DMP på 1-2-3

Introduksjon til datahåndteringsplan for forskere

Ane Møller Gabrielsen

NTNU Universitetsbiblioteket

Innhold

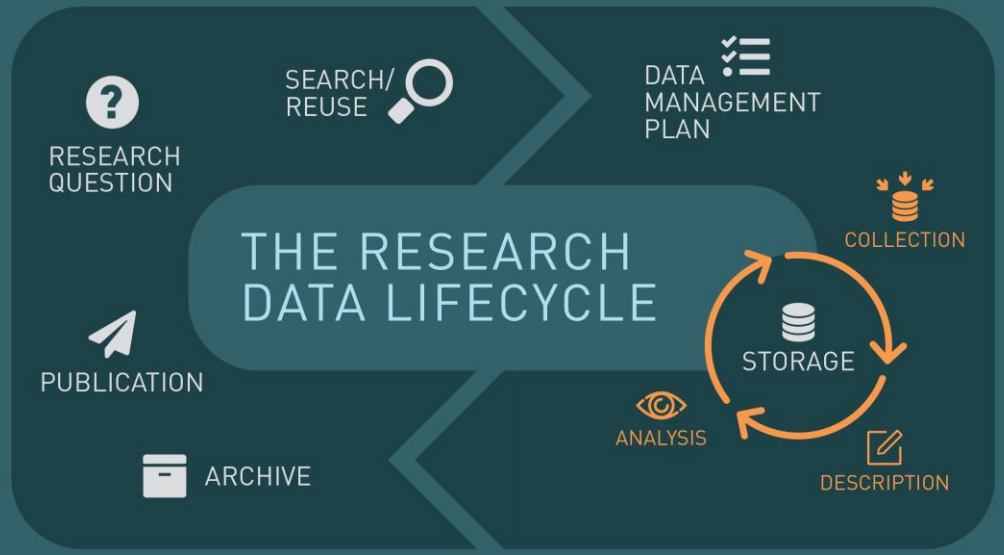
- Introduksjon:
  - Hva er en datahåndteringsplan(DMP)?
  - Hvorfor skal du ha en?
  - Hva skal den inneholde?
- DMP på 1-2-3: DMPOnline
- Hjelp og støtte

- Lenke til denne presentasjonen:

OPEN
SCIENCE
NTNU

Hva er en datahåndteringsplan (DMP)?

- Et dokument som beskriver hvordan forskningsdata skal behandles fra start til slutt i et prosjekt
- Både et formelt krav og "levende" dokument (skal revideres og oppdateres)
- Et redskap for å nå et mål, ikke et mål i seg selv!



NTNU NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

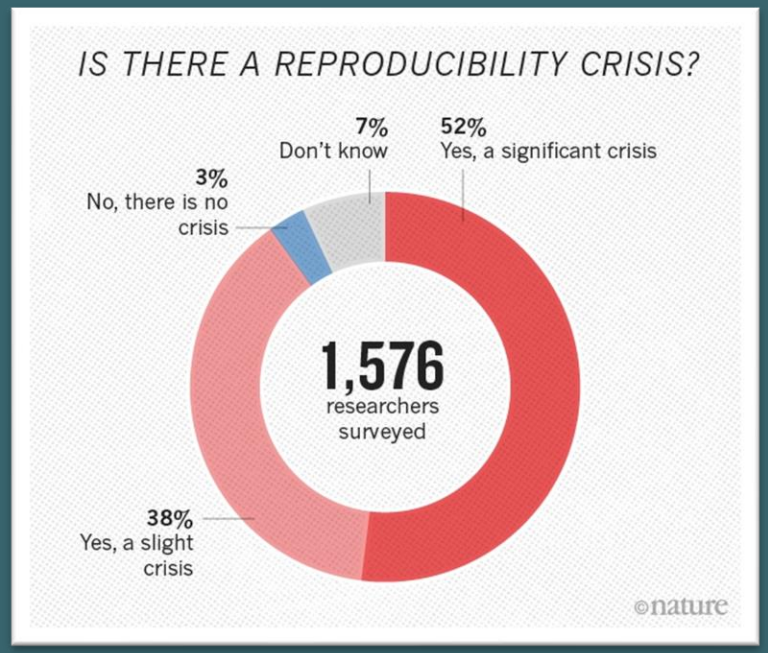
Hvorfor lage en datahåndteringsplan?

- Etterprøvbarhet og gjenbruk!
- Åpen vitenskap!

«Med åpen vitenskap menes et sett prinsipper for mer åpenhet og transparens i forskningsprosessen, samt åpen tilgang til resultatene av forsknings- og undervisningsaktivitet» (NTNU, Politikk for åpen vitenskap)

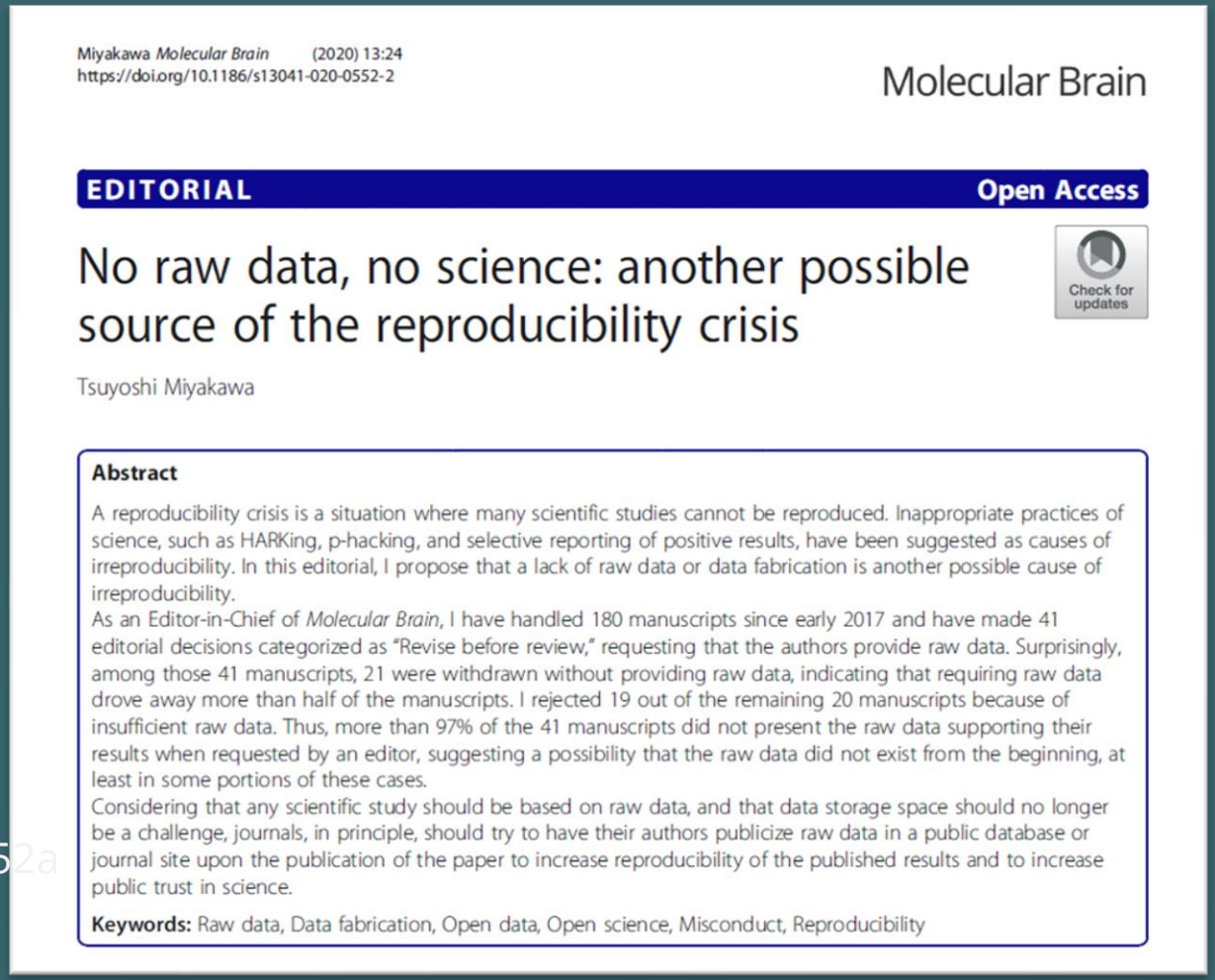
NB: «Så åpent som mulig, så lukket som nødvendig»

Åpne data = transparent, reproduserbar og tilgjengelig forskning



1,500 scientists lift the lid on reproducibility
Survey sheds light on the 'crisis' rocking research.
Monya Baker

Nature 533, 452–454 (26 May 2016) doi:10.1038/533452a



Miyakawa *Molecular Brain* (2020) 13:24
https://doi.org/10.1186/s13041-020-0552-2

Molecular Brain

EDITORIAL **Open Access**

No raw data, no science: another possible source of the reproducibility crisis

Tsuyoshi Miyakawa

Abstract

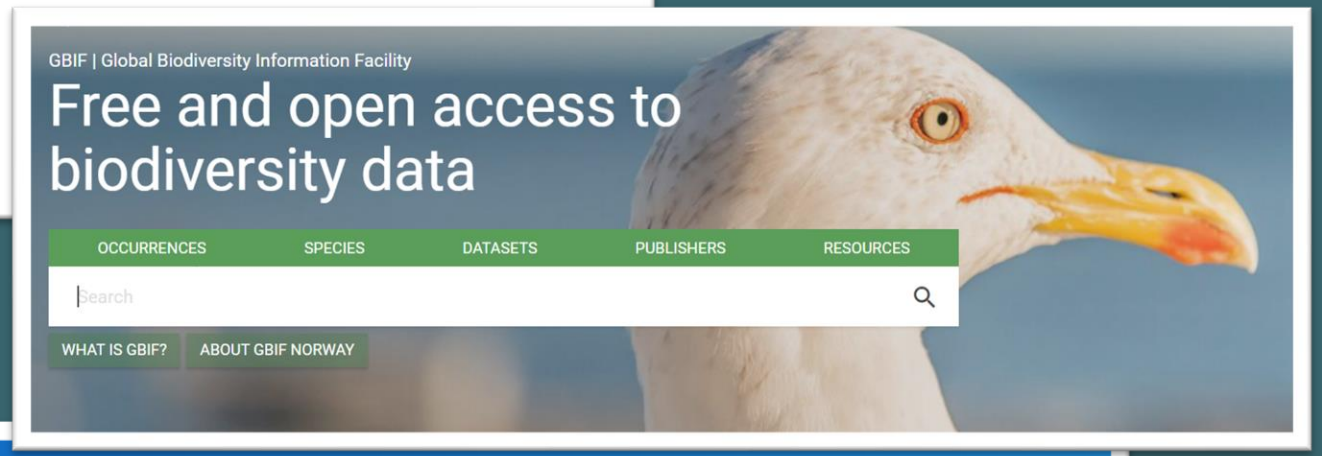
A reproducibility crisis is a situation where many scientific studies cannot be reproduced. Inappropriate practices of science, such as HARKing, p-hacking, and selective reporting of positive results, have been suggested as causes of irreproducibility. In this editorial, I propose that a lack of raw data or data fabrication is another possible cause of irreproducibility.

As an Editor-in-Chief of *Molecular Brain*, I have handled 180 manuscripts since early 2017 and have made 41 editorial decisions categorized as "Revise before review," requesting that the authors provide raw data. Surprisingly, among those 41 manuscripts, 21 were withdrawn without providing raw data, indicating that requiring raw data drove away more than half of the manuscripts. I rejected 19 out of the remaining 20 manuscripts because of insufficient raw data. Thus, more than 97% of the 41 manuscripts did not present the raw data supporting their results when requested by an editor, suggesting a possibility that the raw data did not exist from the beginning, at least in some portions of these cases.

Considering that any scientific study should be based on raw data, and that data storage space should no longer be a challenge, journals, in principle, should try to have their authors publicize raw data in a public database or journal site upon the publication of the paper to increase reproducibility of the published results and to increase public trust in science.

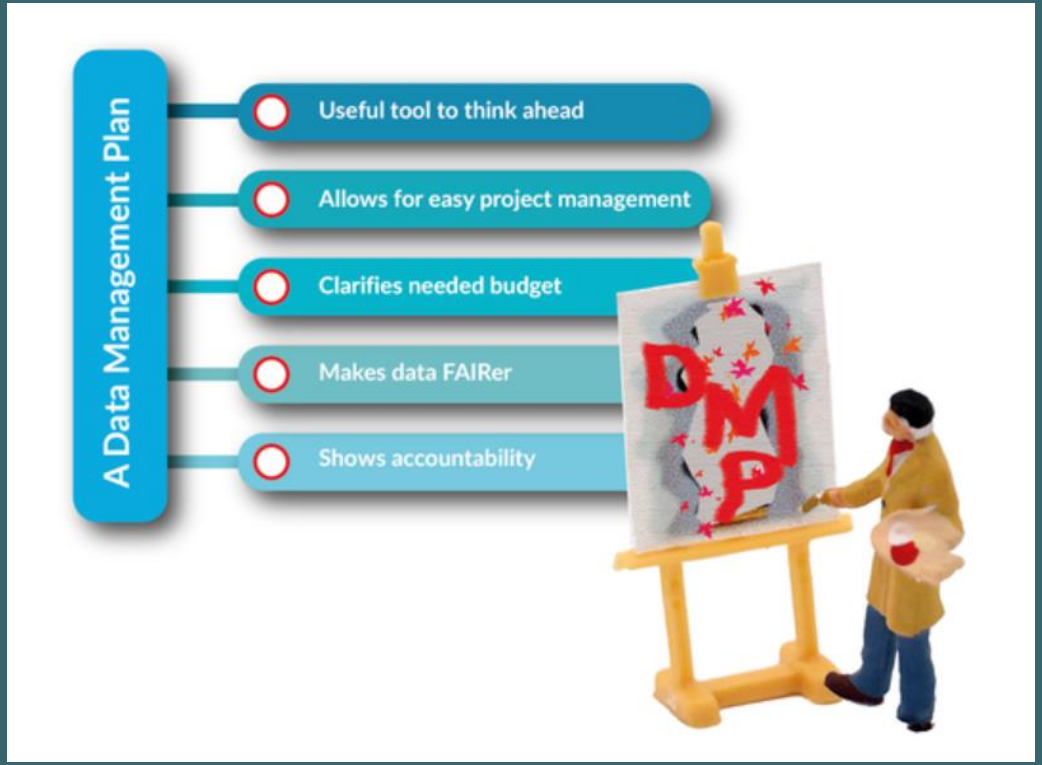
Keywords: Raw data, Data fabrication, Open data, Open science, Misconduct, Reproducibility

Mål: Gjøre datamaterialet tilgjengelig, forståelig og gjenbrukbart



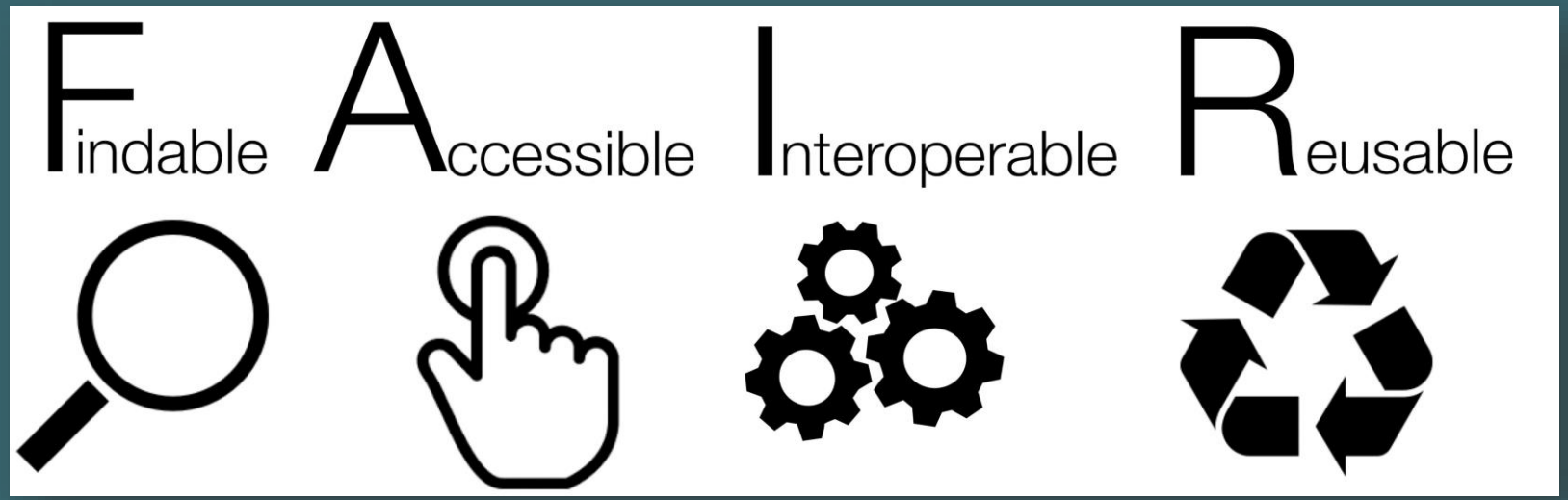
Men: Også et nyttig redskap for prosjektet ditt!

- Hjelper deg med å planlegge og organisere prosjektet ditt
- Hjelper deg å holde oversikt over dataflyten i prosjektet
- Hjelper deg å identifisere potensielle utfordringer og kostnader
- Hjelper deg å sørge for at informasjonssikkerhet og personvern ivaretas
- Hjelper deg å gjøre data **FAIR**

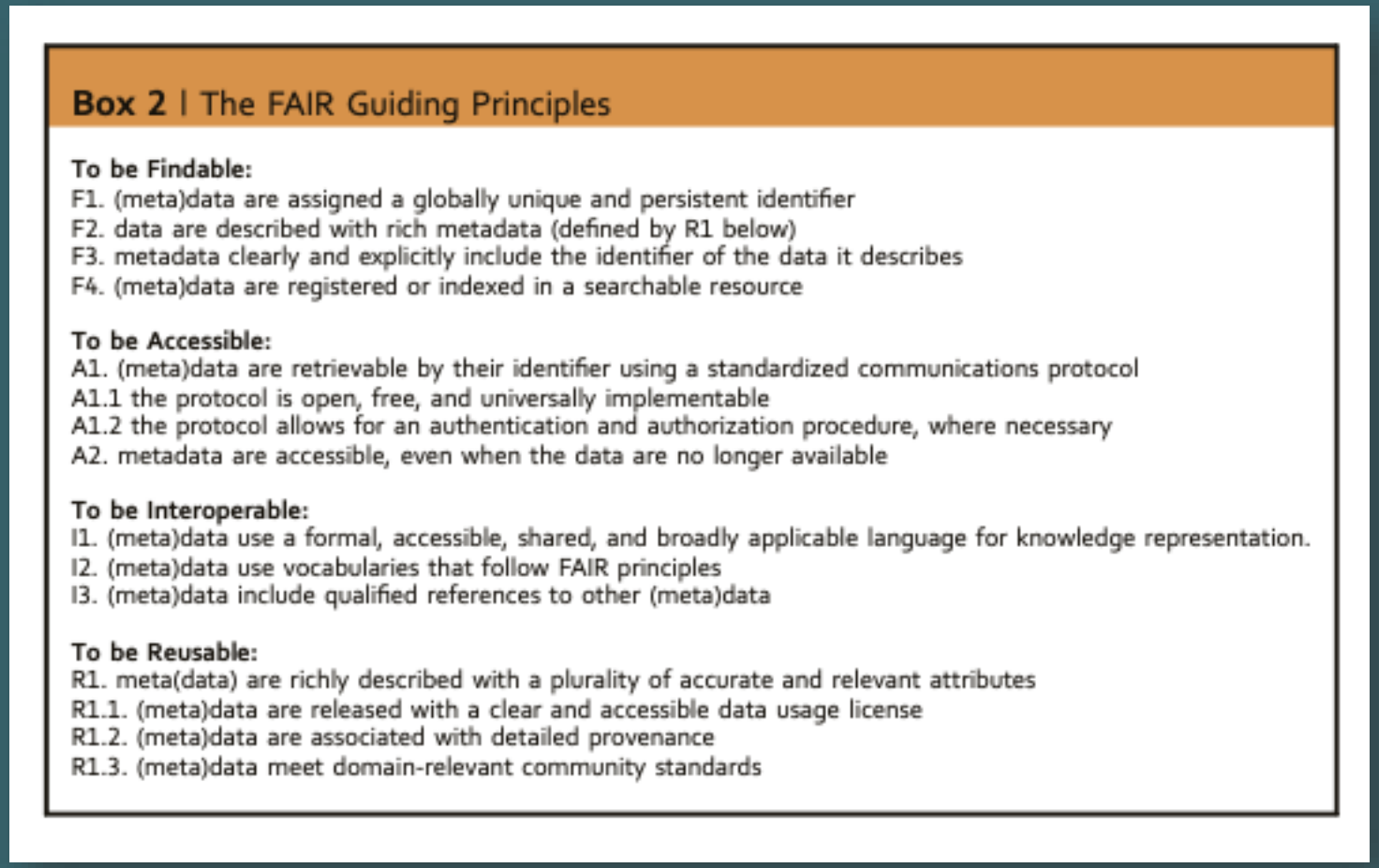


FAIR Data

«The FAIR principles describe how research outputs should be organised so they can be more easily accessed, understood, exchanged and reused» (OpenAIRE)



The FAIR principles



Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

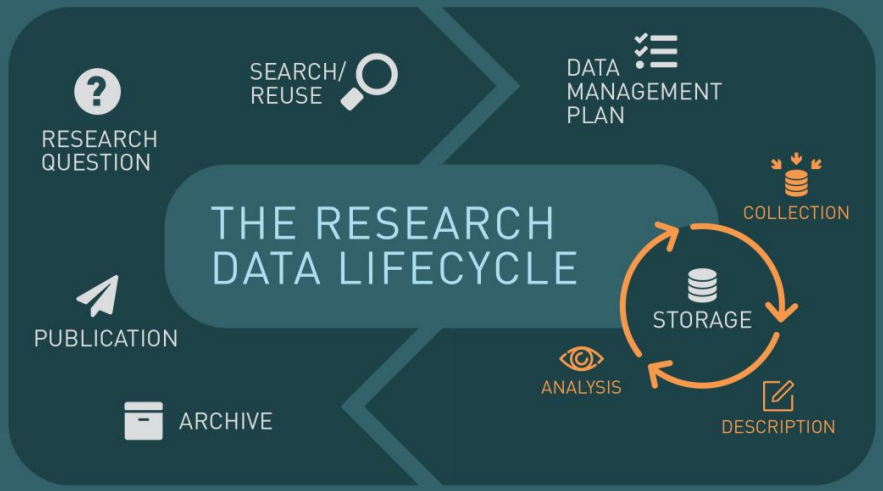
The FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al. 2016)

Hvordan gjøre data FAIR - kortversjon

- Arkiver data og metadata i en søkbar ressurs (forskningsdataarkiv)
- Legg ved all nødvendig informasjon for at framtidige brukere (mennesker og maskiner) skal kunne lese og forstå datasettet
- Bruk tilgjengelige fagstandarder for data og metadata
- Bruk åpne formater og tildel persistente identifikatorer (f.eks. DOI)
- Utstyr datasettet med en egnet lisens

Hva skal planen inneholde?

- «Core Requirements» (Science Europe):
  - Beskrivelse av data, formål og metode for datainnsamling
  - Dokumentasjon og datakvalitet
  - Lagring og backup
  - Etiske og juridiske hensyn
  - Arkivering og deling
  - Kostnader, roller og ansvar



NFR: «Deling av forskningsdata»

Veileder: Dette bør en datahåndteringsplan inneholde

Denne veiledningen er et verktøy for prosjekter som håndte
revidering av søknad og sluttrapportering til Forskningsråde
'Practical Guide to the International Alignment of Research

Forskningsrådet anbefaler å bruke en tjeneste for datahånd
maskinhåndterbar datahåndteringsplan, for eksempel etter
prosjektet laste opp datahåndteringsplanen i format som pc
våre systemer for å tilrettelegge for maskinhåndterbare data
datahåndteringsplanen en vedvarende identifikator, slik som
datahåndteringsplaner tilbyr dette.

- Administrativ informasjon
- Innsamling og/eller bruk av eksisterende data
- Filtyper og format
- Dokumentasjon, metadata og datakvalitet
- Lagring og datasikkerhet underveis i prosjektet
- Rettigheter og delingsbegrensninger: juridiske krav og etiske retningslinjer
- Datadeling og gjenbruk
- Langtidsbevaring
- Datahåndtering – ansvar og ressurser

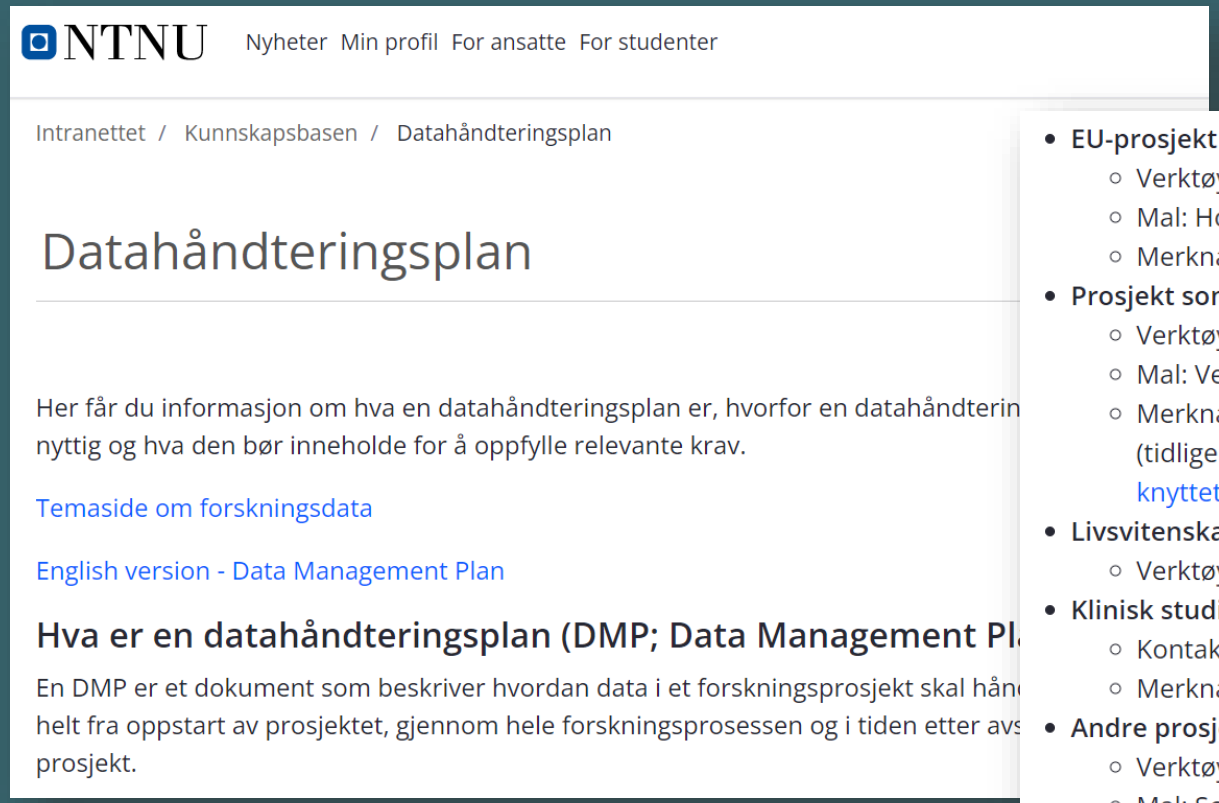
NB: NFR har ingen spesifikke krav til utforming av DMP

Verktøy og tjenestetilbydere for å lage en god datahåndteringsplan

Det finnes flere tilbydere og verktøy som genererer datahåndteringsplaner for forskningsprosjekter. Løsningene gjør det mulig å oppdatere datahåndteringsplanen i løpet av prosjektperioden. Her er eksempler på verktøy og tjenestetilbydere for å generere datahåndteringsplaner:

- DSW ELIXIR-NO
- Norsk senter for forskningsdata (NSD)
- Digital Curation Centre
- easyDMP
- Argos (openaire.eu)

NTNUs anbefalinger for verktøy


NTNU Nyheter Min profil For ansatte For studenter

Intranettet / Kunnskapsbasen / Datahåndteringsplan

Datahåndteringsplan

Her får du informasjon om hva en datahåndteringsplan er, hvorfor en datahåndterin
nyttig og hva den bør inneholde for å oppfylle relevante krav.

Temaside om forskningsdata

English version - Data Management Plan

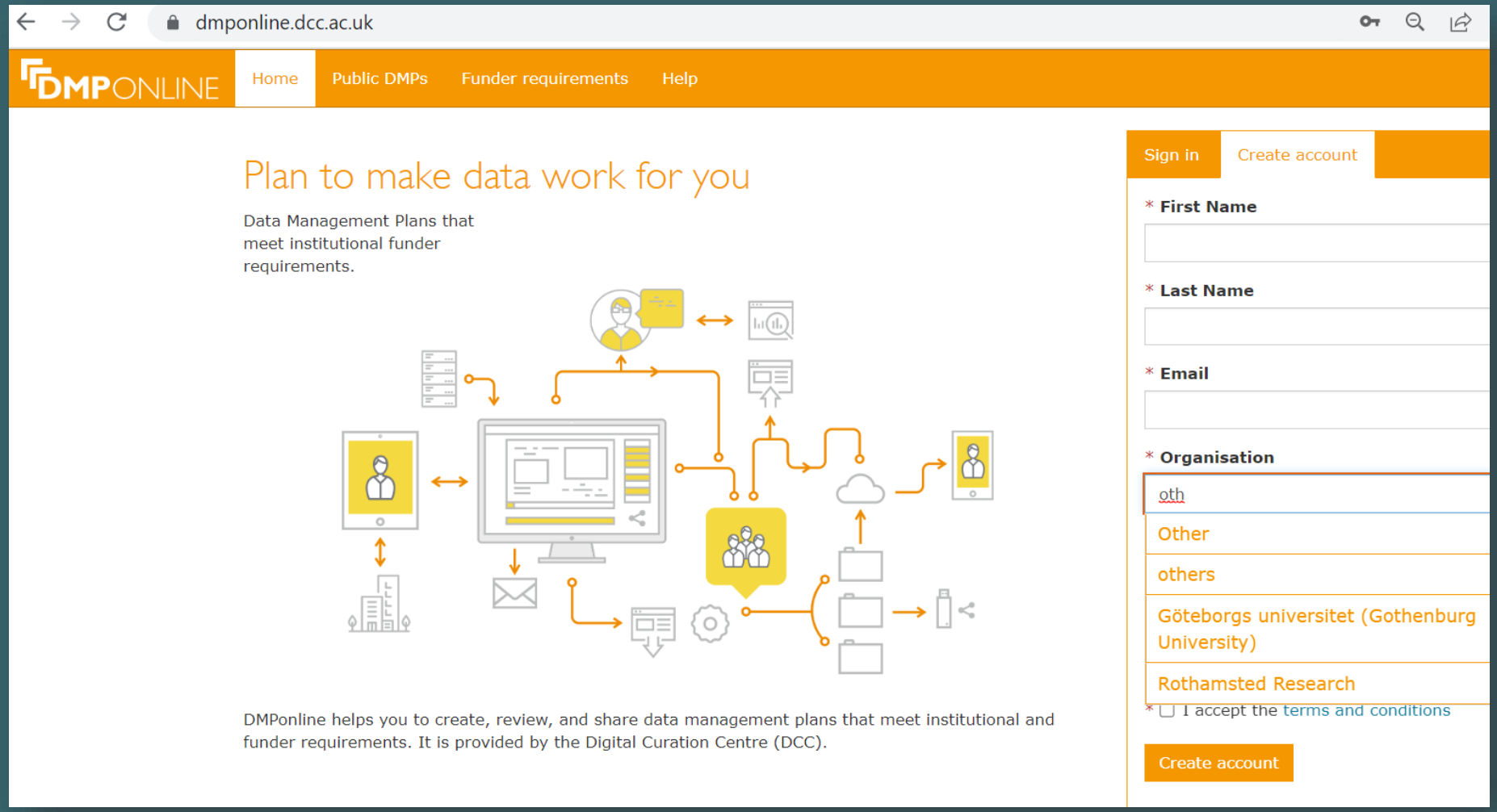
Hva er en datahåndteringsplan (DMP; Data Management Pl

En DMP er et dokument som beskriver hvordan data i et forskningsprosjekt skal hån
helt fra oppstart av prosjektet, gjennom hele forskningsprosessen og i tiden etter avs
prosjekt.

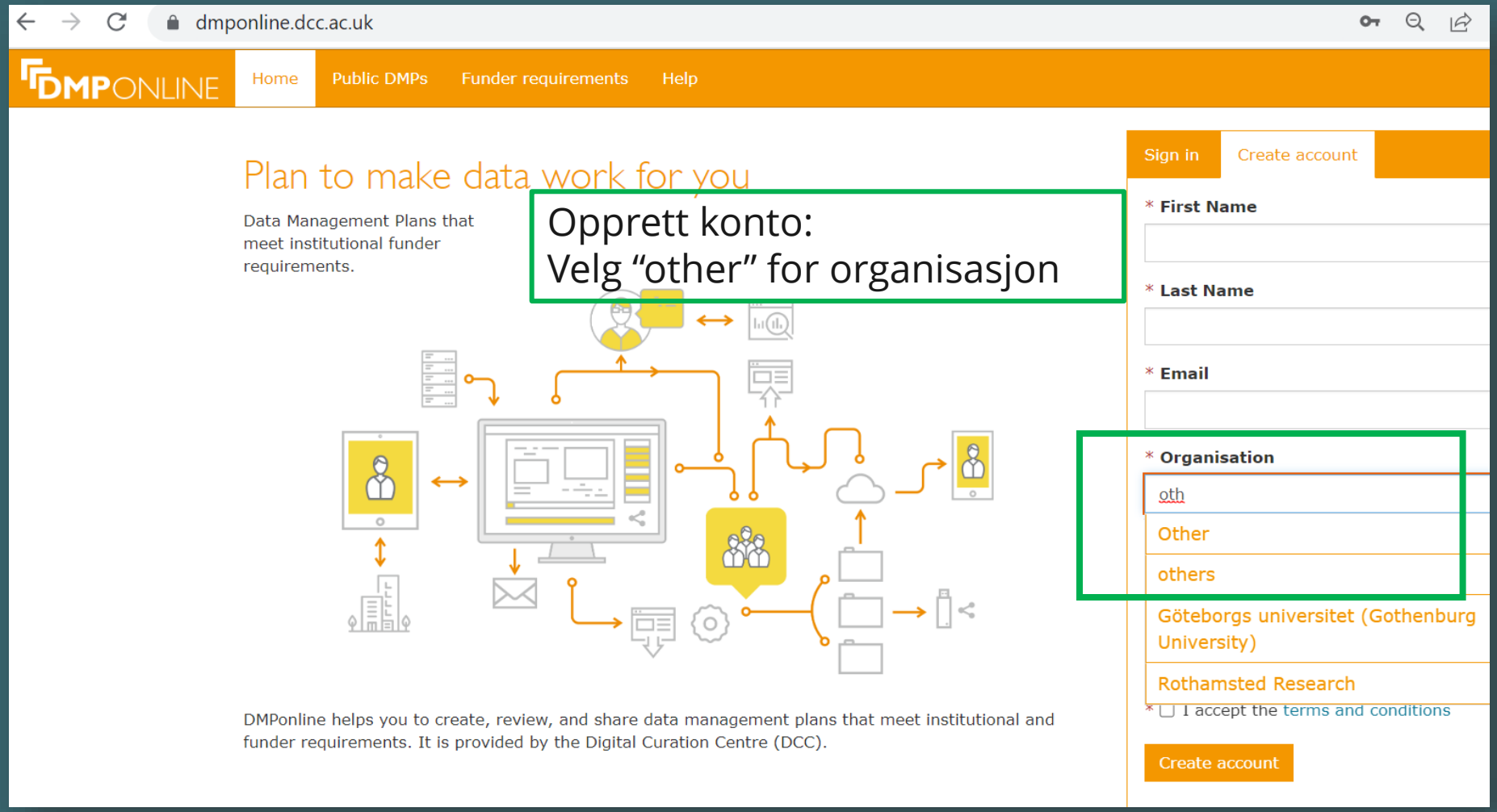
- **EU-prosjekt**
  - Verktøy: DMP Online
  - Mal: Horizon Europe
  - Merknader: Du må opprette egen bruker og velge "other" under Organization.
- **Prosjekt som behandler personopplysninger**
  - Verktøy: NSD
  - Mal: Velg Standard mal
  - Merknader: Logg deg på med Feide (NTNU-bruker). Planen kan deles med andre. Sikt (tidligere NSD) tilbyr også sikker arkivering av forskningsdata og kan gi råd og veiledning knyttet til datainnsamling og personvern etter avtale med NTNU.
- **Livsvitenskap**
  - Verktøy: Elixir Data Stewardship Wizard
- **Klinisk studie**
  - Kontakt Klinisk forskningsenhet Midt-Norge, Klinforsk
  - Merknader: Egen støttetjeneste for NTNU / St. Olav
- **Andre prosjekt: DMPOnline**
  - Verktøy: DMP Online
  - Mal: Science Europe
  - Merknader: Du må opprette egen bruker og velge "other" under Organization. Velg "Science Europe" under "Primary funding organization", eller gå inn på fanen "Reference" > Funder Requirements, og lag ny plan fra Science Europe-malen.

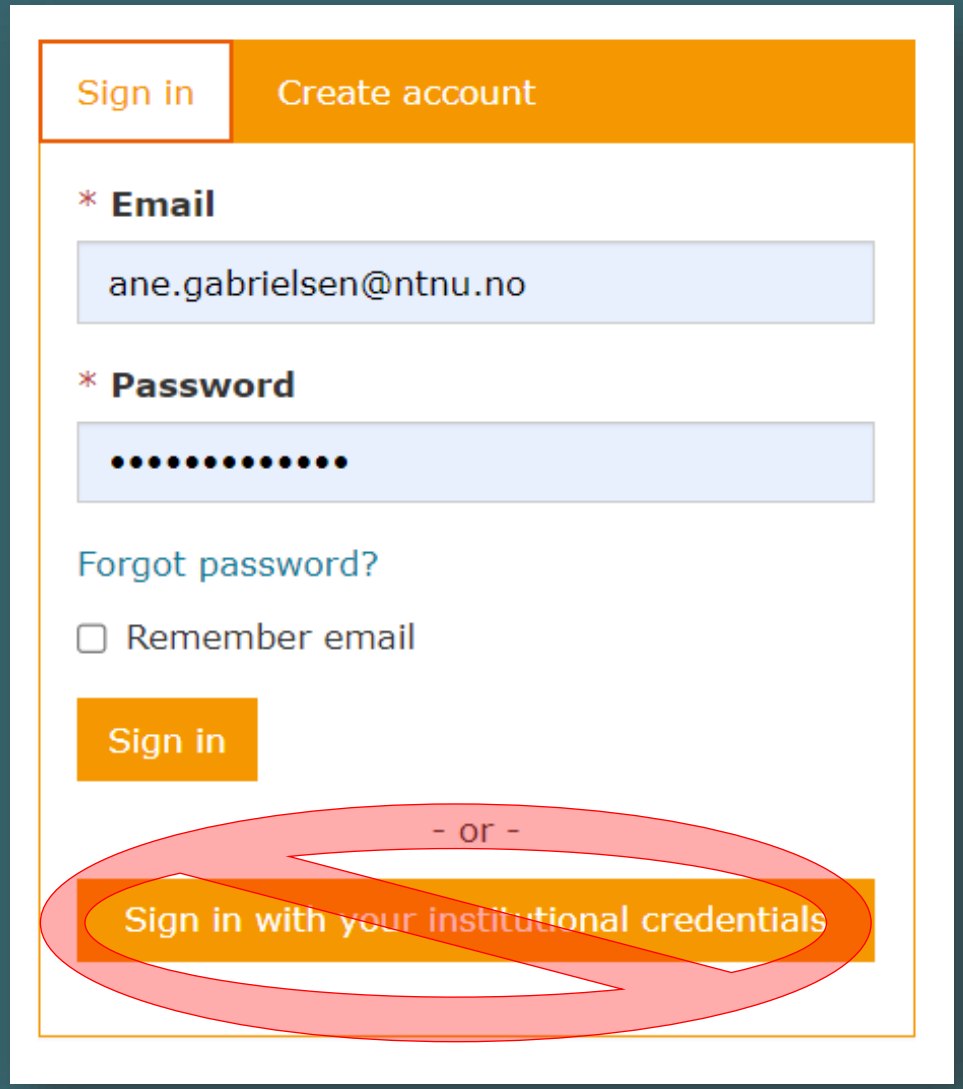
DMP på 1-2-3 med DMPOnline

DMPOnline (Digital Curation Centre)

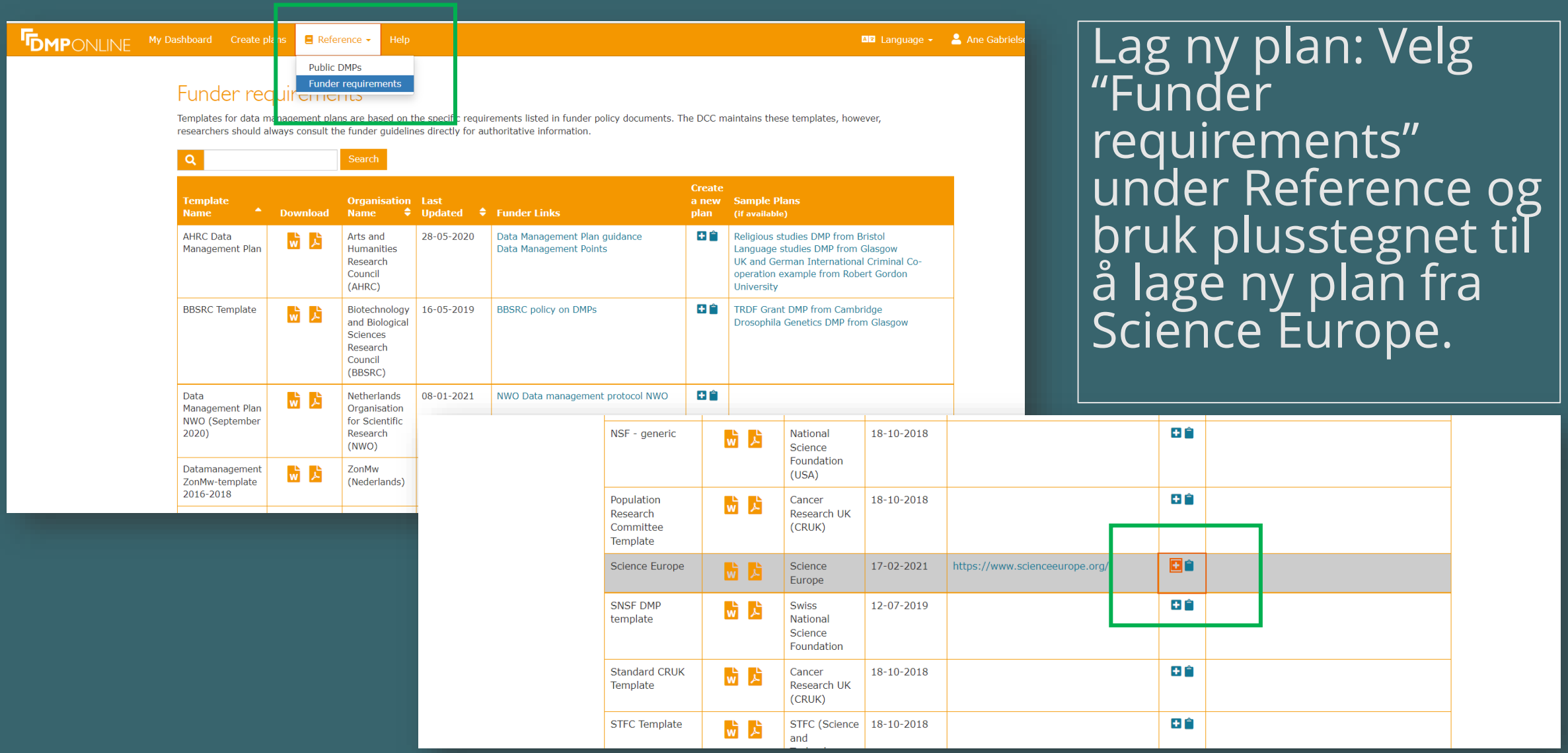


DMPOnline (Digital Curation Centre)



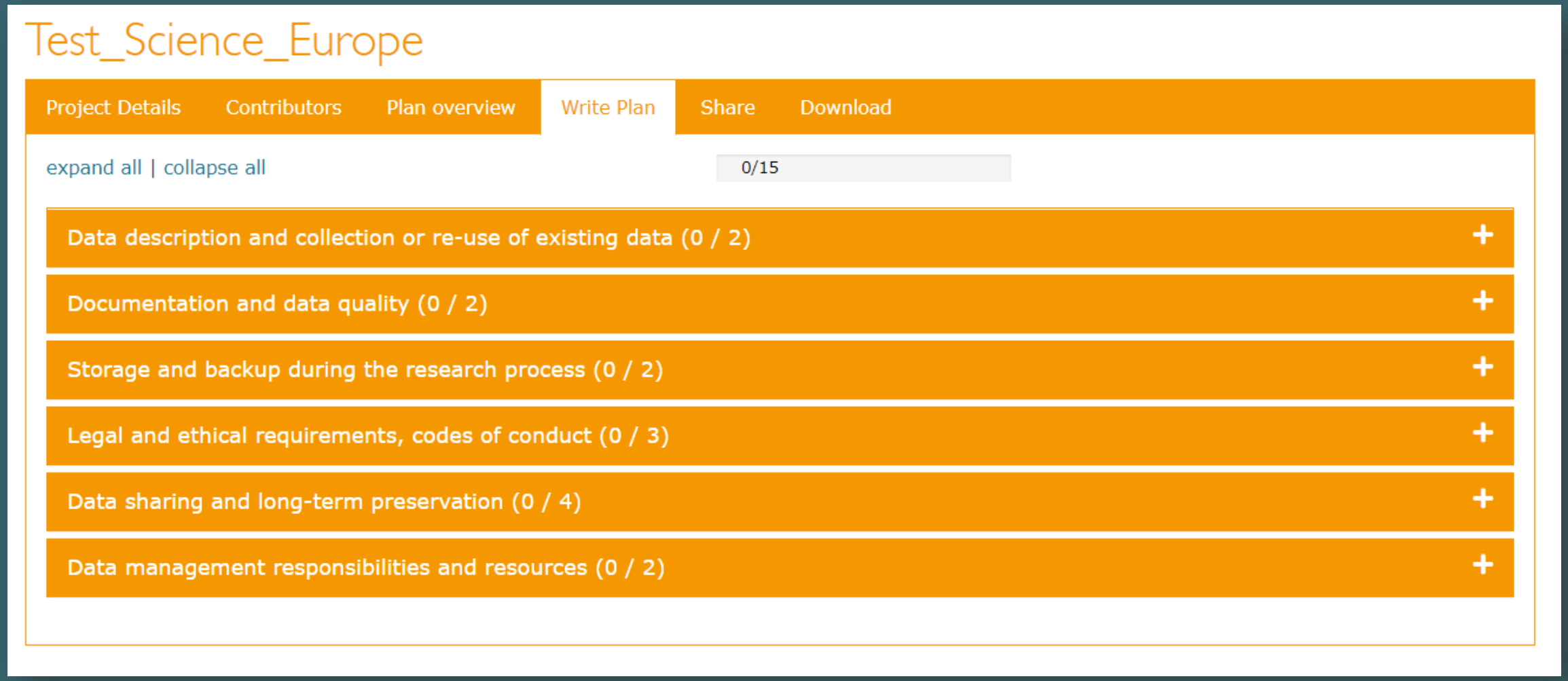


- NB: Logg inn med e-post og password, ikke bruk “institutional credentials”

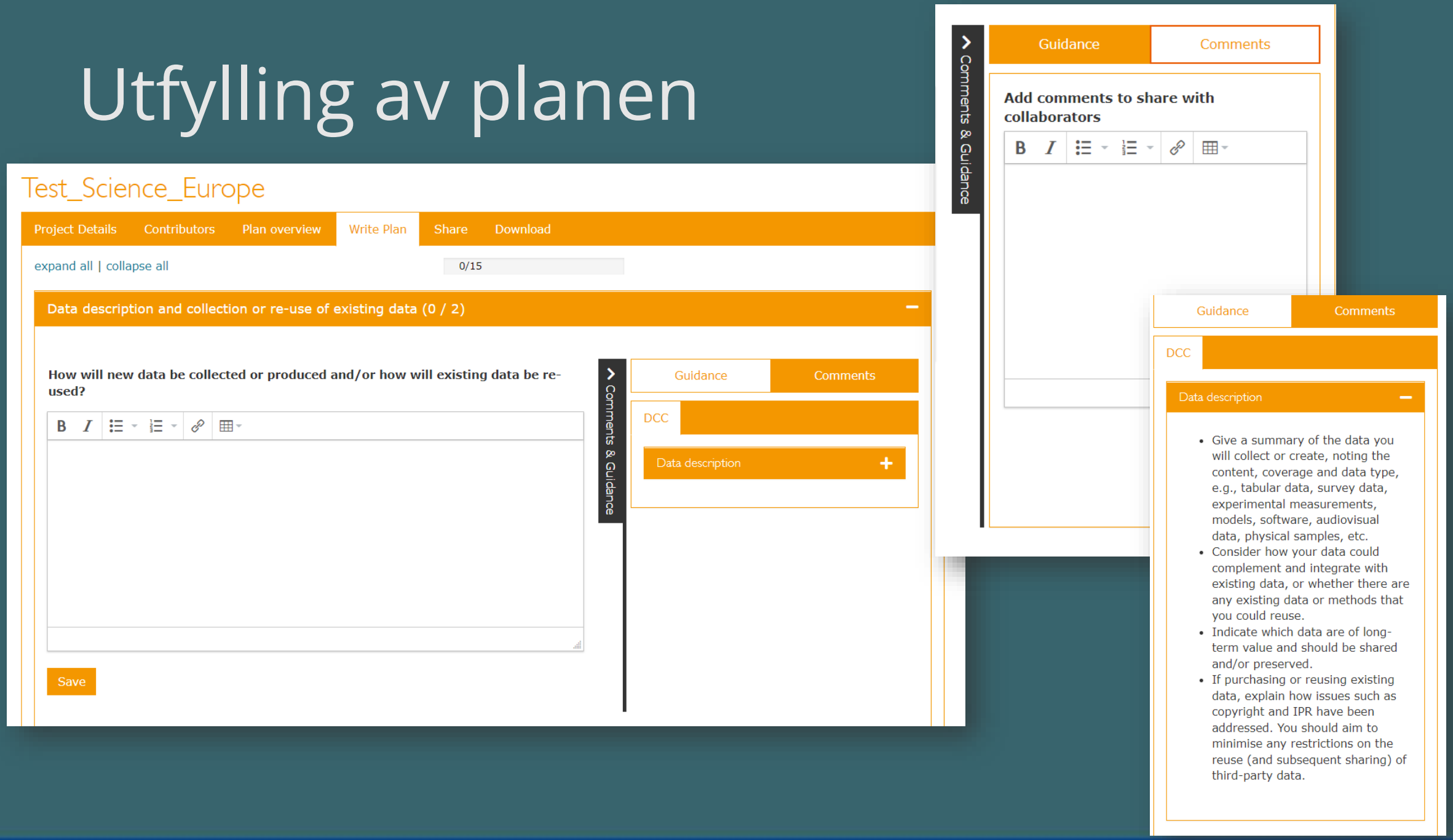


Lag ny plan: Velg "Funder requirements" under Reference og bruk plusstegnet til å lage ny plan fra Science Europe.

Oversikt over planen



Utfylling av planen



Data description and collection or re-use of existing data

- **How will new data be collected or produced and/or how will existing data be re-used?**
- **What data (for example the kinds, formats, and volumes) will be collected or produced?**

How will new data be collected or produced and/or how will existing data be re-used?

- Hvilke metoder/software/utstyr brukes for å samle inn eller produsere data?
- Finnes det eksisterende data som kan gjenbrukes?
- Er det i så fall noen begrensninger for gjenbruk av dette materialet? (Lisenser, opphavsrett etc.)
- Hvordan dokumentere proveniens? (proveniens: opphav, informasjon om opprinnelsen, lokasjonen, eller kilden til et objekt eller en ressurs, inkludert digitale objekter).
- Dersom aktuelt: Hvorfor kan ikke prosjektet gjenbruke eksisterende data?

Gjenbruk av data

- Hvor kan du finne data?
  - BASE (Bielefeld Academic Search Engine)
    - www.base-search.net
  - DataCite
    - search.datacite.org
  - Elsevier Data Search
    - datasearch.elsevier.com
  - Google Dataset Search
    - https://datasetsearch.research.google.com/
- Registry of Research Data Repositories
    - re3data.org

EMBL-EBI: Our impact

We collaborate with scientists and engineers all over the world, and provide the infrastructure needed to share data openly in the life sciences.

390 Petabytes
of raw data storage

81 million
web requests made to our websites on an average day

Data description and collection or re-use of existing data (2 / 2)

How will new data be collected or produced and/or how will existing data be re-used?

Sufficiently addressed DMP according to Science Europe Practical Guide:

- Gives clear details of where the existing data come from and how new data will be collected or produced. It clearly explains methods and software used.
- Explains, if existing data are re-used, how these data will be accessed and any constraints on their re-use.
- Explains clearly, if applicable, why new data must be collected, instead of re-using existing data.

Example Answer ETH Zürich:

This project will work with and generate three main types of raw data.

1. Images from transmitted-light microscopy of giemsa-stained squashed larval brains.
2. Images from confocal microscopy of immunostained whole-mounted larval brains.
3. Western blot data.

What data (for example the kinds, formats, and volumes) will be collected or produced?

- Hva slags type data er det snakk om (bilder, lydopptak, tabeller, tekst, video etc.)
- Hvilke formater? (pdf, xls, doc, txt, rdf, csv etc.).
- Hvilke mengder, sånn ca? (MB, GB, TB ...?)

What data (for example the kinds, formats, and volumes) will be collected or produced?

Sufficiently addressed DMP according to Science Europe Practical Guide:

- Clearly describes or lists what data types will be generated (for example numeric, textual, audio, or video) and their associated data formats, including, if needed, data conversion strategies.
- Explains why certain formats have been chosen and indicates if they are in open and standard format. If a proprietary format is used, it explains why.
- Provides information about the estimated data volume.
- Clearly states, if applicable, that no new data will be produced or generated by the project.

NB. Information derived from previously existing data sources (namely output, processed, and analysed data) are to be considered new data under this question.

Example Answer ETH Zürich:

All data will be stored in digital form, either in the format in which it was originally generated (i.e. Metamorph files, for confocal images; Spectrum Mill files, for mass spectra with results of mass spectra analyses stored in CSV files; TIFF files for gel images; MariaDB SQL dump files for genetics records), or will be converted into a digital form via scanning to create tiff or jpeg files (e.g. western blots or other types of results).

Measurements and quantification of the images will be recorded in excel files (for long term preservation, they will be converted in CSV files.

Micrograph data is expected to total between 100GB and 1TB over the course of the project. Scanned images of western blots are expected to total around 1GB over the course of the project. Other derived data (measurements and quantifications) are not expected to exceed 10MB.

Documentation and data quality

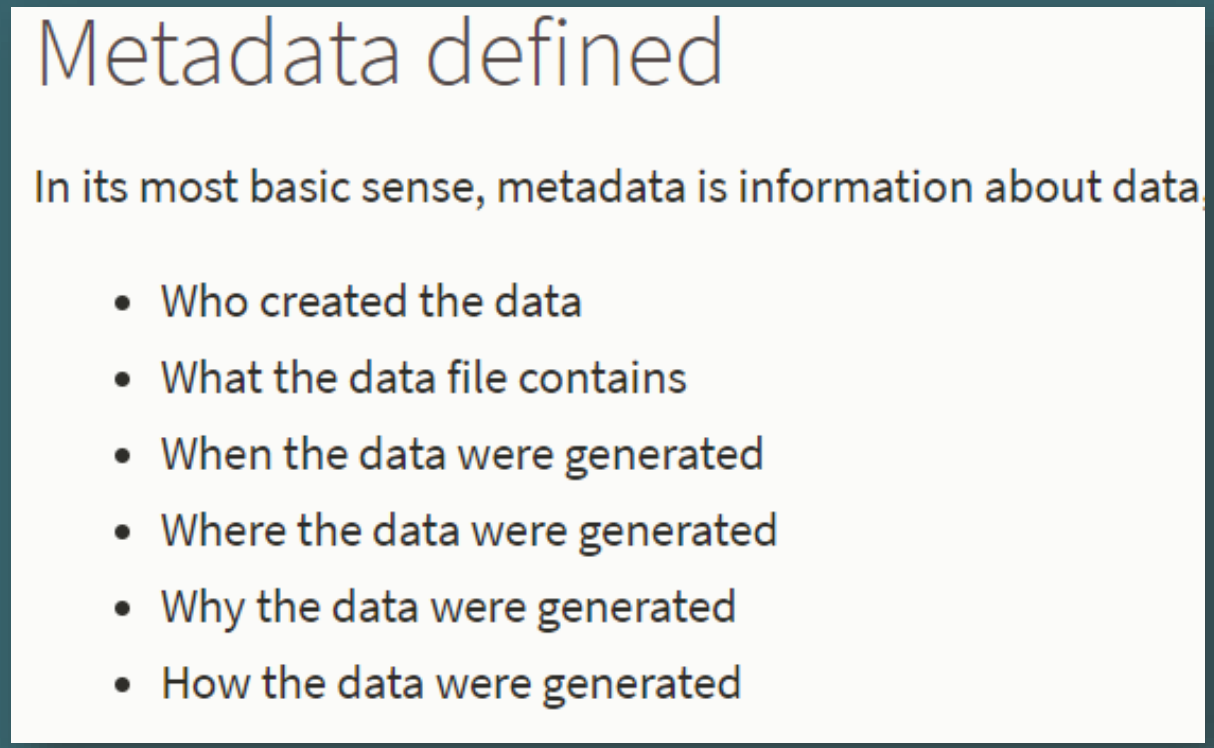
- **What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?**
- **What data quality control measures will be used?**

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

- Hva slags informasjon er nødvendig for at framtidige brukere (inkludert ditt framtidige jeg) skal kunne forstå og bruke dataene i framtiden?
- Hvordan skal du skape og ta vare på dokumentasjon?
- Hvordan skal du organisere, navngi og versjonere filer?
- Metadata og metadatastandarder!

Metadata: data som definerer eller beskriver andre data

- Det står mye nyttig om metadata på nettsidene til Stanford Libraries
- Du finner en katalog over metadatastandarder på sidene til Research Data Alliance (RDA)
- Les *Making a Research Project Understandable* (Fuchs & Kuusniemi 2018)



(Stanford Libraries, Creating Metadata)

Noen vanlige metadatastandarder:

- Dublin Core er en standardisert liste med 15 elementer som beskriver et dataset (eller et annet digital objekt)
- Darwin Core brukes for biodiversitetsdata
- DDI (Data Documentation Initiative) er en standard for samfunnsviten-skapelige data

Element	Description
Title	A name given to the resource, either supplied by the individual assigning metadata or from the object. Example: "A Pilot's Guide to Aircraft Insurance"
Creator	Entity responsible for making the resource. Example: "Duncan, P. A."
Subject	The topic of the resource, typically represented using keywords. Example: "Colonial medicine"
Description	An account of the resource. Example: "Illustrated guide to airport markings and lighting signals for airports with low visibility conditions."
Publisher	An entity responsible for making the resource available. Example: "The University of Texas Press"
Contributor	An entity responsible for making contributions to the resource (e.g. editor, transcriber, illustrator). Example: "Austin Citizen Photograph"

Dublin Core (Wikipedia)

Metadata (How to FAIR)

D.B. Deutz, M.C.H. Buss, J. S. Hansen, K. K. Hansen, K.G. Kjelmann, A.V. Larsen, E. Vlachos, K.F. Holmstr and (2020). How to FAIR: a Danish website to guide researchers on making research data more FAIR

What are metadata?

Metadata are data about data. They play an important role in making your data FAIR. Metadata have to be added continuously to your research data, not just at the beginning or at the end of a project. Metadata can be added manually or automatically, and preferably according to a disciplinary standard. From a FAIR perspective, metadata are more important than your data, because metadata would always be openly available and they link research data and publications in the ***Internet of FAIR Data and Services***. The distinction between data and metadata is not ontological, but it is grounded in use. What is "data" and what is "metadata" is thereby a matter of perspective: Some researchers' metadata can be other researchers' data.

While data documentation is meant to be read and understood by humans, metadata (which are sometimes a part of the documentation) are primarily meant to be processed by machines.

What is a metadata standard?

A metadata standard is a subject-specific guide to your metadata. Metadata elements are grouped into sets designed for a specific purpose and given a standard name and definition. Rules on what content must be included, what syntax must be used, or a controlled vocabulary can also be included in a metadata standard.

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?

Sufficiently addressed DMP according to Science Europe Practical Guide:

- Clearly outlines the metadata that will accompany the data, with reference to good practice in the scientific community (for example uses metadata standards where they exist).
- Clearly outlines the documentation needed to enable data re-use, stating where the information will be recorded (for example a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks).
- Indicates how the data will be organised during the project (for example naming conventions, version control strategy and folder structures).

Example Answer ETH Zürich:

Files will be named according to a pre-agreed convention. The dataset will be accompanied by a README file which will describe the directory hierarchy.

Each directory will contain an INFO.txt file describing the experimental protocol used in that experiment. It will also record any deviations from the protocol and other useful contextual information.

Microscope images capture and store a range of metadata (field size, magnification, lens phase, zoom, gain, pinhole diameter etc.) with each image.

These metadata will be recorded according to the REMBI standard - Reccommended metadata for microscopic images.

What data quality control measures will be used?

- Standardiserte metoder?
- Kalibrering, kontroller?
- Fagfellevurdering, validering av data?
- Repetisjon av målinger/observasjoner?
- Bruk av kontrollerte vokabularer?
- +++

What data quality control measures will be used?

Sufficiently addressed DMP according to Science Europe Practical Guide:

- Clearly describes the approach taken to ensure and document quality control in the collection of data during the lifetime of the project.

Example Answer ETH Zürich:

All samples on which data are collected will be prepared according to published standard protocols in the field [*cite reference*].

Experiments will include appropriate controls to ensure validity [brief description]. Data consistency will be assessed by comparing repeated measures.

Storage and backup during the research process

- **How will data and metadata be stored and backed up during the research process?**
- **How will data security and protection of sensitive data be taken care of during the research?**

Lagring og informasjonssikkerhet

- Forskningsdata skal klassifiseres for å kunne velge riktig lagringsløsning

Mer informasjon: Informasjonsklassifisering (Innsida)

- De fleste typer forskningsdata vil klassifiseres som «interne» mens prosjektet pågår, men det finnes data krever strengere beskyttelse
  - Bedriftshemmeligheter
  - Patentverdig, kommersielle hensyn
  - Personopplysninger – særlige kategorier

NTNU lagringsguide

Lagringsguide

Denne lagringsguiden skal hjelpe deg å velge riktig løsning for hvor du kan lagre og behandle informasjon. Lagringsguiden gir deg en oversikt over hvilke lagringsområder som er vurdert av NTNU, til hvilke formål du kan bruke områdene og informasjon til hvordan du kan ta i bruk områdene. I tillegg gir datainnsamlingsguiden og lagringsguiden deg utfyllende informasjon til bruk i arbeid med datahåndteringsplaner.

English version: Data storage guide

Temaside om informasjonssikkerhet | Sider merket med informasjonssikkerhet

Informasjonssikkerhet og klassifisering

Ved NTNU er det innført retningslinjer for klassifisering av informasjon. All informasjon du behandler og deler har ulike behov for sikring og skjerming, ut fra hvor verdifull den er. For å identifisere verdinivå og hvilke sikkerhetstiltak som er nødvendige, må informasjonen klassifiseres. Konfidensialitetsklassene beskriver hvilken grad av beskyttelse som er nødvendig.

Åpen | Intern | Fortrolig | Strengt fortrolig

Les mer om:

- Informasjonsklassifisering - informasjonssikkerhet
- Klassifisering av filer og informasjon

Hvilke fysiske lagringsmedier kan jeg bruke?

Med fysiske lagringsmedier mener vi behandling og lokal lagring av informasjon, feks lagring på egen maskin (Mac, PC eller harddisk).

Informasjonsklassifisering:	Åpen	Intern	Fortrolig	Strengt fortrolig
Privat-eid bærbar datamaskin	**OK**	NEI	NEI	NEI
Privat-eid hjemmemaskin	**OK**	NEI	NEI	NEI
NTNU-anskaffet hjemmemaskin (egenadministrert)	**OK**	**OK**	NEI	NEI
NTNU-anskaffet bærbar datamaskin (egenadministrert)	**OK**	**OK**	NEI	NEI
NTNU-administrert desktop – kryptert	**OK**	**OK**	**OK**	NEI
NTNU-administrert bærbar datamaskin – kryptert	**OK**	**OK**	**OK**	NEI
Minnepinne/ekstern harddisk	**OK**	**OK**	NEI	NEI
Minnepinne/ekstern harddisk – kryptert	**OK**	**OK**	**OK**(1)	**OK**(2)

(1) Datoene må ligge kryptert på lagringsmediet og passordet oppbevart et annet sted. Les mer om kryptering.

(2) Hele disken må være kryptert med et godt passord (les mer om hvordan lage passord). Passordet må være oppbevart et annet sted.

Lagringstjenester og samhandlingsplattformer

Med lagringstjenester og samhandlingsplattformer mener vi lagring i skytjenester eller servere på NTNU. Les mer om de ulike tjenestene og plattformene ved å klikke på dem.

Informasjonsklassifisering:	Åpen	Intern	Fortrolig	Strengt fortrolig
Personlig skytjeneste (dropbox, google drive ++)	**OK**	NEI	NEI	NEI
NTNU Hjemmeområde («M:-disk»)	**OK**	**OK**	**OK**(1)	NEI
NTNU Fellesområde (T:-enhet, gruppe, prosjekt, osv)	**OK**	**OK**	NEI	NEI
NTNU-administrert Dropbox (kontakt orakel)	**OK**	**OK**	NEI	NEI
NTNU-Box	**OK**	**OK**	NEI	NEI
Office 365 (SharePoint, Teams, Onedrive)	**OK**	**OK**	**OK**(1)	NEI
NTNU NICE-1 - Lagringsområde med økt sikkerhet	**OK**	**OK**	**OK**	NEI
HUNT Cloud	**OK**	**OK**	**OK**(2)	NEI
UiO TSD	**OK**	**OK**	**OK**	**OK**
NIRD (tidligere Norstore, driftes av Uninett)	**OK**	**OK**	NEI	NEI

NTNUs datainnsamlingsguide

Intervju med opptak av lyd og/eller video

Informasjonsklassifisering:	Åpen	Intern	Fortrolig	Strengt fortrolig
Zoom	OK (1)	OK (1)	NEI	NEI
Teams	OK (1)	OK (1)	NEI	NEI
Nettskjema-Diktafon-app (X)	OK	OK	OK (2)	OK (2)
Ekstern lydopptaker/diktafon	OK	OK (3)	OK (3)	NEI (4)

(X) UiO har tidligere rapportert om teknisk ustabilitet ved lagring av opptak. Vi anbefaler derfor at du tester først, og sjekker i etterkant av hvert opptak at det blir lagret.

(1) Du kan bruke innebygd opptaksfunksjon, men pass på at opptaket lagres og behandle og egnet sted, se også Lagringsguiden.

(2) Diktafon-appen må settes opp med innsamling og lagring direkte i TSD (Tjenester for S Data).

Datainnsamling

Denne guiden er en oversikt over NTNUs verktøy og prosedyrer for innsamling av forskningsdata som inneholder personopplysninger. Oversikten skal hjelpe deg som forsker eller student med å ta riktige valg for det du skal gjøre i ditt forsknings- eller studentprosjekt.

Datainnsamlingsguiden gir deg en oppdatert oversikt over hvilke digitale verktøy for innsamling som er vurdert av NTNU, til hvilke formål du kan bruke verktøyene, og informasjon om hvordan du kan ta i bruk verktøyene. I tillegg gir datainnsamlingsguiden og lagringsguiden deg utfyllende informasjon til bruk i arbeid med datahåndteringsplaner.

Storage and backup during the research process (2 / 2)

How will data and metadata be stored and backed up during the research process?

Sufficiently addressed DMP according to Science Europe Practical Guide:

- Clearly (even if briefly) describes:
  › The location where the data and backups will be stored during the research activities.
  › How often backups will be performed.
  › The use of robust, managed storage with automatic backup (for example storage provided by the home institution).

  or

- Explains why institutional storage will not be used (and for what part of the data) and describes the (additional) locations, storage media, and procedures that will be used for storing and backing up data during the project.

Example Answer ETH Zürich/NTNU:

The data we are generating, processing and storing in this project does not pose a particular data security risk. According to institutional guidelines provided through the NTNU Storage Guide, data will be stored on the T:-drive, a part of the centralized file storage system with automatic backup, managed by the NTNU IT department.

How will data security and protection of sensitive data be taken care of during the research?

Sufficiently addressed DMP according to Science Europe Practical Guide:

- Clearly explains:
  › How the data will be recovered in the event of an incident.
  › Which institutional and/or national data protection policies are in place and provides a link to where they can be accessed.
  › Who will have access to the data during the research.
- Clearly describes the additional security measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe, when sensitive data are involved (for example personal data, politically sensitive information, or trade secrets).

Example answer confidential data:

Confidential data from business partners will be stored on NICE, a server with added security, multifactor authentication and automatic backup managed by the NTNU IT department. Access will be limited to the PI and primary research team members.

Relevant documents:

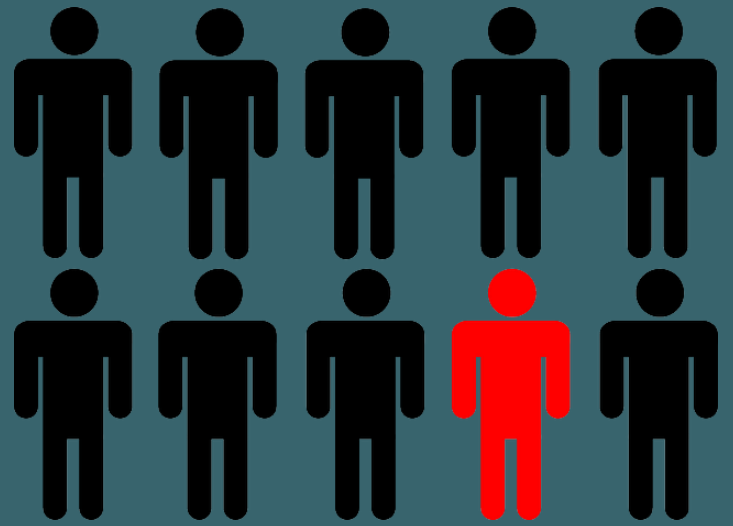
- NTNU Policy for information security
- NTNU Storage Guide

Legal and ethical requirements, codes of conduct

- **If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?**
- **How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?**
- **How will possible ethical issues be taken into account, and codes of conduct followed?**

Personopplysninger/persondata

Personopplysninger: All informasjon som kan identifisere en person, enten direkte eller indirekte



Eksempler:

- Navn, ID-nummer, adresse, telefonnummer, e-postadresse
- IP-adresse
- Bilder, video, lydopptak av stemme
- Bakgrunnsopplysninger som kan spores tilbake til en enkeltperson (Eks: bosted kombinert med kjønn, alder, yrke etc.)

Særlige kategorier av personopplysninger: strengere krav til sikkerhet

1. opplysninger om rasemessig eller etnisk opprinnelse
2. opplysninger om politisk oppfatning
3. opplysninger om religion
4. opplysninger om filosofisk overbevisning
5. opplysninger om fagforeningsmedlemskap
6. genetiske opplysninger
7. biometriske opplysninger (når behandlingsformålet er å entydig identifisere noen)
8. helseopplysninger
9. opplysninger om seksuelle forhold
10. opplysninger om seksuell legning

NB: Opplysninger om straffedommer og lovovertredelser er også underlagt spesielle regler

If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

- Forhåndsvurdering fra Sikt (tidligere NSD)
- NB: Alle prosjekter som behandler personopplysninger skal meldes til Sikt
- (Unntak: Helseforskning ved MH-fakultetet. Annen helseforskning skal meldes til Sikt i tillegg til godkjenning fra REK.)
- Samtykke, pseudonymisering, anonymisering, sletting …
- Se også Innsida: Behandle personopplysninger i student- og forskningsprosjekt

If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

Sufficiently addressed DMP according to Science Europe Practical Guide:

- Clearly indicates if personal data will be collected/used as part of the project, and, if applicable, how compliance with applicable legislation will be ensured (for example by gaining informed consent, considering encryption, anonymisation, or pseudonymisation).
- Describes the procedure to manage access to only authorised users.

Example answer:

A risk assessment concerning data security will be performed before data collection starts.

In order to ensure compliance with GDPR, the project will send a notification form describing all relevant elements of the planned data processing to Norwegian Centre for Research Data (NSD)/SIKT for an assessment.

All participants will be carefully instructed about the aim and nature of our studies prior to participation. Before the study starts, participants will be asked to provide informed consent, including consent for archiving anonymized data. Participation is voluntary. Participants will be told explicitly that they have the right to withdraw from our study without explanation and without penalty.

All personal data will be collected, transferred and stored according to NTNU guidelines. The data will be pseudonymized as soon as possible and the encryption key will be stored separately from the data. Access to the pseudonymized data will be limited to the PI and primary research team members. The PI will assign access rights and ensure that they are up to date. The encryption key will be deleted when the project ends and data for archiving will be further anonymized.

Relevant documents:

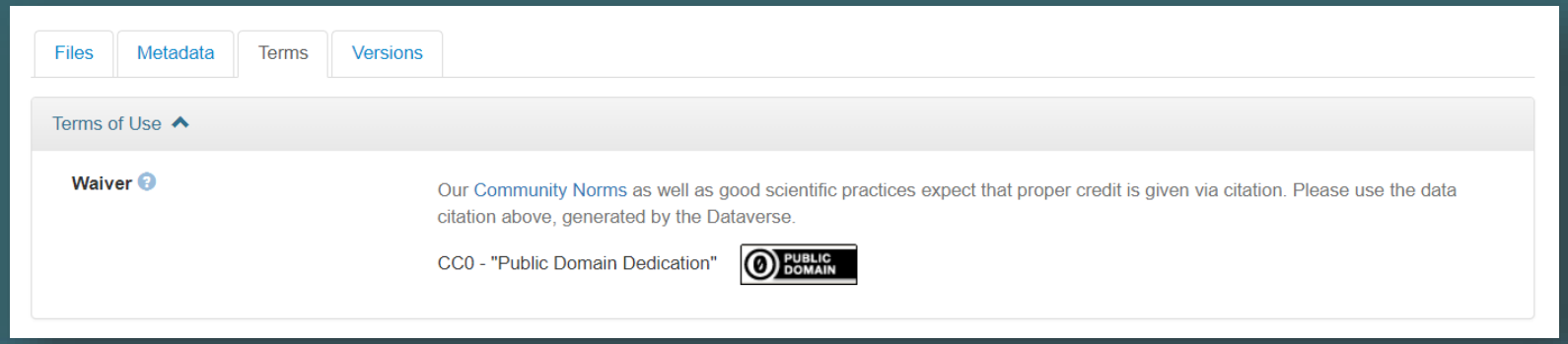
Collection of personal data for research projects (NTNU)

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

- NTNU: Ny IPR-politikk
- Oppdragsforskning/samarbeidsprosjekt: kontrakt regulerer IPR
- Planer om patent? Kontakt NTNU TTO (Technology Transfer Office
- Lovverk som kan være aktuelle:
  - GDPR
  - Eksportkontrollregelverket
  - Sikkerhetsloven

Lisensiering av forskningsdata

- En lisens forteller hva en bruker har lov til å gjøre med et datasett (eller en publikasjon, software etc.)
- Hvordan velge lisens?
  - Sjekk vilkår og konfidensialitet!
  - Standardlisenser: Creative Commons
    - CC0 (Public Domain)
    - NB: Forskere skal alltid oppgi sine kilder og referere til data som gjenbrukes, selv om det skulle være data som er publisert uten formelle krav til sitering



How will possible ethical issues be taken into account, and codes of conduct followed?

- Relevante etiske godkjenninger (REK …)
- Etiske retningslinjer
- Andre etiske hensyn?

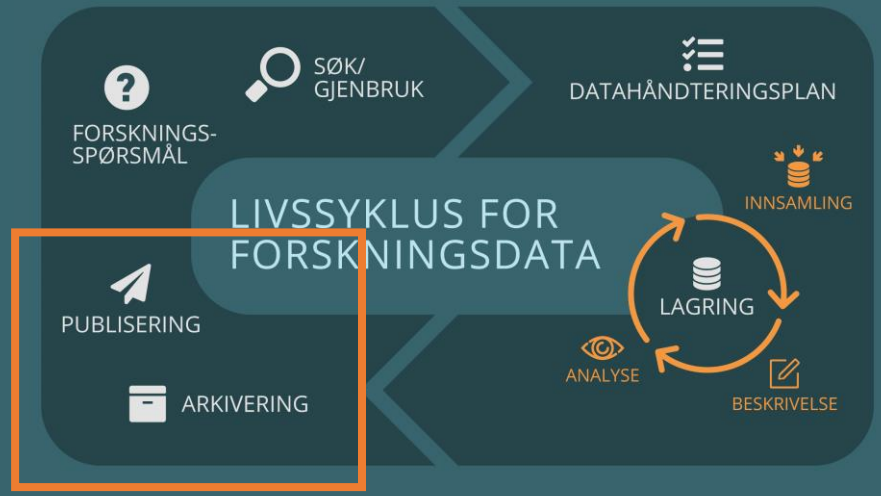
REK

Data sharing and long-term preservation

- **How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?**
- **How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?**
- **What methods or software tools will be needed to access and use the data?**
- **How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?**

Arkivering og publisering av data

- Arkivering kan være lukket (tilgangsbegrenset) eller åpen (publisert)
- NB: Det kan være gode grunner til å ikke publisere data åpent («så åpent som mulig, så lukket som nødvendig»), men disse må spesifiseres
- Sjekk hvilke krav som gjelder (finansiører, tidsskrift, institusjon)
- Publisering av data øker reproduserbarhet, etterprøvbarhet og gjenbruk!
- Vet du hvor du skal arkivere, vet du også mye annet (metadata, lisensmuligheter, tilgang etc.)



Hvordan publisere data?

- Dataarkiv (anbefalt)
  - Generelle arkiv (eks.: Zenodo, Dryad)
  - Fagspesifikke arkiv
  - Institusjonelle arkiv (NTNU Open Research Data)
  - «Data journal» (gjerne i tillegg til arkiv)
    - Eksempel: Scientific Data (Springer Nature)
- Supplement til artikkel
- Blogg, ResearchGate, etc (helst ikke…)

Hvordan finne egnet arkiv?

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

Search

Tips: Velg et arkiv som utsteder DOI (Digital Object Identifier) eller annen persistent identifikator

DataverseNO

Certified by: CORE TRUST SEAL

NTNU Open Research Data

NTNU

DataverseNO > NTNU Open Resear

1 to 10 of 17 Results

Sort

Aqueous glucose measured by NIR spectroscopy

Feb 11, 2021

Fuglerud, Silje Skeide, 2021, "Aqueous glucose measured by NIR spectroscopy", https://doi.org/10.18710/NSHFAK, DataverseNO, V1

Near infrared spectroscopy (NIR) is a promising technique that could be used for continuous blood glucose monitoring in the treatment of diabetic patients. Four interferents (lactate, ethanol, caffeine and acetaminophen) were introduced to study how the glucose predictions varied...

Replication data for : MultiPACK project_An integrated CO2 unit for heating, cooling and DHW installed in a hotel-Data from the field.

Feb 5, 2021

Tosato, Giacomo, 2020, "Replication data for : MultiPACK project_An integrated CO2 unit for heating, cooling and DHW installed in a hotel-Data from the field.", https://doi.org/10.18710/8O9ZV0, DataverseNO, V2

This dataset provides data for a new heat pump unit intended to provide heating, cooling and hot water to a hotel and using CO2 as the working fluid. The heat pump was installed in mid-2018 in a hotel located in a touristic area in North Italy and open nearly all over the year. T...

Replication Data for: Mapping and Modeling Sources of Natural Remanent Magnetization in the Microcline-Sillimanite Gneiss, Northwest Adirondack Mountains: Implications for Crustal Magnetism

Feb 2, 2021

Pastore, Zeudia, 2021, "Replication Data for: Mapping and Modeling Sources of Natural Remanent Magnetization in the Microcline-Sillimanite Gneiss, Northwest Adirondack Mountains: Implications for Crustal Magnetism", https://doi.org/10.18710/OGQWZM, DataverseNO, V1

The dataset includes magnetic inversion results and high-resolution magnetic scan data (before and after alternating field demagnetization to 100 mT) of a 30µm thick section of a Microcline-Sillimanite Gneiss sample from the Russel Belt in the northwest Adirondack Mountains, New...

Hvordan gjøre data FAIR - kortversjon

- Arkiver data og metadata i en søkbar ressurs (forskningsdataarkiv)
- Legg ved all nødvendig informasjon for at framtidige brukere (mennesker og maskiner) skal kunne lese og forstå datasettet
- Bruk tilgjengelige fagstandarder for data og metadata
- Bruk åpne formater og tildel persistente identifikatorer (f.eks. DOI)
- Utstyr datasettet med en egnet lisens

Citation Metadata

Dataset Persistent ID	doi:10.18710/NSHFAK
Publication Date	2021-02-11
Title	Aqueous glucose measured by NIR spectroscopy
Author	Fuglerud, Silje Skeide (NTNU) - ORCID: https://orcid.o
Contact	Use email button above to contact. Fuglerud, Silje Skeide (NTNU)
Description	Near infrared spectroscopy (NIR) is a promising techni treatment of diabetic patients. Four interferents (lactate glucose predictions varied with interferent concentratio when they were not included in the calibration of the m
Subject	Physics; Medicine, Health and Life Sciences; Chemistr
Keyword	NIR Glucose diabetes Near Infrared spectroscopy
Related Publication	Submitted for review. doi:
Producer	Norwegian University of Science and Technology (NTN
Production Place	NOFIMA, Ås, Norway
Contributor	Supervisor : Hjelme, Dag Roar Data Collector : Fuglerud, Silje Skeide Research Group : Artificial Pancreas Trondheim Supervisor : Aksnes, Astrid Related Person : Ellingsen, Reinold
Grant Information	Samarbeidsorganet: 46055510 Norsk Forskningsråd: 248872
Distributor	NTNU Open Research Data (Norwegian University of Science and Technology) (NTNU) https://dataverse.no/dataverse/ntnu
Depositor	Fuglerud, Silje Skeide
Deposit Date	2020-11-10
Kind of Data	Near Infrared absorption spectra

```
NERAL INFORMATION
-----------------
 Title of Dataset: DAS4Microseism - Svalbard distributed acoustic sensing (DAS) strain data for oceanographic study
 DOI: https://doi.org/10.18710/VPRD2H
 Contact Information
   // Name: Kittinat Taweesintananon
   // Institution: NTNU--Norwegian University of Science and Technology
   // Email: kittinat.taweesintananon@ntnu.no
   // ORCID: https://orcid.org/0000-0002-2700-8366

 Contributors: Kittinat Taweesintananon, Martin Landrø, Ståle Emil Johansen, John Robert Potter, Robin André Rørstadb
 Kind of data: Processed strain data in nanostrain unit measured by OptoDAS interrogator.
 Date of data collection/generation: From 2020-06-23 to 2020-08-04.
 Geographic location: Isfjord, Longyearbyen, Svalbard, Norway.
 Funding sources: The Research Council of Norway--Grant no. 309960 for the Centre for Geophysical Forecasting (CGF).

 Description of dataset:
is data set is used to conduct research as presented in a manuscript entitled "Observation of atmospheric and oceanic



------------------------
THODOLOGICAL INFORMATION
------------------------
 Description of sources and methods used for collection/generation of data:
 use a dark fiber of SMF-28 single mode silica type in an existing submarine telecommunication cable, which was trencl
```

NB: Så åpent som mulig, så lukket som nødvendig!

- Hvis du gjenbruker eksisterende data, sjekk lisenser og eventuelle restriksjoner før eventuell viderepublisering!
- Persondata kan som regel ikke arkiveres eller deles uten samtykke fra deltakere.
- Det finnes muligheter for arkivering med tilgangsbegrensning, for eksempel Sikt forskningsdataarkiv.
- Det kan foreligge begrensninger på grunn planer om patentering, kommersielle hensyn etc.
- I samarbeidsprosjekter/oppdragsfinansiert forskning skal det foreligge kontrakter som regulerer rettighetene til resultater og som slår fast hva som eventuelt ikke kan deles.

Data management responsibilities and resources

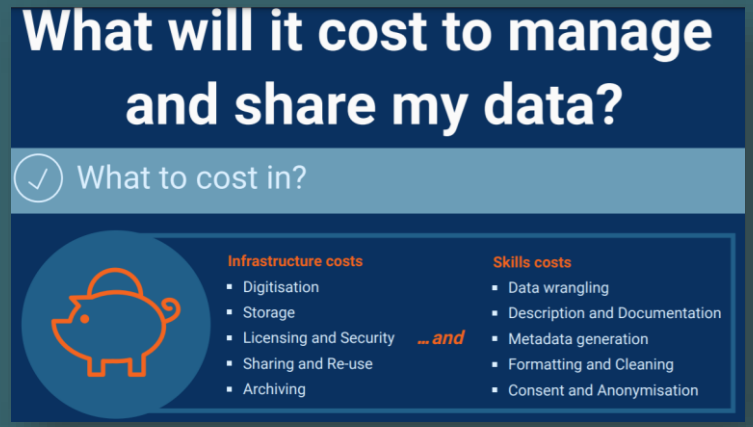
- **Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?**
- **What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?**

Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

- Fordel roller og ansvar for de ulike delene av datahåndteringen
- NFR:
  - Hvilke roller har hvilket ansvar for datahåndteringsaktiviteter i prosjektet? Eksempler på aktiviteter er datafangst, produksjon av metadata, datakvalitet, lagring og sikkerhetskopiering, langtidsbevaring og datadeling. Ansvarlige individer bør oppgis, hvis mulig.
  - For samarbeidsprosjekter; hvordan koordineres ansvar for datahåndtering mellom partnere?
  - Hvem er ansvarlig for å implementere datahåndteringsplanen og for at planen blir gjennomgått og jevnlig oppdatert?

What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

- Tenk over hvilke ressurser (tid, arbeidskraft, penger) som vil kreves for å håndtere data i prosjektet.
  - Eks: Ekstra lagringskapasitet? Må dataene tilrettelegges før publisering?
- Flere finansiører lar deg inkludere kostnader forbundet med datahåndtering i prosjektbudsjettet
  - OpenAire Data Costing Tool



Hva skal du gjøre med planen?

- NFR: planen skal leveres med revidert søknad, endelig versjon i forbindelse med sluttrapportering
- Datahåndteringsplaner bør være offentlige og bør dermed publiseres åpent *[Kan for eksempel publiseres på Zenodo]*.
- Bruk den som et verktøy og oppdater ved behov!

Test_Science_Europe

Project Details | Contributors | Plan overview | Write Plan | Share | Download

Format

pdf

Download settings

Optional Plan Components

- [x] project details coversheet
- [x] question text and section headings
- [x] unanswered questions

PDF formatting

Font

Face: Arial, Helvetica, Sans-Serif

Size (pt): 10

Margin (mm)

Top | **Bottom** | **Left** | **Right**

Download Plan

Plan Overview

A Data Management Plan created using DMPonline

Title: Test_Science_Europe

Creator: Ane Gabrielsen

Affiliation: Other

Funder: Science Europe

Template: Science Europe

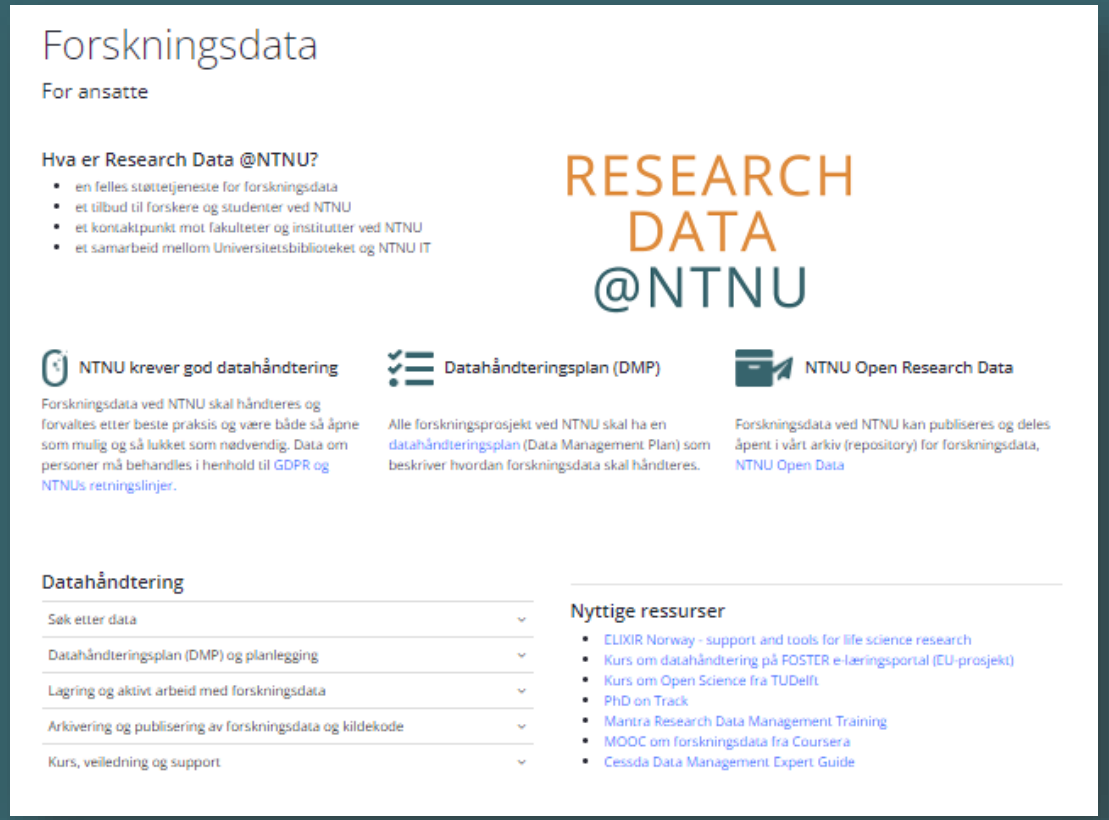
Project abstract:

Short description of project focusing on data.
This DMP includes excerpts from the DMP Evaluation Rubric of Science Europe Practical Guide to the International Alignment of Research Data Management - Extended Edition, DOI doi.org/10.5281/zenodo.4915861
Example answers are from ETH Zürich Data Management Plan SNSF – Guidance on Applications, CC BY-SA ETH Zürich.

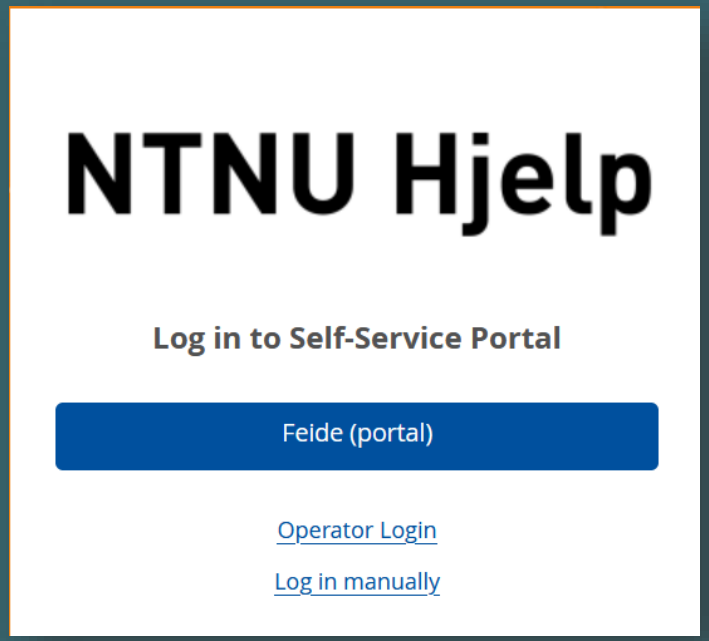
ID: 103561

Hjelp og støtte

Støttetjeneste: Research Data @NTNU



- Kontakt oss gjennom NTNU Hjelp



NTNU IT FORSKNINGSSTØTTE

DIN SAMARBEIDSPARTNER FOR IT-UTFORDRINGER KNYTTET TIL FORSKNING

High Performance Computing (HPC)

→ HPC VED NTNU

IT-støtte for PhD

→ FINN UT MER

Labinstrumentering

→ MER OM LABINSTRUMENTERING

10101
01010

Research Software Development (RSE)

→ MER OM RSE

Science Europe Practical Guide

Core Requirements for Data Management Plans: six aspects that every DMP should cover, with detailed guiding questions.

Criteria for the Selection of Trustworthy Repositories: four topics detailing criteria that every trusted repository should meet.

Guidance for Researchers: more detailed information and examples to support researchers in complying with organisational requirements.

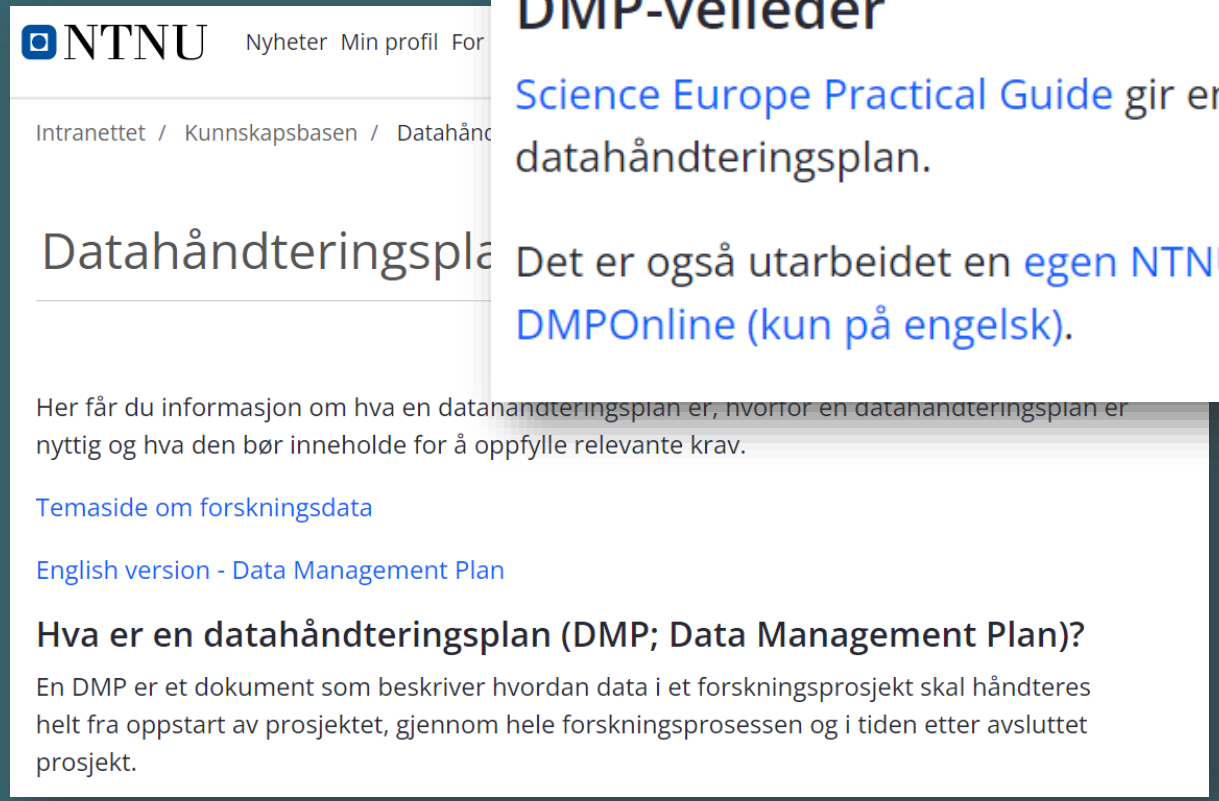
Guidance for Reviewers: guidance to support the evaluation of DMPs by reviewers, aligned with the DMP core requirements presented in previous chapters.

Science Europe Practical Guide

DMP Evaluation Rubric

DMP Question	DMP Guidance	Performance Levels	
GENERAL INFORMATION			
Guidance for Researchers		**Sufficiently Addressed The DMP…**	**Insufficiently Addressed The DMP…**
Administrative information	• Provide information such as name of applicant, project number, funding programme, version of DMP.	• Contains the minimal information required to identify the applicant and the references of the project.	• Provides no or limited information, which makes it hard to identify who is responsible for the project.
1 DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA			
Guidance for Researchers		**Sufficiently Addressed The DMP…**	**Insufficiently Addressed The DMP…**
1 a **How will new data be collected or produced and/or how will existing data be re-used?**	• Explain which methodologies or software will be used if new data are collected or produced. • State any constraints on re-use of existing data if there are any. • Explain how data provenance will be documented. • Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.	• Gives clear details of where the existing data come from and how new data will be collected or produced. It clearly explains methods and software used. • Explains, if existing data are re-used, how these data will be accessed and any constraints on their re-use. • Explains clearly, if applicable, why new data must be collected, instead of re-using existing data.	• Provides little or no details on where the data come from and what data will be collected or re-used. • Does not, if applicable, provide sufficient rationale for generating new data.

NTNUs veiledning



DMP-veileder

Science Europe Practical Guide gir en innføring i kjerneelementene som bør være med i en datahåndteringsplan.

Det er også utarbeidet en egen NTNU-veileder til malen Science Europe i verktøyet DMPOnline (kun på engelsk).

DMP guidance

This guidance document will help you write a DMP for your project based on the Science Europe template in the DMPOnline tool. If you choose other tools or templates, this guide might still be useful, even if topics and questions may not be identical.

The guide provides NTNU specific guidance in addition to excerpts from the DMP Evaluation Rubric of Science Europe Practical Guide to the International Alignment of Research Data Management - Extended Edition (DOI 10.5281/zenodo.4915861) describing the requirements for a sufficiently addressed DMP.

For example answers, see attachment (pdf).

Section: Storage and backup during the research process

Question: How will data and metadata be stored and backed up during the research process?

A sufficiently addressed DMP

- *clearly (even if briefly) describes://*
  - *the location where the data and backups will be stored during the research activities.*
  - *how often backups will be performed.*
  - *the use of robust, managed storage with automatic backup (for example storage provided by the home institution).*

or

- *explains why institutional storage will not be used (and for what part of the data) and describes the (additional) locations, storage media, and procedures that will be used for storing and backing up data during the project.*

«Reviewer's Rubric» fra Science Europe

Describe where your data will be stored during the project period. We recommend using NTNU's standard storage solutions (see NTNU storage guide). For specific information about procedures for back-up for the solution you choose for your project, contact the IT support at NTNU.

Storing data on laptops, external hard drives, or external storage devices such as USB sticks is not recommended. Be sure to consider information security as well as data integrity and accessibility.

NTNU-relevant informasjon

Plan med eksempelsvar (laget i DMPOnline)

Plan Overview

A Data Management Plan created using DMPonline

Title: Example_Answers_NTNU

Creator: Ane Gabrielsen

Affiliation: Other

Funder: Science Europe

Template: Science Europe

Project abstract:

This DMP contains example answers. Some of them are from existing DMPs, others are completely made up. They are only intended as guidance and inspiration and have not been reviewed or otherwise quality controlled.

Note that the answers are not representing one specific project.

Questions regarding research data management at NTNU? Please contact Research Data @NTNU.

Eksempler på DMP

- DMPOnline har mange offentlig tilgjengelige DMPer fra flere fagfelt og basert på ulike maler.
- Flere prosjekt publiserer ulike versjoner av DMPer på plattformer som Zenodo. Her er et eksempel på en første versjon av en DMP fra et EU-prosjekt.

Nyttige lenker

- Research Data @NTNU
- Datahåndteringsplan (NTNU-wiki)
- Science Europe Practical Guide (minimumskrav til datahåndteringsplan med forklaring og veiledning)
- FAIR data principles
- How to FAIR
- Forskningsrådet: Deling av forskningsdata