Western Norway University of Applied Sciences

The Data Management Plan

It’s for your own good

Kjetil Sletteland
Bergen
January 24th 2024

The Research Council of Norway's requirements for data management plans and data sharing

- Data management plans are obligatory for all projects that manage data and receive funding from the Research Council
- Data sharing: As open as possible, as closed as necessary
  - RCN: "Data sets shall not be made openly available if doing so may threaten individual or national security, conflict with applicable data protection regulations or other legal provisions."

- It's easier to share your data and enable reuse if you plan carefully
  - Make sure you have the permissions you need
  - … and that personally identifiable data can be left out without compromising integrity
- If you can't share your data, planning is even more important
  - Secure storage, access control, etc.

https://www.forskningsradet.no/en/research-policy-strategy/open-science/research-data/

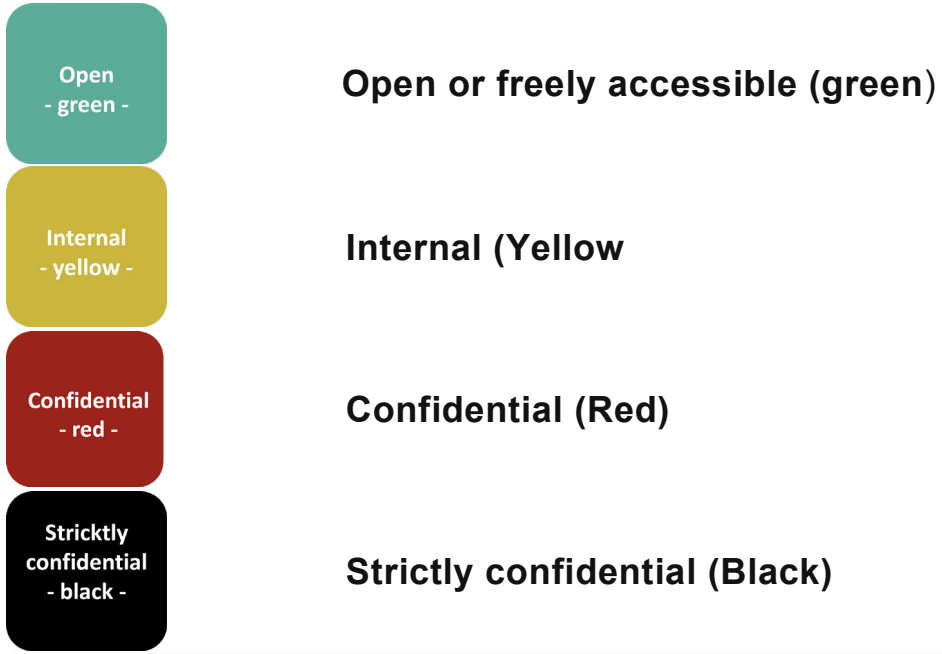
Why share your data?

- RCN: “Better access to research data enhances the quality of research, both because results can be validated and verified in a better way, and because datasets can be used in new ways and in combination with other datasets. Open access to research data also contributes to fewer duplications and unnecessary duplication of effort and will facilitate more interdisciplinary research.”

- And yes, you have to. Unless you can’t.
- There are some advantages for you as a researcher:
  - Increased number of citations (also due to data reuse)
  - Enhanced collaboration opportunities
  - Transparency makes things better when you commit errors

Classification of data

- Green – no personal data
- Yellow – general but not sensitive person-identifiable information. Unpublished research data
- Red – sensitive personal information. Health data,
- Black – sensitive information that can cause considerable harm if it becomes known to unauthorised people
- **Classification of data does not only determine whether data can be shared openly, but also how it can be stored throughout its lifecycle**



https://www.hvl.no/en/research/research-ethics/guideline_personal_identifiable_and_health_data_in_research/guide-for-processing-research-data-at-hvl/

Personal information

Internal (Yellow)

- Anything that can be used to identify a particular person
- Directly:
  - Name, ID, address, phone number, pictures, audio recordings
- Indirectly:
  - Combinations of background information (eg. gender + age + occupation + location)

https://www.hvl.no/en/research/research-ethics/privacy-and-the-use-of-personal-data-in-research/

Sensitive personal information etc.

Confidential (Red)

- Information about race, ethnicity
- Political, philosophical or religious attitudes and convictions
- Union membership
- Genetics, biometrics
- Medical information
- Sexuality

https://www.hvl.no/en/research/research-ethics/privacy-and-the-use-of-personal-data-in-research/

Strictly confidential information

Strictly confidential (Black)

- If you need to know, you've probably already been told

A goal: Sharing **f**indable, **a**ccessible, **i**nteroperable and **r**eusable data

- The FAIR principles – https://go-fair.org

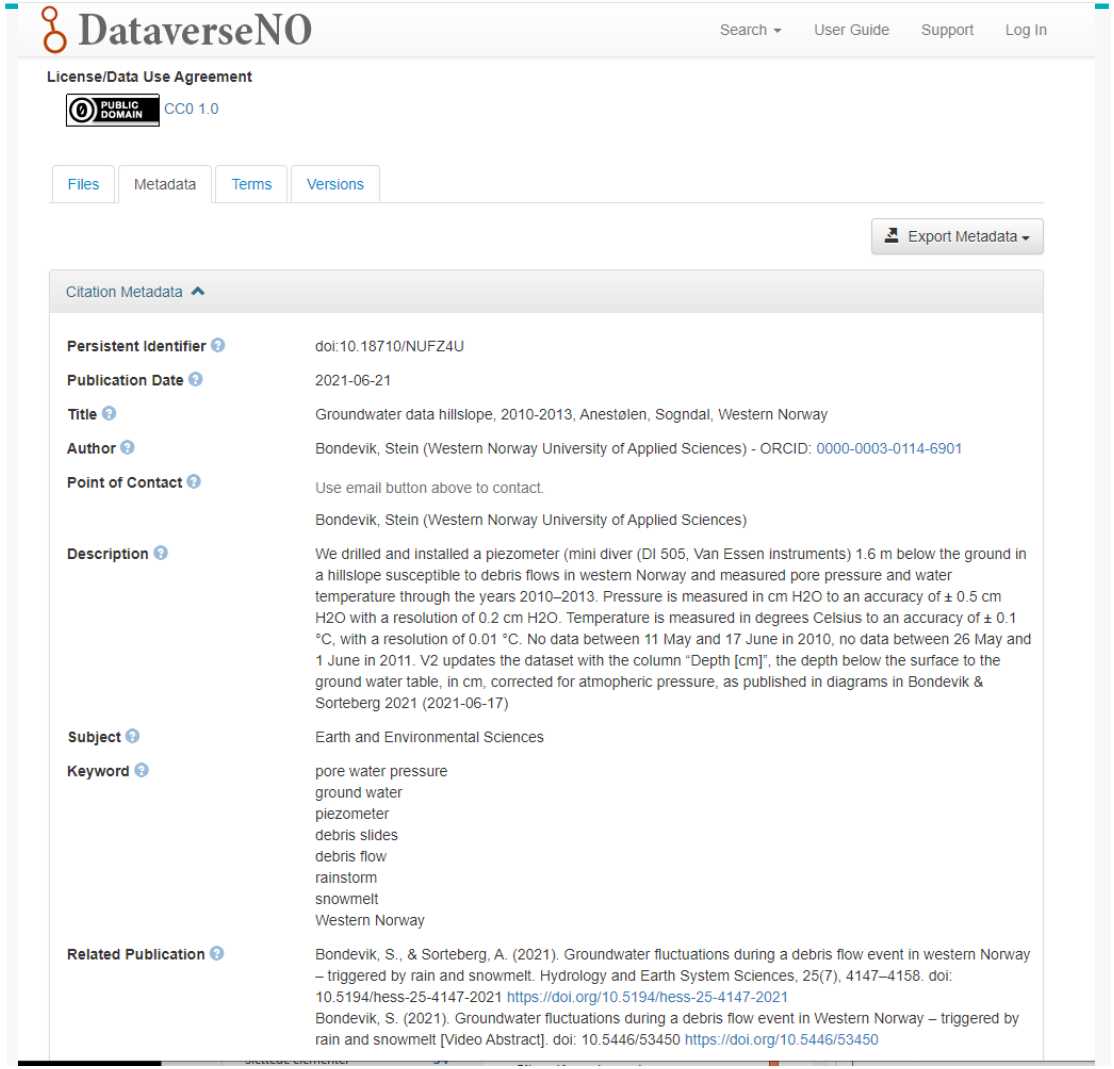
- The principles apply to both data and metadata
- A (lofty) goal is to make (meta)data **machine-actionable** – i.e. the possibility to be found, accessed and reused by computer systems without human intervention
- Standardised and open protocols and formats are key
- … as are documentation and rich metadata
  - Documentation is important for human reuse as well!

- Data can be FAIR even if it isn't shared openly (and openly shared data is not necessarily FAIR)
- Making yellow data green is (is some cases) easy

DataverseNO as a FAIR repository

- Findable:
  - Persistant identifier – DOI
  - Rich metadata
  - Indexed in searchable resources
- Accessible
  - Avaiable over an open protocol (http)
- Interoperable
- Reusable

- ^ The last two depend on some effort from the researcher: Use open formats and languages, document your data
- DataverseNO is purely Open

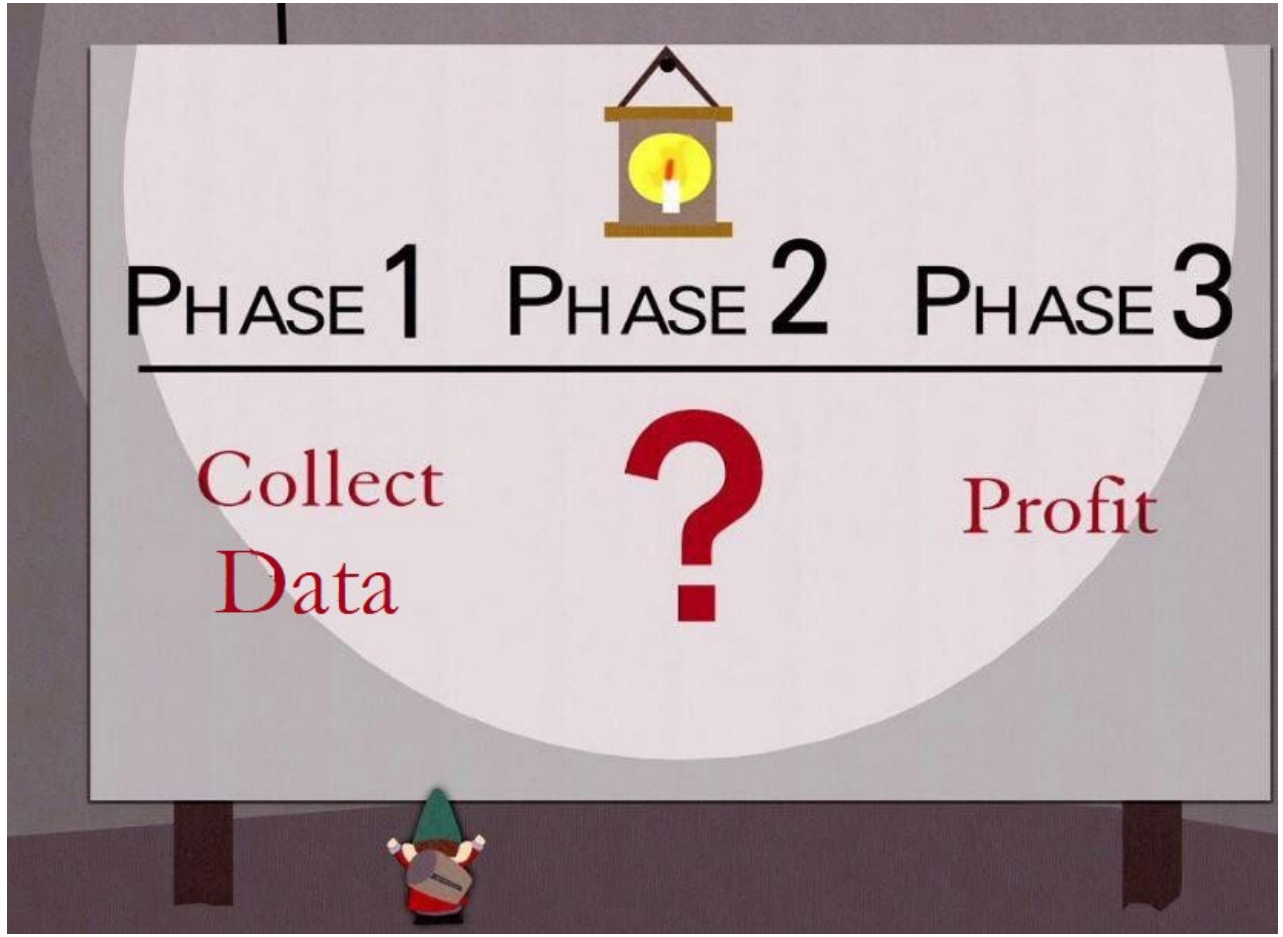


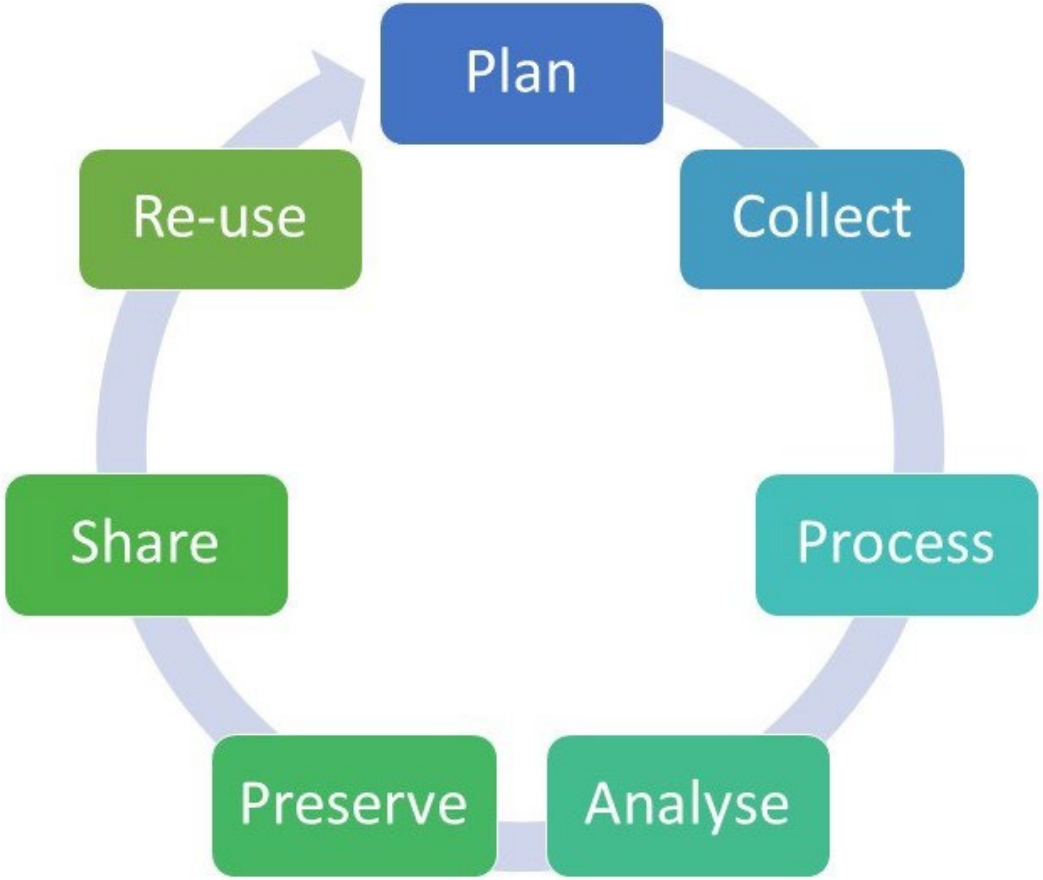
The DMP – what is it?

- A document describing how research data is to be managed throughout its lifecycle
- Both a formal requirement and a **living document**, to be updated and revised
- A tool, not a goal in itself
  - Making a DMP that you can understand and make use of is probably more important than making one that conforms to the Research Council's expectations

- Sikt's DMP tool is intended for personal data

The research data lifecycle





Left: Adapted from T. Parker & M. Stone (1998) “Gnomes”, South Park S02E17
Right: Pilfered from University of Reading (n.d.) The research data lifecycle, https://www.reading.ac.uk/research-services/research-data-management/about-research-data-management/the-research-data-lifecycle

Requirements for DMPs

Sikt:

- Who is responsible for taking care of the data during and after the project, and what resources are needed.
- How you will ensure that the data is well organised and adequately documented.
- The volume and type of data set is generated/used.
- How you will ensure that your data is ethically and legally compatible.
- Where you will securely keep the data during the project's lifetime.
- Plan for long-term preservation and making your data available to others.

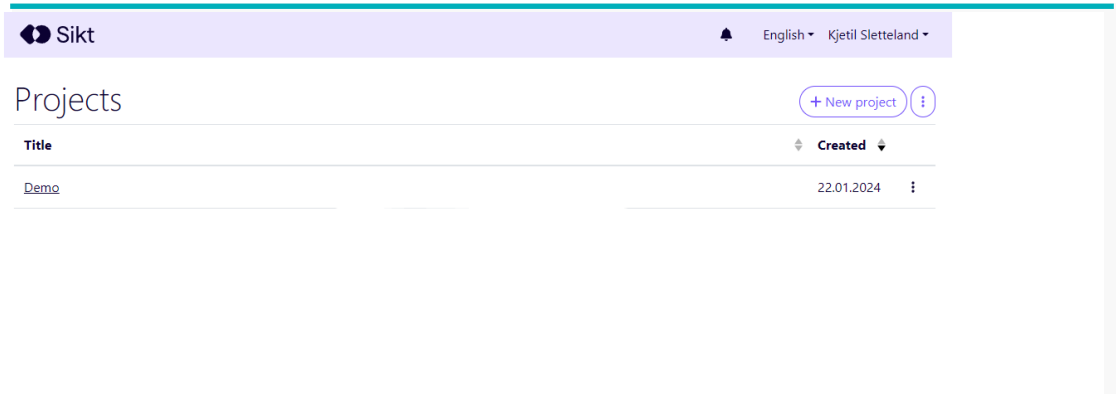
Science Europe:

- Data description and collection or re-use of existing data
- Documentation and data quality
- Storage and backup during the research process
- Legal and ethical requirements, codes of conduct
- Data sharing and long-term preservation
- Data management responsibilities and resources

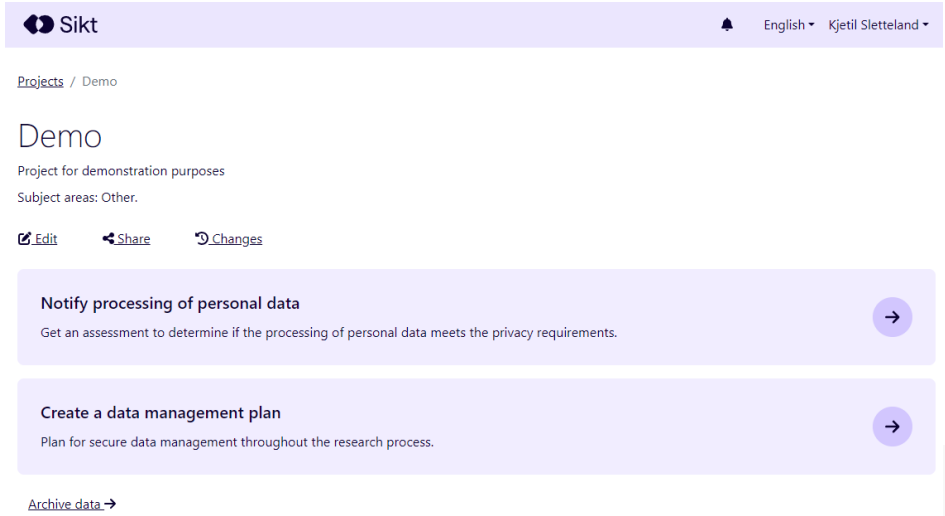
https://sikt.no/en/data-management-plan
https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf

Let’s make a DMP!

- Go to https://minforskning.sikt.no/ and log in
- Make a project (or enter one)

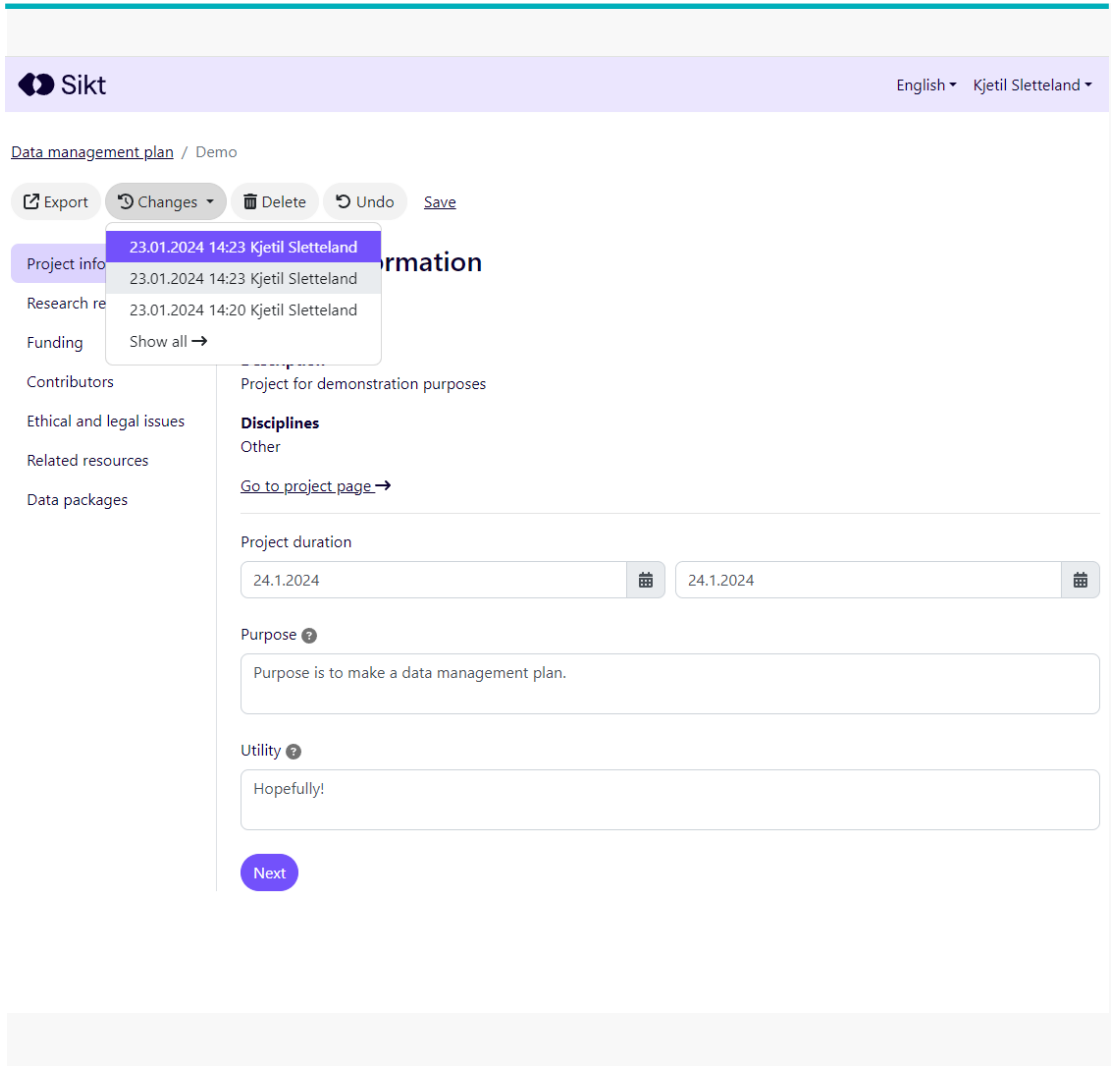


- Click “Notify processing of personal data” if you need to get a privacy assessment or just a reminder of what personal data is – which could be useful!
- “Create a data management plan”

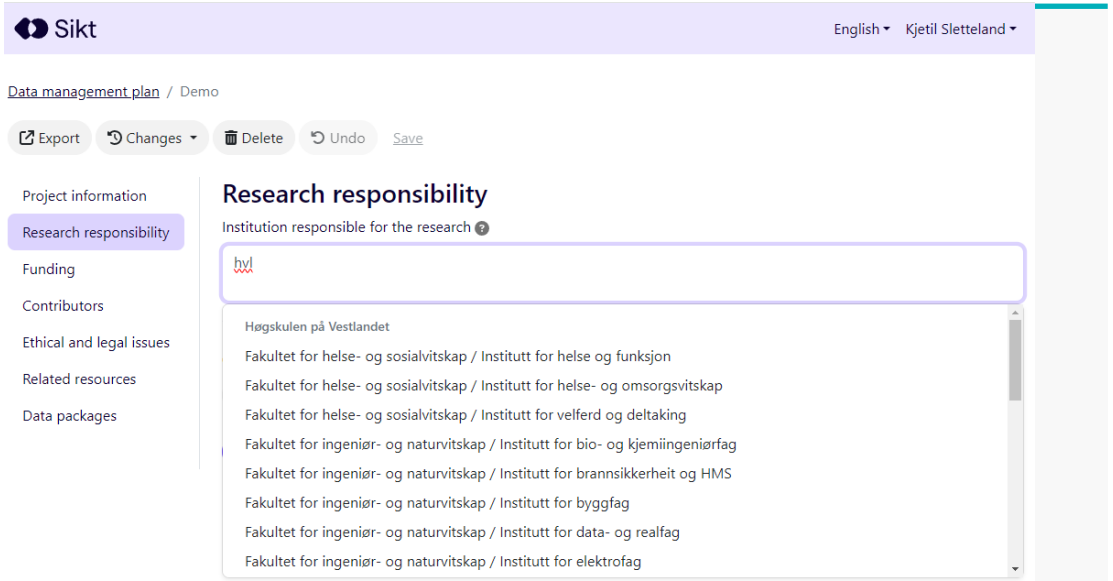


Start

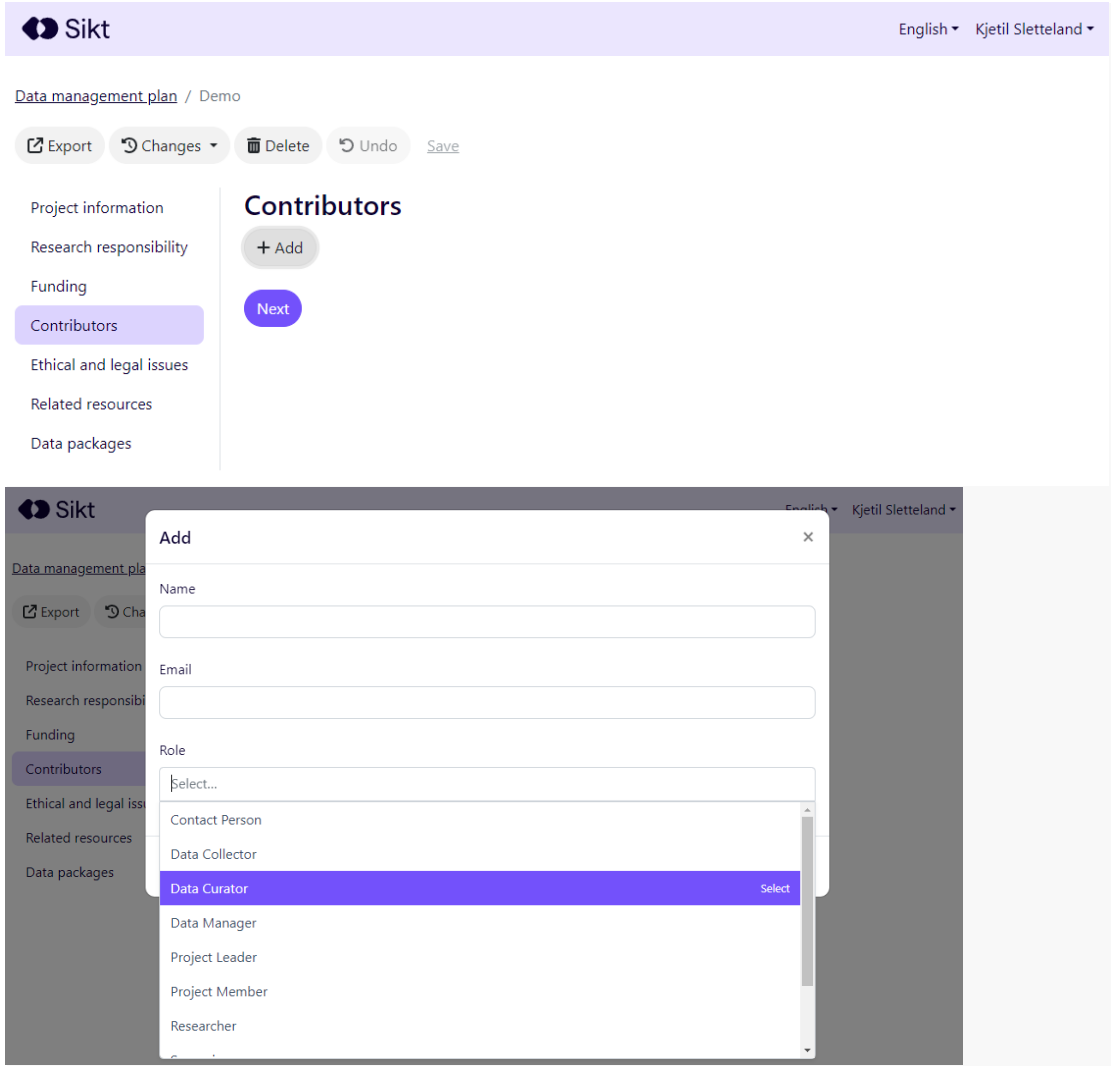
- First page is for general info about the project
- The DMP tool tracks changes, which is nice
- You can hit the question marks for descriptions of fields
- Purpose:
  - “Describe the research questions, issues and hypotheses that are addressed in the project and state how the data collection / generation can shed light on these issues.”
- Utility:
  - “Describe the potential user groups for your data. This may be a subject area or discipline, occupational groups, researchers within specific fields, institutions, organizations, etc.”



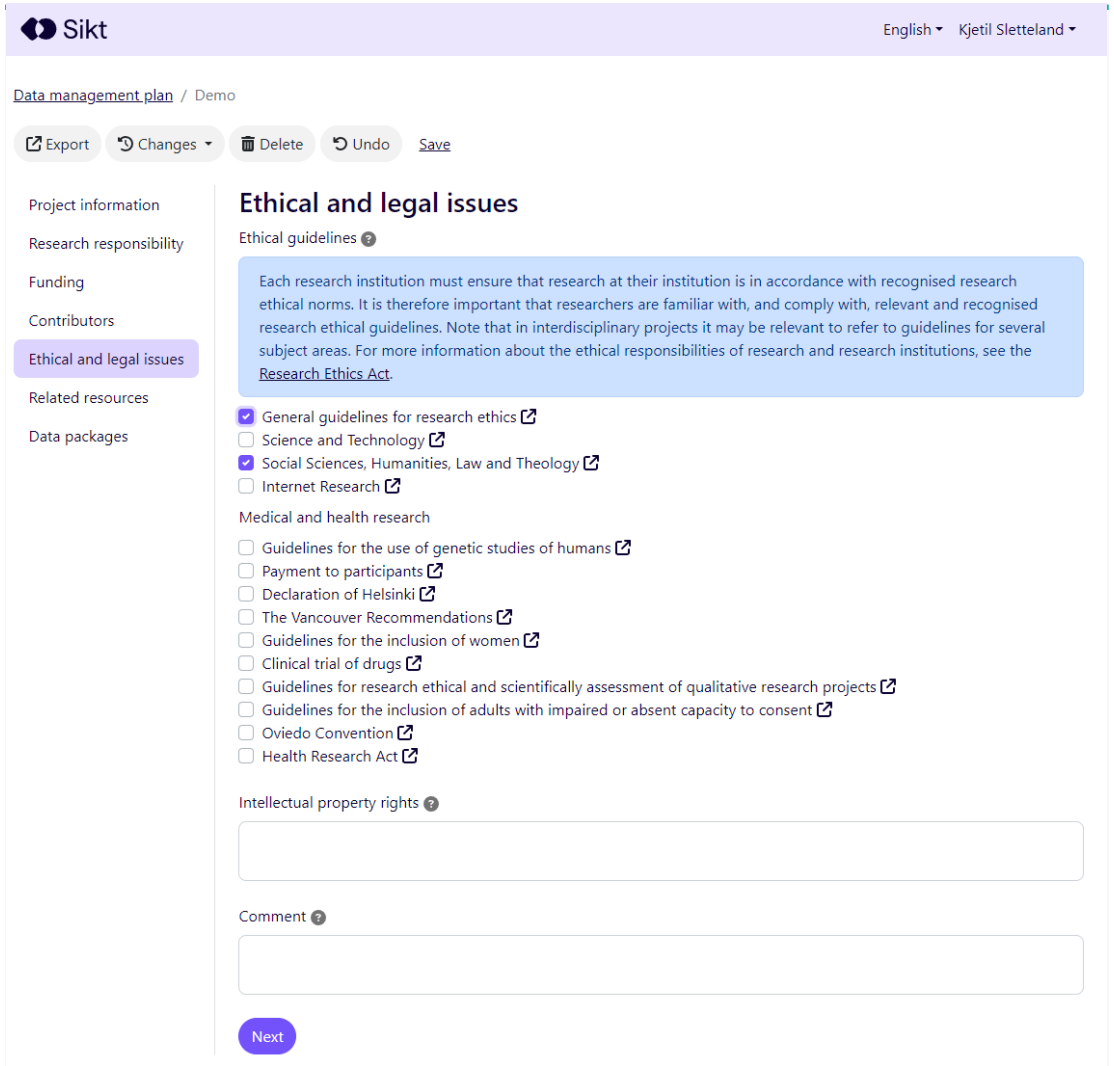
2, 4 …



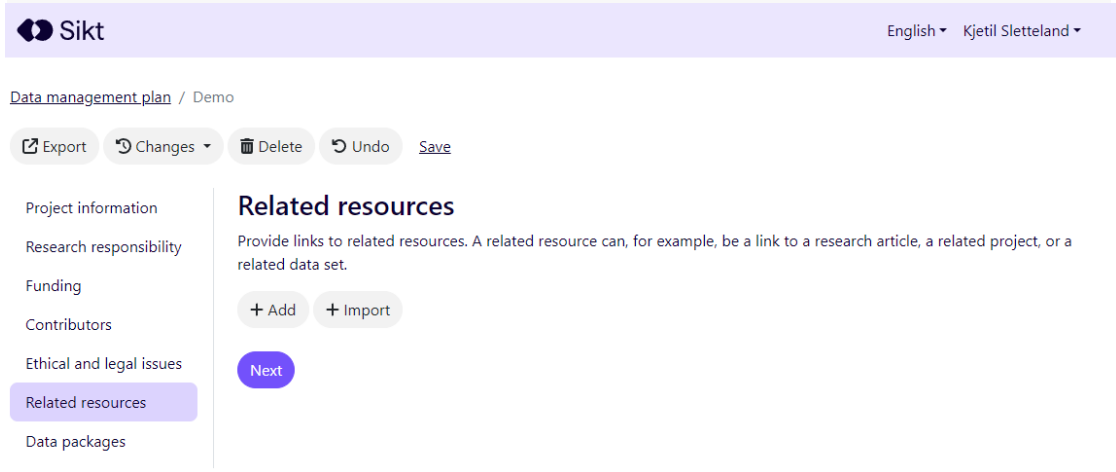
- You can also list additional contributing institutions and unlisted institutions
- I'll skip funding
- Contributors can have multiple roles



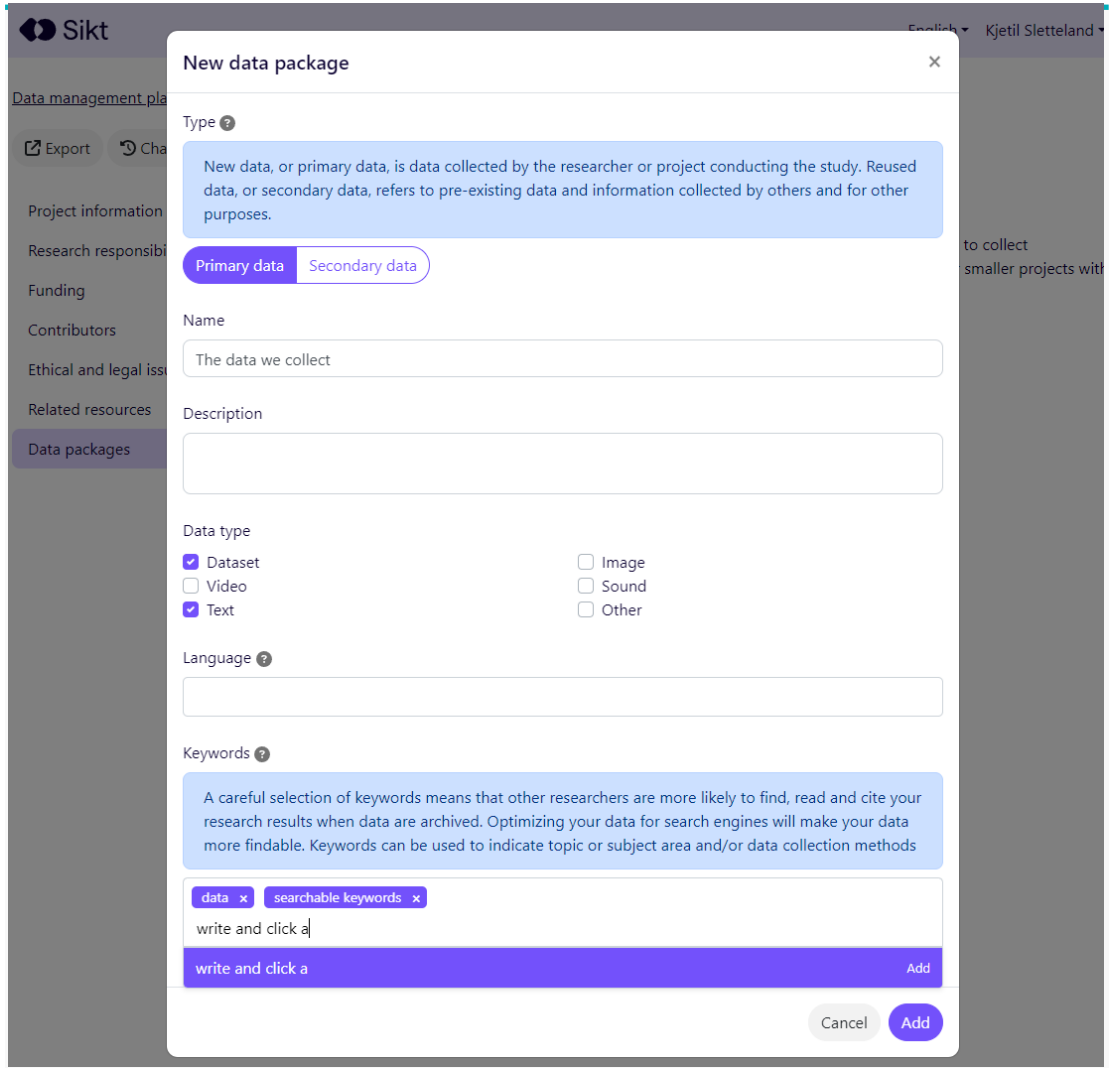
5, 6



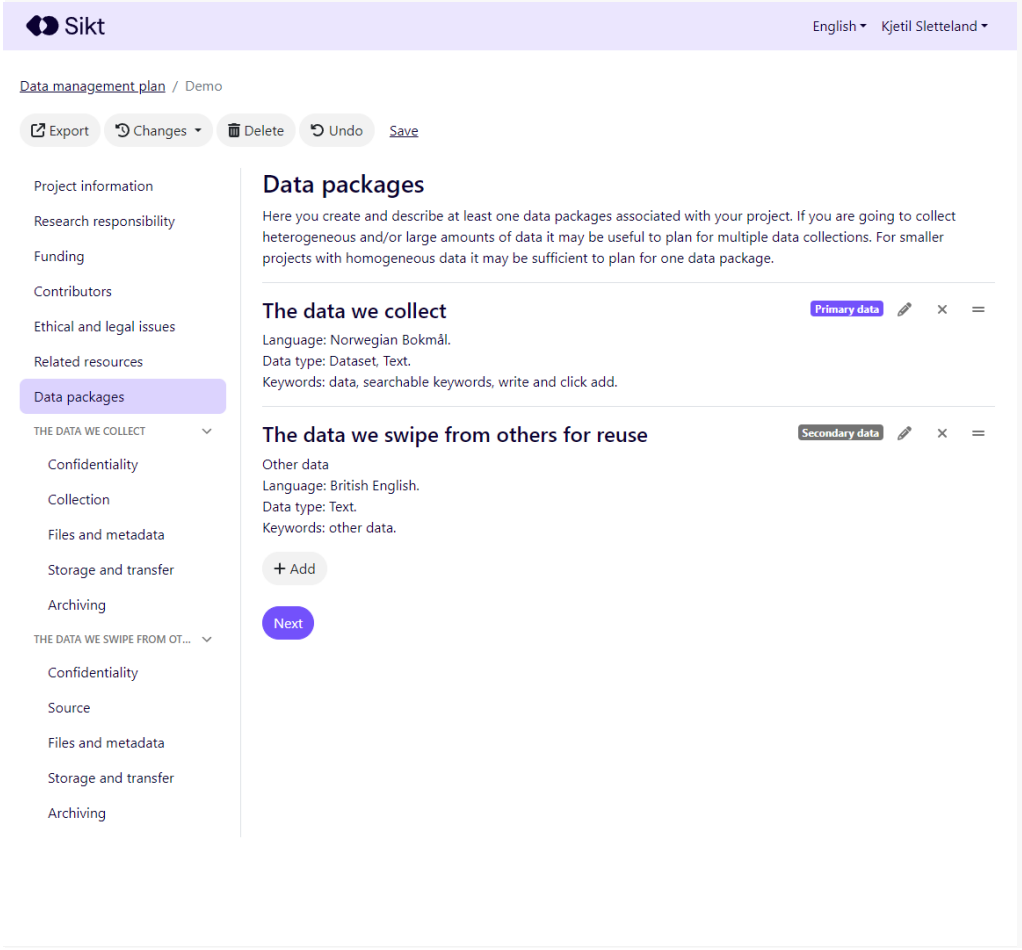
› Related resources = references, links



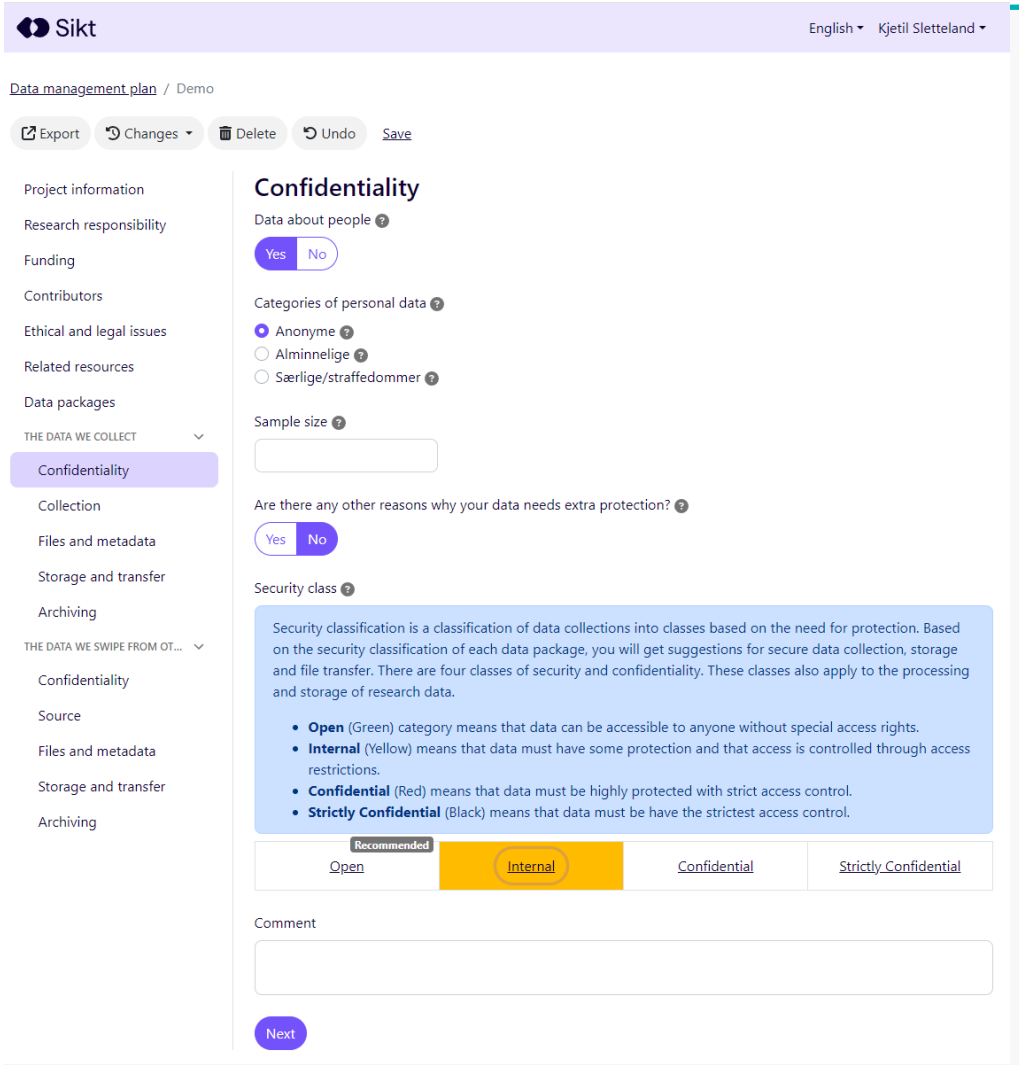
Data packages



› Just when you thought you were done



Confidentiality // Classification of data



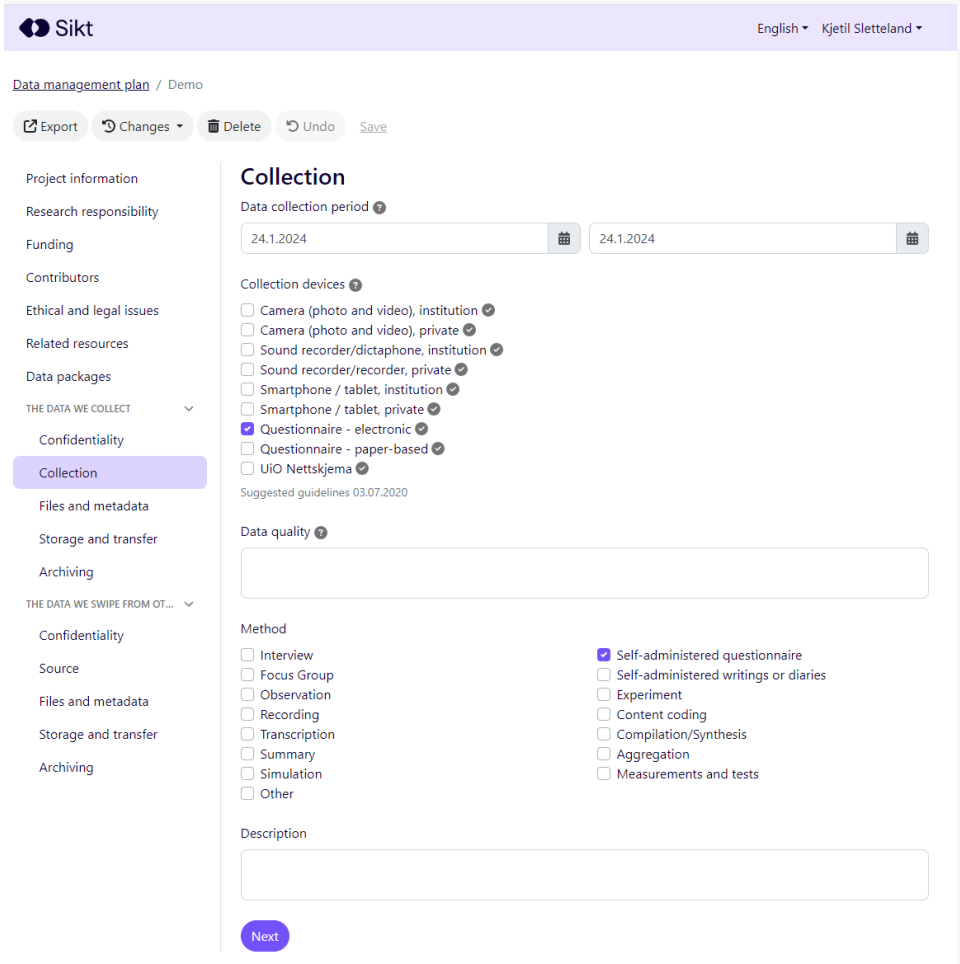
← They forgot to translate a bit, but the help text is available in English

- Remember the security classes from slide 4? The DMP tool makes recommendations based on your selection.
- You can still select a different class than the one recommended
- If your data is confidential, you need to use Secure Storage of Research Data (SILAF)

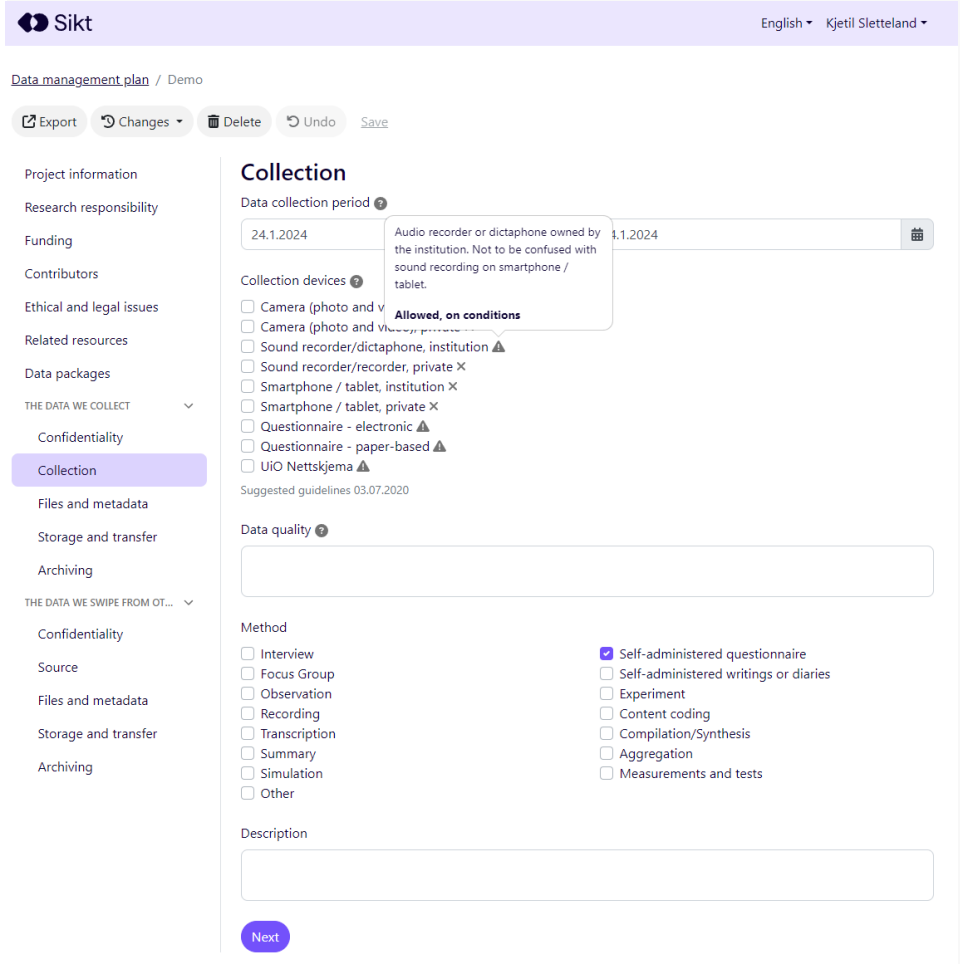
https://www.hvl.no/en/hvl-students/it-tools-and-help-desk/secure-storage-of-research-data-silaf/

Collection devices available will depend on the class selected on the previous screen

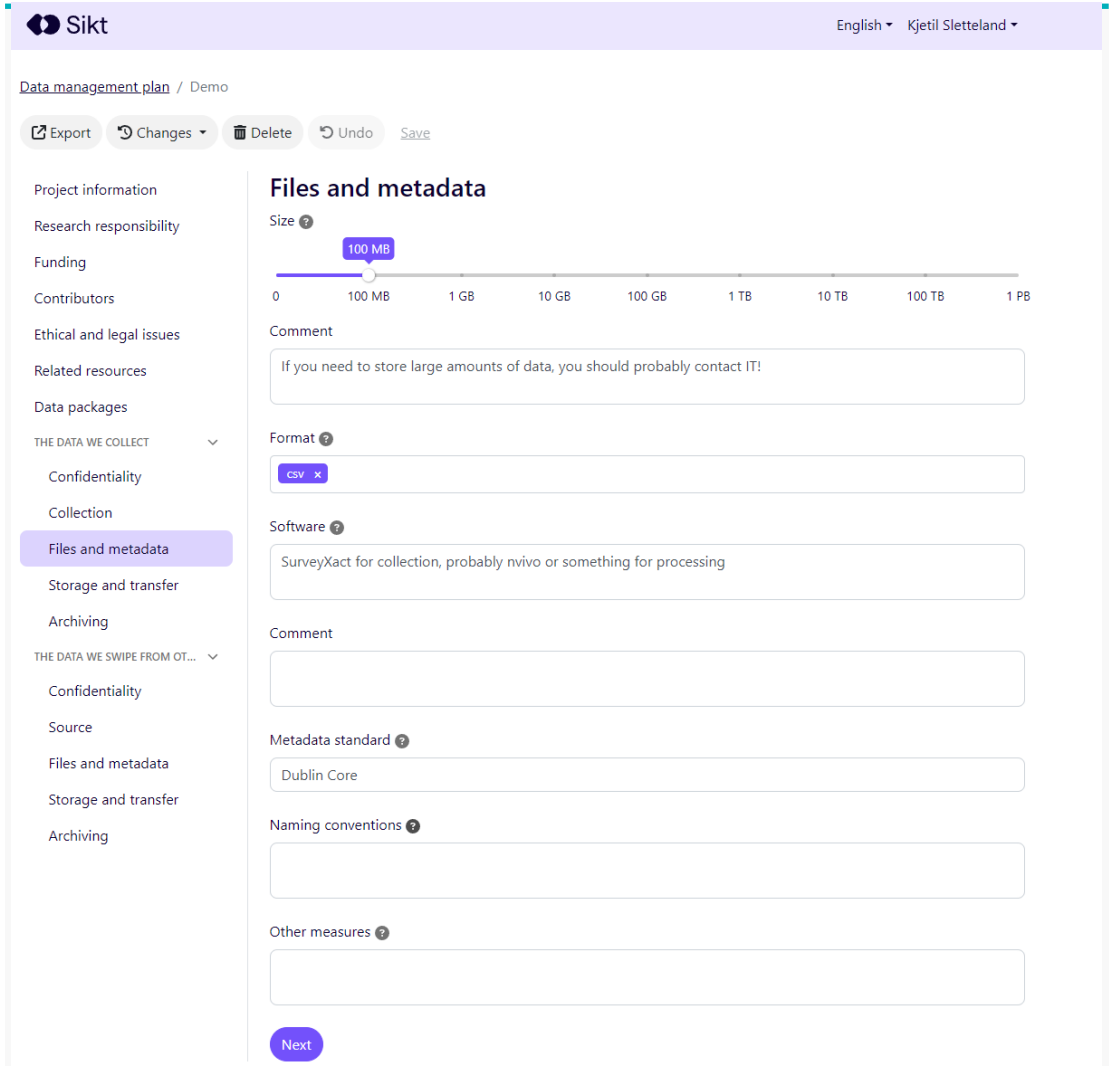
› Green



› Red



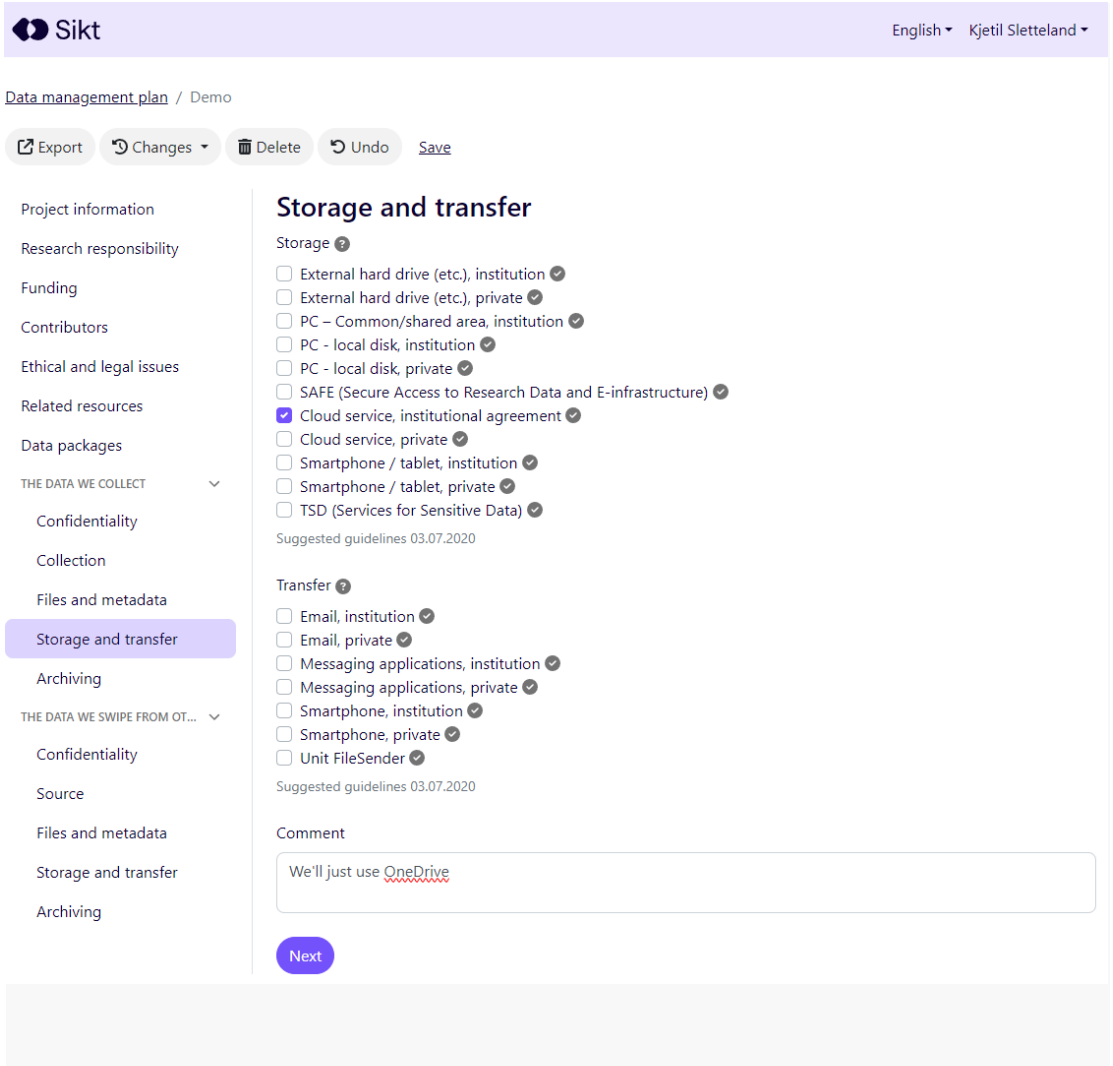
Files and metadata



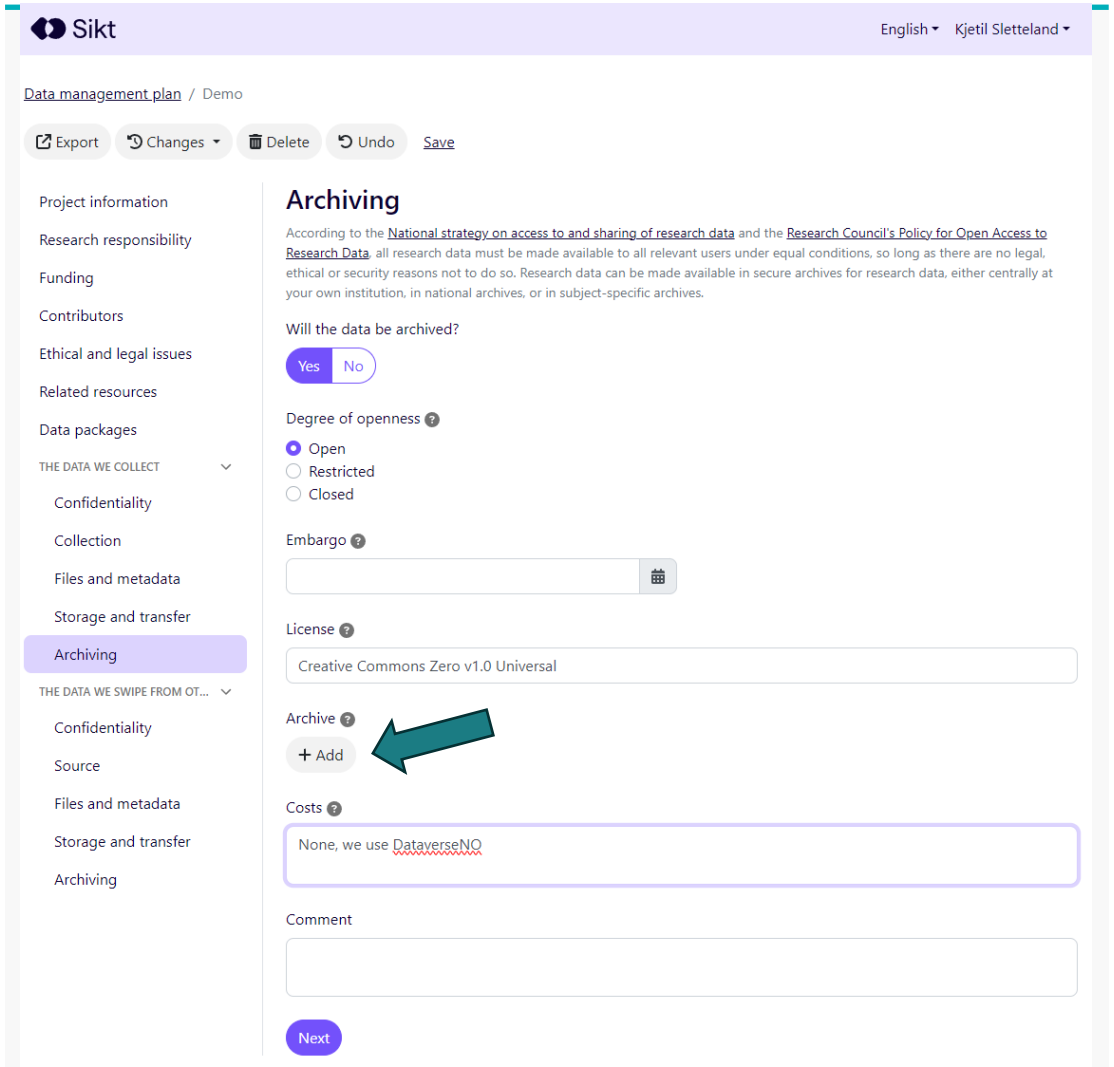
- Many of these are important to think through!
- Do you have big files? Lots of files?
- Which format do you store your files in?
  - Best use something that can be converted to an open format
- Metadata standard?! If you use DataverseNO, you can export to several, so let's select Dublin Core
- Naming conventions
  - If you have lots of files, use a consistent naming scheme and a hierarchical folder structure

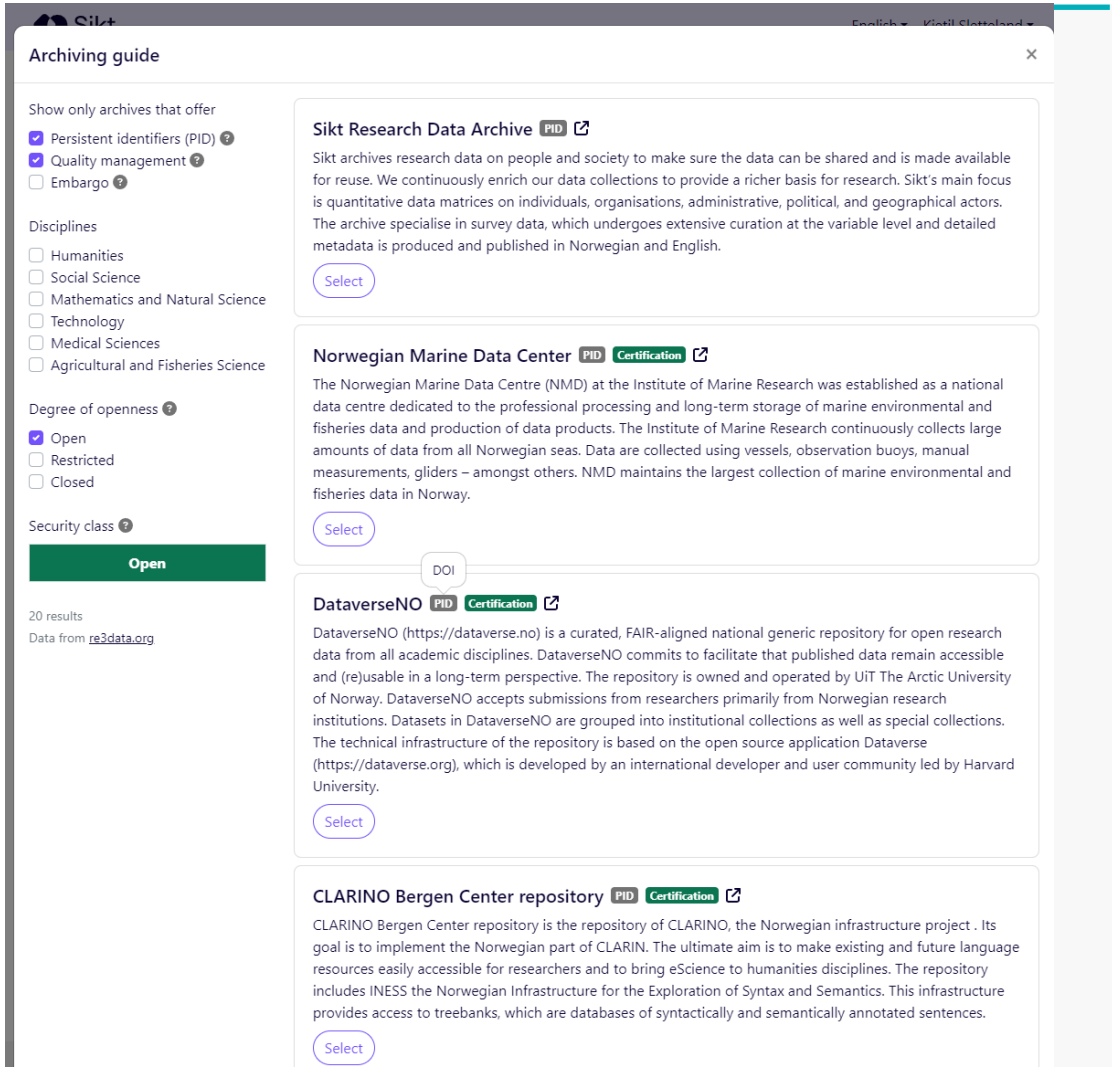
Storage and transfer

- Recommended options will depend on level of confidentiality

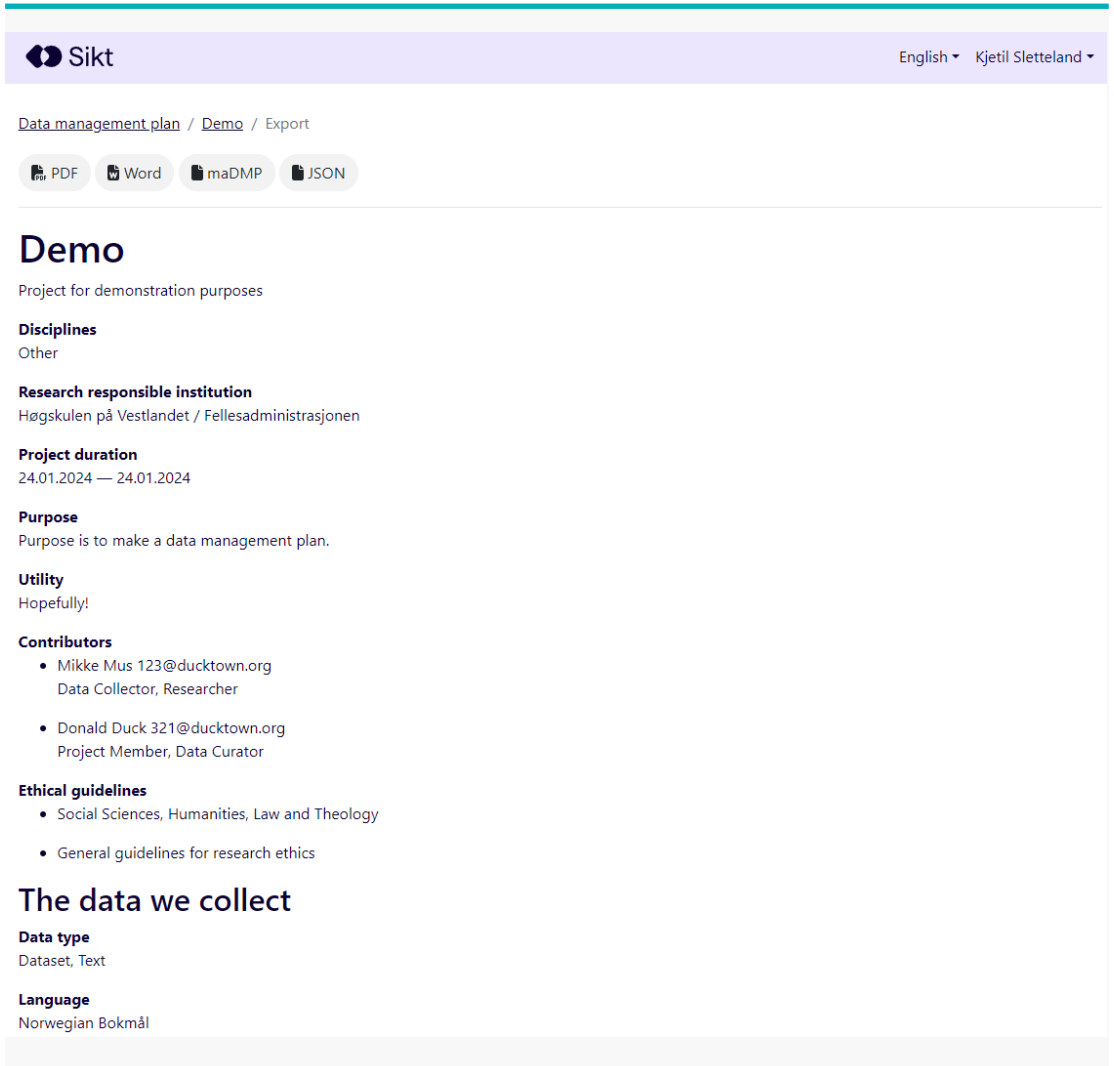


Archiving


Sikt
English
Kjetil Sletteland
Data management plan / Demo
Export
Changes
Delete
Undo
Save
Project information
Research responsibility
Funding
Contributors
Ethical and legal issues
Related resources
Data packages
THE DATA WE COLLECT
Confidentiality
Collection
Files and metadata
Storage and transfer
Archiving
THE DATA WE SWIPE FROM OT...
Confidentiality
Source
Files and metadata
Storage and transfer
Archiving
Archiving
According to the National strategy on access to and sharing of research data and the Research Council's Policy for Open Access to Research Data, all research data must be made available to all relevant users under equal conditions, so long as there are no legal, ethical or security reasons not to do so. Research data can be made available in secure archives for research data, either centrally at your own institution, in national archives, or in subject-specific archives.
Will the data be archived?
Yes
No
Degree of openness
Open
Restricted
Closed
Embargo
License
Creative Commons Zero v1.0 Universal
Archive
+ Add
Costs
None, we use DataverseNO
Comment
Next


Archiving guide
Show only archives that offer
Persistent identifiers (PID)
Quality management
Embargo
Disciplines
Humanities
Social Science
Mathematics and Natural Science
Technology
Medical Sciences
Agricultural and Fisheries Science
Degree of openness
Open
Restricted
Closed
Security class
Open
20 results
Data from re3data.org
Sikt Research Data Archive
PID
Sikt archives research data on people and society to make sure the data can be shared and is made available for reuse. We continuously enrich our data collections to provide a richer basis for research. Sikt's main focus is quantitative data matrices on individuals, organisations, administrative, political, and geographical actors. The archive specialise in survey data, which undergoes extensive curation at the variable level and detailed metadata is produced and published in Norwegian and English.
Select
Norwegian Marine Data Center
PID
Certification
The Norwegian Marine Data Centre (NMD) at the Institute of Marine Research was established as a national data centre dedicated to the professional processing and long-term storage of marine environmental and fisheries data and production of data products. The Institute of Marine Research continuously collects large amounts of data from all Norwegian seas. Data are collected using vessels, observation buoys, manual measurements, gliders – amongst others. NMD maintains the largest collection of marine environmental and fisheries data in Norway.
Select
DOI
DataverseNO
PID
Certification
DataverseNO (https://dataverse.no) is a curated, FAIR-aligned national generic repository for open research data from all academic disciplines. DataverseNO commits to facilitate that published data remain accessible and (re)usable in a long-term perspective. The repository is owned and operated by UiT The Arctic University of Norway. DataverseNO accepts submissions from researchers primarily from Norwegian research institutions. Datasets in DataverseNO are grouped into institutional collections as well as special collections. The technical infrastructure of the repository is based on the open source application Dataverse (https://dataverse.org), which is developed by an international developer and user community led by Harvard University.
Select
CLARINO Bergen Center repository
PID
Certification
CLARINO Bergen Center repository is the repository of CLARINO, the Norwegian infrastructure project . Its goal is to implement the Norwegian part of CLARIN. The ultimate aim is to make existing and future language resources easily accessible for researchers and to bring eScience to humanities disciplines. The repository includes INESS the Norwegian Infrastructure for the Exploration of Syntax and Semantics. This infrastructure provides access to treebanks, which are databases of syntactically and semantically annotated sentences.
Select

Repeat for each data package, and click Export



- Even though you can export in a (hopefully) computer actionable JSON or maDMP format, the RCN needs the DMP in a PDF or Word .docx for now
- DMPs should be published. Zenodo?
- Update it when needed

That’s hopefully all