

National strategi for data management baseret på FAIR-principper

31. august 2021

DeiC

Forord

Forskning skaber resultater, som kan udnyttes af samfundet til at adressere såvel globale som nationale udfordringer. I kraft af den igangværende digitalisering kan forskningsresultater deles og bruges hurtigere og nemmere. Forskningsdata kan anvendes og genbruges af andre forskere, og resultater kan reproducere med henblik på en sammenligning af resultaterne. Det fremmer en mere effektiv og bedre forskning, herunder øger tilliden til forskningen.

Europa-Kommissionen har lanceret en række politiske og strategiske initiativer, der har til formål at understøtte en bedre forsknings- og samfundsmæssig udnyttelse af forskningsdata, herunder blandt andet European Open Science Cloud, der har til formål at etablere et Web of FAIR data og services, der skal gøre det nemmere at finde og genanvende forskningsdata. I 2018 offentliggjorde Uddannelses- og Forskningsstyrelsen en strategi for nationalt samarbejde om digital forskningsinfrastruktur. Denne strategi for data management baseret på FAIR-principperne er et væsentligt initiativ i strategien fra 2018 og skal ses som et første dansk nationalt svar det europæiske åbne data-direktivs (PSI-direktivet) krav om, at der skal udarbejdes nationale politikker for tilgængeliggørelse af forskningsdata i overensstemmelse med FAIR-principperne.

For at bevare forskningsdata og gøre dem tilgængelige for andre er det nødvendigt gennem hele forskningsprocessen at have styr på indhold og på den fysiske placering af de enkelte datasæt. Dette betegnes data management. For at data kan genfindes, tilgås, sammenkobles og genanvendes, skal de leve op til FAIR-principperne (Findable, Accessible, Interoperable and Reusable). Målet med denne strategi er, at forskningsdata genereret via offentlige midler, og eventuelt medfinansieret af private forskningsfinansierende fonde, skal leve op til FAIR principperne.

Målgruppen for denne strategi er forskere og ledelser ved de danske universiteter, bevaringsinstitutioner, som understøtter forskning og andre institutioner, som med offentlige tilskud udfører forskning. Strategien udstikker en række principper og peger på indsatser, som skal fremme processen frem mod en succesfuld implementering af strategien, herunder udvikling af praksis for hvordan forskningsdata håndteres.

Strategien inkluderer mange forskellige forskningsområder og institutioner. Det er derfor nødvendigt med en åben tilgang, hvor ikke alle forventes at kunne bruge samme løsning, og ikke alle kan gå frem med samme hastighed. Det vil kræve ressourcer, som skal findes i et samspil mellem institutionerne og de forskningsfinansierende fonde.

Denne strategi er tilvejebragt af DeiC efter opdrag fra Uddannelses- og Forskningsstyrelsen, og den er et resultat af et bredt samarbejde, herunder bidrag fra mange relevante interessenter og enkeltpersoner. Alene i forbindelse med høringsprocessen indkom der mere end 700 kommentarer og forslag til strategiudkastet. Vi ser frem til at følge de næste skridt på vejen frem mod en national implementering af data management baseret på FAIR-principperne ved de danske forskningsinstitutioner.

Direktør Hans Müller Pedersen

Styrelsen for Uddannelse og Forskning

Professor John Renner Hansen

Bestyrelsesformand for DeiC

Indholdsfortegnelse

Forord.....	i
Indledning	1
1.1 <i>Formål og vision</i>	<i>1</i>
1.2 <i>Baggrund.....</i>	<i>2</i>
1.3 <i>Gyldighedsområde</i>	<i>3</i>
1.4 <i>Værdien af åbne og FAIR data</i>	<i>4</i>
1.5 <i>Målgruppe og struktur</i>	<i>5</i>
2 Principper for en data management-praksis som understøtter FAIR	6
2.1 <i>Data management skal understøtte FAIR-principperne og mulighederne for deling og genbrug af data</i>	<i>6</i>
2.2 <i>Relevansen og vægtningen af de enkelte FAIR-principper varierer mellem forskningsområder og skal defineres inden for disse.....</i>	<i>6</i>
2.3 <i>Det skal være muligt at bevare alle typer af forskningsoutput og filformater som er relevante for forskningen på kort og lang sigt.....</i>	<i>6</i>
2.4 <i>Data management er en kontinuerlig proces gennem hele forskningsforløbet.....</i>	<i>7</i>
2.5 <i>Relevant teknisk infrastruktur og værktøjer skal tilbydes alle forskere uafhængigt af fag- og institutionstilhørsforhold.....</i>	<i>7</i>
2.6 <i>Den fornødne hjælp og de fornødne vidensressourcer skal være tilgængelige uanset fag- og institutionstilhørsforhold.....</i>	<i>7</i>
3 Indsatser til etablering af en data management-praksis som understøtter FAIR-principperne	8
3.1 <i>FAIR data management er en praksisændring inden for mange fagområder</i>	<i>8</i>
3.2 <i>Faggrupper understøttes i at definere egen implementering af FAIR-principper</i>	<i>9</i>
3.3 <i>Understøtte bevaring af forskningsdata på både kort og lang sigt</i>	<i>10</i>
3.4 <i>Understøtte forskningens krav til dokumentation gennem hele forskningsprocessen</i>	<i>11</i>
3.5 <i>Etablere teknisk infrastruktur, der understøtter data management og tilgængeliggørelse af FAIR data</i>	<i>12</i>
3.6 <i>Etablere vidensinfrastruktur som leverer hjælp og vidensressourcer til data management og FAIR</i>	<i>13</i>
4 Finansiering.....	16
4.1 <i>Principper for fordeling af udgifter til data management og fysiske datalagre i forbindelse med fondsfinansierede forskningsprojekter.....</i>	<i>16</i>
4.2 <i>Principper for etablering af omkostningsniveau og placering i de to kategorier.....</i>	<i>19</i>
5 Implementering og styring	21
6 Ordforklaring.....	22
7 Appendiks 1 – Arbejdsgrupper der har bidraget	24
7.1 <i>FAIR Strategi Styregruppe</i>	<i>24</i>
7.2 <i>Arbejdsgruppen FAIR Finans</i>	<i>24</i>
7.3 <i>Arbejdsgruppen FAIR Strategi</i>	<i>24</i>
7.4 <i>Universitetsgruppen for FAIR strategien</i>	<i>25</i>

Indledning

1.1 Formål og vision

Formålet med strategien er at etablere basis for beslutninger om en national 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, sammenkobles² og genbruges, og at betingelserne for brug af data er klart formulerede. Strategien skal være i overensstemmelse med EU's direktiv om åbne data³ og den danske implementering af dette i PSI-loven⁴, samt bidrage til at opfylde forskningens og samfundets forventning til stigende tilgængelighed af forskningsresultater.

Strategien støtter således op om direktivets krav om den størst mulige offentlige tilgængelighed af offentligt finansierede data på forskningsområdet. Ud over at forøge mulighederne for genanvendelse, kan efterlevelse af principperne i strategien bidrage til højere reproducerbarhed af forskningen og dermed til opfyldelse af *Den danske kodeks for integritet i forskning*⁵.

Gennem strategien fokuseres primært på "FAIR" mere end på "åben", idet data godt kan være FAIR uden at være åbne dvs. frit tilgængelige. Data management efter FAIR-principperne er således også relevant, i forhold til data som af etiske, juridiske, kommercielle eller andre årsager ikke kan stilles åbent til rådighed. FAIR skal dog i denne strategi forstås som »så åbent som muligt, så lukket som nødvendigt«, altså at data gøres åbne, med mindre eksplicitte grunde taler imod det.

FAIR data kan i sig selv ses som en proces, snarere end som et enten-eller med mulighed for efterlevelse af flere eller færre af principperne. Strategien skal således bidrage til at der tages væsentlige skridt til, at forskningsdata bliver mere FAIR, og at forskningsdata behandles og anerkendes som en ressource, der deles i nationale og globale datainfrastrukturer på måder, som stimulerer åben forskning og åben innovation.

Overordnet er det visionen med data management i forskningen, at forskere og øvrige aktører løbende foretager kvalificerede beslutninger og handlinger i forhold til vedligeholdelse og adgang til deres data (og evt. andre resultater/produkter som software, protokoller o. lign) under hensyntagen til forskningens fremme, forskningsinstitutionernes politikker, eksisterende aftaler, samfundets behov og gældende lovgivning. Intentionen er, at strategien skal bidrage til at gøre det lettere for forskerne at dele og genbruge data førende til flere FAIR data af høj kvalitet og dermed til bedre muligheder for forskningen.

Strategien skal komme med anbefalinger til, hvorledes håndteringen af forskningsdata efter de opstillede principper skal finansieres samt opstille forslag til politikker for data management i overensstemmelse med FAIR-principperne for projekter finansieret af de forskningsfinansierende fonde. Man kan samlet sige, at

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 I FAIR termer anvendes begrebet interoperabilitet. 'Sammenkobling' anvendes synonymt med interoperabilitet i strategien.

³ Europa-Parlamentets og Rådets direktiv (EU) 2019/1024 om åbne data og videreanvendelse af den offentlige sektors informationer (2019)

⁴ Den gældende *Lov om videreanvendelse af den offentlige sektors informationer* er under revision for at implementere den seneste udgave af direktivet.

⁵ Uddannelses- og Forskningsministeriet, <https://ufm.dk/publikationer/2015/filer/file> (2014)

målet med strategien er at opstille principper og de tilhørende indsatser for en FAIR data management-praksis og dennes finansiering, samt indsatser for strategiens realisering.

Selv om FAIR-principperne fra 2016 er relativt nye, kan de i en vis forstand forstås som en opsummering og strukturering af allerede etableret praksis. Digitalt stærke forskningsmiljøer, som fx. de europæiske ESFRI-projekter⁶, har allerede i en årrække organiseret sig og bygget løsninger som i høj grad lever op til strategiens mål og principper. For faggrupper, som allerede nu benytter primært fagspecifikke internationale infrastrukturel løsninger eller etablerede institutionelle løsninger, vil nærværende strategi formentlig ikke betyde nogen væsentlige ændringer; de vil fortsat kunne bruge eksisterende infrastrukturer og forskerstøtteorganisering.

For andre fagområder, som måske i mindre grad er databundne, eller som ikke i samme grad har organiseret sig i internationale infrastrukturer, kan strategien byde på en forandring. De enkelte fagområder spiller i den forbindelse en væsentlig rolle i selv at definere hvordan og i hvilket omfang, de vil implementere FAIR principperne i forskningspraksis. Det er forventningen, at implementering af strategien vil strække sig over en årrække med en stigende andel FAIR data i takt med tilvejebringelsen af støttestrukturerne.

Strategien gør det ikke ud for at være en udtømmende behandling af alle relevante problemstillinger omkring FAIR data, men holder sig til nogle hovedspor, der kan hjælpe med at etablere og udbrede en data management-praksis, der understøtter FAIR principperne. Flere detaljer vil kunne indgå i en kommende implementeringsplan. Dog vil der givetvis i praksis løbende vise sig både muligheder og begrænsninger, som kan udnyttes og skal håndteres. Det vil være op til alle de involverede parter i strategiens implementering at støtte op og samarbejde om dette i de kommende år.

1.2 Baggrund

Samfundet og forskningen benytter i stadig stigende grad forskellige typer data som basis for afdækning, analyser og modellering, og det er derfor vigtigt, at der skabes muligheder for at finde, tilgå, sammenkoble og genbruge data. 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 forskningsinfrastruktur⁷.

Strategien tjener også som den danske politik for tilgængelighed iht. EU's direktiv om åbne data (tidligere PSI-direktivet), som i sin nyeste udgave også omfatter forskningsdata og kræver, at medlemslandene opstiller politikker for tilgængelighed af forskningsdata efter FAIR-principperne⁸.

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 tilgængelige, direktivet nævner således, at der skal tages hensyn til problemstillinger i forbindelse med intellektuelle ejendomsrettigheder, beskyttelse af personoplysninger og privatlivets fred, sikkerhed og legitime kommercielle interesser. Data,

⁶ European Strategy Forum on Research Infrastructures, se <https://www.esfri.eu/>

⁷ Strategi for nationalt samarbejde om digital forskningsinfrastruktur, Styrelsen for Forskning og Uddannelse (2018)

⁸ Europa-Parlamentets og Rådets direktiv (EU) 2019/1024 om åbne data og videreanvendelse af den offentlige sektors informationer (2019)

der ikke kan gives åben adgang til, kan dog ofte 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, dvs. enten datasættet i sin helhed eller afledte versioner heraf.

Danmark er medlem af European Open Science Cloud (EOSC)⁹, og det er forventningen, at forskningsdata fremover kan 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. FAIR spiller dog en væsentlig rolle i visionerne om EOSC, og en opfyldelse af FAIR-principperne vil i sig selv være et skridt hen imod tilgængeliggørelse gennem EOSC. Herudover læner strategien sig i nogen grad op ad anbefalingerne fra EOSC FAIR Working Group, som fokuserer på etablering af FAIR praksis¹⁰. I forhold til etablering af infrastrukturer og tjenester bør sådanne tage udgangspunkt i de kendte standarder såvel som i de kendte krav til arkitektur og interoperabilitet i EOSC¹¹.

Det fremgår af bemærkningerne til udkastet til en revideret PSI-lov, at EU's databeskyttelsesforordning (GDPR) og den danske databeskyttelseslov altid vil have forrang over bestemmelserne i åbne data-direktivet og PSI-loven¹². Selv personhenførbare data vil dog med fordel ofte kunne gøres FAIR, selv om de ikke gøres åbent tilgængelige. Arbejdet med data management i henhold til denne strategi kan herudover berøre en række problemstillinger, som er reguleret af anden lovgivning og lokale politikker. Herunder kan nævnes arkivloven, sikkerhedspolitikker m.v. Sådanne forhold skal naturligvis tages i betragtning i det praktiske data management-arbejde, uanset at de som udgangspunkt ikke er behandlet i denne strategi.

1.3 Gyldighedsområde

Nærværende strategi gælder i overensstemmelse med Direktivet om åbne datas artikel 10 og forslaget til den danske PSI-lov for offentligt finansierede forskningsdata forstået som resultatet af (eller output fra) 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 repositorium, så vidt muligt skal kunne videreanvendes til både kommercielle og ikke-kommercielle formål¹³.

Forskningsdata defineres i Direktivet om åbne data¹⁴ 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.*

⁹ <https://eosc.eu/>

¹⁰ Se <https://eoscsecretariat.eu/eosc-fair-wg-outputs>

¹¹ Se bl.a. <https://eoscsecretariat.eu/eosc-architecture-wg-outputs>

¹² Det erkendes at de nærmere detaljer om afklaring af lovgrundlag kan være vanskeligt i internationale samarbejder mellem lande med forskellige fortolkninger af GDPR, ophavsret og lign.

¹³ Åbne data-direktivet, artikel 10 stk. 2.

¹⁴ Åbne data-direktivet, artikel 2, stykke 9

Gennem strategien benyttes begrebet *data*, forstået som det digitale indhold af enhver art, man kan referere til i form af datasæt, filer og databaser m.m. Heri kan indgå alle typer af digitale forskningsoutput, som indgår i forskningen. Det være sig logbøger, software, workflows og scripts m.v, i overensstemmelse med EOSC-rapporten *Turning FAIR into reality*¹⁵, som i stedet for data benytter begrebet *FAIR digitale objekter* for at tydeliggøre bredden af forskningens output.

Definitionen ovenfor har fokus på, at data skal være indsamlet eller produceret i forbindelse med forskningsaktiviteter uanset organiseringen eller de nærmere omstændigheder af disse. Således vil fx data, der indsamles eller produceres ifm. myndighedsbetjening være omfattet, så længe myndighedsbetjeningen baserer sig på forskning. Når det gælder data fra forskningsunderstøttende samlinger som fx. biblioteker, museer og registre, er denne strategis gyldighed begrænset til, de data som direkte bliver gjort til genstand for forskningen, som angivet i definitionen ovenfor¹⁶. Uanset dette vil en forøget tilgængeliggørelse af sådanne samlinger efter FAIR-principperne givetvis kunne bidrage væsentligt til forskningen i Danmark.

Strategiens fokus er på håndteringen af nyproducerede forskningsdata og sigter dermed ikke på en omfattende retrospektiv proces. Dermed udelukkes selvfølgelig ikke, at der allerede kan ligge endog meget værdifulde data hos enkeltforskere eller institutioner, som med fordel ville kunne håndteres efter strategiens principper.

1.4 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 på linje med øvrige aktiviteter i samfundet i højere og højere grad bliver digital, er det nødvendigt, at ikke blot publikationer, men også data gøres nemt tilgængelige, så det bliver muligt at inddrage andres data i egen forskning. Et større datagrundlag kan give nye muligheder for analyser og modeller, fx. klimamodeller, analyser af store tekstkorpora m.v. God data management praksis efter principperne i denne strategi kan bidrage til at gøre dette muligt i praksis. FAIR data skal både være til gavn for forskningen og for samfundet som helhed. For nogle forskningsfelters vedkommende er dette allerede en etableret praksis, hvor fx. beregningskraft og datainfrastrukturer på tværs af lande og organisationer er en forudsætning for at kunne gøre sig gældende inden for området. Dette forhold forventes at brede sig til flere forskningsfelter, og strategiens mål er at bidrage til at kunne opfylde forskernes behov i den forbindelse.

FAIR data vil desuden gøre forskningens outputs 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 troværdighed af forskningen, at datagrundlaget gøres tilgængeligt i den udstrækning, det er muligt.

Forudsætningen, for at disse visioner kan realiseres og komme forskningen til gavn, er bl.a., at data kan findes og identificeres, og at der skal være tilknyttet relevante, meningsgivende og fyldestgørende metadata, herunder at betingelserne for genbrug er angivet i form af en tilknyttet licens. Dette er nogle af formå-

¹⁵ *Turning FAIR into reality. European Commission, DOI:10.2777/1524 (2018)*

¹⁶ Hermed tages der ikke stilling til de pågældende institutioners eventuelle forpligtelser i henhold til andre artikler af Direktivet om åbne data og PSI-loven.

lene 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.

1.5 Målgruppe og struktur

Strategien opstiller en række principper for data management-praksis som understøtter FAIR data 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, men er her listet under det område, hvor de spiller den største rolle.

Principperne retter sig bredt mod alle interessenter i en sådan praksis, herunder forskere, data stewards, forskningsledere og forskningsfinansierende fonde. Forslagene til indsatsområder retter sig derimod primært mod beslutningstagere på forskningsudførende og forskningsfinansierende institutioner. Det vil sige personer, som kan påtage sig ansvar for strategiens implementering og finansiering, herunder for hvordan arbejdet tilrettelægges og gennemføres samt for tilvejebringelsen af de nødvendige politikker og den nødvendige organisering i de enkelte organisationer.

Principperne og indsatsområderne er ikke prioriteret, idet der er underforstået en indbyrdes afhængighed imellem dem.

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

2 Principper for en data management-praksis som understøtter FAIR

2.1 Data management skal understøtte FAIR-principperne og mulighederne for deling og genbrug af data

1. Når data af forskerne eller af forskningsinstitutionen vurderes at kunne have værdi for andre end forskerne selv, skal det overvejes hvordan og i hvilket omfang, data kan gøres tilgængelige.
2. Metadata, som beskriver data kan som hovedregel gøres åbne, også i tilfælde, hvor data ikke er åbent tilgængelige.
3. Data sikres et minimumsniveau i forhold til FAIR ved, at de er identificeret entydigt ved en identifikator (*Persistent Identifier*, PID) og er beskrevet med metadata.
4. En opfyldelse af FAIR-principperne er en trinvis proces, hvor opfyldelsen af enkelte FAIR-principper kan ske gradvist.

2.2 Relevansen og vægtningen af de enkelte FAIR-principper varierer mellem forskningsområder og skal defineres inden for disse

1. Niveauet for opfyldelsen af FAIR ud over de minimale krav som beskrevet ovenfor defineres af de enkelte forskningsområder i relation til internationale normer og standarder for området.
2. Forskere skal arbejde hen imod konsensus inden for deres respektive forskningsområder om, hvilket niveau af FAIRness de tilstræber, herunder hvilke FAIR-principper de vægter højest.
3. De enkelte forskningsprojekter fastlægger processen for at opnå det ønskede niveau af FAIRness som del af deres data management plan.

2.3 Det skal være muligt at bevare alle typer af forskningsoutput og filformater som er relevante for forskningen på kort og lang sigt

1. Data som af forskerne eller af forskningsinstitutionen skønnes at kunne have værdi i fremtiden skal lagres og tilgængeliggøres i en teknisk og organisatorisk sikker datainfrastruktur i overensstemmelse med FAIR-principperne.
2. I tilfælde hvor det fx. af økonomiske grunde er nødvendigt at slette allerede tilgængeliggjorte data, identificeres og håndteres særligt værdifulde data med henblik på langtidsbevaring.
3. 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.

2.4 Data management er en kontinuerlig proces gennem hele forskningsforløbet

1. Forskningsprojekter planlægger og dokumenterer planen for datahåndtering gennem forskningsforløbet i en data management plan som løbende vedligeholdes.
2. Metoder og værktøjer til datahåndtering skal være tilgængelige til at kunne hjælpe forskerne med, at data og metadata bliver indsamlet og lagret undervejs i forskningsforløbet.
3. Betingelserne (licensering) for videreanvendelse af data skal gøres tilgængelige som del af metadata på både menneske- og maskinlæsbar form.

2.5 Relevant teknisk infrastruktur og værktøjer skal tilbydes alle forskere uafhængigt af fag- og institutionstilhørsforhold

1. Infrastruktur til datalagring og tilgængeliggørelse efter FAIR-principperne skal stå til rådighed for alle forskere, som er ansat ved danske forskningsinstitutioner.
2. Den tekniske infrastruktur skal bedst muligt hjælpe med at opfylde forskningsinstitutionernes forpligtelser i forhold til Rigsarkivet, således at dobbeltregistrering kan undgås.
3. De valgte tekniske værktøjer og platforme, hvad enten disse er kommercielle eller open source, skal muliggøre, at data kan trækkes ud og gøres FAIR på åben, standardiseret form.

2.6 Den fornødne hjælp og de fornødne vidensressourcer skal være tilgængelige uanset fag- og institutionstilhørsforhold

1. Det skal være muligt for forskere at arbejde med deres forskningsdata ud fra FAIR-principperne, uden at de behøver at være eksperter i data management ved, at de tilbydes tilstrækkelig og kompetent hjælp fra institutionerne.
2. Kompetencer omkring data management og FAIR udvikles i en proces, der involverer forskere og støttefunktioner.
3. Data stewards skal stå til rådighed på flere niveauer og med tilstrækkelig faglig forankring i fagområderne.

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

3.1 FAIR data management er en praksisændring inden for mange fagområder

En forskningsverden, hvor forskere ikke kun deler deres metoder og data, men også stiller disse til rådighed på måder, så data kan genbruges, vil være med til at forandre 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 nye arbejdsområder i forbindelse med denne praksisændring
- For bevillingssystemet, som skal anerkende, at det at gøre metadata og data genanvendelige er tidskrævende og dermed udgør en yderligere omkostning
- For rekrutteringssystemet, som skal anerkende behovet for nye kompetencer.

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

Det vil være nødvendigt, at forskningsråd og fonde ikke bare indarbejder støtte til denne form for aktivitet, men også anerkender indsatser for FAIR data management som meritgivende aktiviteter.

Forskningsinstitutioner bør stimulere og facilitere denne praksisændring, 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 inden for et stadigt stigende antal videnskaber er i dag helt beroende på at data indsamles, systematiseres, kategoriseres, analyseres (fx. af computere), gemmes (fx. i repositorier eller databaser) og gentilgås af internationale forskningskollegaer.
- At facilitere en forståelse for at FAIR-principperne ikke nødvendigvis er ensbetydende med åben adgang til data, men at også lukkede data kan gøres FAIR ved, at de som minimum bliver tilknyttet en identifikator (PID) samt beskrivende metadata.

Forskningsinstitutioner bør i samarbejde med Uddannelses- og Forskningsministeriet arbejde nationalt og internationalt med meriteringssystemet, således at forskeres arbejde med at publicere, vedligeholde og understøtte deling af FAIR data anerkendes og belønnes.

Forslag til konkrete indsatser:

- Styrkelse af kommunikation og dialog om FAIR data management.

- Opstart af en proces med fokus på at gøre flere outputs i forskningen meritgivende og at sikre en anerkendelse af data management med henblik på FAIR. Dette skal gøres på nationalt niveau med inddragelse af internationale initiativer. Meriteringen bør tage højde for forskelligheder i forskningsområder.

3.2 Faggrupper understøttes i at definere egen implementering af FAIR-principper

De enkelte fagområder har forskellige traditioner for, hvordan data deles og struktureres, ligesom data indgår på forskellige måder og i forskelligt omfang under hensyn til fagets metodiske tilgange og etableret praksis. Det er derfor vigtigt, at det er forskerne i de faglige miljøer som definerer, hvordan og hvor langt FAIR kan implementeres inden for deres respektive områder. De skridt, der tages i retning af tilvejebringelse af FAIR data, skal bygge på god videnskabelig praksis, for at resultatet skal kunne blive af værdi for andre.

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. Desuden kan der være etiske grunde til, at data ikke kan offentliggøres fx i form af et særligt fortrolighedsforhold mellem forsker og informant. Mange forskningsdata vil dog uden store juridiske og etiske udfordringer kunne gøres åbne.

Forskergrupper bør lave en plan for deres understøttelse af FAIR data. En sådan kan fastlægge fx hvilke metadata, man vil eksponere samt hvilke datatyper, ontologier, standarder, PID-systemer m.m., som er relevante at benytte for den pågældende forskning.

Forskningsinstitutioner skal gennem deres støttefunktioner understøtte fagområdernes arbejde med at udarbejde planer og, under hensyntagen til omkostningerne, planlægge hvilke principper, de vil arbejde med, og hvor langt de vil nå med FAIR inden for de næste tre år.

Forskningsinstitutioner og DeiC¹⁷ skal samarbejde om at sikre, at de rette kompetencer og værktøjer er til stede til at understøtte forskerne i formulering og definition af deres FAIR planer.

Forskningsinstitutioner skal støtte forskerne i at kræve, at data som offentliggøres som del af publicering i tidsskrifter eller andre typer publikationer, skal gøres FAIR, sådan at dobbelt-publicering kan undgås.

Forslag til konkrete indsatser:

Forskningsinstitutionerne skal sikre, at der udarbejdes beskrivelser af god data management-praksis inden for forskellige fagområder gerne med afsæt i international praksis. Disse skal som minimum indeholde beskrivelse af, hvordan metadata kan indsamles som en integreret del af arbejdsprocessen. Dette kan indebære:

¹⁷ DeiC er etableret som et samarbejde mellem de otte danske universiteter og UFM. Grundlaget for DeiC er beskrevet i "En national strategi for forskningsinfrastruktur".

- Udarbejdelse af kortsigtede mål for tilgængeliggørelse og deling af data fx med udgangspunkt i metoder omkring modenhedsmodeller¹⁸.
- Hjælpe faggrupper med at identificere, udvikle og/eller publicere standarder, metoder, værktøjer og infrastruktur til at gøre deres data FAIR, som har udspring i eller som i hvert fald spiller sammen med internationale løsninger.
- Arbejde med data management praksis for et udvalg af data og med metoder til kvalitetssikring af data og metadata.
- Udarbejde principper for korrekt (gen)brug af data.

3.3 Understøtte bevaring af forskningsdata på både kort og lang 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 løbende kan tilgås og benyttes efter de til enhver tid gældende forskningsmæssige standarder. Dette kræver stabil finansiering og organisering og er en forudsætning for, at FAIR ambitionerne kan indfris.

Det primære sigte med denne strategi er, at forskningsdata tilgængeliggøres efter FAIR-principperne. Dette kan i visse tilfælde dreje sig om, at data kun gøres tilgængelige, mens et forskningsprojekt er aktivt, men vil ofte indebære, at data bevares og kurateres efter, at forskningen er afsluttet, typisk i en infrastruktur eller et datasystem som er særligt indrettet til dette. Sådanne datasystemer benævnes i Direktivet for åbne data for institutionelle eller emnebaserede datasamlinger (repositorier).

Bevaring i 5-10 år, svarende til fx. kravene i *Den danske kodeks for integritet i forskning*, vil som hovedregel bestå i drift af et repository med det nødvendige niveau af governance, informationssikkerhed og dokumentation til at kunne sikre tilliden til indholdet som forskningsoutput. Dette vil i nogle, men ikke alle, tilfælde indebære en certificering efter international standard. Der bør sigtes på så vidt muligt at skabe et internationalt ophæng for systemer, hvor data deles, fx i regi af internationale forskningsinfrastrukturer. Driften af datarepositorier kræver betydelige ressourcer foruden den tekniske infrastruktur og IT-systemer også datakyndig bemanning og derfor stabil finansiering.

Det kan være vanskeligt at afgøre, hvad der potentielt vil være værdifuldt for eftertiden at bevare, og der kendes adskillige tilfælde, hvor upåagtede forskningsoutput eller publikationer har fået ny betydning mange år senere. Ikke desto mindre vil det fortsat i mange sammenhænge være nødvendigt at vælge ud, hvad der skal bevares på både kort og længere sigt. Overvejelserne kan knytte sig til begrænsninger i forhåndenværende ressourcer af både menneskelig, økonomisk og teknisk art.

Udvælgelse må nødvendigvis ske på nutidens betingelser og kriterier. I hvert fald den indledende udvælgelse bør foretages af forskerne selv, primært efter kriterier omkring forventet forskningsmæssig og samfundsmæssig relevans. Kriterierne for udvælgelse kan påvirkes af politikker for de repositorier eller arkiver som benyttes til fortsat bevaring og tilgængeliggørelse. Procedurer og kriterier for udvælgelsen kan med fordel indarbejdes i data management-plan for det givne forskningsprojekt.

¹⁸ Fx CMMI Institutes *Data Management Maturity (DMM)*, (<https://cmmiinstitute.com/data-management-maturity>) eller RDAs FAIR Data Maturity Model: <https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0>

Bevaring ud over en 5-10 års periode kan i nogle tilfælde kræve særlige tiltag, ligesom der vil være behov for ekspertise og stabil organisation over længere tid. Det vil desuden over tid formentlig stille større krav til data at forblive FAIR og forståelige uden for den oprindelige forståelsesramme. Hovedreglen bør være, at data bevares og gøres tilgængelige i holdbare, åbne formater, hvor specifikationen af formatet er veldokumenteret og offentlig tilgængelig.

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

Det er 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. Spørgsmålet om bevaringshorisont vil dog afhænge af fagområder, og for nogle fagspecifikke repositorier kan bevaring i længere perioder allerede nu være en del af politikken.

Forskningsinstitutioner bør arbejde sammen med de forskellige faggrupper for at afdække, hvad der allerede findes af stabile repositorie-løsninger og identificere, hvor sådanne mangler. Desuden skal der udvikles principper og metoder til at beslutte hvilke data, som skal gemmes og tilgængeliggøres, såvel som retningslinjer og kriterier for, hvornår data har mistet deres relevans og kan slettes.

Forskningsinstitutioner skal medvirke til at der etableres praksis omkring og ressourcer til en kontinuerlig forvaltning af data også efter forskningens afslutning. Herunder skal spørgsmål omkring ansvaret for den fortsatte forvaltning og tilgængeliggørelse af data efter forskere har forladt den institution, hvor data er blevet produceret, afklares.

Forskningsinstitutioner og DeiC skal samarbejde om sikring af driften af institutionelle og nationale data-lagre og repositorier, der skal have tilstrækkelig langsigtet finansiering og vedvarende, løbende sikring og kuratering af data, herunder identificering af data til langtidsbevaring.

Rigsarkivet skal sikre, at de forskningsdata de modtager til arkivering, så vidt muligt tilgængeliggøres efter FAIR-principperne.

Forslag til konkrete indsatser:

- Samarbejde, også med internationale aktører, om udvikling af metoder og infrastruktur til kort- og langtidsbevaring af forskningsdata, som er tilpasset forskningens behov for diversitet i forskningsoutput- og filtyper, såvel som skalering i forhold til datamængder.
- Samarbejde om at definere kriterier for datas værdi ift. genanvendelighed og langtidsopbevaring.

3.4 Understøtte forskningens krav til dokumentation gennem hele forskningsprocessen

Alle informationer, strukturer, metadata etc., som er nødvendige for, at data og andet forskningsoutput kan deles med andre, bør dokumenteres som en integreret del af forskningsprocessen. Det kan ske ved at sikre systematisk opsamling, som beskrives i en data management-plan, der opdateres gennem hele forskningsforløbet. Flere og flere forskningsinstitutioner og fonde kræver, at der udarbejdes en data manage-

ment-plan fx i forbindelse med ansøgninger om fondsmidler. God data management er med til at sikre transparens af forskningen.

Ved afslutningen af et forskningsforløb bør der være en opfølgning på data management-planen, hvor det genovervejes, hvilke data og forskningsoutput, der skal gemmes, hvor, hvor længe og iht. hvilke licenser, herunder om der er scripts og metoder, der understøtter data, som med fordel kan deles med andre.

Fonde bør gå i dialog med forskningsinstitutionerne for at undersøge, hvordan de gennem deres politikker kan understøtte og stimulere, at data (og andet output) gøres FAIR.

Forskningsinstitutioner skal støtte forskerne i at opbygge velfungerende arbejdsgange til løbende opsamling af metadata baseret på eksisterende god dokumentationspraksis i de forskellige discipliner.

Forskningsinstitutioner skal sikre, at der skabes klarhed omkring ansvarsfordelingen i forhold til alle aspekter af data management til understøttelse af FAIR samt aktivt understøtte data management-praksis. Dialog og kommunikation med forskerne bør understøtte et samlet billede af god data management-praksis, herunder fx informationssikkerhed, GDPR, licenser, tilgængeliggørelse, langtidsbevaring mm.

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

Forslag til konkrete indsatser:

- Der skal inden for fagområderne 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.
- Der bør på nationalt plan stimuleres og faciliteres en diskussion om håndtering af risikoen ved FAIR data. Risikofaktorer er eksempelvis fortrolighed, ophavsret, anden immaterialretlig regulering, *dual use*, GDPR samt nationale sikkerhedsinteresser.
- Institutionelle retningslinjer for brugen af datalicenser skal udarbejdes og udbredes, eventuelt i nationalt samarbejde, primært baseret på internationale standarder.

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

Forskere ved universiteter og andre offentlige forskningsinstitutioner skal have adgang 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 EOSC. Vigtigt er, uanset om infrastrukturen bygges lokalt af institutionerne, som fælles nationale løsninger eller som del af internationale samarbejder, at de tilbudte løsninger forankres i forskningens behov og bedst muligt imødekommer disse.

Infrastrukturløsninger skal respektere, at forskningen er international, og skal derfor overholde internationale standarder. Hvor anerkendte internationale løsninger allerede findes, fx open source software, bør

sådanne foretrækkes til implementering frem for egenudvikling. Dette gælder for implementering af infrastruktur og tekniske værktøjer, såvel som for valg af metadatastandards, vokabularer mm.

Infrastrukturen skal så godt som muligt støtte forskerne i at undgå dobbeltarbejde i forbindelse med anmeldelse og eventuel aflevering til Rigsarkivet. I første omgang sigtes især på at understøtte workflows i fx repositorie-systemer sådan at Rigsarkivet adviseres om tilføjelser af nye data og derefter selv henter de metadata, der skal til, for at data anses for værende korrekt anmeldt. På længere sigt bør workflows mellem forskningsinfrastrukturen og Rigsarkivet helt automatiseres, så dobbeltarbejde i forbindelse med både anmeldelse og aflevering kan undgås fuldstændigt.

Forskningsinstitutionerne skal sikre, at der etableres og vedligeholdes den basale infrastruktur som stilles til rådighed for alle institutionens forskere, herunder datalagring, netværk mm. svarende til nutidige behov.

Forskningsinstitutionerne skal samarbejde om at identificere og komme med forslag til hvilke dele af den tekniske infrastruktur, som med fordel kan løses nationalt - under hensyntagen til den internationale udvikling og de nationale og lokale behov og erfaringer.

Forskningsinstitutionerne, Rigsarkivet og DeiC skal arbejde sammen om at etablere workflows, der kan føre til, at registrering af nødvendige metadata i institutionernes datasystemer af Rigsarkivet kan accepteres som opfyldelse af anmeldelsespligten.

Forskningsinstitutionerne, Rigsarkivet og DeiC skal på længere sigt, og senest i forbindelse med revision af Anmeldelsesbekendtgørelsen i 2025, samarbejde om at sikre sammenhæng og interoperabilitet mellem deres respektive tekniske infrastrukturer, sådan at processer kan gøres fuldautomatiske.

Forslag til konkrete indsatser:

Specificér og byg nationale dele af den tekniske infrastruktur:

- Etabler et generelt tilgængeligt nationalt trusted data repository, hvor forskere kan gemme og eksponere metadata og data samt tilknytte persistente identifikatorer (PID) og licens.
- Specificér og tilbyd komponenter nationalt, som efterspørges af forskerne, fx PID-systemer, ontologidatabase og andet, der kan lette arbejdet med at gøre data FAIR.
- Udbyg den nationale tjeneste for data management-planer med fagtilpassede skabeloner samt tilpasninger, der kan gøre data management-planer anvendelige i forbindelse med anmeldelse til Rigsarkivet.
- Sørg for at anbefalede tjenester og infrastrukturer bliver certificeret efter internationale standarder i nødvendigt omfang, og at governance og politikker sikrer korrekt administration af adgangen til data, også efter at forskere har forladt institutionen.

3.6 Etablere vidensinfrastruktur som leverer 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 blive en aktiv del af infrastrukturer og løbende skabe de nødvendige metadata. Der er blevet kaldt på etablering af en ny profession, "data stewards", som bidrager til at facilitere data management-processerne, og at data gøres og forbliver FAIR.

"Data stewards" med indgående kendskab til data managements problemstillinger, metoder, jura og tekniske værktøjer, skal kunne 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.

Opbygningen og organisering af data stewardship-kompetence, som kan indgå i tæt samarbejde med forskerne, er en væsentlig forudsætning for, at FAIR data ikke kommer til at betyde en betydelig belastning og merkomkostning for forskerne. Samtidig med at data stewards forventes at være forankret lokalt i forskningsmiljøerne, kan der med fordel være en national koordinering, som sikrer god praksis på tværs af institutionerne.

Kompetencer inden for Data Stewardship har således flere dimensioner, og kan forstås som:

1. fagligt niveauopdelt, fra opkvalificerede bibliotekarer eller IT-medarbejdere, til særlige bachelor-, master- eller PhD-uddannede specialister. Sidstnævnte kan eventuelt være integreret i forskningsgrupperne som forskningskompetence på lige fod med forskeres andre faglige kompetencer (dvs. integreret på peer niveau som medforfatter til artikler mv.).
2. havende forskelligt teknisk og vidensmæssigt indhold, der spænder over et kontinuum fra generel (generisk) data management til det fagspecifikke.
3. væsensforskellige, i såvel form som indhold, i de forskellige hovedområder (teknisk, naturvidenskab, sundhedsforsikring, samfundsfag og humaniora).

Forskningsinstitutioner og DeiC skal samarbejde om at sikre, at data stewardship kompetencer tilvejebringes gennem opkvalificering og nye uddannelser.

Forskningsinstitutionerne og DeiC skal sikre national koordinering og facilitere vidensudveksling på tværs af forskningsinstitutioner og fagområder.

Forskningsinstitutioner og DeiC skal støtte brugen af modenhedsmodeller og metrikker i forhold til FAIR i henhold til fremvoksende international standardisering¹⁹.

Forslag til konkrete indsatser

- Forskningsinstitutionerne støtter opbygning af lokal data stewardship support til understøttelsen af institutionernes forskere.

¹⁹ Se fx RDA's FAIR Data Maturity Model: specification and guidelines, DOI: 10.15497/RDA0050 og Recommendations on FAIR metrics for EOSC, DOI: 10.2777/70791

- Der etableres koordinering og samarbejde på tværs af forskningsinstitutionerne om tilvejebringelsen af data stewardship service, hvor dette er gavnligt ud fra eksempelvis et ressourcemæssigt hensyn, fx til større tværgående projekter. Større nationale satsninger koordineres under DeiC.
- Relevant efteruddannelse i FAIR, data management og data stewardship skal udvikles og tilbydes til forskere og data stewards på relevante faglige niveauer.
- Viden om data management og FAIR-principperne skal stilles til rådighed på relevante trin af forskernes uddannelse, fx på PhD-skolerne.

4 Finansiering

4.1 Principper for fordeling af udgifter til data management og fysiske data-lagre i forbindelse med fondsfinansierede forskningsprojekter

4.1.1 Forskningsfinansiering

At forberede og vedligeholde data efter FAIR-principperne og dernæst at gøre data åbent tilgængelige, hvis der ikke er noget, der eksplicit 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 data forventes det, at forskningsinvesteringer udnyttes bedre som konsekvens af bedre data management og bedre genudnyttelse af data. Men det er ikke gratis at implementere strategien. Ekstraudgiften må dækkes af dem, som allerede betaler for forskning, dvs. hovedsageligt forskningsinstitutionerne og de private og offentlige forskningsfinansierende fonde.

4.1.2 Institutionernes forpligtelser

Institutionerne er forpligtede til at servicere forskerne med en basal e-Infrastruktur, jf. det der kategoriseres som **kategori 2** nedenfor, som bl.a. gør det muligt at skabe, bruge og gemme forskningsdata på betryggende vis. Den basale infrastruktur skal kunne håndtere forskellige typer data og overholde nationale og internationale love og regler. Kompetent brugervejledning og kompetenceopbygning er en grundlæggende del af e-Infrastrukturen.

4.1.3 Finansiering af og ansvar for etablering af den basale e-Infrastruktur

En betydelig del af e-Infrastrukturen etableres gennem DeiC, som en national infrastruktur med datalagre og større computersystemer. Det hele bindes sammen af Forskningsnettet, som også gennem NORDUnet forbinder de danske institutioner med de internationale netværk. Det giver adgang til bl.a. European Open Science Cloud (EOSC), hvor de internationale forskningsinstitutioner og ESFRI-projekter i fremtiden vil gøre deres data tilgængelige. Finansiering af e-Infrastrukturen kan være placeret internationalt som del af et forskningssamarbejde, som fx i ESFRI-projekterne, som finansieres gennem de internationale medlemskaber.

Udgifterne for den e-Infrastruktur som udbydes nationalt betales af de otte universiteter og Ministeriet for Uddannelse og Forskning efter en nærmere aftalt nøgle.

Institutionerne har det fulde ansvar for, at der etableres lokale IT-løsninger, jf. tabel 2a og 2b, der sikrer forskere og studerende adgang til de nationale og internationale ressourcer samt at etablere hjælpefunktioner, hvor brugerne kan få hjælp.

Udgifterne til håndtering af data efter de principper, som er beskrevet i denne strategi, vil variere i betydeligt 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 e-Infrastruktur. Det er en forudsætning for strategien, at den basale infrastruktur med passende tilførsel af ressourcer kan udvides, og dermed kan levere den nødvendi-

ge service til store og komplekse datasæt. 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.

Basal IT-infrastruktur til data management vil i høj grad blive finansieret gennem det nationale samarbejde i regi af DeiC. Herudover findes der i dag allerede på visse forskningsområder en betydelig datainfrastruktur på internationalt niveau, hvorfor yderligere tiltag på sådanne områder ikke bliver nødvendigt på nationalt eller lokalt niveau.

Udgifter til større projektspecifikke udvidelser af den basale infrastruktur, både mht. hardware, datahåndtering og øget træk på hjælpefunktioner, skal kunne inkluderes i projekternes budgetter. Udgifter, som institutionerne har til etablering af e-Infrastrukturen, kan evt. efter aftale med den enkelte fond medtages som institutionernes medfinansiering af projekter.

4.1.4 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 tre hovedkategorier, med nogle underkategorier:

Kategori 1 - International e-Infrastruktur

Denne kategori er de internationalt forankrede datalagre, som betales gennem internationale medlemskaber og vedligeholdes af værtsinstitutionerne, fx EMBL, CERN, ESO og ESFRI initierede forskningsinfrastrukturer. I mange tilfælde vil det være gratis at lagre data ved disse institutioner, eventuelt også uden dansk medlemskab og betaling.

Kategori 2 - Basal e-Infrastruktur

Basal IT-infrastruktur som forskningsinstitutionen under alle omstændigheder skal etablere, fx netværk, 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 dataopbevaring som eksemplificeret i tabel 1. Den nærmere afgrænsning af basal IT-infrastruktur forventes foretaget i samarbejde mellem forskningsinstitutionerne og fondene.

Disse udgifter betegnes **Kategori 2.A** og 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 e-Infrastruktur, og som derfor er en del af institutionernes basisbudget.

Nogle projekter vil dog trække på den basale IT-infrastruktur i et sådant omfang, at det er nødvendigt at tilføre ekstra ressourcer for at dække projekternes e-Infrastruktur behov. Der kan fx være tale om generelle hjælpefunktioner og licenser samt opjustering af lagerkapaciteten. Udgifter, som relaterer til projektets forøgede brug på den basale e-Infrastruktur og databehandlingen, der gør data FAIR (men ikke nødvendigvis åbne) betegnes som **Kategori 2.B**, og kan inkluderes i projektansøgningens budget.

Kategori 3 - Særegen e-Infrastruktur

Der findes dele af e-Infrastrukturen, som alene knytter sig til et givet projekt, fx data stewards tilknyttet projektet samt etablering af større nye langtidslagre (repositorium eller databaser) og som ikke er en udvi-

delse af eksisterende basale e-Infrastruktur. Disse betegnes **Kategori 3** og udgifterne inkluderes i projektansøgningens budget.

Tabel 1 nedenfor beskriver i mere detalje med eksempler på de enkelte elementer af de nævnte kategorier.

Tabel 1 – FAIR finansieringskategorier

Kategori	Fysisk Infrastruktur	Årsværk
1 International e-Infrastrukturen Basal IT-infrastruktur internationalt, finansieret gennem internationalt samarbejde hos institutionerne.	Infrastruktur tilvejebragt gennem internationale samarbejdsaftaler Fx Det Europæiske Bioinformatikinstitut (EBI), CERN, de europæiske ESFRI projekter samt andre internationale infrastrukturer med eller uden dansk medfinansiering.	Udstationeringer eller ansættelser i internationale organisationer.
2A Almenforbrug af basal e-Infrastruktur Basal IT-infrastruktur nationalt eller lokalt, finansieret af institutionerne, eventuelt gennem samarbejde, fx nationalt i regi af DeiC.	Almindelig IT og lagringskapacitet under og efter bevillingsperiodens udløb (data fra projektet). Basis e-Infrastrukturen kan eventuelt være del af nationale eller internationale samarbejdsstrukturer, evt. i regi af EOSC-infrastruktur. ESFRI eller andet internationalt samarbejde. Software og services, basis licenser fx MATLAB, Microsoft, Amazon, Dropbox mv.	Drift, vedligeholdelse og udvikling. Basal rådgivning og support. Forskerstøttefunktioner, fx teknisk IT, Data stewards, GDPR. Harmonisering på tværs af institutioner. Monitorering af FAIR-principper.

2B Overforbrug af basal e-Infrastruktur Ekstra IT-infrastruktur hos institutionerne koblet til forskningsprojekt el. bevilling.	Meget stort krav til eller træk på datalagre (repositorium eller databaser). Specifikke licenser. Special software.	Særligt store træk på institutionens IT, forskerstøtte eller på Data Steward mandskab, til at udvikle infrastruktur og foretage data-management mhp. at gøre data FAIR.
3 Særegen e-Infrastruktur IT tilknyttet en bevilling eller projekt.	Helt særlige behov for datalagre (repositorium eller databaser). Specifikke software licenser. Special software.	Egne dedikerede Data Stewards til projektet og forskergruppe.

4.2 Principper for etablering af omkostningsniveau og placering i de to kategorier

Det skal være både transparent og let for ansøgerne at bestemme hvilke udgifter, 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. Institutionerne kan eventuelt tilbyde e-infrastruktur på markedslignende vilkår.

Forslag til konkrete indsatser:

- Der etableres en liste over udgifter til FAIR data management-relaterede udgifter (software, lagerplads, årsværk), som løbende opdateres. Listen er vejledende og kan anvendes af fonde og universiteter, som udgangspunkt for udarbejdelse og vurdering af budgetter.
- Listen udarbejdes af et udvalg med repræsentanter fra Danske Universiteter og forskningsfondene.

4.2.1 De forskningsfinansierende fondes krav til håndtering af forskningsdata

Kravene til datahåndtering sidestilles med andre lov- og aftalebestemte regler, som gælder for fondsfinansierede projekter. Det gælder, at værtsinstitutionerne har ansvaret for, at reglerne overholdes. Forskerne kan dermed ikke tilsidesætte princippet om, at man skal tage stilling til i, hvilken grad hvilke 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. Dette i henhold til loven om åbne data, så længe det ikke strider mod anden lovgivning.

5 Implementering og styring

Strategien indeholder en lang række anbefalinger som skal udfoldes i en implementering. Uddannelses- og Forskningsministeriet nedsætter en følgegruppe til at sikre fremdrift i implementeringen. Følgegruppen skal bestå af beslutningskompetente repræsentanter for strategiens primære modtagere, dvs. forskningsinstitutioner og forskningsfinansierende fonde. Følgegruppen kan efter behov nedsætte et antal arbejdsgrupper til at udarbejde konkrete implementeringsplaner. Sådanne arbejdsgrupper kan, ud over fra forskningsinstitutioner og fonde, indeholde medlemmer fra andre typer af institutioner.

Følgegruppen rapporterer til Uddannelses- og Forskningsstyrelsen og holder desuden Danske Universiteter orienteret om sit arbejde.

6 Ordforklaring

Ord	Forklaring
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.
Data stewardship	Indsamling, annotering, strukturering, metadatering, arkivering og kuratering af data, med inddragelse af politikker, lovgivning, standarder, vokabularer mm., ofte med en forankring i et bestemt fagområde
FAIR	Findable, Accessible, Interoperable, Reusable
FAIR Data Point	Datasystem som eksponerer data og metadata i overensstemmelse med FAIR-principperne
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.
FAIRness	Graden af opfyldelsen af FAIR-principperne.
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.
Metadata	Data som beskriver andre data: identifikation, beskrivelse/dokumentation, tilblivelseshistorie, licens, mm.
Modenhedsmodeller	Modenhedsmodeller kan bruges af et forskningsprojekt eller et fagområde til at definere det ønskede mål for FAIR. Man kan så bruge FAIR metrikker til at evaluere opfyldelsen, således at man vurderer resultaterne i forhold til de opstillede mål.
PID (Persistent)	Entydig identifikation af digital ressource. Skal normalt kunne omsættes til en konkret

Identifier)	webadresse via en PID tjeneste (fx. DOI, ORCID).
Repositoryum	Database over digitale objekter/forskningsoutput omfattende data, metadata og PID, ofte med en søgbar brugergrænseflade og computergrænseflade (API). Repositorier er som oftest organiseret efter fagområder (<i>disciplinary repository</i> , gerne internationalt) eller institutionelt (<i>institutional repository</i> , typisk af et universitet).
Trusted repository	Et repository som er certificeret efter en anerkendt standard, fx. Core Trust Seal, som omhandler bl.a. organisatoriske, tekniske, finansielle og juridiske forhold ifm. repositoryet.

7 Appendiks 1 – Arbejdsgrupper der har bidraget

Dokumentet er udarbejdet af arbejdsgrupper nedsat af DeiC's bestyrelse, efter opdrag fra Uddannelses- og Forskningsstyrelsen.

7.1 FAIR Strategi Styregruppe

- John Renner Hansen, Københavns Universitet, DeiC
- Kirsten Winther Jørgensen, Copenhagen Business School, DeiC
- Anne-Sofie Jensen, Rigsarkivet
- Kira Stina Hansen, Det Kgl. Bibliotek
- Lars Christensen, Styrelsen for Forskning og Uddannelse
- Anders Sparre Conrad, DeiC

7.2 Arbejdsgruppen FAIR Finans

- John Renner Hansen, Københavns Universitet, DeiC
- Ole Skøtt, Syddansk Universitet, DeiC
- Thomas Midtgaard, Styrelsen for Forskning og Uddannelse
- Johanne Thorup Dalgaard, Danmarks Frie Forskningsfond
- Birgit Pedersen, Danmarks Frie Forskningsfond
- Maja Horst, Danmarks Frie Forskningsfond
- Steen Markus, Grundforskningsfonden
- Søren Degn-Pedersen, Innovationsfonden
- Thomas Brandt Nielsen, Innovationsfonden
- Ulrik Nicolai de Lichtenberg, Novo Nordisk Fonden
- Jan Egebjerg, Lundbeck Fonden
- Lars Hansen, Velux Foundations (Observatør)
- Thomas Buchvald Vind, Syddansk Universitet
- Rene Belsø, DeiC

7.3 Arbejdsgruppen FAIR Strategi

- Anders Sparre Conrad, DeiC
- Birte Christensen, Aarhus Universitet, DeiC
- Mette Hall-Andersen, Rigsarkivet
- Anne-Sofie Jensen, Rigsarkivet
- Bjarne Andersen, Det Kgl. Bibliotek
- Thomas Midtgaard, Styrelsen for Forskning og Uddannelse
- Bartłomiej Wilkowsk, Statens Serum Institut
- Ebbe Villadsen, Nationale Forskningscenter for Arbejdsmiljø (NFA)
- Martin Sønderholm, De Nationale Geologiske Undersøgelser for Danmark og Grønland (GEUS)
- Adam Arndt, Digitaliseringsstyrelsen
- Ivan Thaulow, Danmarks Statistik

- Niels Ploug, Danmarks Statistik
- Solveig Vibe-Petersen, Danmarks Statistik
- Rene Belsø, DeiC

7.4 Universitetsgruppen for FAIR strategien

- Birte Christensen-Dalsgaard, Aarhus Universitet
- Susanne den Boer Beckers, Københavns Universitet
- Thomas Schlichting, Københavns Universitet
- Marieke Buss, Copenhagen Business School
- Sacha Zurcher, Roskilde Universitet
- Kirsten Ohm Kyvik, Syddansk Universitet
- Karsten Kryger Hansen, Aalborg Universitet