

Planlegging av forskningsdatahåndtering

Matthew Good, Agata Bochynska

Senter for digital forskerstøtte, Universitetsbibliotek, Universitetet i Oslo

13.11.2024

Spørsmål? Kontakt oss på research-data@uio.no

Agenda:

- Forskjellige typer forskningsdata
- Forskningsdatahåndtering
- Datahåndteringsplaner

Pause

- Dataklassifisering og lagring
- Dataorganisering

Pause

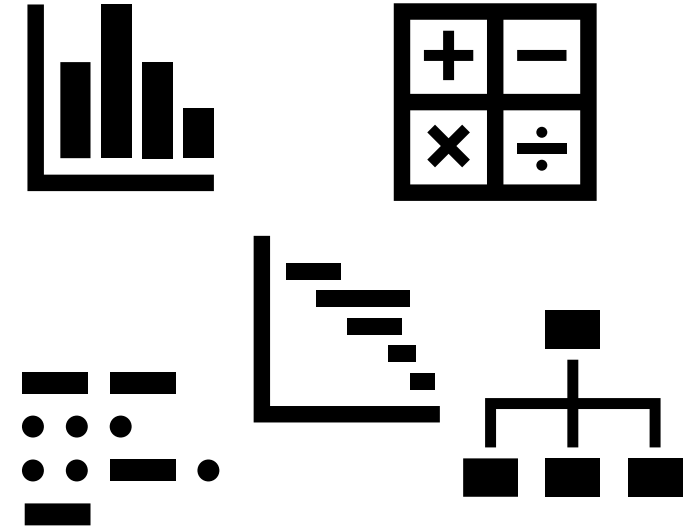
- Dokumentasjon og metadata
- Oppsummering og endelige spørsmål
- Ressurser og slutt

Hva er forskningsdata?





Kvalitativ
(deskriptiv)



Kvantitativ
(numerisk)

Forskjellige typer forskningsdata

Research data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

Ref: Borgman, Christine L. 2015. *Big Data, Little Data, No Data : Scholarship in the Networked World*. Cambridge, MA: MIT Press.

- Primære (ny) eller sekundære (eksisterende)
- Observasjoner, eksperimentell, beregninger, kilder
- Tekst, lyd, video, bilde
- Persondata
 - Generelle persondata, e.g. navn, adresse, stemmen
 - Sensitive persondata, e.g. helseinformasjon, etnisitet, religion

Hva er persondata?

Alle typer vurderinger og opplysninger som kan knyttes til en identifisert eller identifiserbar fysisk person

Med andre ord, alle data om levende personer som de kan identifiseres fra.

Direkte og indirekte identifiserende opplysninger

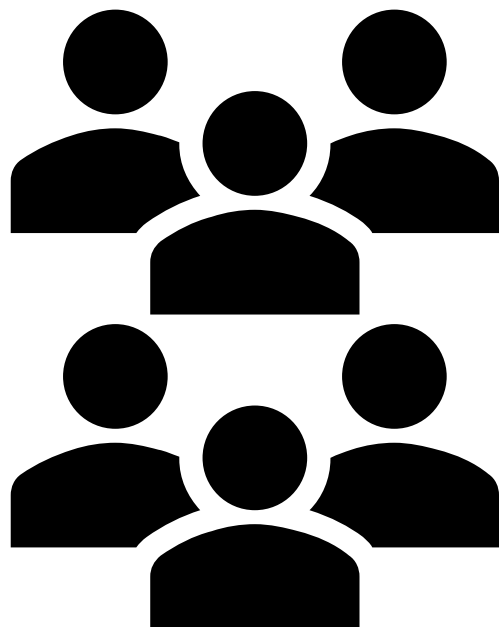
Et direkte identifikator er informasjon som er tilstrekkelig for å identifisere en person uten andre informasjon

e.g. en persons navn, fødselsnummer, et ansiktsbilde, fingeravtrykk

Et indirekte identifikator er informasjon som kan føre til identifisering når den kombineres med annen tilgjengelig informasjon

e.g. lokaliseringsdata, nettidentifikatorer, uvanlig jobbtittel, en stilling som til enhver tid kun innehas av én person, alder, utdanning, etnisk bakgrunn

Anonyme data eller ikke?



Hva er en **generell personopplysning**?

- navn, adresse, alder, telefonnummer, e-postadresse og fødselsnummer
- innhold i saksdokumenter, utredninger eller vurderinger
- innholdet i e-postkommunikasjon
- video- og lydopptak hvor enkeltpersoner kan gjenkjennes
- bilder hvor enkeltpersoner kan gjenkjennes
- logging av aktivitet i datasystemer hvor loggene kan knyttes til bestemte personer, for eksempel registrering av hvem som til enhver tid er pålogget ulike datasystemer

Særlige kategorier av personopplysninger (sensitive data)



- helseopplysninger og helserelaterte forhold
- genetiske og biometriske opplysninger med det formål å entydig identifisere en fysisk person
- etnisk eller rasemessig opprinnelse
- politiske, filosofiske eller religiøse oppfatninger og livssyn
- seksuell orientering eller seksuelle forhold
- fagforeningsmedlemskap

Forskjellige typer forskningsdata

Research data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

Ref: Borgman, Christine L. 2015. *Big Data, Little Data, No Data : Scholarship in the Networked World*. Cambridge, MA: MIT Press.

- Primære (ny) eller sekundære (eksisterende)
- Observasjoner, eksperimentell, beregninger, kilder
- Tekst, lyd, video, bilde
- Persondata
 - Generelle persondata, e.g. navn, adresse, stemmen
 - Sensitive persondata, e.g. helseinformasjon, etnisitet, religion

Oppgave 1: Datatyper

Tenk – Diskuter – Del

Tenk på hvilke typer data du samler inn
(eller skal samle inn) i ditt prosjekt.

Del og diskuter med sidemann!

5 min

4.0

4th revolution

Cyber physical
systems



3.0

3rd revolution

Electronic and IT
systems, automation



2.0

2nd revolution

Mass production and
electricity



1.0

1st revolution

Mechanization, steam
and water power



“**Open Science** has the potential of making the scientific process more **transparent, inclusive** and **democratic**. It is (...) a true game changer in bridging the science, technology and innovation gaps and fulfilling the **human right to science**.”



UNESCO Recommendation on Open Science

<https://youtu.be/I3Wkvx> ZaFo

<https://www.unesco.org/en/natural-sciences/open-science>

Gjenfinnbare

Tilgjengelige

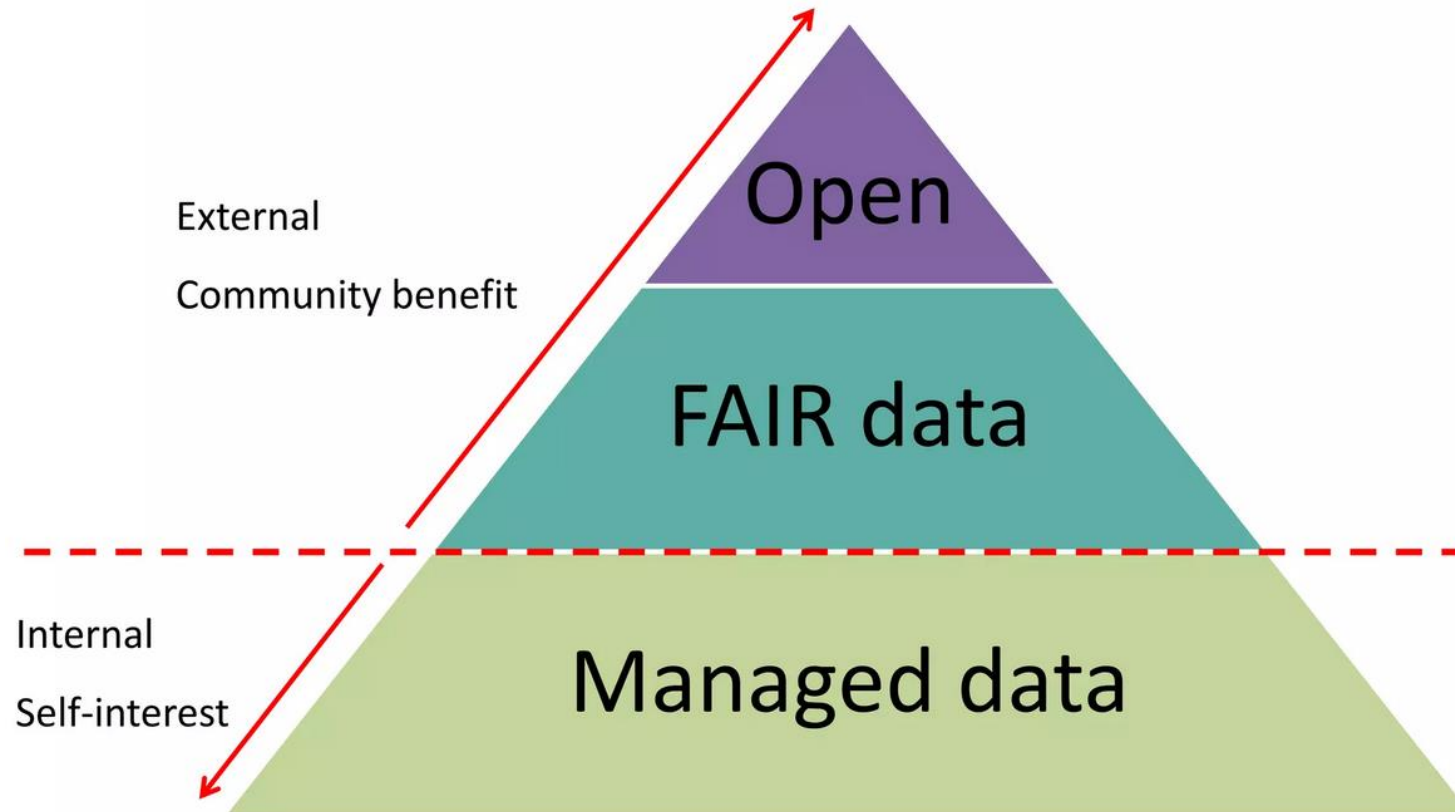


Scriberia 

Samhandlende

Gjenbrukbare

Så **åpent** som mulig,
så **lukket** som nødvendig





Beskyttelse
Organisering
Dokumentasjon (og metadata)
Klassifisering
Korttidslagring
Langtidslagring
Lisensiering
Deling
Formidling (publisering)
Gjenbruk

Datahåndteringsplan (DHP)

- Et **levende dokument**
- Spesifiserer hvilke **type** data behandles i prosjektet
- Beskriver hvordan du skal **håndtere** data i prosjektet (organisere, dokumentere, klassifisere, lagre)
- Informasjon om **teknisk behov**
- Beskriver om og hvordan data kan **deles**
- **Avtale** mellom prosjektdeltakere
- Et **verktøy** for å holde oversikt over dataene



Datahåndteringsplan (DHP)

Tekst-baserte maler

- UiOs mal (på [Norsk](#) og [Engelsk](#)) - et enkelt Word-mal med lenker til UiOs ressurser
- [Horizon Europe mal](#) - anbefalt for bruk med Horizon Europe prosjekter

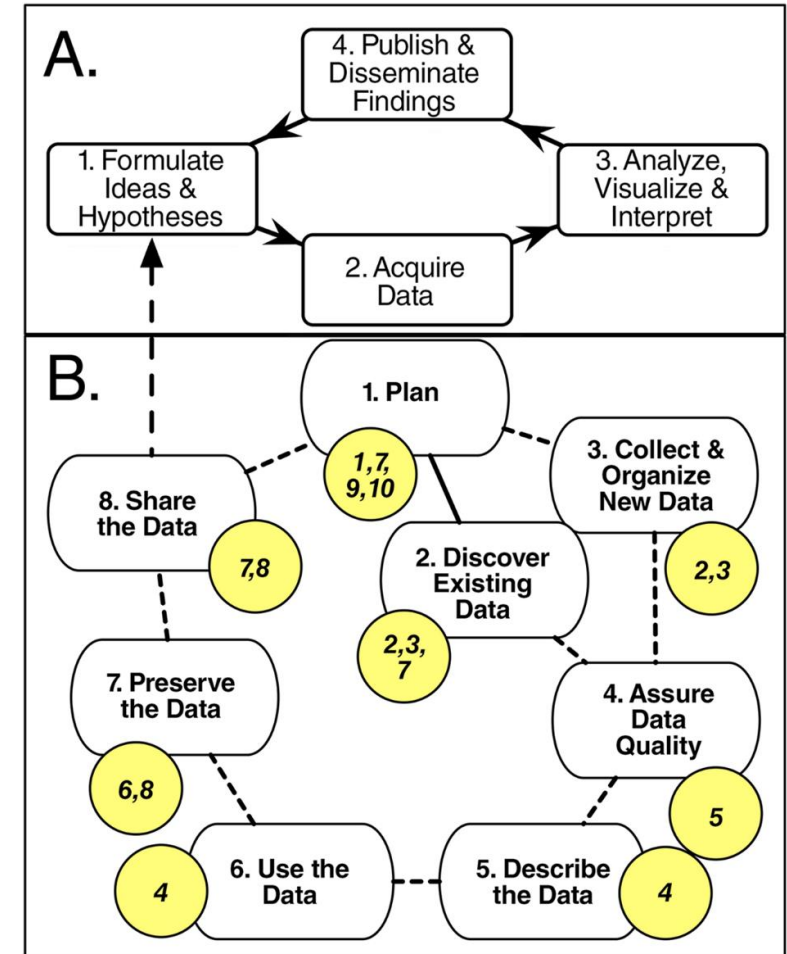
Interaktiv verktøy og nettskjemaer som støtter skriving

- [Sikt \(tidligere NSD\)](#) - enkelt og lett å bruke mal med begrenset funksjonalitet
- [DMPonline](#) og [DMPTool](#) - generisk verktøy som støtter flere maler; kan brukes med DHP-er som skal deles åpent i fremtiden



10 Simple Rules to Writing a Data Management Plan

1. Forstår **krav** fra forskningssponsoren
2. Identifiser hvilke **data** som skal samles inn
3. Definer hvordan dataene skal **organiseres**
4. Forklar hvordan dataene skal **dokumenteres**
5. Beskriv hvordan **kvaliteten** skal opprettholdes
6. Utvikle en solid **lagrings- og arkiveringsstrategi**
7. Definer prosjektets **datapolitikk**
8. Beskriv hvordan dataene skal **formidles**
9. Tildel **roller og ansvar**
10. Utarbeid et realistisk **budsjett**



La oss se på et eksempel!

Trinity College Botanic Garden long-term monitoring program
(DCC template)

Flere eksempler finnes i "Public DMPs" seksjon på DMPonline:

https://dmponline.dcc.ac.uk/public_plans

Oppgave 2: DHP

Skriv og diskuter

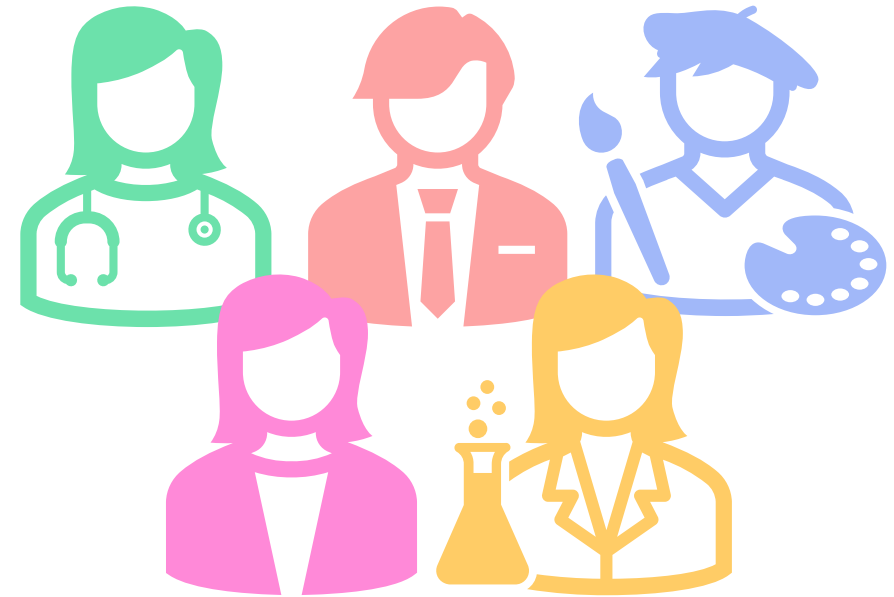
Skriv ned stikkord for de viktigste elementene du vil inkludere i en datahåndteringsplan

2 min

Dataklassifisering og valg av lagringsløsning

Data klassifisering

- All informasjon skal ha en entydig og identifiserbar eier
- Det skal være mulig å finne ut hvem som er ansvarlig for at informasjonen er vedlikeholdt, oppdatert og riktig merket.
- **Eier av informasjonen skal klassifisere informasjonen. Eier er også ansvarlig for vurderingene som ligger bak klassifiseringen**



Data klassifisering

Dataklassifisering ved UiO er basert på hvor skadelig det ville være dersom informasjonen blir kjent for uvedkommende.



<https://www.uio.no/tjenester/it/sikkerhet/isis/tillegg/lagring/infoklasser.html>

Hvor kan data i de ulike klassene lagres?

Lagringsguiden

<https://www.uio.no/tjenester/it/sikkerhet/isis/tillegg/lagringsguide.html>



Grønn data

Åpen eller fritt tilgjengelig til allmennheten

Benyttes dersom **det ikke forårsaker noen skade for institusjonen**, eller samarbeidspartner hvis informasjonen blir kjent for uvedkommende.



Grønn data

Åpen eller fritt tilgjengelig til allmennheten

Eksempler er:

- en web-side som presenterer en avdeling, et kurs eller en enhet som legges åpent ut på internett
- studiemateriell som ligger åpent, men som er merket med en gitt lisens og/eller opphavsrett
- forskningsdata som ikke trenger noen beskyttelse (forskeren står ansvarlig for denne vurderingen)
- undervisningsmaterieell som ikke trenger noen beskyttelse (underviseren står ansvarlig for denne vurderingen)



Grønn data

Åpen eller fritt tilgjengelig til allmennheten

Anbefalte lagringsløsninger:

Hvor du vil, men **helst ikke** bare på din private bærbare pc (svært sårbar!)



Gul data

Begrenset informasjon som ikke er åpen for alle

Denne klassen benyttes dersom det vil kunne forårsake **en viss skade** hvis informasjonen blir kjent for uvedkommende.

Informasjonen må ha en viss beskyttelse og kan være tilgjengelig for både eksterne og interne, med kontrollerte tilgangsrettigheter.



Gul data

Begrenset informasjon som ikke er åpen for alle

Eksempler på slik informasjon kan være:

- enkelte arbeidsdokumenter
- informasjon som er unntatt offentlighet
- mange typer av personopplysninger
- karakterer
- studentarbeider
- eksamensbesvarelser
- upubliserte forskningsdata og -arbeider



Gul data

Begrenset informasjon som ikke er åpen for alle

Anbefalte lagringsløsninger:

- **Ikke** din bærbare pc, med mindre du overholder [retningslinjene for bruk av private maskiner](#)
- Kryptert minnepinne eller ekstern harddisk
- Kryptert bærbare pc driftet av UiO
- UiO skytjenester (UiO OneDrive, UiO Google Drive)
- UiO lagringshotell
- UiO personlig forskerlagring
- UiO fellesområde (for enhet eller gruppe)
- UiO epost; UiO Teams; UiO Zoom; UiO Nettskjema; UiO Canvas, UiO Vortex, ...



Rød data

Fortrolig informasjon som universitetet er pålagt å begrense tilgangen til i lov, forskrift, avtaler, reglementer og annet regelverk

Benyttes hvis det vil forårsake skade for offentlige interesser, universitetet, enkeltperson eller samarbeidspartner hvis informasjonen blir kjent for uvedkommende.



Rød data

Fortrolig informasjon som universitetet er pålagt å begrense tilgangen til i lov, forskrift, avtaler, reglementer og annet regelverk

Eksempler kan være:

- [særskilte kategorier av personopplysninger \(tidligere kalt «sensitive personopplysninger»\)](#)
- data underlagt [eksportkontroll](#)
- personalmapper
- en del informasjon om f. eks. sikring av bygninger og IT-systemer
- helseopplysninger



Rød data

Fortrolig informasjon som universitetet er pålagt å begrense tilgangen til i lov, forskrift, avtaler, reglementer og annet regelverk

Anbefalte lagringsløsninger:

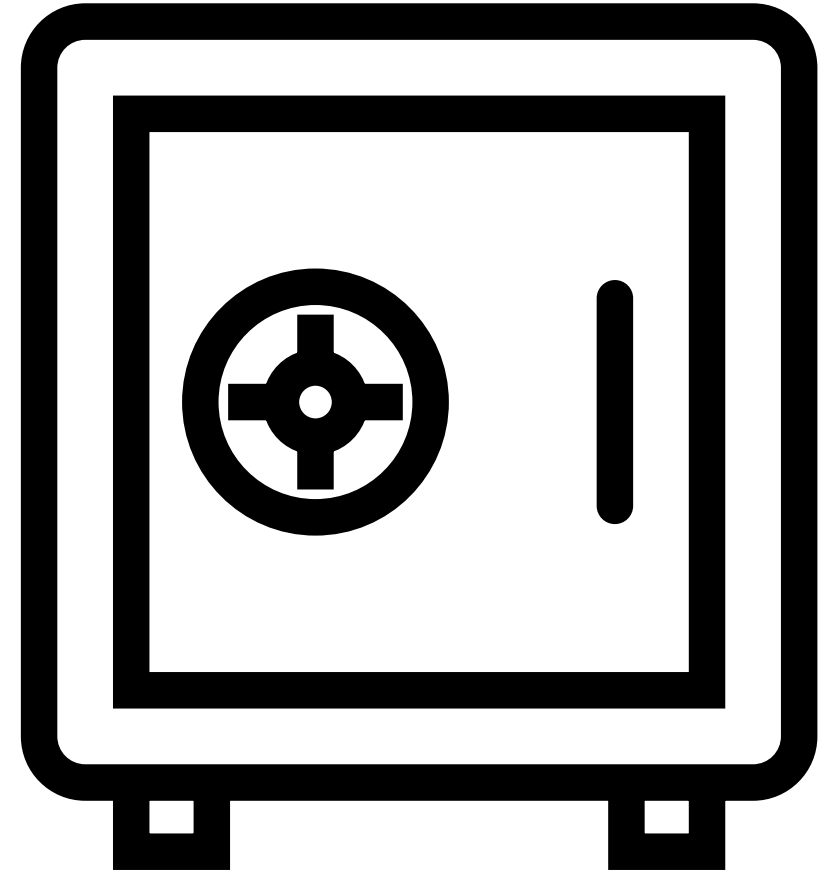
- Bærbare pc med fullkryptert disk, kryptert minnepinne, kryptert ekstern harddisk
- UiO TSD Tjenester for sensitive data
- *Educloud Research*
- Vortex, på fellesområde, eller på UiOs lagringshotell etter en egen vurdering og tilrettelegging av spesielt tilgangskontroll



Svart data

Strengt fortrolig omfatter samme type informasjon som Fortrolig (rød), men der spesielle hensyn gjør at man ønsker å beskytte dataene ytterligere.

Benyttes dersom det **vil kunne forårsake betydelig skade** for offentlige interesser, universitetet, enkeltperson eller samarbeidspartner at informasjonen blir kjent for uvedkommende.

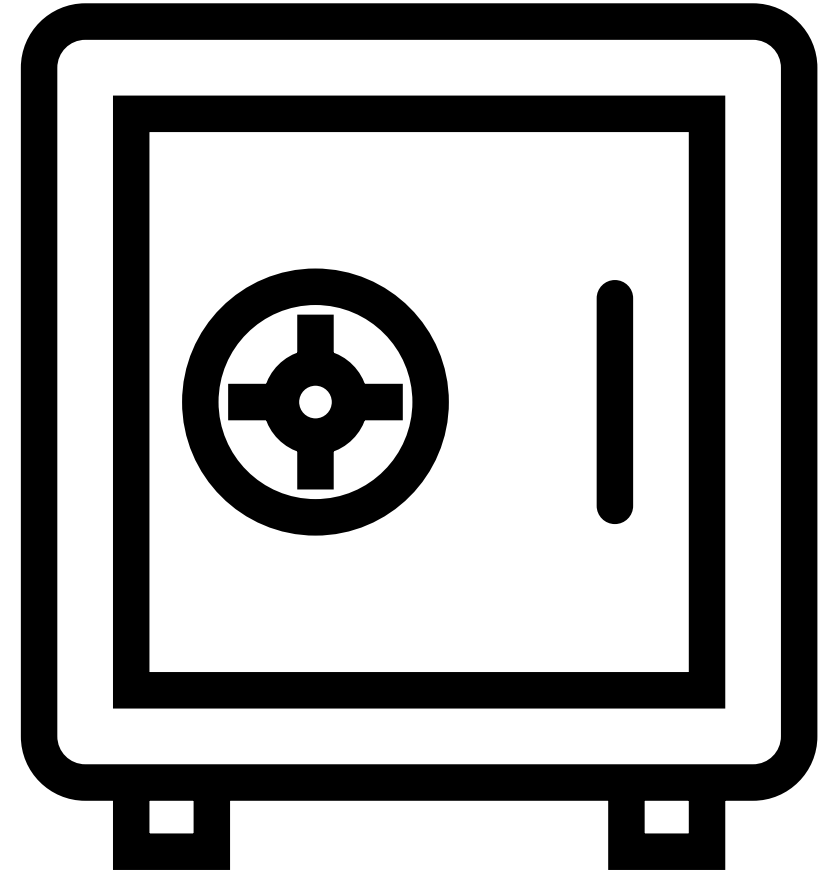


Svart data

Strengt fortrolig omfatter samme type informasjon som Fortrolig (rød), men der spesielle hensyn gjør at man ønsker å beskytte dataene ytterligere.

Eksempler kan være:

- store mengder sensitive personopplysninger
- store mengder helseopplysninger
- forskningsdata og datasett av stor økonomisk verdi



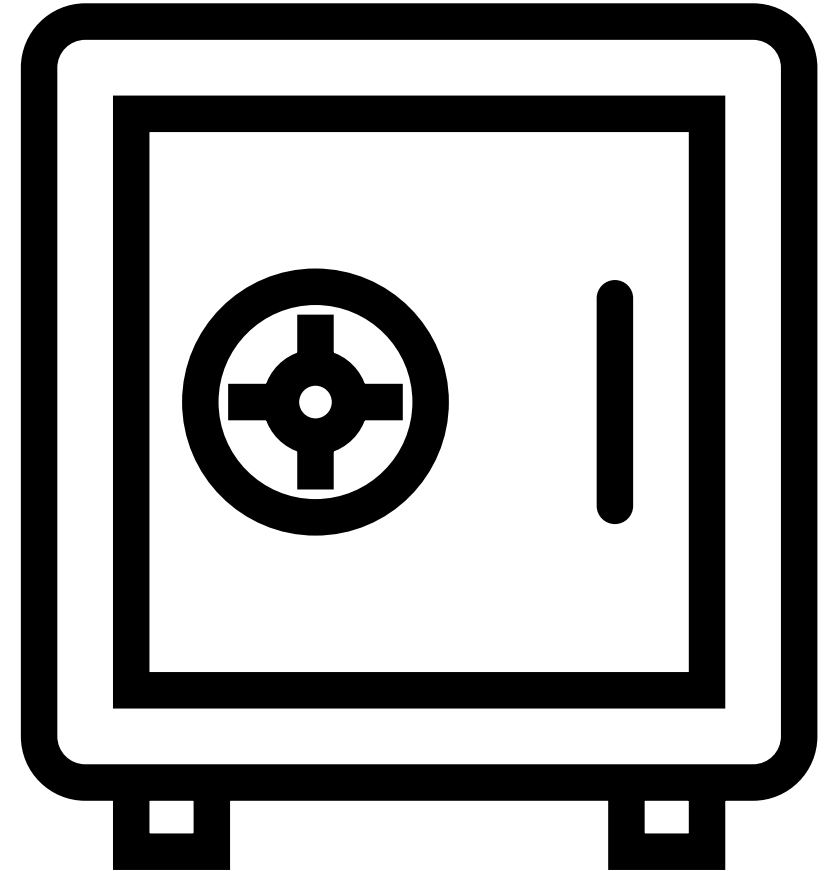
Svart data

Strengt fortrolig omfatter samme type informasjon som Fortrolig (rød), men der spesielle hensyn gjør at man ønsker å beskytte dataene ytterligere.

Plassering av data og informasjon i denne kategorien gjøres i samarbeid med USITs jurister og IT-sikkerhetssjefen.

Anbefalte lagringsløsninger:

- UiO TSD Tjenester for sensitive data



Data klassifisering

Plasser alltid informasjonen i en tilstrekkelig sikker klasse

Hvis du er usikker om noe er rød eller gul, velg rød



Oppgave 3: Dataklasser

Tenk – Diskuter – Del

Tenk på hvilke dataklasser du jobber med (eller skal jobbe med). Hvorfor?

Del og diskuter med sidemann!

5 min

Skytjenester

Skytjenester som er vurdert eller under vurdering

https://www.uio.no/tjenester/it/ny-tjeneste/vurderte_skytjenester.html

Tjenester med UiO avtale

<https://www.uio.no/tjenester/it/alle/>



Backup / Sikkerhetskopiering



- Bruk sikre lagringsløsninger der data/filer blir sikkerhetskopierte
- Oppbevar alltid en kopi av ditt rådatamateriale trygt gjemt bort (som i en mappe du aldri redigerer)
- Å lagre filer kun på din egen datamaskin er ikke trygt!
- En minnepinne eller en ekstern harddisk er ikke trygg!
- Hvor mye ekstra arbeid tar det å gjenskape alle dine data eller ditt arbeid?

<https://www.uio.no/tjenester/it/lagring-samarbeid/backup/index.html>

Dataorganisering

Lag en oversiktlig mappesystem

1. Lag en tydelig **mappestruktur**
2. Bruk **beskrivende** og **informative** mappenavn
3. Unngå mapper som er for generelle, bruk flere **undermapper** istedenfor
4. Hold **aktiv** og **avsluttet** deler av prosjektet separat
5. Sett av litt tid for å rydde i filene av og til



SvalbardEx271020

File Home Share View

SvalbardEx271020

Quick access

- Documents
- Downloads
- Pictures
- GEO1010
- Masterseminar 1_al

ite

ur

(V

ev

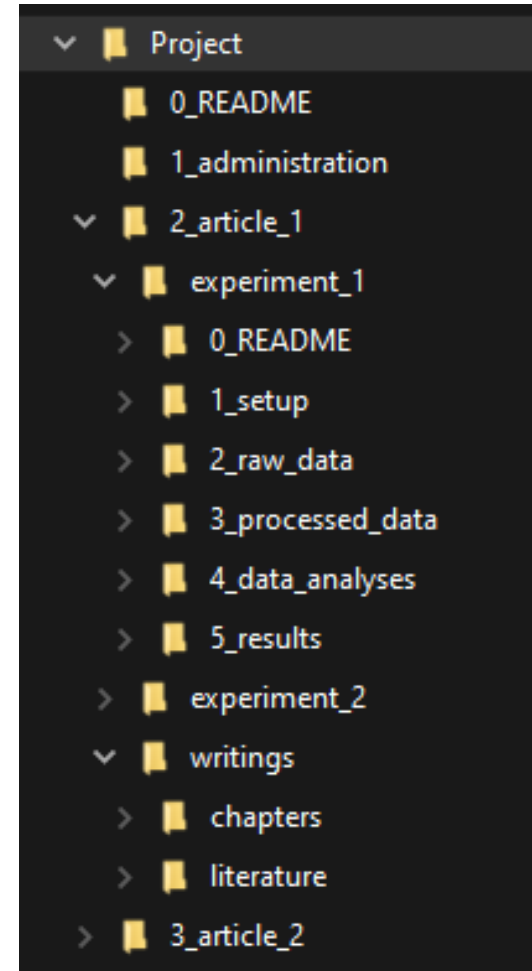
Network

Name	Date modified	Type	Size
P-10-01	26.10.2020 16:29	File folder	
SEM 2012	26.10.2020 16:25	File folder	
P-10-03-01.tif	11.06.2012 11:17	TIF File	1 880 KB
P-10-03-02b.tif	11.06.2012 11:25	TIF File	1 880 KB
P-10-03-02c.tif	11.06.2012 11:26	TIF File	1 880 KB
P-10-03-05.tif	11.06.2012 11:38	TIF File	1 880 KB
P-10-03-79b.tif	11.06.2012 12:59	TIF File	1 880 KB
P-10-03-92b.tif	11.06.2012 13:26	TIF File	1 880 KB
P-10-03-109.tif	11.06.2012 13:49	TIF File	1 880 KB
P-10-03-119b.tif	11.06.2012 14:02	TIF File	1 880 KB
P-10-03-193b.tif	11.06.2012 15:45	TIF File	1 880 KB
P-10-03-222b.tif	11.06.2012 16:29	TIF File	1 880 KB
P-10-03-222c.tif	11.06.2012 16:30	TIF File	1 880 KB
P-10-03-226b.tif	12.06.2012 09:33	TIF File	1 880 KB
P-10-03-226c.tif	12.06.2012 09:33	TIF File	1 880 KB
Gasser_2014.pdf	27.01.2015 13:11	Adobe Acrobat D...	5 127 KB
Gernigon_and_Bronner_2012.pdf	30.01.2015 15:07	Adobe Acrobat D...	3 445 KB
Gernigon_et_al_2014.pdf	26.01.2015 16:02	Adobe Acrobat D...	54 447 KB
Glorstad_Clark-2010.pdf	15.09.2014 15:06	Adobe Acrobat D...	10 269 KB
Gronlie et al. 1980 Seismic inversion of Bj...	09.10.2012 13:28	Adobe Acrobat D...	643 KB
Gudlaugsson_et_al_1998.pdf	17.10.2014 12:50	Adobe Acrobat D...	3 411 KB
Harland_and_Gayer_1972.pdf	17.10.2014 11:47	Adobe Acrobat D...	1 474 KB
Høy_and_Lundchien_2011_NBarenstSea_...	17.09.2014 14:10	Adobe Acrobat D...	13 600 KB
Isaksen 1996 Organic geocem Bjornoya.p...	09.10.2012 13:24	Adobe Acrobat D...	1 439 KB
Klausen et al 2015 Triassic Snadd in Baren...	30.03.2015 13:15	Adobe Acrobat D...	14 911 KB
Klausen_et_al_2014_Triassic_Snadd_Fm.pdf	08.09.2014 13:20	Adobe Acrobat D...	5 069 KB
Klitzke_et_al_Barents_Sea_Region_2014-pr...	21.01.2015 15:22	Adobe Acrobat D...	2 951 KB
Gabrielsen et al. 1990_Structural_E_BS_NP...	11.03.2013 15:45	Adobe Acrobat D...	9 099 KB
Gac_et_al_2013_ultra_deep_EBB.pdf	20.01.2015 12:15	Adobe Acrobat D...	1 008 KB
Abstract.doc	15.09.2014 16:04	Microsoft Word 9...	234 KB
Copy of NuAge_090214-120115-run1.xls	22.01.2015 13:06	Microsoft Excel 97...	1 390 KB
CPP_Svalbard_Bjornoya_Copy of CUMUL...	17.12.2014 15:35	Microsoft Excel 97...	2 429 KB
2013 Wintermeeting poster-1_Edina_Final...	07.01.2013 15:12	Adobe Acrobat D...	130 KB
ICPMS data	26.10.2020 16:31	File folder	
Franz Josef presentation.pptx	09.01.2013 12:06	Microsoft PowerP...	11 334 KB

Eksempel på et prosjekt
uten mappestruktur

Hvordan det kan se ut:

```
project_name/
├── README.md           # overview of the project
├── data/               # data files used in the project
│   ├── README.md      # describes where data came from
│   └── sub-folder/     # may contain subdirectories
├── processed_data/    # intermediate files from the analysis
├── manuscript/        # manuscript describing the results
├── results/           # results of the analysis (data, tables, figures)
├── src/               # contains all code in the project
│   ├── LICENSE        # license for your code
│   ├── requirements.txt # software requirements and dependencies
│   └── ...
└── doc/               # documentation for your project
    ├── index.rst
    └── ...
```



Research project with a proper data file structure. Image taken from CodeRefinery, Lesson on Reproducible Research.
Shared under CC-BY 4.0 <https://coderefinery.github.io/reproducible-research/organizing-projects/>

Oppgave 4: Mappestruktur

Tenk – Del – Sammenlign

Tenk på den ideelle mappestruktur for ditt forskningsprosjekt. Diskuter deretter i par eller en mindre gruppe.

4 min

Filnavn anbefalinger

- Korte navn (men lang nok at det fortsatt har mening)
- Mest generelle informasjon først, så legge til detaljer
- Bruk understrek mellom separate ord, IKKE bruk mellomrom i filnavn!
- Skriv datoer i omvendt rekkefølge (YYYYMMDD)
- Tall (e.g. versjon nummer) bør ha samme antall siffer (e.g. 01, ikke bare 1)
- Versjon nummer helt til sist
 - Ikke bruk \ / ~ ! @ # \$ % ^ & * () ` ; : < > ? . , [] { } ' " |

20220503_DOEProject_DesignDocument_Smith_v2-01.docx



The diagram shows the filename '20220503_DOEProject_DesignDocument_Smith_v2-01.docx' with several annotations: a blue arrow points down to the 'v2-01' version number; four green arrows point up to the underscores separating the date, project name, document type, and author; and the filename is divided into five colored segments: purple for the date, yellow for the project name, red for the document type, orange for the author, and a light orange segment for the file extension.

Versjonerte data

Versjonering betyr å lagre **nye kopier** av filene dine når du gjør endringer, slik at du senere kan gå tilbake og **hente** spesifikke **versjoner** av filene dine.

DataFilNavn_1.0 = original dokument

DataFilNavn_1.1 = original dokument med mindre endringer

DataFilNavn_2.0 = dokument med vesentlige endringer



Versjonskontroll

Ideelt for **samarbeid**

Sporing og håndtering endringer til en fil eller gruppe filer over tid

Du kan hente frem **hvilken som helst versjon** når som helst

Det kan brukes for dokumenter, programvareutvikling, store nettsider, osv.

Versjonskontroll tabell

Manuell versjonskontroll av endringer i filene

Version	Date	Changes
v01	2023-02-21	KM created the presentation
v02	2023-04-05	JS added the section on informed consent
v03	2023-05-17	KM deleted the section on file formats

Change log:

- Added
- Changed
- Removed

Versjonskontroll system

Git: Gratis og åpenkilde versjonskontroll system



<https://git-scm.com/>
<https://github.com/>
<https://youtu.be/gY2JwRfin1M>

Versjonskontroll system

Git: Gratis og åpenkilde versjonskontroll system

GitHub: er en nettbasert tjeneste for programvareutvikling og versjonskontroll ved hjelp av Git.



<https://git-scm.com/>
<https://github.com/>
<https://youtu.be/gY2JwRfin1M>

Standard (åpen) fil formatter

Regnet som det beste med tanke på gjenbrukbarhet, tilgjengelighet, og bærekraftighet

Eksempler:

- | | | |
|-----------------------|-----------------------------|-----------------------------------|
| • Tekst: | .txt; .pdf (PDF/A) | not .docx |
| • Regneark: | .tsv, .csv | not .xlsx |
| • Lyd: | .wav, .aiff, .flac .mp3 | not .m4a, .ape, .ogg, .wma |
| • Bilde: | .tiff, .png, .jpg | not .psd, .gif, .bmp |
| • Video: | .mp4 | not .avi, .mov |
| • Statistisk analyse: | .R, .Rdata, .dat, .sps, .do | not .por, .sav, .dta |
-
- Trusted data archives often have guidelines and recommendations for this:
 - DataverseNO: <https://site.uit.no/dataverseno/deposit/prepare/#preferred-file-formats>
 - Data Archiving and Networked Services (DANS): <https://dans.knaw.nl/en/file-formats/>

Datadokumentasjon

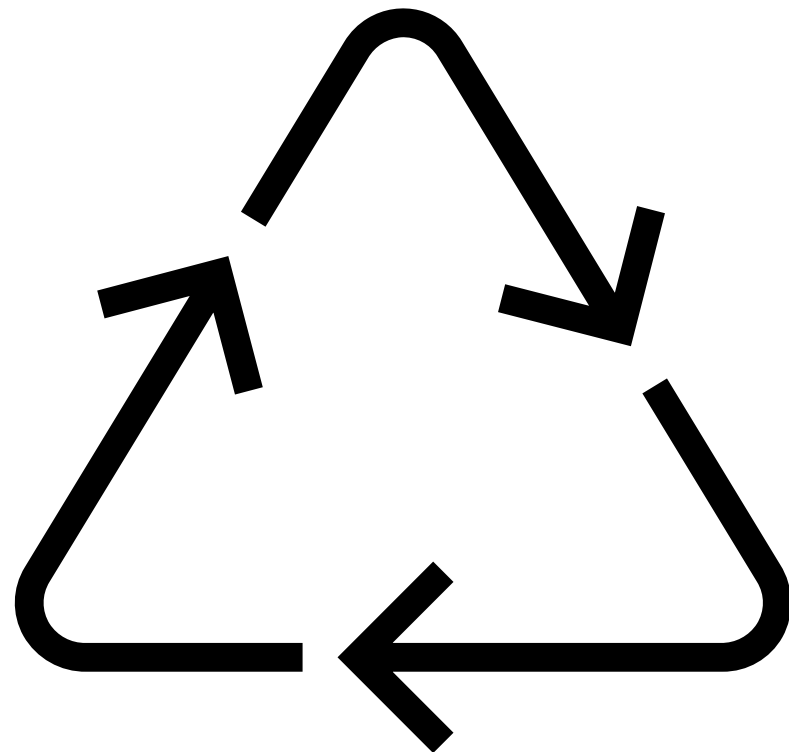
Dokumentasjon

Systematisk dokumentasjon av forskningsdata er avgjørende for å kunne publisere, gjenfinne, gjenbruke og sitere dataene (og FAIR).

Klar og detaljert dokumentasjon kan bidra til å **forbedre datakvaliteten generelt**.

Hvorfor dokumentasjon?

- Hjelper andre (og «fremtidige deg») med å forstå prosjektet ditt og gjenbruke dataene dine:
 - alt du trenger å vite om dataene dine
 - alt dine samarbeidspartnerne trenger å vite om dine data
 - alt du trenger å vite om samarbeidspartnerne data
 - alt andre trenger å vite om dine data for å kunne gjenbruke dem
- Du må ha dokumentasjon når du arkivere dine data



Hvilken type dokumentasjon?

Prosjekt-nivå dokumentasjon forklarer målet med studien, hvilke forskningsspørsmålene/hypotesene de har, hvilke metoder brukes, hvilke instrumenter og målinger brukes, osv.

Data- eller objekt-nivå dokumentasjon gis informasjon om bestemt objekter som bilder, variabler in en database/tabell, transkripsjoner, osv.

Eksempler på dokumentasjon

- Feltjournal
- Labjournal og eksperimentelle protokoller
- Skripter
- Spørreundersøkelse, kodebøker, data dictionaries
- Syntaks og output filer
- Metodologisk rapporter
- Geolokasjon, topografi, retning (e.g. når du samle en prøve)
- Instrumentinnstillinger og kalibrering



Data-nivå dokumentasjon: **Kodebok**

Beskriver alle variabler og deres koder i datasettet

Inneholder ofte:

- variabelnavn, variabelkoder, variabelformatter, manglende data (i kvantitativ forskning)
- koder, code definisjoner, eksempler på hva som skal inkluderes med en gitt kode (i kvalitativ forskning)

Kan også kalles en **dataordbok** (data dictionary på engelsk)

Data-nivå dokumentasjon: Kodebok

Eksempel

sex: biologisk kjønn selv-rapportert
(0 – kvinne, 1 – mann)

gender: selv-rapportert kjønn (0 –
kvinne, 1 – mann, 2 – annen)

age: alder i måneder på testdatoen
(automatisk beregnet i systemet
som forskjellen mellom testdatoen
og fødselsdatoen)

	A	B	C	D	E	F	G
1	subjid	sex	gender	recruitment	condition	group	age
2	10001	1		2	9999	9999	272
3	10002	1	1	1	2	2	187
4	10003	1		1	9999	9999	291
5	10004	1	1	1	1	1	300
6	10005	1	1	1	1	1	288
7	10006	1		6	0	9999	254
8	10007	1	1	4	1	2	233
9	10008	0	2	9999	0	0	266
10	10009	1	1	2	1	1	195
11	10010	1	2	2	1	1	207
12	10011	1	1	1	1	2	234
13	10012	1		2	1	2	231
14	10013	1	1	0	1	2	257
15	10014	0	2	0	0	0	309
16	10015	0	2	1	2	2	254
17	10016	0	2	5	0	9999	161
18	10017	1		6	0	9999	
19	10018	1	1	1	1	2	314
20	10019	1	1	5	0	0	188
21	10020	1		0	1	9999	
22	10021	1	1	4	9999	9999	
23	10022	1		1	1	2	294
24	10023	1	1	0	9999	9999	
25	10024	1	1	1	1	1	291
26	10025	1	1	2	1	1	214

Kontrollerte vokabularer

Standardiserte og organiserte lister av foretrukne termer og fraser brukt for å indeksere og deretter hente ut data og annet innhold

Sikrer en konsekvent måte å beskrive data og deres variabler på

Vanlige typer inkluderer termlister, autoritetsregister og tesauruse

Kontrollerte vokabularer

	A	B	C	D	E	F	G	
1	participant number	age	lifestyle	religion	sex	salary group	occupation group	
2	1	45	1	0	1	A	B	
3	2	54	2	1	1	A	D	
4	3	36	1	0	2	B	A	
5	4	47	1	NA	2	C	C	
6	5	30	2	0	0	B	A	
7	6	32	3	0	1	D	B	
8	7	52	3	0	1	A	NA	
9	8	34	1	1	2	A	A	
10								

	A	B	C	D	E	F	G	
1	ID	yob	habits	beliefs	gender	income	profession	
2	101	1971	2	1	1	A	1	
3	102	1983	4	0	2	B	2	
4	103	1977	3	0	1	C	4	
5	104	1974	1	2		A	2	
6	105	1987	1	1	2	A	1	
7	106	1994	3	0	1		1	
8	107	1990	2	2	2	C	2	
9	108	1988	1	3	1	B	3	
10	109	1982	4	1	0	C	4	
11								

Hvordan lage dokumentasjon?

Mer informasjon er bedre enn lite

Strukturert informasjon er bedre enn ustrukturert

Gjenbruk maler for god dokumentasjon

Det har også blitt vanlig å **opprette flere README-filer** for dokumentasjon på prosjekt- og data-nivå

Hvordan lage dokumentasjon?

- README.txt-filer:

- **Enkelt tekstfiler** som gir informasjon om alle elementene i samme mappe (eller om arbeidsflyt eller prosesser hvor disse elementene har blitt brukt)
- **Den første filen som bør åpnes**
- **Gir et kart** som kan brukes til å utforske filene
- Opprett **en** README.txt fil **per mappe**
- Innholdet bør **støtte andre i å orientere seg** i mappen
- En README.txt-fil på **prosjektnivå** bør gi **generell prosjektinformasjon** og en svært grov oversikt over innhold i filene og deres plasseringer
- En README.txt-fil på **data- (eller objekt-)nivå** ville være **mer spesifikk** med hensyn til hva hver fil inneholder

Dokumentasjon vs. metadata?

"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**."

Oppgave 5: Dokumentasjon

Shut Up and Write

Skriv om dataenes opprinnelse i 2 min.

Skriv da om datalagring med fokus på klasse,
format, størrelse, osv. i 2 min

4 min



Beskyttelse
Organisering
Dokumentasjon (og metadata)
Klassifisering
Korttidslagring
Lisensiering
Deling
Langtidslagring
Formidling (publisering)
Gjenbruk

Kurs i forsknings- datahåndtering og deling

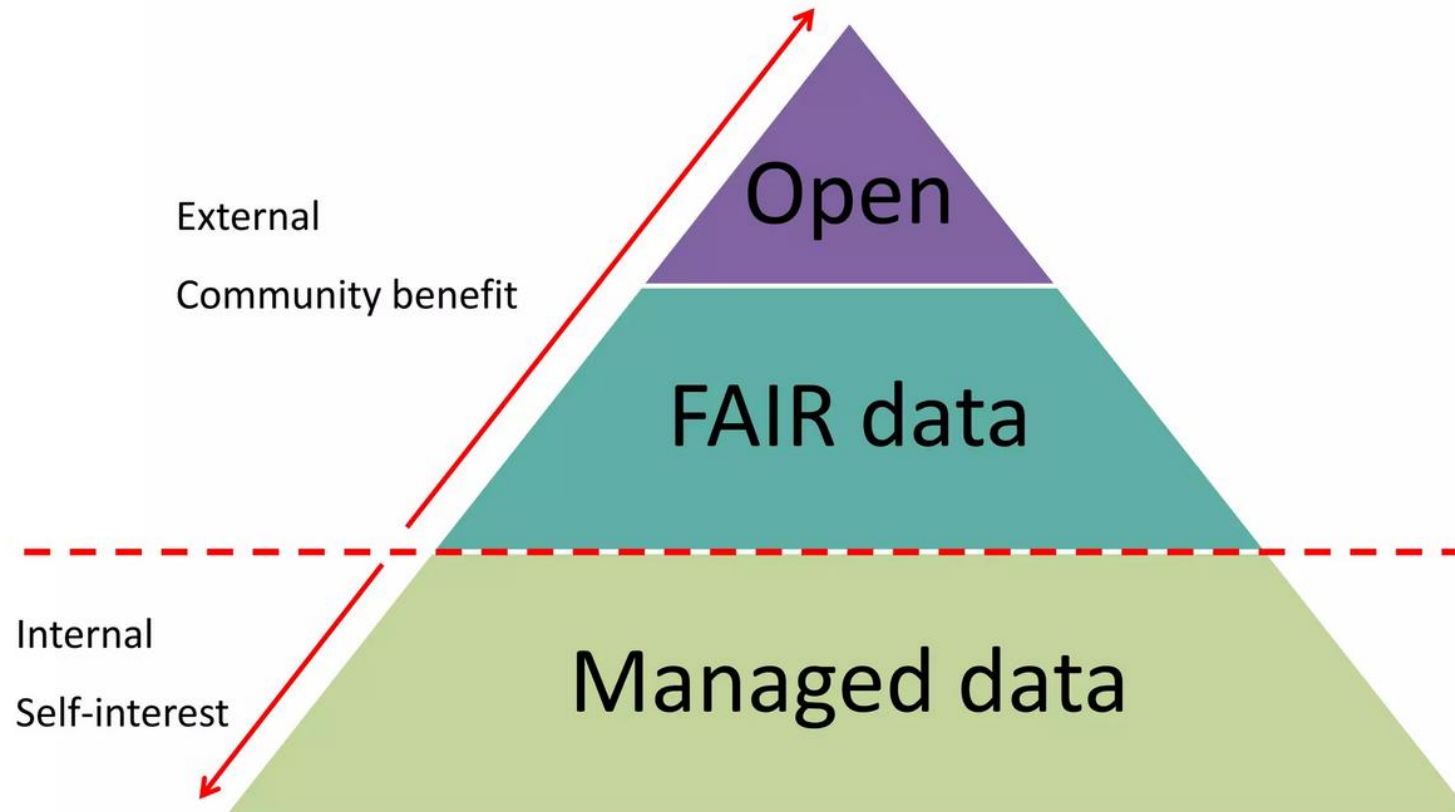
Høst 2024

Time and place: Nov. 14, 2024 9:00 AM – 12:00 PM, UiO, Blindern, Georg Sverdrups Hus, Senter for digital forskersøtte (DSC)

Del, arkiver og gjenbruk forskningsdata

I dette halvdagskurset i forskningsdatahåndtering modul 2, vil deltakerne få kjennskap til potensialet ved deling av forskningsdata i tråd med FAIR prinsippene om gjenfinnbare, tilgjengelige, interoperable og gjenbrukbare data.

Så **åpent** som mulig,
så **lukket** som nødvendig



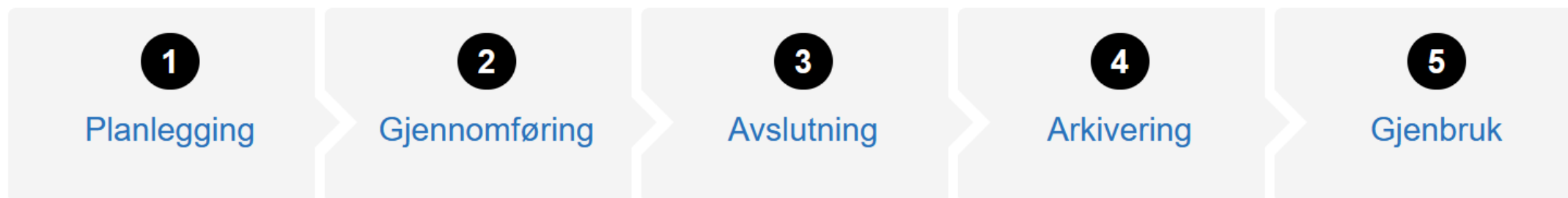
Trenger du hjelp
med
forskningsdata?

Send en epost til
research-data@uio.no

Flere ressursar?

Håndtering av forskningsdata inkluderer dokumentasjon, organisering, lisensiering, deling, arkivering og gjenbruk.

Forskningsdata gjennom forskningsprosjektet



Tema

- Datahåndteringsplan
- Dokumentasjon av forskningsdata
- Forskningsdata (definisjon)
- Lagring av forskningsdata
- Metadata
- Organisering av forskningsdata
- Sitering av forskningsdata
- Åpen deling eller begrensninger?

Prinsipper

- CARE-prinsippene
- FAIR-prinsippene

Senter for digital forskerstøtte (DSC)

<https://www.ub.uio.no/bibliotekene/dsc/index.html>

Senter for digital forskerstøtte (DSC) er et ressurs- og støttesenter for forskere. Vi tilbyr veiledning, kurs og workshops i bruk av digitale metoder og verktøy i forskning og formidling.

Vi kan hjelpe deg med:

Åpen forskning	+
Forskningsdata	+
Vitenskapelig visualisering	+
Digitale verktøy	+

Åpen og reproduserbar forskning

<https://www.ub.uio.no/bibliotekene/dsc/apen-repro-forskning/index.html>

Lær om hvordan du gjør forskningen din mer åpen og reproduserbar og engasjer deg i initiativer og miljøer som fokuserer på å dele og forbedre forskningen ved UiO.

Flere forskere og studenter på tvers av disipliner implementerer åpen forskningspraksis, forhåndsregistrerer hypoteser, metoder og analyseplaner og deler forskningsmateriell, data og analyseskript. Senter for digital forskerstøtte kan hjelpe deg med å lære om og implementere denne praksisen i din egen forskning, samt gi råd om retningslinjer og krav fra finansiører.

Open Science Lunch



Hver siste torsdag i måneden møtes vi kl 12:00 for å diskutere temaer knyttet til åpen forskning..

ReproducibiliTea@UiO



Bli med oss i en Journal Club hvor vi leser og diskuterer artikler om forbedring av vitenskap, reproduserbarhet og åpen forskning.

Norwegian Reproducibility Network



Bli med i et bredere fellesskap som har som mål å fremme og muliggjøre streng, robust og transparent forskningspraksis i Norge.

Kurs og workshops



Klikk her for listen over kommende og tidligere kurs og workshops om åpen og reproduserbar forskning ved UiO.



Open Science Lunch

Each last Thursday of the month at 12:00 we invite you to join us for a lunch seminar to hear about how to make your research more open. We will discuss research transparency and visibility, open publishing, data sharing, and more!

Time and place: Dec. 5, 2024 12:00 PM – 1:00 PM, [Zoom](#)

Who owns data about your research?

Join us for this Open Science Lunch to get a closer look at the Barcelona Declaration and OpenAlex and to learn more about ongoing initiatives to promote open research information.

ReproducibiliTea

Journal Club

**JOIN IN AND DISCUSS WITH FELLOW
STUDENTS AND RESEARCHERS**

**OPEN RESEARCH, REPRODUCIBILITY
and RESEARCH IMPROVEMENT**



Join us

Everyone is welcome to join us - whether you are an enthusiast of open and reproducible research, a skeptic, or a cautious explorer. Currently, all meetings are hybrid with the possibility of joining on-site at Blindern or via Zoom. Grab a cup of tea (coffee?) and join us!

Subscribe to our mailing list





Carpentry@UiO

Carpentry@UiO is a community of people who are passionate about learning, teaching, and sharing best practices and digital skills for making the research process more reproducible and effective. If you want to get involved, or join one of our workshops, check us out!



The Unix Shell

Shell speeds up repetitive and tedious processes. It is also essential skills needed to use high-performance computing (HPC) resources.



Version Control with Git

Git helps you to keep track of what you've done, for a better collaboration and for yourself in future. In the workshop we use GitHub as well.



Programming in Python

Python is now widely used in scientific computing with various powerful packages. Carpentry@UiO runs workshops for participants with no programming experience ("Plotting and Programming in Python" lesson) and for participants at intermediate level ("Programming with Python" lesson, episodes 10-12).



R for Reproducible Scientific Analysis

R is commonly used for statistical analysis, but it is also a powerful programming language. Workshops on R focuses on teaching best practices for scientific computing: breaking down analyses into modular units, task automation, and encapsulation. Workshops on R may use lessons from Data Carpentry instead.



Using Databases and SQL

Databases include powerful tools for search and analysis, and can handle large, complex data sets. The lesson will show how to use a database to explore research data by using SQL.



Carpentry@UiO

Carpentry@UiO is a community of people who are passionate about learning, teaching, and sharing best practices and digital skills for making the research process more reproducible and effective. If you want to get involved, or join one of our workshops, check us out!



Learn, teach, and share digital skills and best practices

Be a part of an interdisciplinary community

Make use of and contribute to community-built teaching materials

Det senteret for digitalforskerstøttes nyhetsbrev,
en del av Universitetsbiblioteket i Oslo

The Digital Scholarship Centre's Newsletter,
part of the University of Oslo Library

DSC NEWS

Senter for digitalforskerstøtte
Digital Scholarship Centre



<https://sympa.uio.no/ub.uio.no/subscribe/dsc-news/subscribe>

Takk!

Matthew Good, Agata Bochynska

Senter for digital forskerstøtte, Universitetsbibliotek, Universitetet i Oslo

13.11.2024

Spørsmål? Kontakt oss på research-data@uio.no

Gi oss din tilbakemelding!



<https://nettskjema.no/a/468976>