speech, …
- Frequencies
- …

700 mill. words
+ 1 mill words/week

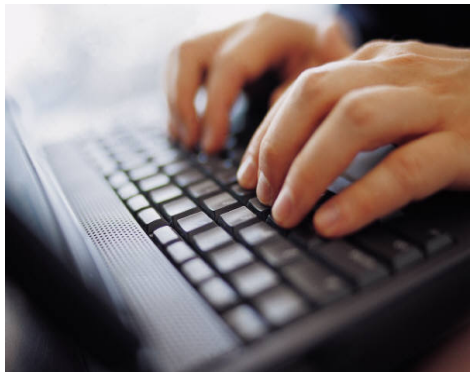


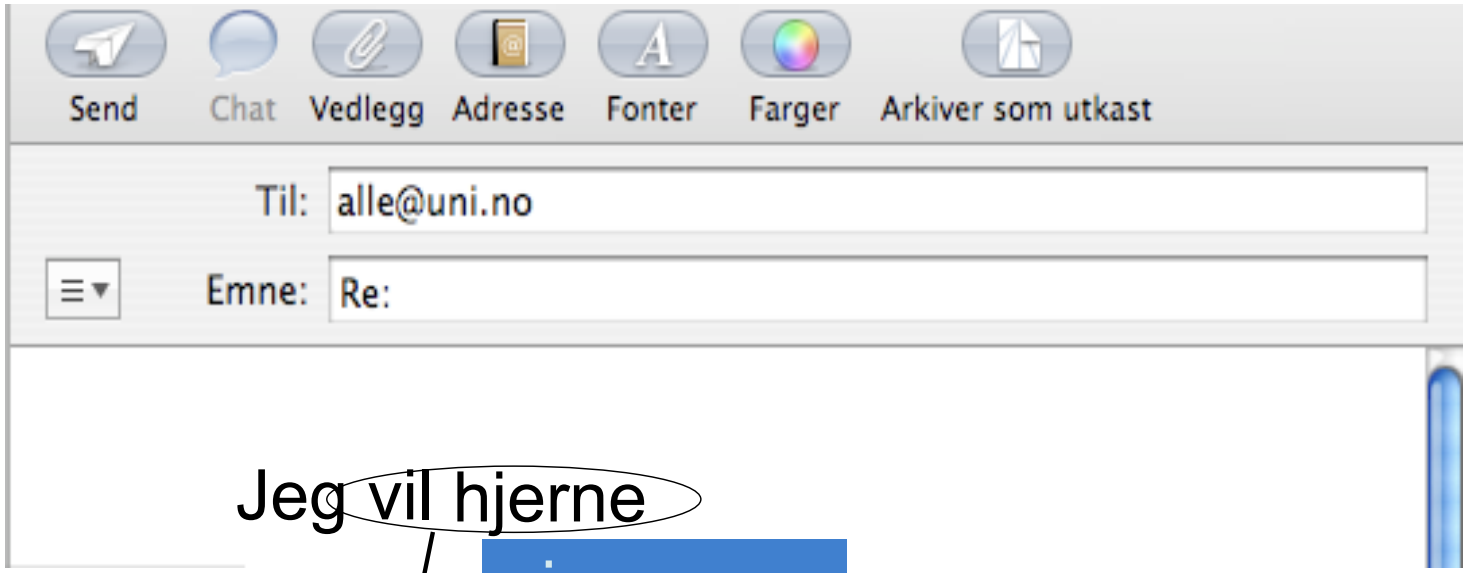
Research and development
- Neologisms
- Foreign words in Norwegian
- Media research
- Reading and writing aid
- …

NSD NORSK SENTER FOR FORSKNINGSDATA

Reference: Newspaper corpus (http://hdl.handle.net/11495/D9B5-0349-4330-0)

Example 2: SHARING DATA FOR RE-USE (3/3)





Ord	frekvens
vil gjerne	17942
vil hjerne	0

NSD
NORSK SENTER FOR FORSKNINGSDATA

Basic concepts of research data management

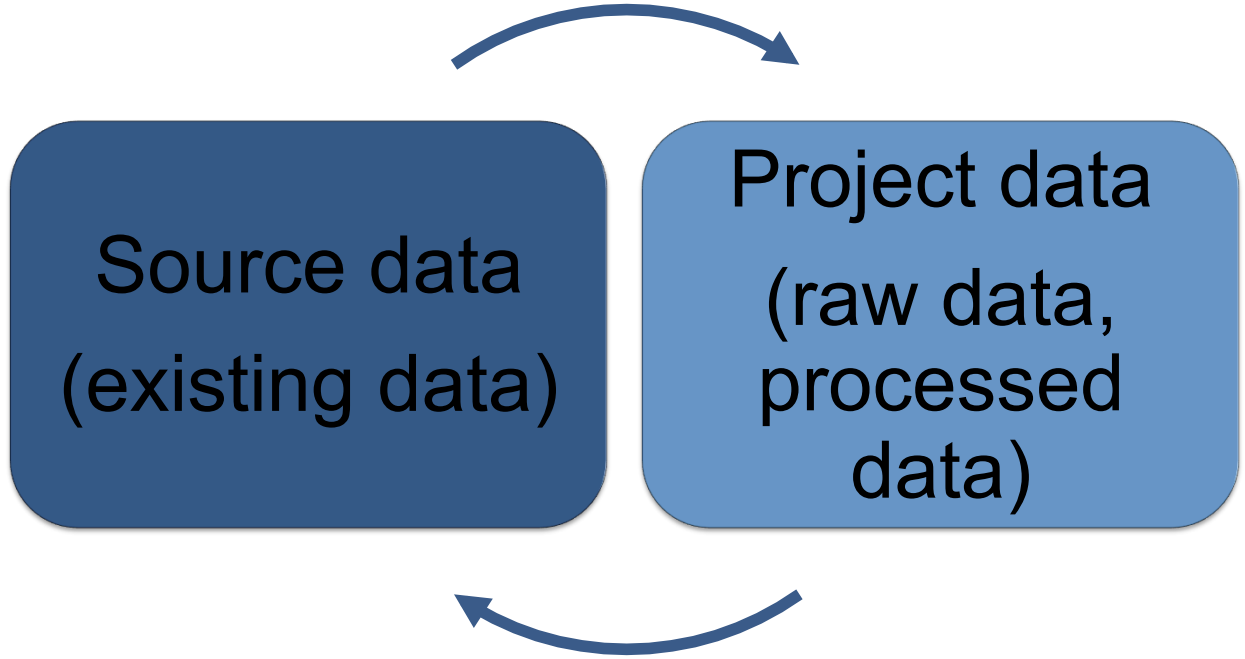
- What is research data?
- The problem: data loss
- FAIR data principle
- Data life cycle
- Data Management Plan (DMP)
- Store vs. archive data
- Share data/make data available- what does it mean?
- Funder/institutional requirements

NSD
NORSK SENTER FOR FORSKNINGSDATA

What is research data?

All data created by researchers in the course of their project

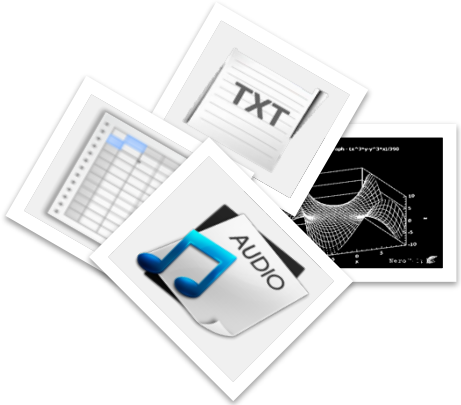
Source data = already existing data, independently of your research
Research data = data resulting from your project (from raw data to processed data)



NORSK SENTER FOR FORSKNINGSDATA

What are your data?

- If your memory stick/laptop/hard drive is corrupt, how much can you afford to loose?
- Project with multiple partners: who has (access to) which data?
- What if somebody questions your findings, what would you use to back up your claims?
- Would you be able to reproduce your figures in 5 years, or understand the variable abbreviations effortlessly?



NSD
NORSK SENTER FOR FORSKNINGSDATA

The problem: data loss/missing access

”Mostly due to current methods capture and data malpractice, approximately **50% of all research data and experiments is considered not reproducible**, and the vast majority (likely over 80%) of data never makes it to a trusted and sustainable repository.”

NSD NORSK SENTER FOR FORSKNINGSDATA

Source: ”Realising the European Open Science Cloud”. Commission High Level Expert Group on the European Open Science Cloud. 2016. p.9. DOI:10.2777/940154.

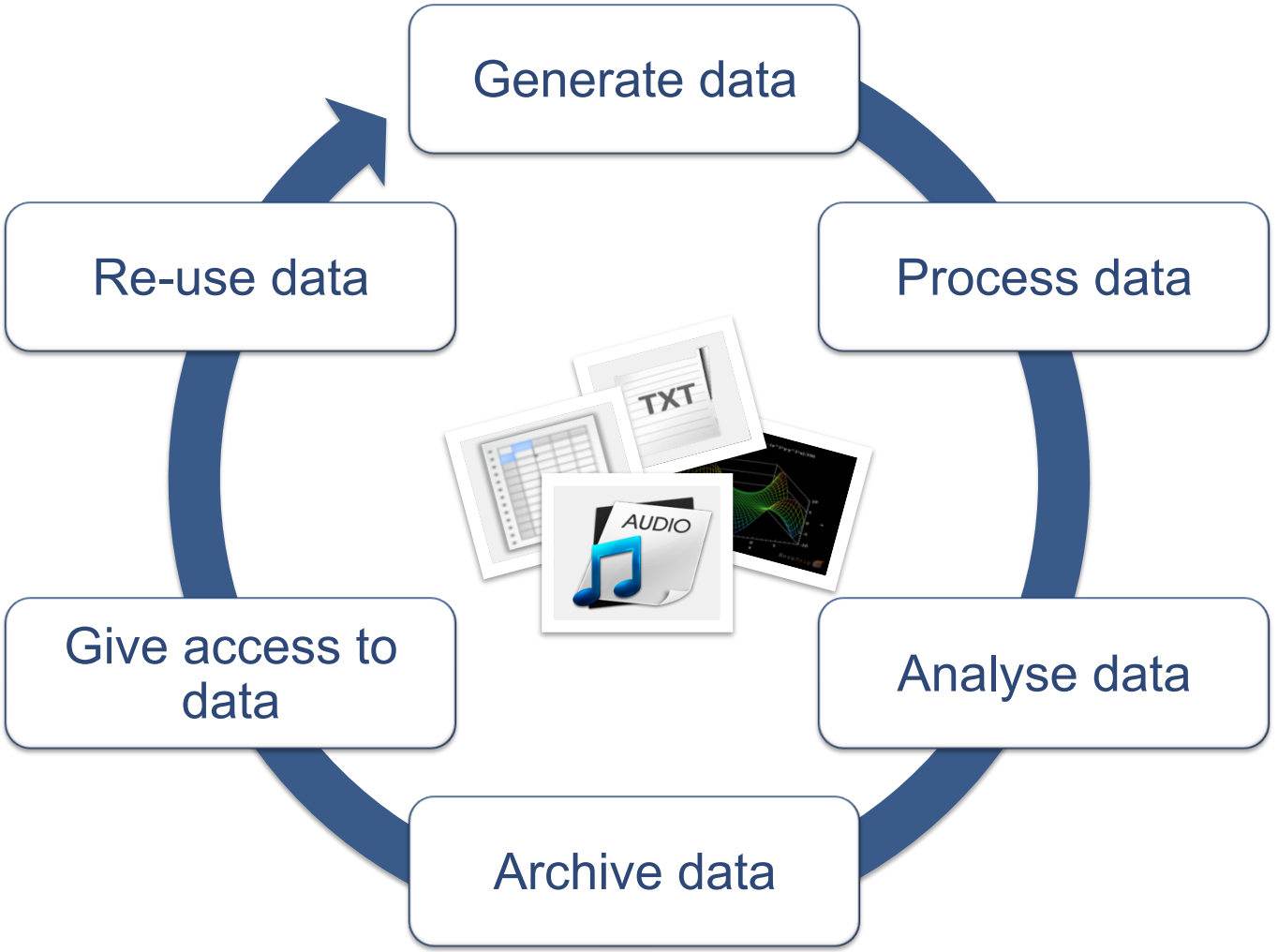
FAIR

Findable
Accessible
Interoperable
Reusable

NSD NORSK SENTER FOR FORSKNINGSDATA

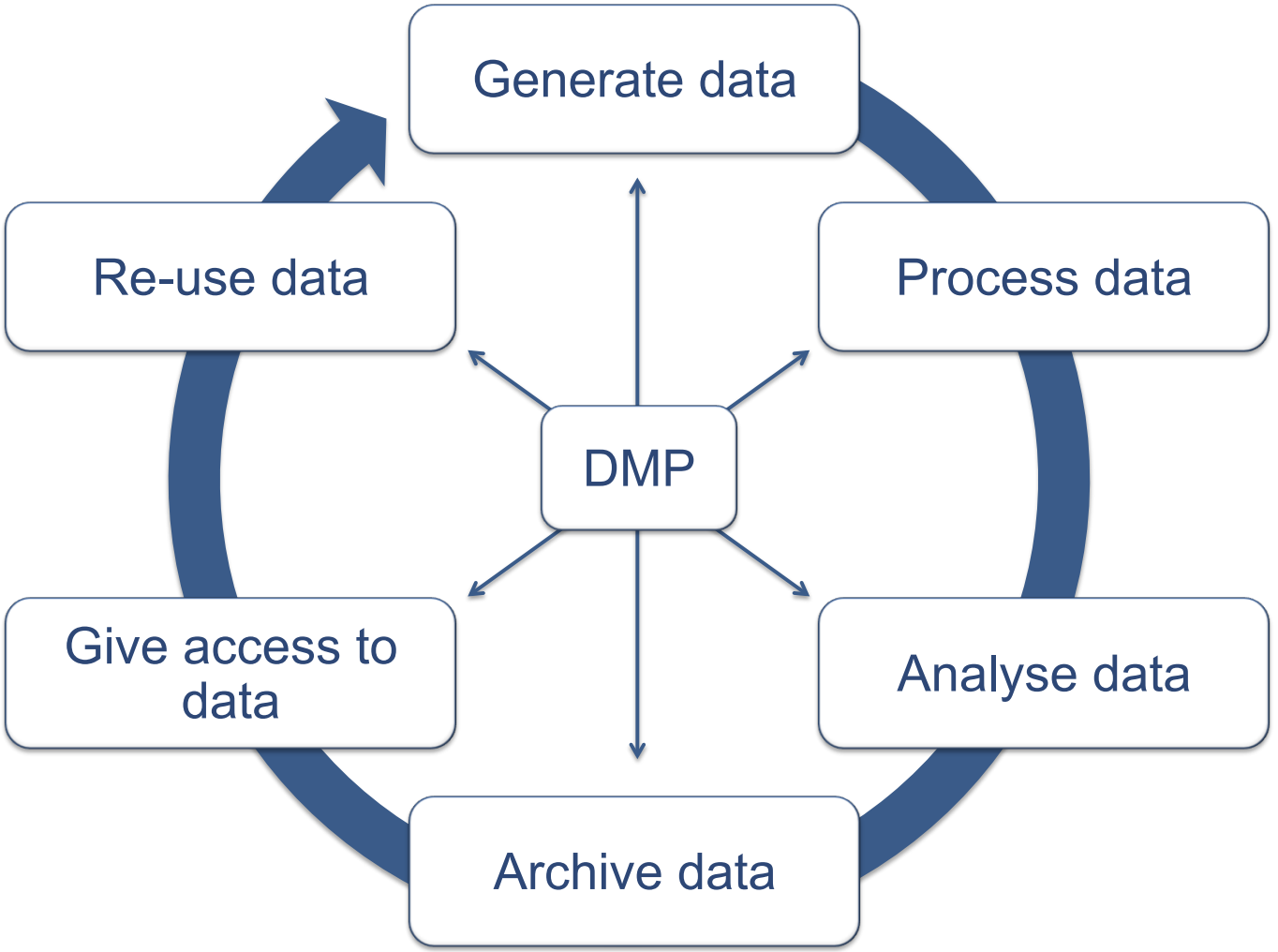
Source: ”Realising the European Open Science Cloud”. Commission High Level Expert Group on the European Open Science Cloud. 2016. p.9. DOI:10.2777/940154.

Data lifecycle



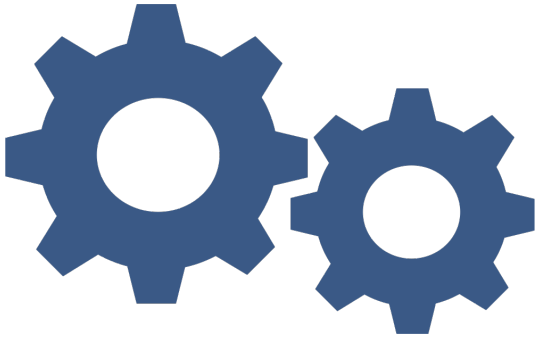
NORSK SENTER FOR FORSKNINGSDATA

Data management plan (DMP)



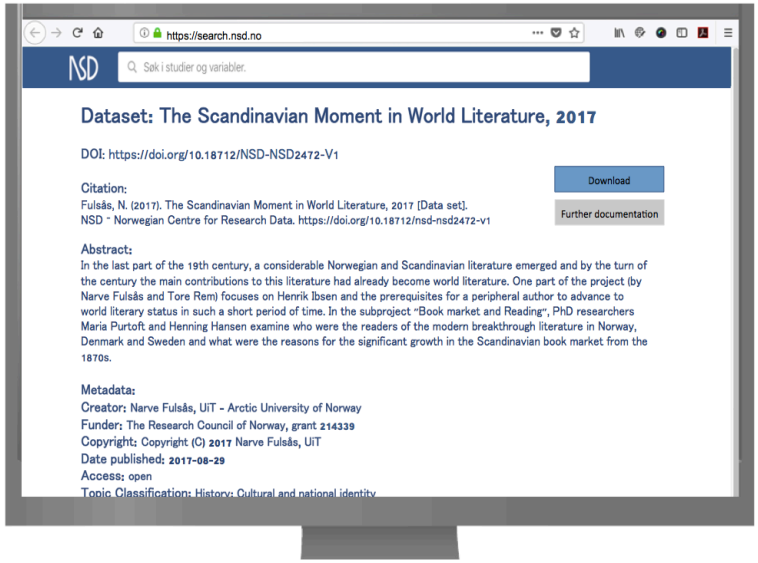
NORSK SENTER FOR FORSKNINGSDATA

Store data vs. archive data



During your project: **store** your data.

- Safe, automated backup and version control
- User-friendly solution to work with them, share and manage access rights for project partners



When (parts of) data is no longer in daily use: **archive** your data.

- **F**indable (searchable, PID, metadata)
- **A**ccessible (authentication&authorization)
- **I**nteroperable (open/persistent file formats, standardized metadata)
- **R**e-usable: Documentation, maintenance/curation of data now and in the future

NSD
NORSK SENTER FOR FORSKNINGSDATA

Sharing data

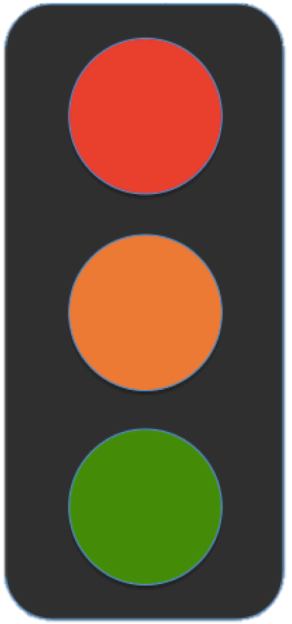
Several modes of sharing, e.g.:

- Downloadable (raw data, processed data)
- Searchable (e.g. online search interfaces)

Several degrees of access:

- Open
- Open w/authentication
- Restricted (conditioned by users/use)
- Closed (e.g. for privacy reasons)

+ embargo



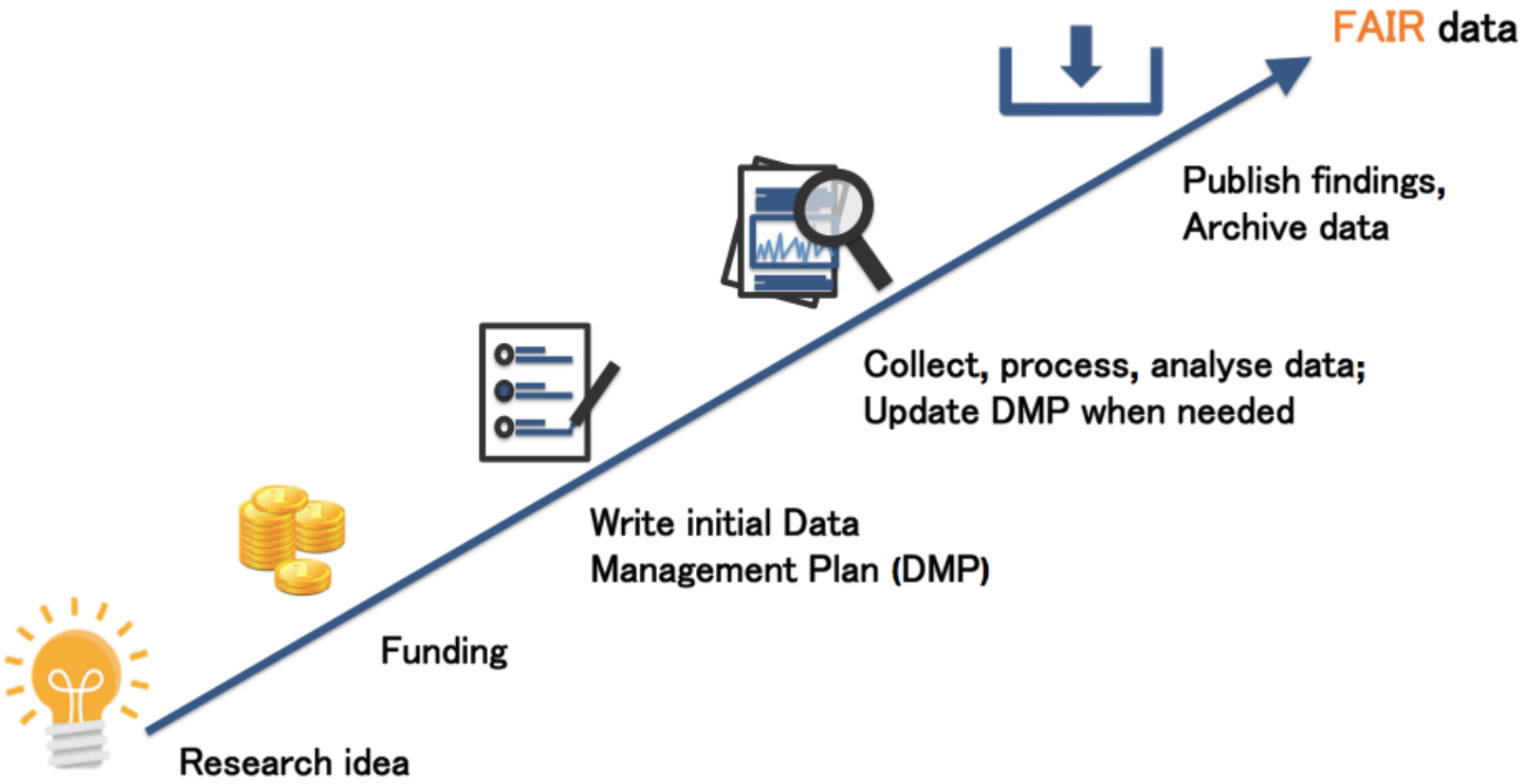
NSD
NORSK SENTER FOR FORSKNINGSDATA

Requirements

- EU (Horizon 2020): data management plan + make data available ”as open as possible, as restricted as necessary” [link];
- Research Council of Norway (RCN) (from 2018): data management plan + “open by default” policy [link];
- Journals: Increasing requirement to share data [link];
- Institutions: data management plan + open access to data (NMBU [link]), University of Oslo [link], UiT – Arctic Univ. of Norway [link], NTNU [link], …);
- GDPR - General Data Protection Regulation 25. May 2018

NORSK SENTER FOR FORSKNINGSDATA

Requirements in practice - NMBU



NSD
NORSK SENTER FOR FORSKNINGSDATA

But…

- *Should all data be shared openly?*

No. But the fact that they exist should be public information.

- *When can data be shared?*

Typically in connection with publications, or at the end of the project (embargo if necessary).

- *Under which conditions & with whom?*

As open as possible, as closed as necessary.

- *Can others find, understand and use my data?*

Yes. With proper documentation and long-term curation.

NSD
NORSK SENTER FOR FORSKNINGSDATA

NORSK SENTER FOR FORSKNINGSDATA

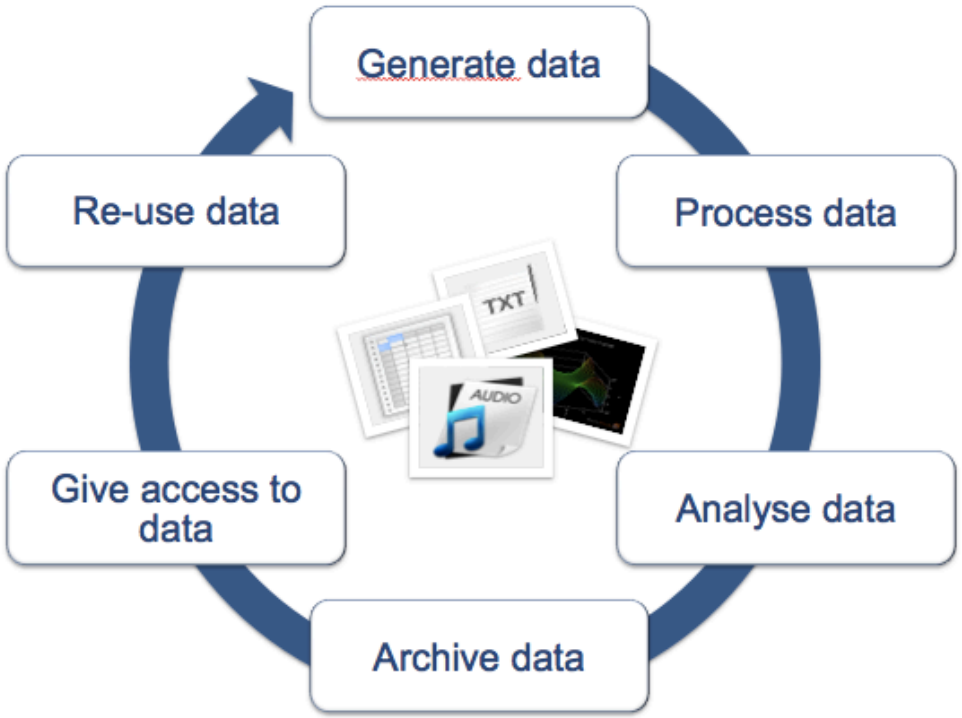
Source: https://www.cartoonstock.com/cartoonview.asp?catref=rman9976

WHY – What's in it for you?

- Visibility
- Documentation
- Security (long-term)
- Integrity and trust
- Network, collaboration

NSD
NORSK SENTER FOR FORSKNINGSDATA

HOW - Research data management



Project start

- What data do you need? Data search?
- Identify possible legal & ethical issues
- Costs (e.g. for archiving data)

During project

- Safe storage (for working data)
- How/where to share efficiently w/project partners?
- Backup, versioning
- Make conventions for naming of files/folders/code
- Document what you do -> readme.txt, log.txt, remember.txt, link to the DMP (e.g how are files and code organized?)

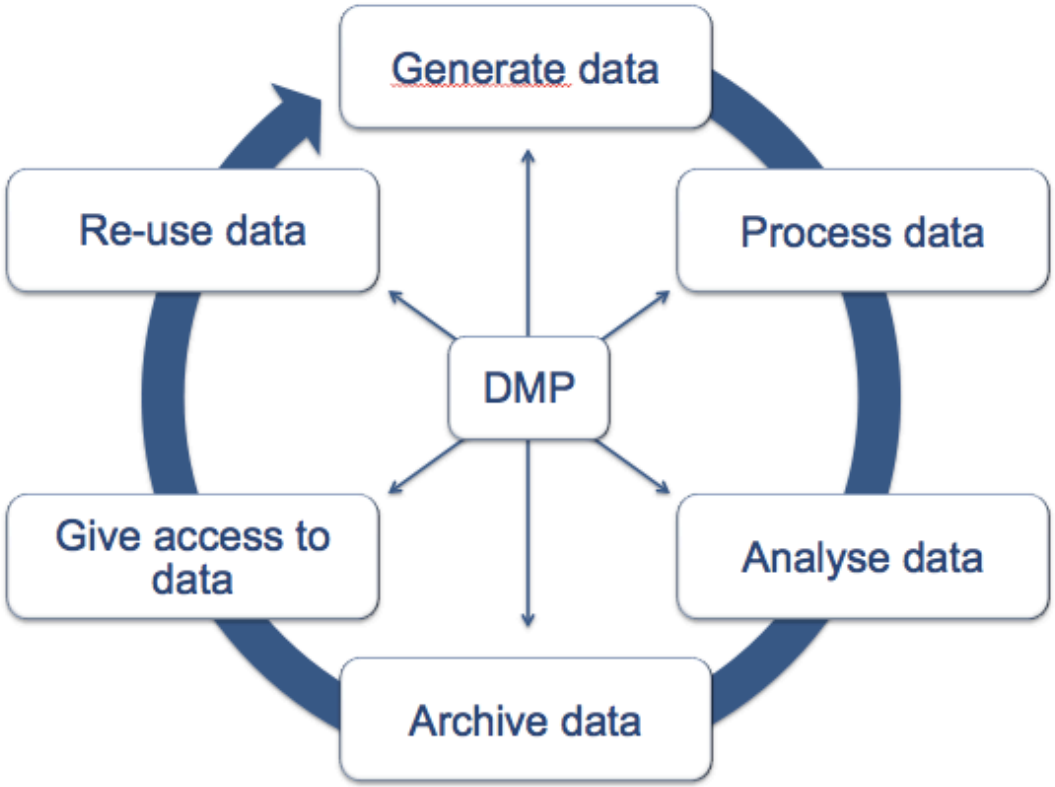
End of project

- Select data for archiving (for DOI assignment etc)
- Are the file formats ok?
- Final documentation and metadata

→ **Time saved if you documented from the start**

NSD
NORSK SENTER FOR FORSKNINGSDATA

Make a data management plan (DMP)

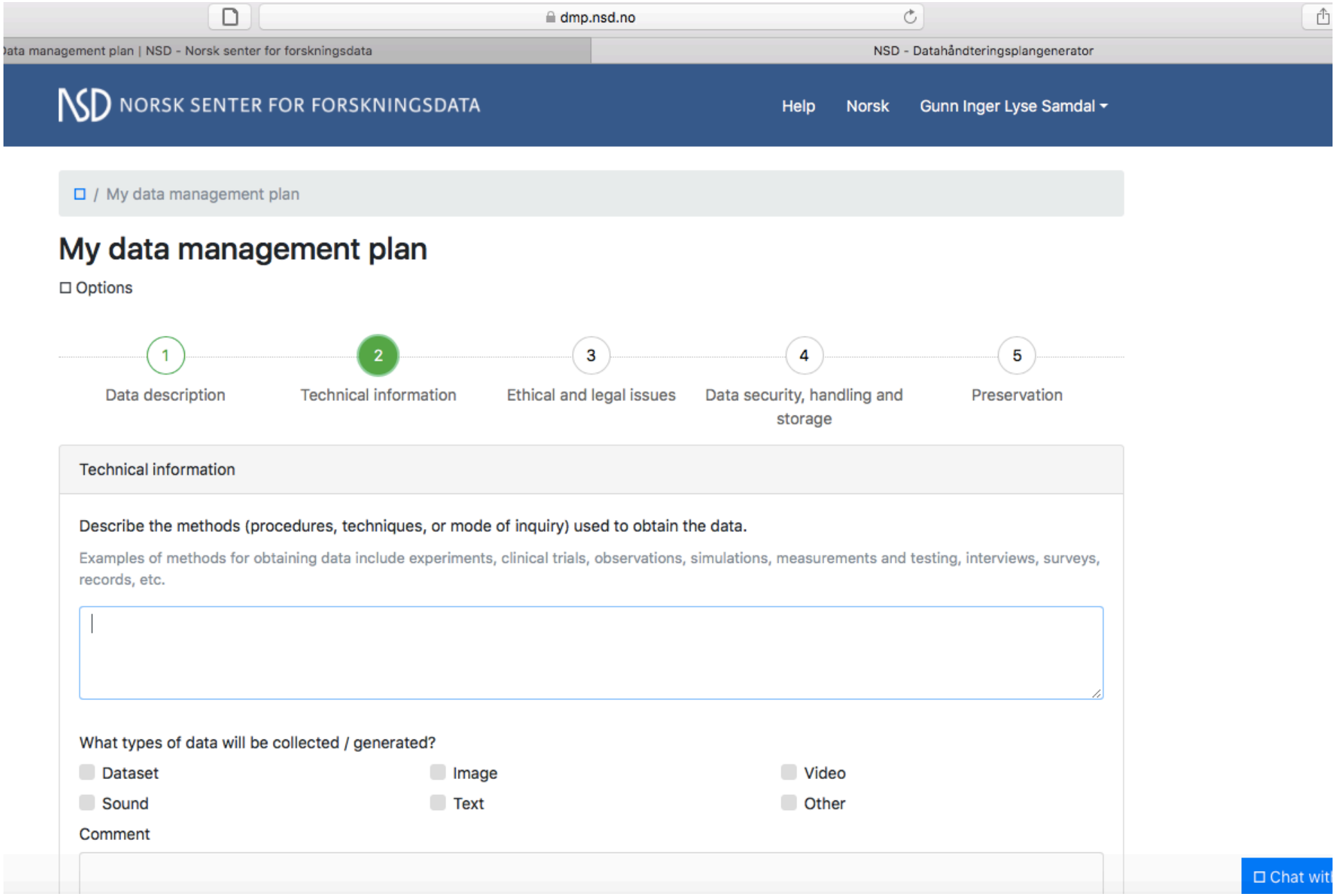


DMP as a document that covers the whole cycle

A checklist with questions to make you aware of what you need to consider and when

A dynamic document for continuous updating.

NORSK SENTER FOR FORSKNINGSDATA



NSD NORSK SENTER FOR FORSKNINGSDATA

How to choose a data archive/ repository?

-National? Community repository? International?

-Provide persistent identifier (PID)? DOI, handle,…

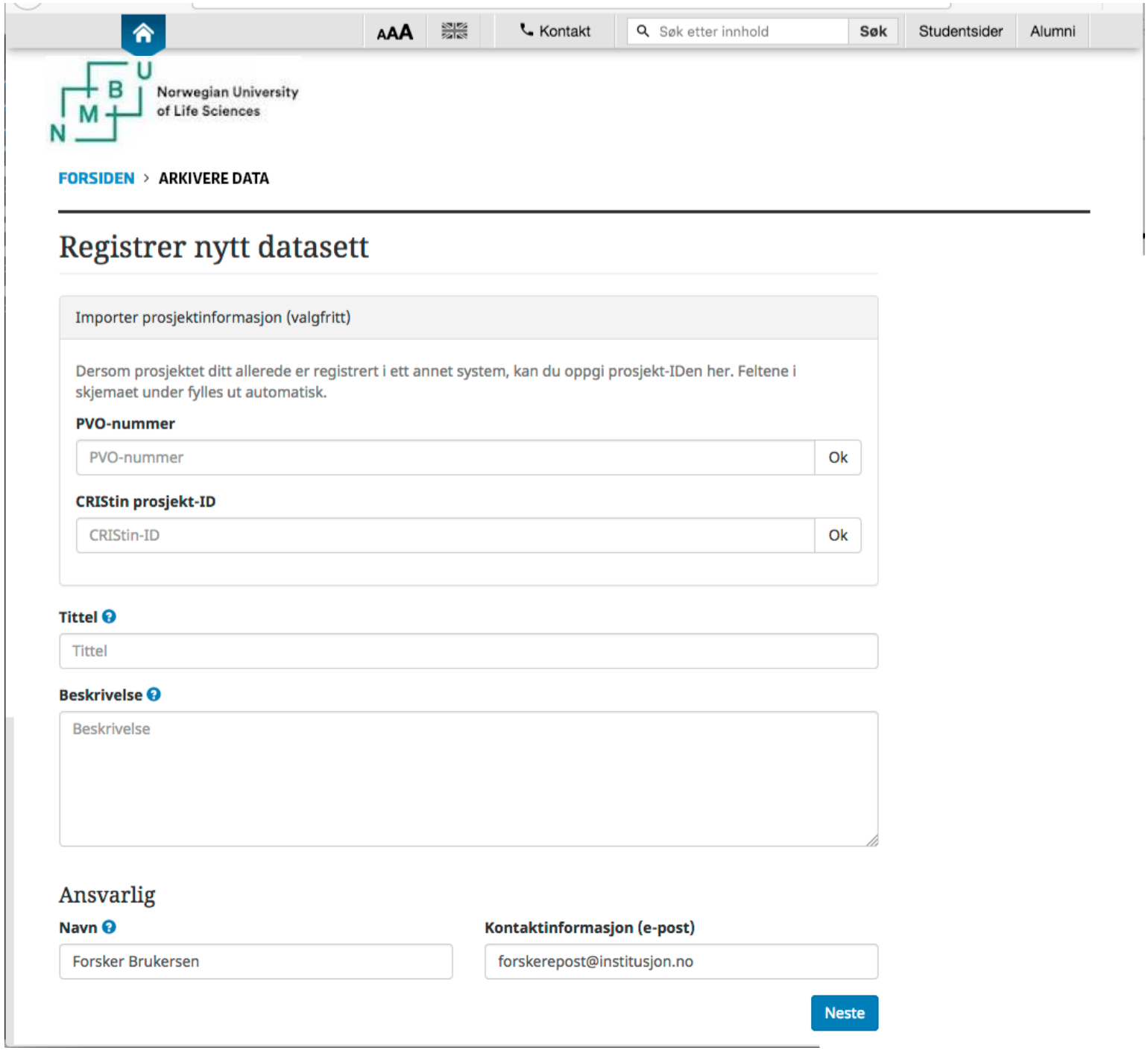
-Certified?

-Standardised metadata (ensure “F” in FAIR)

-Capacity, guidance and costs? E.g. how many GB, TB are you allowed to archive?

=> The largest and most comprehensive registry of data repositories available

NSD
NORSK SENTER FOR FORSKNINGSDATA


AAA
Kontakt
Søk etter innhold
Søk
Studentsider
Alumni
Norwegian University of Life Sciences
FORSIDEN › ARKIVERE DATA
Registrer nytt datasett
Importer prosjektinformasjon (valgfritt)
Dersom prosjektet ditt allerede er registrert i ett annet system, kan du oppgi prosjekt-IDen her. Feltene i skjemaet under fylles ut automatisk.
PVO-nummer
PVO-nummer
Ok
CRIStin prosjekt-ID
CRIStin-ID
Ok
Tittel
Tittel
Beskrivelse
Beskrivelse
Ansvarlig
Navn
Forsker Brukersen
Kontaktinformasjon (e-post)
forskerepost@institusjon.no
Neste

FAIR research data at NSD

Findable:

- Archived datasets are given a Persistent ID (DOI);
- Indexed and searchable metadata (e.g. via DataCite).

Accessible:

- Open access protocol;
- Clear procedure for authentication and authorization.

Interoperable:

- NSD metadata follows the DDI standard, controlled vocabulary;
- Data available in different formats.

Re-usable:

- Clear user conditions (decided by you, not NSD)
- NSD curates your data and can return them in formats relevant to your research community.

NSD
NORSK SENTER FOR FORSKNINGSDATA

Support for research from NSD

Data management and planning:

NSD webpages on data management and archiving (English):
http://www.nsd.uib.no/arkivering/en/data-management-plan.html
FAQs about data management: http://www.nsd.uib.no/arkivering/en/005_faq.html
Find data? Search portal: www.search.nsd.no
Archive data? Archiving portal: https://arkiveringsportalen.nsd.no/
Data management plan: www.dmp.nsd.no (independent of where you will archive your data)
For questions, contact: dataarkivering@nsd.no

Personal data:

FAQs about personal data: http://www.nsd.uib.no/personvernombud/en/help/faq.html
Unsure whether you have personal data? Try NSD's notification test:
http://www.nsd.uib.no/personvernombud/en/notify/index.html
For questions, contact: personvernombudet@nsd.no

NSD
NORSK SENTER FOR FORSKNINGSDATA

Support for research from NSD

Coming:

Digital training (e-course, videos)

NORSK SENTER FOR FORSKNINGSDATA

Other useful webpages

International data search: www.datacite.org
(all NSD data are also findable here)

Search for repositories: www.re3data.org

Online training on data management in Europe, developed by the infrastructure CESSDA:
https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management

NSD
NORSK SENTER FOR FORSKNINGSDATA