

National strategi for data management baseret på FAIR principper

Indholdsfortegnelse

Indholdsfortegnelse	1
Indledning	2
Formål og vision	2
Baggrund	3
Gyldighedsområde	4
Værdien af åbne og FAIR data	5
Målgruppe og struktur	5
Principper for en data management-praksis som understøtter FAIR	6
Data management skal understøtte FAIR-principperne og mulighederne for deling og genbrug af data.	6
Relevansen og vægtningen af de enkelte FAIR principper varierer mellem forskningsområder og skal defineres inden for disse	6
Det skal være muligt at bevare alle typer af forskningsobjekter og filformater som er relevante for forskningen, på både kort og langt sigt	6
Data management er en kontinuerlig proces i datas hele livscyklus	7
Relevant teknisk infrastruktur og værktøjer skal tilbydes alle forskere, uafhængigt af fag- og institutionstilhørsforhold	7
Den fornødne hjælp og de fornødne vidensressourcer skal være tilgængelige uanset fag- og institutionstilhørsforhold	7
Indsatser til etablering af en data management-praksis som understøtter FAIR	7
Tilgængeliggørelse og FAIR data er en kulturforandring inden for mange fagområder	8
Forslag til konkrete indsatser	8
Understøttelse af faggrupper i at definere egen implementering af FAIR-principper	9
Forslag til konkrete indsatser	9
Understøtte bevaring af forskningsdata på både kort og langt sigt	10
Forslag til konkrete indsatser	11
Understøtte at forskningen kan gennemføre løbende dokumentation i projektførelsen gennem hele livscyklus	11
Forslag til konkrete indsatser	11
Etablere teknisk infrastruktur, der understøtter data management og tilgængeliggørelse af FAIR data	12
Forslag til konkrete indsatser	12

Etablere vidensinfrastruktur som leverer nødvendig hjælp og vidensressourcer til data management og FAIR	13
Forslag til konkrete indsatser	14
Finansiering	14
Principper for fordeling af udgifter til data management og fysiske datalagre i forbindelse med fondsfinansierede forskningsprojekter	14
Forskningsfinansiering	14
Institutionernes forpligtigelser	14
Finansiering og ansvar for etablering af den basale eInfrastruktur	15
Udgifter til Data Management og fysiske datalagre	15
Principper for etablering af omkostningsniveau og placering i de to kategorier	17
Forslag til konkrete indsatser	17
De forskningsfinansierende fondes krav til håndtering af forskningsdata	17
Governance og implementering	17
Ordforklaring	18

Indledning

Formål og vision

Formålet med strategien er at etablere basis for beslutninger om implementering og finansiering af en data management-praksis som er baseret på FAIR-principperne¹ (*Findable, Accessible, Interoperable, Reusable*, dvs at data skal kunne findes, tilgås og genbruges, og at betingelserne for brug af data er klart formuleret). Strategien skal være i overensstemmelse med EU's Åbne data-direktiv² og den danske implementering af dette i PSI-loven³, samt bidrage til at opfylde forskningens og samfundets forventning til stigende tilgængelighed af forskningsresultater.

Strategien skal bidrage til at der tages væsentlige skridt til, at forskningsdata bliver mere FAIR, og at forskningsdata i sig selv behandles og anerkendes som forskningsresultater der deles i nationale og globale datainfrastrukturer på måder, som stimulerer åben forskning og åben innovation - under hensyntagen til nødvendige sikkerhedsforanstaltninger omkring såvel persondata som omkring rettigheder, kommercielle interesser og nationale sikkerhedsinteresser.

¹ Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016)

² Europa-parlamentets og Rådets direktiv (EU) 2019/1024 om åbne data og videreanvendelse af den offentlige sektors informationer (2019)

³ Den gældende *Lov om videreanvendelse af den offentlige sektors informationer* er under revision for at implementere den seneste udgave af direktivet.

Overordnet er det visionen med data management i forskningen at forskere løbende foretager kvalificerede beslutninger og handlinger i forhold til vedligeholdelse og adgang til deres data (og evt. andre resultater som software, protokoller o. lign), under hensyntagen til forskningens fremme, samfundets behov og gældende lovgivning.

Løsninger skal være internationalt forankret og leve op til internationale standarder. Hvor anerkendte internationale løsninger allerede findes, fx. open source software, allerede findes, bør sådanne foretrækkes til implementering frem for egenudvikling. Dette gælder for implementering af infrastruktur og tekniske værktøjer, såvel som for valg af metadatastandarder, vokabularer mm.

Baggrund

Samfundet og forskningen benytter i stadig stigende grad data som basis for beslutninger og analyser, og det er derfor vigtigt, at der skabes muligheder for at finde, tilgå og genbruge data, herunder at betingelserne for at genbruge data bliver klart formuleret. Planen om at udarbejde en ny national data management-strategi, som skal understøtte, at flere forskningsdata bliver FAIR, har sit udspring i Strategi for nationalt samarbejde om digital forskningsinfrastruktur⁴, og dermed i en ide om, at data management efter FAIR-principperne har sin plads i det nationale e-infrastrukturlandskab.

Strategien tjener i den forbindelse også som den danske politik for tilgængelighed iht. EU's Åbne data-direktiv (tidligere PSI-direktivet), som i sin nyeste udgave også omfatter forskningsdata og kræver at medlemslandene opstiller politikker for tilgængeliggørelse af forskningsdata efter FAIR principperne.

Det er målet, at data gøres så åbne som muligt i overensstemmelse med FAIR-principperne, sådan at data så vidt muligt kan tilgås og genbruges af andre. Det skal dog understreges, at FAIR data ikke er det samme som åbne data. Der kan være gode grunde til at data ikke kan gøres åbent tilgængelig. Med ganske få undtagelser kan data, der ikke kan gives åben adgang til, dog alligevel gøres FAIR, idet forskeren kan publicere metadata der beskriver selve eksistensen af datasættet, samt betingelser for, hvordan andre eventuelt kan få adgang til dette, enten datasættet i sin helhed, eller afledte versioner heraf.

Danmark er medlem af European Open Science Cloud (EOSC)⁵, og det er forventningen at nyproducerede forskningsdata gøres tilgængelige gennem EOSC. I skrivende stund er en infrastruktur til dette endnu ikke på plads, og en arkitektur for EOSC er fortsat under udvikling. Den bedste måde at opfylde dette mål på er derfor at følge anbefalingerne fra EOSC⁶ og

⁴ Strategi for nationalt samarbejde om digital forskningsinfrastruktur, Styrelsen for Forskning og Uddannelse (2018)

⁵ <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

⁶ Rapporten *Turning FAIR into reality*, EU Kommissionen, doi: 10.2777/1524 (2018), såvel som politikker og rapporter under publicering fra den igangværende EOSC FAIR Working Group

førende internationale organisationer som Research Data Alliance (RDA)⁷ og GO FAIR initiativet⁸ omkring implementering af FAIR principperne. Samtidig skal det sikres at infrastruktur og tjenester tager udgangspunkt i kendte standarder og løbende tilpasses eventuelle fremtidige krav til arkitektur og interoperabilitet i EOSC.

Strategien skal endvidere komme med anbefalinger til, hvorledes håndteringen af forskningsdata efter de opstillede principper skal finansieres samt opstille forslag til politikker for FAIR data management for projekter finansieret af de forskningsfinansierende fonde. Man kan samlet sige, at målet med strategien er at opstille principper og de tilhørende indsatser for en FAIR data managementpraksis og dennes finansiering samt mulige indsatser for dens realisering.

Arbejdet med data management i henhold til denne strategi kan berøre en række problemstillinger, som er reguleret af anden lovgivning og lokale politikker. Herunder kan nævnes persondataloven (GDPR), Arkivloven, sikkerhedsforanstaltninger m.v. Specielt bemærkes det i bemærkningerne til udkastet til revideret PSI-lov at EU's databeskyttelsesforordning og den danske databeskyttelseslov altid vil have forrang over bestemmelserne i åbne data-direktivet og PSI-loven.

Gyldighedsområde

Nærværende strategi gælder, i overensstemmelse med Åbne data-direktivets artikel 10 og forslaget til den danske PSI-lov, for offentligt finansierede forskningsdata, forstået som forskning finansieret af det offentlige og forskning udført af offentlige institutioner, hvor der kan være privat finansiering. Strategien retter sig, i overensstemmelse med dette juridiske grundlag, mod forskningsinstitutioner og forskningsfinansierende organisationer. Direktivet understreger specielt, at forskningsdata som af forskere, forskningsinstitutioner eller forskningsfinansierende fonde er blevet tilgængeliggjort gennem et institutionelt eller emnebaseret repository, som grundregel skal kunne videreanvendes til både kommercielle og ikke-kommercielle formål⁹.

Forskningsdata defineres i Åbne data-direktivet¹⁰ som *dokumenter i digital form, dog ikke videnskabelige publikationer, som er indsamlet eller produceret som led i videnskabelige forskningsaktiviteter og anvendes som evidens i forskningsprocessen, eller som generelt er accepteret i forskerkredse som nødvendige til validering af forskningskonklusioner og -resultater.*

I overensstemmelse med denne definition opfattes data i nærværende strategi primært i traditionel forstand som det digitale indhold, man kan referere til i form af datasæt, filer og databaser af forskellig art. Det skal dog understreges at metoderne omkring FAIR med fordel kan bruges til alle typer af digitale objekter som indgår i forskningen, det være sig logbøger,

⁷ <https://www.rd-alliance.org/>

⁸ <https://www.go-fair.org/>

⁹ Åbne data-direktivet, artikel 10 stk. 2.

¹⁰ Åbne data-direktivet artikel 2, stykke 9

software, workflows og scripts, digitale surrogater for fysiske objekter, organisationer, begreber og meget andet. Strategien omfatter desuden publikation af data der er offentliggjort i deres egen ret, uden nødvendigvis at være basis for en artikel.

Strategien erstatter eller tilsidesætter ikke anden gældende lovgivning på området, fx anmeldelsespligten ifølge Arkivloven.

Værdien af åbne og FAIR data

Det er et bærende princip i videnskaberne at forskere bygger videre på hinandens resultater. I en verden hvor forskning i højere og højere grad bliver digital, er det nødvendigt at data gøres nemt tilgængelige, så det bliver muligt at inddrage andres data i egen forskning. Et større datagrundlag giver nye muligheder for analyser og modeller, fx klimamodeller.

FAIR data vil gøre forskningsresultater mere tilgængelige for samfundet generelt, lette virksomheders forsknings- og udviklingsarbejde og have betydning for borgeres udøvelse af citizen science aktiviteter. Samtidig vil det bidrage til transparens og dermed legitimitet af forskningen.

Forudsætningen for at disse visioner kan realiseres og komme forskningen til gavn, er bl.a. at data kan findes og identificeres, og at betingelserne for genbrug er kendt. Dette er nogle af formålene med FAIR principperne, som fuldt udfoldet sigter på at muliggøre genbrug af data på en forsvarlig måde, til størst mulig gavn for forskningen og for samfundet i bredere forstand.

Målgruppe og struktur

Dokumentet opstiller en række principper for data management-praksis som understøtter FAIR, efterfulgt af anbefalinger til indsatser, der kan styrke gennemførelsen af de enkelte principper. Disse indsatser er forsøgt struktureret efter hvem, der er de primære aktører for iværksættelsen. De anbefalede indsatser understøtter ofte mere end ét princip.

Principperne retter sig bredt mod alle interessenter i en sådan praksis: forskere, data managers/data stewards, forskningsledere og forskningsfinansierende fonde. Forslagene til indsatsområder retter sig derimod primært mod beslutningstagere på forskningsudførende og forskningsfinansierende institutioner, som har kapacitet og midler til at iværksætte og finansiere tiltagene. Det vil sige personer, som kan påtage sig ansvar for strategiens implementering, herunder for hvordan denne lægges til rette, og hvordan ansvaret fordeles i de enkelte organisationer.

Idet der i kommissoriet imødeses en egentlig implementeringsplan omfatter strategien ikke forslag til konkret implementering, men derimod alene forslag til *indsatsområder*.

Principper for en data management-praksis som understøtter FAIR

- Data management skal understøtte FAIR-principperne og mulighederne for deling og genbrug af data.
 - Data som tilgængeliggøres, skal gøres FAIR ved som minimum at blive identificeret entydigt ved datareference (*Persistent Identifier*, PID) og tilføjes dokumentation (metadata).
 - Implementering af FAIR er en trinvis proces, ikke alle principperne skal opfyldes fra starten.
 - Når data vurderes at have blivende værdi og bevares, skal data gøres åbent tilgængelige, hvis ikke specifikke grunde taler imod det
 - Beskrivelser af data (metadata) kan som hovedregel gøres åbne, også i tilfælde, hvor data ikke er åbent tilgængelige.
- Relevansen og vægtningen af de enkelte FAIR principper varierer mellem forskningsområder og skal defineres inden for disse
 - Implementeringen af FAIR ud over de minimale krav som beskrevet ovenfor, defineres af de enkelte forskningsområder på nationalt plan, i relation til internationale standarder.
 - Forskningsområderne skal forholde sig til hvor langt de vil gå med implementeringen, og hvilke FAIR principper de vægter højest.
- Det skal være muligt at bevare alle typer af forskningsobjekter og filformater som er relevante for forskningen, på både kort og langt sigt
 - Data som skønnes at kunne have værdi i fremtiden, skal lagres og tilgængeliggøres i en teknisk og organisatorisk sikker datainfrastruktur, i overensstemmelse med FAIR principperne
 - Særligt værdifulde data til fortsat bevaring i en længere årrække, identificeres og håndteres løbende med henblik på eventuelle særlige skridt til langtidsbevaring.
 - I tilfælde hvor tilgængeliggjorte data slettes, bør PID-en fortsat være tilgængelig og kunne føre til metadata om de slettede data samt oplysning om at de nu er utilgængelige.

- **Data management er en kontinuerlig proces i datas hele livscyklus**
 - Metoder og værktøjer bør være tilgængelige og kunne sikre, at data og dokumentation bliver indsamlet og lagret undervejs.
 - Betingelserne (licensering) for videreanvendelse skal være tilgængelige som del af metadata, i enten menneske- eller maskinlæsbar form.

- **Relevant teknisk infrastruktur og værktøjer skal tilbydes alle forskere, uafhængigt af fag- og institutionstilhørsforhold**
 - Datalagring og tilgængeliggørelse efter FAIR principperne skal stå til rådighed for alle forskere
 - Data i relevante infrastrukturer skal kunne tjene som opfyldelse af forskningsinstitutionernes forpligtelser i forhold til Rigsarkivet.
 - Valgte tekniske værktøjer skal muliggøre at data kan trækkes ud og gøres FAIR på åben, standardiseret form. Brugen af data må ikke hindres ved at de kun kan tilgås med licensbelagt software.

- **Den fornødne hjælp og de fornødne vidensressourcer skal være tilgængelige uanset fag- og institutionstilhørsforhold**
 - Det skal være muligt for forskere at arbejde med at gøre data FAIR, uden at de behøver at være eksperter i alle dele af data management, ved at de tilbydes hjælp fra institutionerne.
 - Den nødvendige viden om data management og FAIR skal være tilgængelig for forskere på alle forskningsinstitutioner.

Indsatser til etablering af en data management-praksis som understøtter FAIR

Tilgængeliggørelse og FAIR data er en kulturforandring inden for mange fagområder

En forskningsverden, hvor forskere ikke kun er åbne om deres metoder og data, men også stiller disse til rådighed på måder så data kan genbruges, vil være med til at udvide opfattelsen af, hvad der er nyttigt og brugbart output i forhold til forskning. Dette har implikationer ikke bare for forskernes arbejdsmetoder, men også for en række af de omkringliggende strukturer, herunder:

- for meriteringssystemet, som skal anerkende disse alternative outputs
- for bevillingssystemet, som skal anerkende at det at gøre metadata og data genbrugbare, er en yderligere omkostning
- for rekrutteringssystemet, som skal anerkende behovet for nye kompetencer.

Det vil kræve en indsats af forskningsinstitutionerne, som skal medvirke til en kulturændring i flere fagområder i forhold til en digital forskningspraksis og arbejde aktivt i forhold til de metrikker, som bruges ved evalueringer og ansættelser.

Det vil være nødvendigt, at forskningsråd og fonde inkorporerer ikke bare støtte til denne form for aktivitet, men også anerkender disse som meritgivende aktiviteter.

Forskningsinstitutioner bør stimulere og facilitere denne kulturændring, særligt i fagområder, som i dag ikke har en tradition for at dele data uden for en sluttet gruppe, herunder

- At den forskningsstrategiske betydning af god data management anerkendes. Forskning for et stadig stigende antal videnskaber er i dag helt beroende på at data indsamles, systematiseres, kategoriseres, analyseres (fx af computere), gemmes (fx i repositories eller databaser) og gen-tilgås løbende af internationale forskningskollegaer, eller deres maskiner.
- Ændre holdningen til, hvad der er publicerbart og delbart,
- at facilitere en forståelse af at FAIR-principperne ikke nødvendigvis er ensbetydende med åben adgang til data, men at også lukkede data kan gøres FAIR ved, at de bliver tilknyttet en identifikator (PID) samt beskrivende metadata.

Forskningsinstitutioner bør arbejde nationalt og internationalt med meriteringssystemet, således at forskeres arbejde med at publicere, vedligeholde og understøtte deling af FAIR data anerkendes og meriteres.

Forslag til konkrete indsatser

Styrkelse af kommunikation og dialog - omkring Open Science og herunder FAIR.

Arbejde med at gøre data FAIR gøres meritgivende.

Understøttelse af faggrupper i at definere egen implementering af FAIR-principper

De enkelte fagområder har forskellige traditioner for, hvordan data deles og struktureres. Desuden ser forskellige fagområder meget forskelligt på betydningen og den praktiske implementering af de enkelte FAIR-principper. Det er derfor vigtigt, at det er de faglige miljøer som definerer, hvordan og hvor langt FAIR kan implementeres inden for deres område.

Betydningen og strukturen af data kan variere meget både inden for og mellem fagområder. I nogle tilfælde handler det om personoplysninger, i andre om at data kan være skabt i et privat-offentligt samarbejde, som kan være underlagt fortrolighedsaftaler, eller data kan være copyrightbeskyttede for at nævne eksempler. De fleste forskningsdata vil dog uden store juridiske og etiske udfordringer kunne gøres FAIR.

Forskningsinstitutioner skal understøtte de forskellige fagområder i at udarbejde/tydeliggøre deres profil i forhold til internationale aktiviteter og, under hensyntagen til omkostningerne, udarbejde en plan for, hvor langt de vil nå med FAIR inden for de næste tre år.

Forskningsinstitutioner og DeiC¹¹ skal sikre, at de rette kompetencer og værktøjer er til stede til at komplementere forskerne i deres formulering og definition af deres profil.

Forslag til konkrete indsatser

Forskningsinstitutionerne skal på nationalt plan sikre, at der udarbejdes beskrivelser af god data management praksis inden for forskellige fagområder. Disse skal som minimum indeholde beskrivelse af, hvordan metadata kan indsamles som en integreret del af arbejdsprocessen samt være et krav i forbindelse med afslutning af projekter/aktiviteter. Dette kan indebære:

- *Udarbejdelse af kortsigtede mål for tilgængeliggørelse og deling af data fx med udgangspunkt i metoder omkring modenhedsmodeller¹²*
- *Hjælpe faggrupper med at identificere, udvikle og/eller publicere standarder, metoder, værktøjer og infrastruktur til at gøre deres data FAIR.*
- *At det accepteres at vurderingen af eksisterende datasæt i forhold til genbrug (kildekritik) er en vigtig disciplin.*

¹¹ DeiC er etableret som et samarbejde mellem de otte danske universiteter og UFM. Grundlaget for DeiC er beskrevet i "En national strategi for forskningsinfrastruktur"

¹² Fx. CMMI Institutes *Data Management Maturity (DMM)*, (<https://cmmiinstitute.com/data-management-maturity>) eller DCC's *RISE Model*, (<https://www.dcc.ac.uk/guidance/how-guides/RISE>)

Understøtte bevaring af forskningsdata på både kort og langt sigt

Det koster ressourcer, både menneskelige og tekniske, at bevare data efter et projekts ophør. Den bedste måde at bevare data på er at holde dem levende, hvorved det sikres at de er relevante og kan tilgås efter dagens standarder. Dette kræver stabil finansiering og organisering.

Det primære sigte med denne strategi er at forskningsdata tilgængeliggøres efter FAIR principperne. Dette kræver i sagens natur at data bevares, ofte i en infrastruktur eller et datasystem som drives af en forskningsinstitution, DeiC, bevaringsinstitutioner eller af forskergrupper. Sådanne datasystemer benævnes i Åbne data-direktivet for institutionelle eller emnebaserede datasamlinger (repositories). Driften af sådanne repositories kræver, foruden den tekniske infrastruktur og IT-systemer, datakyndig bemanning og derfor stabil governance og finansiering.

Bevaring på kort sigt (5-10 år) vil som hovedregel bestå i drift af et repository med det nødvendige niveau at governance og sikkerhed til at kunne sikre tilliden til indholdet som forskningsobjekter. Dette vil i nogle, men ikke alle, tilfælde indebære en certificering efter international standard.

Mange data vil efter nogle år miste relevans og kunne slettes, men nogle data med særlig stor videnskabelig eller samfundsmæssig værdi vil skulle bevares på lang sigt. Dette kan i nogle tilfælde kræve særlige tiltag, såsom vurdering af formater og konverteringer, såvel som behov for stabil organisation over lang tid. Det vil over tid formentlig stille større krav til data at forblive FAIR og forståelige uden for den oprindelige forståelsesramme.

Hvor Rigsarkivet efter samråd med forskerne modtager data med henblik på langtidsbevaring efter arkivloven og tilgængeliggør disse efter FAIR principperne, kan denne strategis krav fortsat siges at være opfyldt.

Herudover er det anbefalingen i de første 5 år efter strategiens ikrafttræden at fokusere på korttidsbevaring, for at få taget væsentlige skridt til at data i større skala bliver bevaret og tilgængeliggjort i det hele taget. Spørgsmålet om særlige krav til infrastruktur, organisation og finansiering af langtidsbevaring bør tages op igen inden 2025.

Forskningsinstitutioner bør arbejde sammen med de forskellige faggrupper for at afdække hvad der allerede findes af stabile datalagringsløsninger og identificere hvor sådanne mangler. Desuden bør der udvikles principper og metoder til at beslutte hvilke data, som skal gemmes og hvilke, der kan slettes.

Forskningsinstitutioner skal medvirke til at der etableres praksis omkring en kontinuerlig forvaltning af data, også efter projekters afslutning.

Forskningsinstitutioner og DeiC skal sikre at driften af institutionelle og nationale datalagre har en tilstrækkelig langsigtet finansiering og vedvarende, løbende sikring af data, herunder identificering af data til langtidsbevaring.

Rigsarkivet skal medvirke til at sikre, at de langtidsbevarede og FAIRificerede forskningsdata kan indgå i en fremadrettet fælles national datainfrastruktur.

Forslag til konkrete indsatser

Samarbejde, også med internationale aktører, om udvikling af metoder og infrastruktur til langtidsbevaring af forskningsdata, som er tilpasset forskningens behov for diversitet i objekt- og filtyper, såvel som skalering i forhold til datamængder.

Understøtte at forskningen kan gennemføre løbende dokumentation i projektførelsen gennem hele livscyklus

Alle informationer, strukturer, metadata etc., som er nødvendige for, at data og andet forskningsoutput kan deles med andre skabes som en integreret del af forskningsprocessen. De relevante overvejelser gennemføres ved starten af et projekt, og flere og flere forskningsinstitutioner og fonde kræver nu, at der udarbejdes en data management-plan f.eks. i forbindelse med ansøgninger. God data management er også med til at sikre reproducerbarhed af forskningen.

Ved afslutningen af et projekt bør der være en opfølgning på data management-planen, hvor det genovervejes, hvad der skal gemmes, hvor og hvor længe. Hvis data bliver afleveret til Rigsarkivet, anses dette for at opfylde strategiens krav. Er der scripts og metoder, som med fordel kan deles med andre - og hvordan?

Forskningsinstitutioner og fonde skal overveje, hvilke krav de stiller i forhold til dokumentation af forskningsprocessen.

Rigsarkivet skal udarbejde procedurer således at forskningsinstitutionerne ved at levere standardiserede FAIR metadata kan opfylde deres anmeldelsespligt.

Forskningsinstitutioner skal sikre at alle forskere anerkender deres ansvar i forbindelse med data management og understøtte udøvelsen af dette.

Forskningsinstitutioner og DeiC skal facilitere at brugen af datalicenser bliver udbredt.

Forslag til konkrete indsatser

Der skal udarbejdes anbefalinger af metoder, som sikrer løbende dokumentation af forskningsdata, i takt med at disse indsamles og genereres, således at alle nødvendige informationer er til stede når resultater skal gøres tilgængelige.

I forbindelse med håndtering af fortrolige data skal der støttes op om, at det er nødvendigt at foretage en risikovurdering i forhold til valg af dataopbevaringsløsninger.

Universiteterne bør stimulere og facilitere en diskussion om risikoen ved åbne data - eksempelvis i forhold til Intellectual Property Rights, patenter eller i forhold til misbrug.

Nationale retningslinjer for brugen af datalicenser skal udarbejdes og udbredes, på basis af og i samarbejde med internationale aktører.

Etablere teknisk infrastruktur, der understøtter data management og tilgængeliggørelse af FAIR data

Forskere ved universiteter og andre offentlige forskningsinstitutioner skal have tilgang til den nødvendige tekniske infrastruktur til indsamling, lagring og tilgængeliggørelse af data. Hvad der er relevant, vil afhænge af projektet, projektets størrelse og niveau af internationalisering. Da forskning ofte er tværinstitutionel og international, kan denne infrastruktur med fordel ses som del af et globalt økosystem, herunder i EOSC.

DeiC skal identificere og komme med forslag til hvilke dele af den tekniske infrastruktur som med fordel kan løses nationalt - under hensyntagen til den internationale udvikling, de nationale og de lokale behov og erfaringer. Kald for løsninger bør gennemføres hurtigst muligt derefter.

Rigsarkivet, forskningsinstitutionerne og DeiC skal samarbejde om at sikre sammenhæng og interaktion mellem deres respektive tekniske infrastrukturer og processer i forbindelse med anmeldelse og mulig aflevering af data til Rigsarkivet.

Forslag til konkrete indsatser

Specificer og byg danske dele af den tekniske infrastruktur:

- *Etabler et generelt tilgængeligt nationalt trusted data repository. Ved at deponere data i dette system kan forskeren gemme og eksponere metadata og data, tilknytte en PID til data og metadata samt tilknytte licens¹³.*
- *Specificer og tilbyd eventuelt andre komponenter nationalt som efterspørges af forskerne, fx. PID-systemer, ontologi-database eller andet der kan lette arbejdet med at gøre data FAIR.*
- *Udbyg den nationale tjeneste for data management-planer med fagtilpassede skabeloner samt med maskinlæsbarhed, der gør data management planer FAIR og dermed eventuelt også anvendelige i forbindelse med anmeldelse til Rigsarkivet.*
- *Sørg for at anbefalede tjenester og infrastrukturer bliver certificeret efter internationale standarder i nødvendigt omfang.*

¹³ Denne indsats er foreslået i notatet til DeiC's bestyrelse, Datamanagement i Danmark, 2019

Etablere vidensinfrastruktur som leverer nødvendig hjælp og vidensressourcer til data management og FAIR

Forandringerne i retning af at forskningen bliver mere datadreven og mere international, påvirker forskningsmetoderne og kræver, ud over nye tekniske værktøjer, ny ekspertise og specialiserede støttefunktioner. I mange forskningsfelter er disse traditioner allerede etablerede. For andre vil der være et stort skridt fra at erkende vigtigheden af data management til at være en aktiv del af infrastrukturen og løbende skabe de nødvendige metadata. Der er i mange sammenhænge blevet kaldt på etablering af en ny profession, "data stewards", som bidrager til at facilitere data management-processerne og at gøre data FAIR.

"Data stewards" med indgående kendskab til data managements problemstillinger, metoder og tekniske værktøjer, skal kunne stilles til rådighed og eventuelt indgå i forskningsprojekter på linje med eksempelvis laboranter, programmører, ingeniører eller bibliotekarer. Dette kompetencefelt kan findes i varianter fra det meget generelle på et mere rådgivende niveau, til det fagspecifikke på højeste akademiske niveau, som er fuldt integreret i forskningen.

Forskningsinstitutioner og DeiC skal sikre, at de nødvendige kompetencer til FAIR data management udvikles og er til rådighed i løbet af forskningsprocessen.

Forskningsinstitutioner og DeiC skal gennem nationalt og internationalt samarbejde sikre, at de rette kompetencer og værktøjer er til stede, til at støtte forskerne i deres implementering af FAIR data. Kompetencer indenfor Data Stewardship har flere dimensioner, og kan:

- A. forstås som fagligt niveauopdelt, ved at spænder fra opkvalificeret IT medarbejdere eller bibliotekarer, til særligt bachelor- eller master-uddannet specialister, til også at blive set som faglighed, der integreres i og af forskningsgrupperne, som forskningskompetence på lige fod med forskeres andre faglige kompetencer (dvs. integreret på peer niveau som medforfatter til artikler mv.),
- B. forstås som havende forskelligt indhold, der spænder sig i et kontinuum fra generel (generisk) data management til det mere fagspecifikke,
- C. forstås som væsensforskellige, i såvel form som indhold, i de forskellige hovedområder (teknisk, naturvidenskab, sundhedsforsikring, samfundsfag og humaniora)

DeiC skal sikre national kritisk masse og koordinering og facilitere vidensudveksling på tværs af universiteter og fagområder.

Forskningsinstitutioner og DeiC skal støtte udvikling og brug af modenhedsmodeller og metrikker i forhold til FAIR, i henhold til fremvoksende international standardisering, primært i RDA og EOSC.

Forslag til konkrete indsatser

Som del af DeiC's foreslåede¹⁴ back office-funktion etableres et nationalt data stewardship kompetencecenter til koordinering af lokale initiativer og til understøttelse af kompetenceudvikling og af større tværgående projekter.

Som del af universiteternes front office-funktion etableres lokale data stewardship kompetencecentre til understøttelsen af institutionernes forskere.

Relevant efteruddannelse i FAIR, data management og data stewardship skal udvikles og tilbydes.

Tværfagligt samarbejde omkring værktøjer, metoder og udveksling af erfaring skal koordineres nationalt.

Viden om data management og FAIR principperne skal stilles til rådighed på relevante trin af forskernes uddannelse, fx. på PhD-skolerne.

Finansiering

Principper for fordeling af udgifter til data management og fysiske datalagre i forbindelse med fondsfinansierede forskningsprojekter

Forskningsfinansiering

At forberede og vedligeholde data efter principperne i FAIR og dernæst at gøre data åbent tilgængelige, hvis der ikke er noget der taler for det modsatte, stiller krav til nye processer i forskningen, til ny teknisk infrastruktur og værktøjer, nye kompetencer og supportfunktioner. Ved at bruge principperne om FAIR og åbne data forventes det, at forskningsinvesteringerne udnyttes bedre. Men det er ikke gratis at implementere. Ekstraudgiften må dækkes af de, som allerede betaler for forskningsprojekterne, dvs. hovedsageligt forskningsinstitutionerne og de private og offentlige forskningsfinansierende fonde.

Institutionernes forpligtigelser

Institutionerne er forpligtigede til at servicere forskerne med en basal eInfrastruktur, som bl.a. gør det muligt at skabe, bruge og gemme forskningsdata på betryggende vis. Den basale infrastruktur skal kunne håndtere alle typer data, og overholde nationale og internationale love og regler. Kompetent brugervejledning og kompetenceopbygning er en grundlæggende del af eInfrastrukturen.

¹⁴ Foreslået i notatet *Datamanagement i Danmark*, 2019

Finansiering og ansvar for etablering af den basale eInfrastruktur

En betydelig del af eInfrastrukturen leveres i fremtiden af Dansk eInfrastruktur Cooperation (DeiC), som en national infrastruktur med datalagre og større computersystemer. Det hele bindes sammen af Forskningsnettet, herunder forbindelse til European Open Science Cloud (EOSC) og adgang til regnekraft ved EuroHPC, herunder centret LUMI, som Danmark har en særlig interesse i.

Udgifterne betales af de otte universiteter og Ministeriet for Uddannelse og Forskning efter en nærmere aftalt nøgle.

Lokalt har universiteterne det fulde ansvar for, at der etableres IT-løsninger, der sikrer forskere og studerende adgang til de nationale og internationale ressourcer, samt at etablere hjælpefunktioner, front office, hvor brugerne kan få hjælp.

Udgifterne til håndtering af data efter de principper, som er beskrevet i denne strategi, vil variere i betydelig omfang, alt efter projekternes størrelse og fokusområde. Nogle vil naturligt kunne håndteres i den basale infrastruktur andre vil, på grund af omfang og kompleksitet nødvendiggøre et betydeligt træk, der overstiger, hvad man kan forvente fra den basale eInfrastruktur. Det er en forudsætning for strategien, at den basale infrastruktur med passende tilførsel af resurser kan udvides og dermed kan levere den nødvendige eService til store og komplekse data projekter. Det skal kun i yderst sjældne tilfælde være nødvendigt at opbygge nye systemer fra grunden og have udgifter til vedligehold og drift af sådanne ved institutionerne.

Udgifter til større udvidelser af den basale infrastruktur, både mht. hardware, datahåndtering og øget træk på hjælpefunktioner, skal kunne inkluderes i projekternes budgetter ligesom udgifter, som institutionerne har til etablering af eInfrastrukturen, skal kunne medtages som institutionernes evt. medfinansiering af projekter.

Udgifter til Data Management og fysiske datalagre

Udgifter til data management infrastruktur, som værtsinstitutionerne må etablere og vedligeholde for at understøtte projekterne, kan groft opdeles i to kategorier:

1. Basal IT-infrastruktur som universitetet under alle omstændigheder skal etablere, fx netværker, print faciliteter, basis software, sikkerhed og de årsværk, som er nødvendige for at holde den basale IT-infrastruktur fungerende og sikker. Med til den basale infrastruktur hører også datahåndtering og almindelig data opbevaring som eksemplificeret i tabel 1.

Udgifter afholdes af institutionerne, idet det må anses som en nødvendighed for opretholdelse af uddannelser og forskning på niveau med internationale partnere, at der findes en basal eInfrastruktur og derfor er en del af institutionernes basisbudget.

Nogle projekter vil trække på den basale IT-infrastruktur i et sådan omfang, at det er nødvendigt at tilføre ekstra ressourcer for at dække projekternes eInfrastruktur behov. Der kan fx være tale om generelle hjælpe funktioner og licenser samt opjustering af lagerkapaciteten.

Udgifter, som relaterer til projektets forøgede brug på den basale eInfrastruktur, **kan inkluderes i projektansøgningens budget.**

2. Dele af eInfrastrukturen, som alene knytter sig til et givet projekt, fx data stewards tilknyttet projektet samt etablering af større langtidslagre (arkiver), som bevarer data ud over projektperioden.

Udgifterne **inkluderes i projektansøgningens budget.**

Tabel 1. beskriver i mere detalje de enkelte elementer af de to kategorier.

Kategori	Software	Lagerplads	Årsværk
1.a Almen IT-infrastruktur hos institutionerne	Basis	Basis EOSC-infrastruktur EMBL-netværk Lagerkapacitet efter bevillingsperiodens udløb (Data fra projektet)	Vedligeholdelse Udvikling Data managers GDPR Harmonisering på tværs af institutioner Kontrol af FAIR principper
1.b Ekstra IT-infrastruktur hos institutionerne koblet til forskningsprojekt el. bevilling	Specifikke licenser Special software	Meget stort træk på datalagre	Større træk på institutionens mandskab til at udvikle data-management og gøre data FAIR
2. IT tilknyttet en bevilling eller projekt	Specifikke licenser Special software	Helt særlige behov for langtidssarkiver	Egne Data stewards

Principper for etablering af omkostningsniveau og placering i de to kategorier

Det skal være både transparent og let for ansøgerne, at bestemme hvilke udgiftsposter, der kan medtages i ansøgningsbudgettet. Der kan være forskel på, hvad der kan søges om ved de forskellige fonde, fx på grund af forskellig praksis mht. overhead.

Forslag til konkrete indsatser

Der etableres en liste over udgifter til eInfrastruktur, som løbende opdateres. Listen er vejledende, men accepteres af fonde og universiteter, som udgangspunkt ved kontraktforhandlinger.

Listen udarbejdes af et udvalg med repræsentanter fra Danske Universiteter og forskningsfondene.

De forskningsfinansierende fondes krav til håndtering af forskningsdata

Kravene til datahåndtering sidestilles med andre lov- og aftalebestedte regler, som gælder for fondsfinansierede projekter. Det gælder at værtsinstitutionerne har ansvaret for at reglerne overholdes. Forskerne kan dermed ikke tilsidesætte regler om, at data skal gøres FAIR, under henvisning til at data er produceret i forbindelse med et fondsfinansieret projekt. Dette gælder for projekter med støtte fra såvel private som offentlige fonde.

Afrapporteringen til fondene indeholder normalt en liste over publikationer med relation til projektet. I fremtiden skal listen også indeholde en angivelse af de datasæt, som er etableret under projektet angivet ved PID (Persistent Identifier), og med angivelse om data er gjort tilgængelige for andre forskere efter at resultaterne er offentliggjort, dvs. om de opfylder loven om åbne data.

Governance og implementering

Styregruppen for udarbejdelsen af nærværende strategi udarbejder, så snart strategien er trådt i kraft, en implementeringsplan. Som et første punkt i denne skal der nedsættes en styregruppe for implementeringen med repræsentation fra de institutionstyper og ministerområder der er omfattet af strategien. Denne gruppe får ansvar for den løbende opfølgning og afrapportering af implementeringen. Implementeringsplanen fastlægger styregruppens sammensætning, såvel som frekvensen for opfølgning og afrapportering.

Implementeringsplanen skal desuden indeholde retningslinjer for hvorledes der vil blive målt på implementeringen i opfølgningssøjled.

Ordforklaring

FAIR	Findable, Accessible, Interoperable, Reusable
FAIR Data Point	Datasystem som eksponerer data og metadata i overensstemmelse med FAIR principperne
Metadata	Data som beskriver andre data: identifikation, beskrivelse/dokumentation, tilblivelseshistorie, licens, mm.
PID (Persistent Identifier)	Entydig identifikation af digital ressource. Skal normalt kunne omsættes til en konkret webadresse via en PID tjeneste (fx. DOI, ORCID)
Repository	Database over "digitale objekter", omfattende data, metadata og PID, ofte med en søgbar brugergrænseflade og computergænseflade (API)
Trusted repository	Repository som er certificeret efter en anerkendt standard, fx. Core Trust Seal, som omhandler bl.a. organisatoriske, tekniske, finansielle og juridiske forhold ifm. repositoret
API	Application Programming Interface (API) er en teknisk grænseflade direkte til en computers software program. Den definerer de typer opkald eller anmodninger, der kan foretages, hvordan man foretager dem, de dataformater, der skal bruges, hvilke data der leveres tilbage, m.m.
FAIR metrikker	Metoder til undersøgelse af i hvilket omfang FAIR principperne er opfyldt. Der arbejdes internationalt, bl.a. i EOSC, på at etablere fælles standarder for FAIR metrikker. FAIR metrikker skal primært ses som et værktøj i forskernes hænder til at gøre data mere FAIR.
Modenhedsmodeller	Modenhedsmodeller kan bruges af et forskningsprojekt eller et fagområde til at definere det ønskede mål for FAIRificering. Man kan så bruge FAIR metrikker til at evaluere opfyldelsen, således at man vægter resultaterne i forhold til de opstillede mål.
Fortrolige data	Data som ikke kan stilles til rådighed efter en åben licens, men skal holdes fortrolige af hensyn til intellektuelle ejendomsrettigheder, beskyttelse af personoplysninger, privatlivets fred, sikkerhed og/eller legitime kommercielle interesser.