



# Introduksjon til kurset, FAIR-prinsippa og datakuratering

RDA-kurs i FAIR forskingsdatakuratering

UiT, 20.-21. januar 2020

Philipp Conzett, UiT

# Kva er RDA?

- Johanne har allereie sagt det viktigaste om RDA og den norske RDA-node.
- Eit nasjonalt konsortium har fått pengar frå RDA Europa/EU til å **etablere den norske RDA-noden**.
- **Prosjektgruppa** skal til sommaren koma med eit **forslag** til kva den norske RDA-noden skal vera, og korleis den skal organiserast.
- Prosjektgruppa består av: Katrine Utaaker Segadal (NSD, koordinator), Trond Kvamme (NSD, koordinator), Margaret Fotland (UiO), Philipp Conzett (UiT), Maria Francesca Iozzi (Uninett Sigma2), Anne Fjellbirkeland (UiB), Ingrid Heggland (NTNU), Jan Erik Garshol (Unit).
- Fleire av dei er her i dag og i morgon. Snakk med oss viss du ønskjer å vita meir om RDA og den norske RDA-noden.

# Kva skal den norske RDA-noden gjera?

Nokre hovudoppgåver:

- Formidla kunnskap om god handtering og deling av forskingsdata, t.d. gjennom nasjonale kurs som dette
- Få forskarar og støttepersonale til å engasjera seg i dette arbeidet

Kva kan du gjera?

- Meld deg inn i RDA! Ope og gratis medlemskap for enkeltpersonar.
- Bli med i den [norske RDA-noden](#), og hald deg oppdatert på RDA-aktivitetar i Noreg!
- Sprei informasjon om RDA på din institusjon!

# Kva er målet med dette kurset?

## Kurs i **FAIR** forskingsdatahandtering

Kurset skal gje deltakarane opplæring i **fire sentrale område** innanfor handtering og kuratering av forskingsdata:

- Bolk 1: **dokumentasjon og metadata**
- Bolk 2: **filorganisering og filformat**
- Bolk 3: **opphavsrett og eigarskap**
- Bolk 4: **lisensar**

Avgrensingar: Lagt mest vekt på data som kan gjerast **ope tilgjengelege** (handtering av personidentifiserande og sensitive data må venta til eit anna kurs).

# Kven er målgruppa for dette kurset?

Kurset er retta mot personar som jobbar med **forskningsstøtte** på norske forskingsinstitusjonar.

Deltakarane er ei god blanding:

- Frå **16 institusjonar**
- Variasjon i kor mykje **erfaring** med handtering og kuratering av forskingsdata deltakarane har

Kor godt kjend er du med FAIR-prinsippa?



■ Godt kjend ■ Noko kjend  
■ Lite kjend ■ Ikkje kjend

Kor mykje erfaring med kuratering har du?



■ Mykje erfaring ■ Noka erfaring  
■ Lita erfaring ■ Inga erfaring

# Nokre viktige omgrep

Sjølv om dei fleste er kjende med **FAIR-prinsippa** og har noka er litt erfaring med **kuratering**, kan det vera greitt å **friska opp** begge **omgrepa** før vi gå laus på den første temabolken.

# Kva er FAIR?

- Eit sett med generelle prinsipp for god handtering og tilgjengeleggjering av forskingsdata
- Data som er FAIR,
  - kan gjenfinnast,
  - er tilgjengelege,
  - er interoperable, og
  - kan gjenbrukast.

På engelsk:

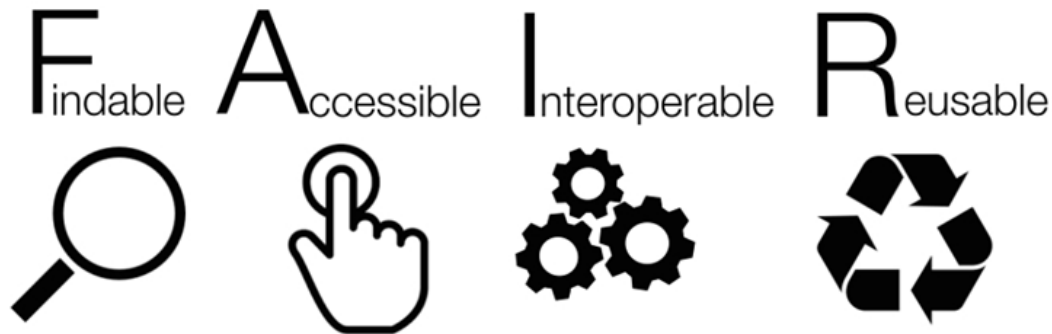
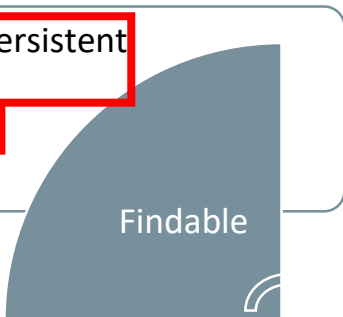


Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0

# Findable

- Publisert med persistent identifikator
- Gode metadata
- Indeksert



Metadata = beskriving av data

Døme på metadata:

- Nøkkelord
- Geografisk informasjon

Keyword ?

Trolldomsprosess  
Finnmark  
1600-tallet  
Tidlig nytid  
Rettshistorie  
Vardø  
Tingbok

Geospatial Metadata ^

Geographic Coverage ? Norway Finnmark

Hagen, Rune Blix, 2019, "Rettsforfulgte trollfolk i Finnmark, 1593-1692", <https://doi.org/10.18710/OWP5IP>, DataverseNO, V1

DOI = Digital Object Identifier = ein type persistent identifikator ~ varig lenkje/URL

**DataCite**  
FIND, ACCESS, AND REUSE DATA

**Google Dataset Search** Beta

**BASE**  
Bielefeld Academic Search Engine

**UIT Norges arktiske universitet**

NEW SEARCH DATABASES

My Library Search Oria

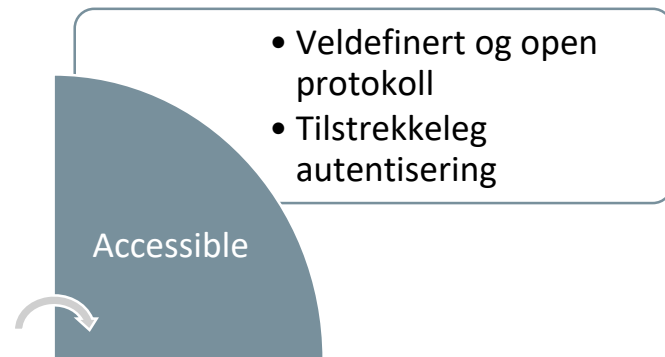


# Accessible

*Accessible* har mest med tekniske aspekt ved dataarkiv å gjera. Men:

- Når ein publiserer dataa sine, bør ein velja eit arkiv som gjer dei tilgjengelege gjennom eit **veldefinert og open protokoll** (t.d. https).
- Ein bør også velja eit arkiv som gjer dataa tilgjengelege i tråd med innhaldet. Døme:

Visse typar data kan ikkje gjerast ope tilgjengelege (t.d. ikkje-anonymiserte persondata), men dei kan kanskje likevel delast i eit arkiv der dei som ønskjer å lasta ned data, må registrera seg og logga inn. Då treng ein **tilstrekkeleg autentisering**.



# Interoperable

- Opne metadata-format
- Felles standardar
- Konsistente vokabular

- Bruk felles metadatastandardar. Det gjeld både
  - generelle metadata, t.d. internasjonalt datoformat (t.d. ISO-8601): ÅÅÅÅ-MM-DD (2019-12-09), og
  - fagspesifikke metadata, t.d. Data Documentation Initiative (DDI) = internasjonal standard for beskriving av data brukte i spørjeskjema og andre observasjonsmetodar i samfunnsfag og helsefag.
- Bruk konsistente metadatavokabular, t.d. DDI-vokabularet for aggregeringsmetode (Aggregation Method); utdrag:
- Interoperabilitet mogleggjer søk og gjenbruk på tvers av datasett og arkiv.

<b>Value of the Code</b>	<b>Descriptive Term of the Code</b>	<b>Definition of the Code</b>
<b>Maximum</b>	Maximum	The highest value attained or recorded.
<b>Minimum</b>	Minimum	The lowest value attained or recorded.

# Re-usable

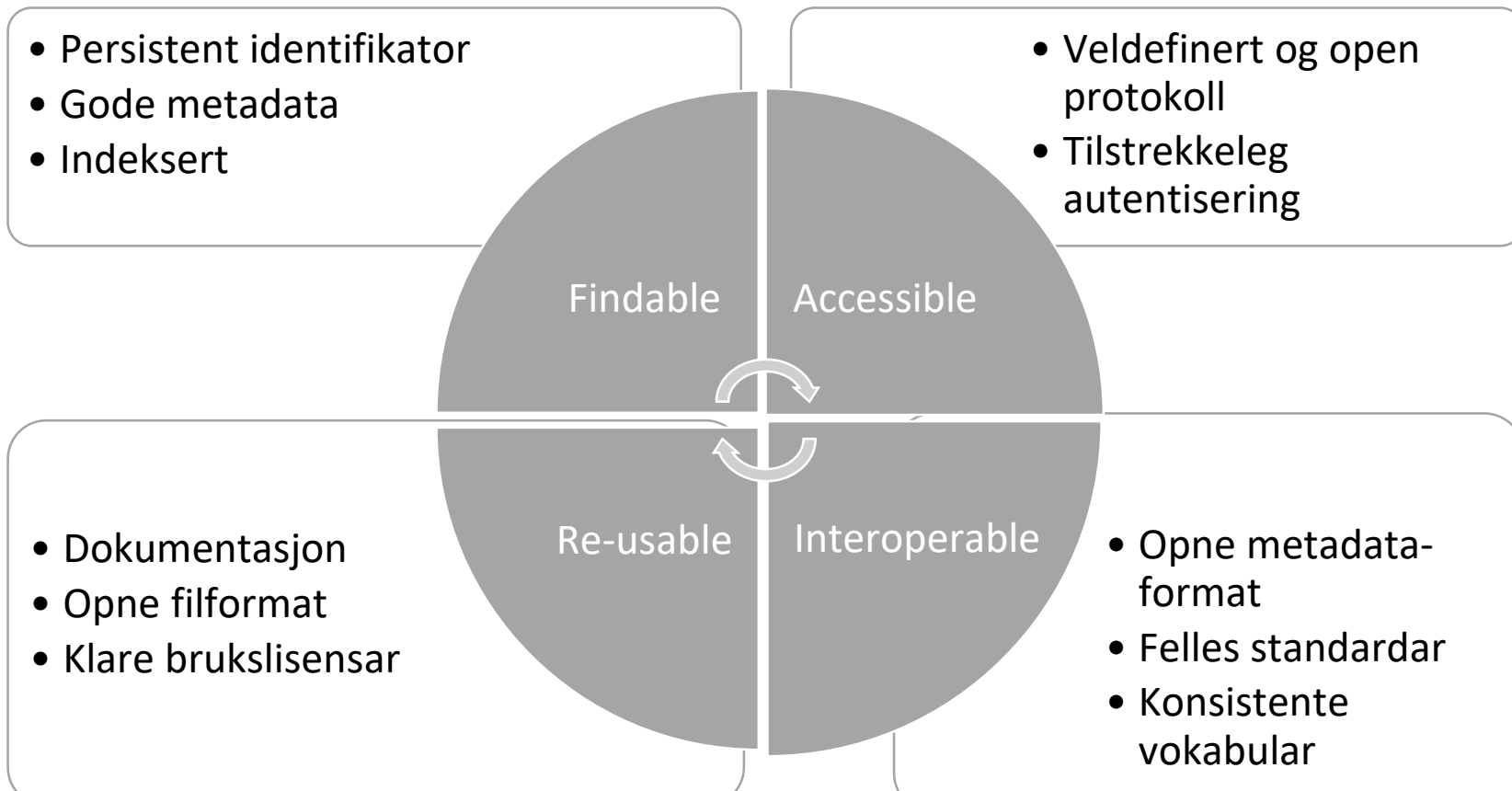
- Dokumenter data, slik at dei er forståelege og kan gjenbrukast av fagellar.
- Arkiver data i føretrekte/arkivverdige filformat slik at filene kan opnast og lesast på lang sikt, t.d. rein tekst (.txt) i tillegg til Excel (.xlsx).
- Definer ein klar brukslisens for dataa dine slik at dei som ønskjer å bruka dei, veit kva dei har lov til å gjera med dei. Døme: Creative Commons (CC)-lisensar



Re-usable

- Dokumentasjon
- Opne filformat
- Klare brukslisensar

# FAIR-prinsippa oppsummert



# Om kuratering

Kva?

- Sørgja for at data er lagde til rette i tråd med FAIR-prinsippa

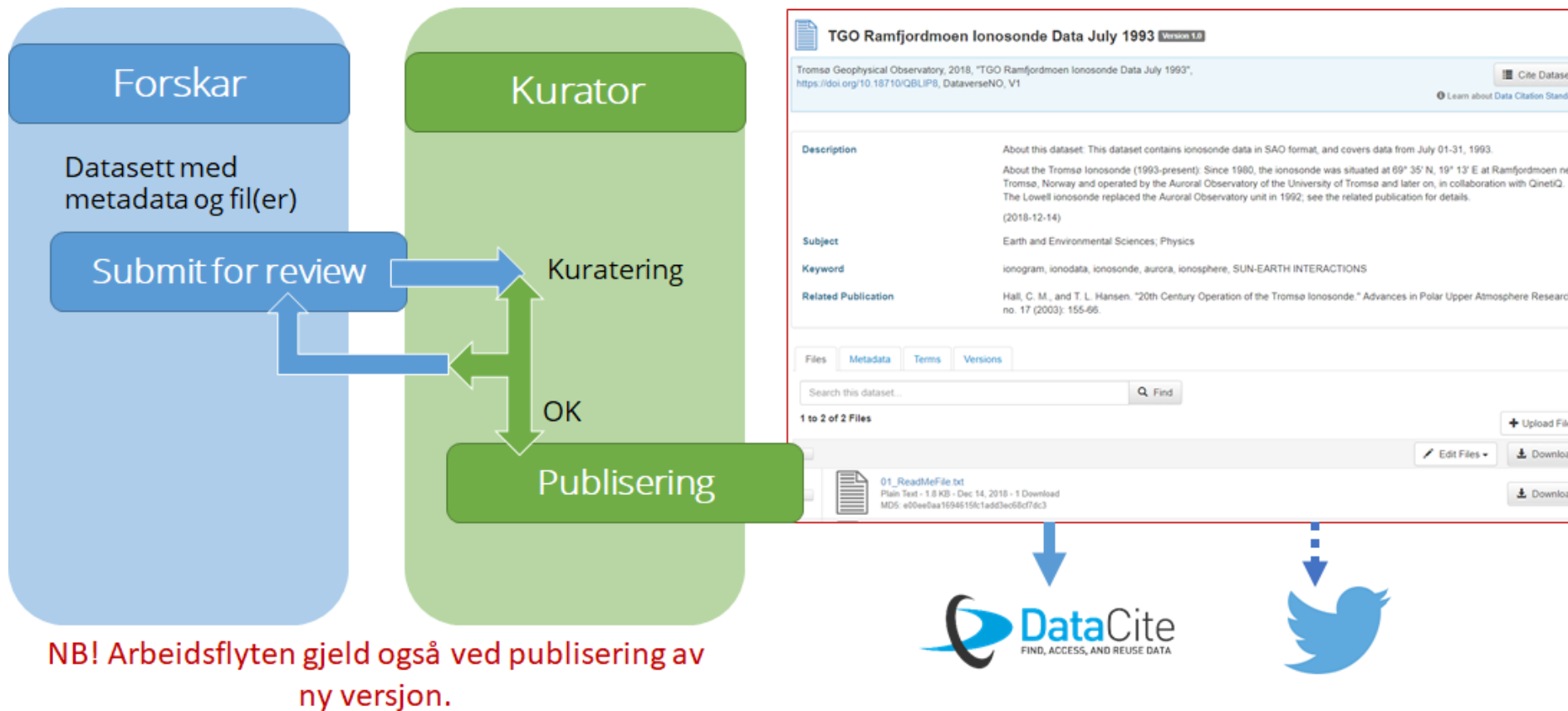
Når?

- I heile livsløpet til forskingsdata = god forskingsdatahandtering
- Før arkivering = tilrettelegging for arkivering, ofte meir arbeidskrevjande

Kven?

- Forskaren sjølv (i framtida?)
- Meir spesialisert på handtering av forskingsdata: kurator, datarøktar, ...

# Eit døme: Kuratering i DataverseNO



# Litt meir informasjon om øvingsoppgåvene

- Gruppearbeid med øvingsoppgåver etter kvar temabolk (bortsett frå denne introduksjonen; felles oppgåvebolk etter temabolk 3 og 4 i morgon)
- 5 grupper
  - Kvar gruppe har ein **gruppeleiar/sekretær**
  - Gruppene er fordelt på **grupperom her på UB**; sjå kartlenkjer
- I den delte Google-mappa (**[tinyurl.com/rdano2020](https://tinyurl.com/rdano2020)**) finn dokker:
  - **Gruppeinndeling** med lenkjer til kart over romma (jf. kursannonsen)
  - **Øvingsoppgåvene**
  - **Oppskrift** på korleis dokker kan **logga inn på testbasen** til DataverseNO der vi har lagt ut 4 **øvingsdatasett** som oppgåvene til dels tek utgangspunkt i.
  - Eitt delt dokument per gruppe i **mappa “Svar”** der dokker skal skriva inn svara og ev. spørsmåla dokkar.

# Neste post på programmet:

Bolk 1: Dokumentasjon og metadata – del 1 v/Iris Alfredsson (SND)