

Foranalyse: indførelse af FAIR data i Danmark

Udarbejdet for Styrelsen for Forskning og Uddannelse

Foranalyse: indførelse af FAIR data i Danmark

Formål med rapporten

Denne foranalyse ser på potentialet for at implementere FAIR data i Danmark ved at estimere gevinster og omkostninger samt undersøge barrierer og muligheder for at implementere FAIR data i Danmark.

Analyseteam

Oxford Research

- Adm. direktør Jakob Stoumann
- Chefanalytiker Morten Larsen
- Senioranalytiker Jakob Falk
- Analytiker Anders Gøgsig Randrup

Højbjerg Brauer Schultz

- Partner Esben Anton Schultz
- Analytiker Sarah Kildahl Nielsen

Om Oxford Research

Knowledge for a better society

Oxford Research er en specialiseret vidensvirksomhed med fokus på velfærdsområderne og erhvervs- og regionaludvikling. Oxford Research gennemfører skræddersyede analyser, implementeringsevalueringer og effektevalueringer for offentlige myndigheder, fonde og organisationer i civilsamfundet. Vi rådgiver også om strategiudvikling, faciliterer udviklingsprocesser og formidler vores viden på undervisningsforløb og seminarer. Vi kombinerer akademisk fordybelse, strategisk forståelse og god kommunikation – på den måde skaber vi anvendelsesorienteret viden, der kan gøre en forskel.

Om Højbjerg Brauer Schultz

Højbjerg Brauer Schultz er et af Nordens førende samfundsøkonomiske konsulenthuse. Højbjerg Brauer Schultz er skabt for at levere viden, der kan udvikle og fremtidssikre velfærdssamfundet. Vi rådgiver offentlige myndigheder, interesseorganisationer, private virksomheder og internationale organisationer. Ved at bygge bro mellem faglig viden, empiriske resultater og den politiske virkelighed leverer vi anvendelsesorienterede analyser, som er veldokumenterede og klart formidlet.

Indhold

EXECUTIVE SUMMARY	3
SAMMENDRAG	5
1. INDLEDNING	13
2. DE SAMFUNDSØKONOMISKE KONSEKVENSER VED AT INDFØRE FAIR FORSKNINGSDATA I DANMARK	21
2.1 Analysemodel	22
2.2 Resultater i grundscenariet	25
2.3 Følsomhedsberegninger	28
3. MULIGHEDER OG BARRIERER FOR AT INDFØRE FAIR DATA I DANMARK	37
3.1 Vurdering af FAIR-principperne	37
3.2 Elementerne i økosystemet	40
4. DE TYSKE ERFARINGER MED AT INDFØRE FAIR DATA	53
5. METODE OG DATAINDSAMLING	58
6. REFERENCER	69

Executive summary

The Danish Agency for Science and Higher Education has commissioned Oxford Research and Højbjerg Brauer Schultz to carry out a preliminary analysis of the potential for implementing FAIR data in Denmark. FAIR is a set of guiding principles to make research data Findable, Accessible, Interoperable, and Reusable.

The purpose of the analysis is two folded, 1) to estimate the benefits and costs of the implementation of FAIR, 2) to examine the barriers and opportunities in implementing FAIR data at the Danish research institutions. Specifically, the analysis provides answers to:

- What costs and what benefits can be expected from the implementation of FAIR data – is it socio-economically a good case?
- How do the main Danish stakeholders assess the potential of the FAIR data concept? What opportunities and barriers do they see in terms of policy, organisation, technology, new habits/culture, time (for researchers) and service development to implement FAIR data in Denmark?
- What are the key learnings from the implementation process of FAIR data in Germany in terms of policy, organisation, technology, infrastructure and change in research culture, and what benefits are achieved/foreseen in Germany?

Two main methodological approaches have been used throughout the study. 1) A systematic literature study of costs and benefits of implementing FAIR data in Denmark. 2) A systematic desk research of relevant studies, reports, articles and 36 in-depth qualitative interviews that focus on key stakeholders' experiences and knowledge of the FAIR principles and their assessment of the potential for introducing FAIR data.

Status – how far is Denmark with FAIR?

Overall, there is no systematic approach to the FAIR principles in Denmark. There is no national Danish policy in this area nor any common consensus on binding guidelines and practices. However, the Danish key stakeholders, including universities, research libraries and political authorities, generally support the FAIR principles and see an immense potential for reaping the benefits in terms of better research and socioeconomic impact. Currently, the Danish approach to FAIR data is mainly locally based. This means that each university, including its different research departments and faculties have developed their own practice to data management and thus comply differently with the FAIR principles. Some universities have developed a research data management policy addressing the FAIR principles and invested in data infrastructure that support a system based on the FAIR principles. Other universities have just recently learned about the FAIR principles and are to varying degrees in a process of developing policies that can support the FAIR principles.

Potential for a positive socio-economic scenario

The literature study indicates that there is a positive socio-economic value of introducing FAIR data in Denmark. This result is based on some of the most conservative assumptions found in the literature. In this scenario we find that the socio-economic net present value is about 2 billion DKK over a 40-year period, if we assume that 50 percent of all research data in Denmark will comply with the FAIR principles. This corresponds to an annual socio-economic gain of about 50 million DKK.

However, it is important to be aware that the calculations in this study are based on several assumptions and the size of the socio-economic value depends to a large extent on how successful and to what extent the FAIR principles are implemented. Therefore, the socio-economic values should not be regarded as accurate estimates, but rather as an assessment under the chosen assumptions.

Key factors to implement FAIR data in Denmark

The analysis finds that the implementation of FAIR data – meaning an ecosystem of research data that foster the findability, accessibility, interoperability and reusability principles – requires that each element in the ecosystem can be linked together and support each other. To have a well-functioning ecosystem to foster the FAIR principles, each element must be handled and balanced. For example, it is a precondition that there should be the right skills to handle and share research data, that there should be a basic infrastructure for storing data, and for researchers to have easy and simple access to data. Since researchers are required to comply with the FAIR principle in practice, it is crucial to stimulate researchers' demand and incentive.

The key factors to implement FAIR data in Denmark are firstly a national coordination and cooperation across research actors, libraries and research funding actors, because it is necessary for creating coherence in the system. Competence and culture are another key area that must be supported to give the researchers the necessary skills and incentives for sharing data. Access to research data is a third crucial element, because a precondition for encouraging researchers to take part in sharing and using digital research data is to give them easy access to data. In addition, there is a need to address a wide range of technical and legal barriers – not only at national level but also internationally. Finally, the analysis stresses that further investments in specialised infrastructure for storage, handling, processing and dissemination of research data is needed in Denmark.

Recommendation for a Danish FAIR data solution

Overall, we recommend building on existing capacity and expertise. The analysis shows that many of the elements needed to realise FAIR data already exist, but they are fragmented and dispersed. Coordination and collaboration are crucial for developing a common approach and understanding of the FAIR data principles. However, a FAIR data solution should be based on a 'bottom-up' approach, meaning that it should be demand driven and based on the researchers' practices and the international context of research. A FAIR data solution should build on local solutions with a cohesive national superstructure – not a “one-size-fits-all” solution, but a solution that can grow based on local, academic and research environments.

Sammendrag

Forskningsdata er i den grad kommet på dagsordenen. Det skyldes for det første, at der genereres stadig større mængder digitale forskningsdata. Graham Pryor taler om, at der finder en ”dataoversvømmelse” sted, og har fundet frem til, at den årlige globale vækst i produktionen af forskningsdata i elektronisk form er på ca. 30 pct.¹ Sideløbende har der både politisk og i forskningsmiljøerne været en bevægelse mod at åbne data op, og den har udmøntet sig i ambitionerne om ’Open Access’ og ’Open Science’.

Under det hollandske formandskab i EU blev European Open Science Cloud (EOSC) introduceret. EOSC sigter mod at udbygge digital deling af forskningsdata i Europa. FAIR-principperne (Findable, Accessible, Interoperable, Reusable) har sidenhen opnået international opmærksomhed og vundet indpas som principper i Holland og Tyskland. Man er i fuld gang med at realisere og implementere FAIR i de to lande, men erfaringerne viser, at det er en kompleks og omfattende proces, der kræver fokus på mange elementer, herunder kulturændringer, infrastrukturudvikling, kompetenceudvikling, reformering af videnskabelige praksisser og meriterings-systemer.

De fire principper, som FAIR består af, indebærer følgende:

- **Findable:** Data tildeles en unik og konsistent identifikator for at kunne genfindes uafhængigt af videnskabeligt forskningsområde.
- **Accessible:** Data beskrives ud fra en standardiseret kommunikationsprotokol for at gøre data tilgængelige – selv efter at data ikke længere er til rådighed.
- **Interoperable:** Data er kompatible med et fælles sprog på tværs af forskningsfelter.
- **Reusable:** Data lever op til en række kvalitetsstandarder for at være læselige – både for mennesker og maskiner.

Principperne bygger overordnet på visionen om et fælles ’data-økosystem’, som har til hensigt at gøre digitale forskningsdata tilgængelige på tværs af forskningsaktører og faglige discipliner. Potentialet ved at indføre FAIR virker umiddelbart stort, idet det ikke alene vil forbedre betingelserne for den individuelle forsker og øge forskningskvaliteten, men formentlig også vil kunne skabe en samfundsmæssig gevinst, idet effektiviteten af både offentlige og private investeringer i videnproduktionen forøges. Som vores analyse viser, vil den samfundsøkonomiske nutidsværdi være på ca. 2 mia. kr. over en 40-årig periode, såfremt 50 pct. af alle forskningsdata i Danmark gøres FAIR.

At få indført og implementeret FAIR kan være med til at give Danmark et forspring internationalt som viden- og forskningsnation. Danmark klarer sig allerede i international sammenhæng godt på centrale FoU-indikatorer og er et af de europæiske lande, der har den højeste FoU-intensitet (målt på FoU’s andel af BNP), men andre lande – på nuværende tidspunkt Tyskland og Holland – er interesserede i mulighederne og potentialerne ved FAIR, hvilket yderligere understreger behovet for, at Danmark ser nærmere på mulighederne for at integrere

¹ Graham Pryor (2012): Why manage research data?

FAIR-principperne i håndteringen af forskningsdata. Imidlertid er det også vigtigt at afveje gevinsterne og omkostningerne ved at implementere en sådan løsning, da den, som analysen viser, er forbundet med markante udfordringer og relativt store omkostninger.

ANALYSENS FORMÅL

Denne foranalyse er sat i søen for at se på potentialet for at implementere FAIR data i Danmark ved at afdække to hoveddele: 1) at estimere gevinster og omkostninger ved indførelsen af FAIR forskningsdata, 2) at afdække og undersøge, hvilke barrierer og muligheder der er for at implementere FAIR forskningsdata på forskningsinstitutionerne i Danmark. Analysen fokuserer konkret på at besvare nedenstående spørgsmål:

- **Hvilke omkostninger og hvilke gevinster kan forventes ved indførelse af FAIR data – er det samfundsøkonomisk en god forretning?**
- **Hvordan arbejder man med implementeringen af FAIR forskningsdata i Tyskland, herunder hvordan organiseres det, hvad er barriererne, hvilke investeringer er der foretaget/skal der foretages, og hvilke gevinster er opnået/forudses?**
- **Hvordan forholder det danske landskab for produktion, lagring og opbevaring af forskningsdata sig til FAIR data-konceptet? Hvilke muligheder og barrierer ser de for sig, og hvad skal der til – i form af politik, organisation, teknik, nye vaner/kultur, tid (for forskere) og serviceudvikling. Kort sagt, hvad er de kritiske faktorer for at implementere FAIR data i Danmark?**

Analysen bygger på en række forskellige metoder og datakilder. Denne metodiske triangulering sikrer, at potentialet for at indføre FAIR data i Danmark belyses fra flere relevante vinkler og sikrer valide og anvendelsesorienterede analyseresultater. Estimeringen af omkostninger og gevinster ved indførelsen af FAIR data i Danmark bygger i denne analyse primært på et grundigt litteraturstudie af gevinster og omkostninger ved indførelsen af FAIR data. Vi har forholdt os kritisk til litteraturen, og hvor det er muligt, har vi anvendt danske datakilder eller justeret antagelserne, så de passer bedre ind i en dansk kontekst. Indhentningen af tyske erfaringer og vurderingen af de danske muligheder og barrierer bygger på en kombination af en grundig desk research af studier, rapporter og artikler samt en bred interviewundersøgelse, der kommer rundt om de centrale interessenters erfaringer og kendskab til FAIR-principperne og deres vurderinger af potentialet for at indføre FAIR forskningsdata. I alt har vi gennemført 36 interviews med danske og tyske nøgleinteressenter.

STATUS – HVOR LANGT ER DANMARK MED FAIR?

De danske nøgleinteressenter, herunder universiteterne, forskningsbibliotekerne og de politiske myndigheder, bakker generelt op om FAIR-principperne. Flere fremhæver, at der er tale om fornuftige principper, som det er svært at erklære sig uenig i. De ser et stort potentiale for at høste gevinster i form af bedre forskning og samfundsøkonomiske effekter, såfremt FAIR-principperne efterleves. De fleste repræsentanter fra forskningsinstitutionerne fremhæver imidlertid, at det er vanskeligt at finde ud af, hvad der helt konkret ligger under de enkelte principper, og hvordan de kan omsættes til praksis.

Manglende systematik og tværgående koordination i den danske FAIR-indsats

Overordnet arbejdes der endnu ikke systematisk med FAIR-principperne i Danmark. Der er ikke en national dansk politik på området og heller ingen fælles konsensus om forpligtende retningslinjer og praksisser for FAIR data. Der er dog en spirende interesse for FAIR i Danmark, set i lyset af at Holland og Tyskland med GOFAIR-initiativet er begyndt at arbejde aktivt på at fremme dagsordenen for at gøre forskningsdata FAIR. Bl.a. er centrale, koordinerende danske aktører som DeIC (Danish e-Infrastructure Cooperation), DM Forum (Nationalt Forum for Forskningsdata Management), DEFF (Danmarks Elektroniske Fag- og Forskningsbibliotek) begyndt at se nærmere på FAIR-principperne, og hvilken rolle de kan spille i en dansk løsning for åbne forskningsdata.

Den danske FAIR-indsats er, set i lyset af den manglende nationale politik, relativt ukoordineret, og mange af aktørerne ved ikke, hvad der generelt foregår på området i Danmark. Arbejdet med FAIR forskningsdata er derfor i høj grad lokalt forankret. De få eksisterende initiativer, der harmonerer med FAIR, er igangsat af de enkelte universiteter (nogle gange i samarbejde med hinanden) og forskningsbiblioteker. De fleste institutioner har på lederniveau et overordnet kendskab til FAIR, men har ikke udarbejdet interne, formaliserede retningslinjer for deling og håndtering af data på tværs af institutter og fakulteter. Denne manglende interne formalisering skyldes bl.a. en stor usikkerhed om, hvad forskningsdata er, og hvordan de skal opbevares. Det betyder i praksis, at ansvaret for deling og håndtering af data de fleste steder er placeret decentralt på fakultets- og institutniveau, hvor det så er op til de enkelte forskere at tage stilling til den konkrete håndtering og datadeling. Enkelte universiteter er dog længere i arbejdet med FAIR, bl.a. DTU, der opererer med FAIR-principperne i sin forskningsdata management-politik for opbevaring af primære materialer og data.

Yderligere behov for investeringer i forsknings- og datainfrastrukturen

Desuden er der få danske universiteter, der stiller fælles systemer til databevaring til rådighed for deres forskere. Universiteterne stiller lagerplads til rådighed for forskerne – ofte i form af netværksdrev – men har ikke fået foretaget de nødvendige centrale investeringer i hverken dataopbevaring eller integrering af data i et system, så de kan deles på tværs af universitetets fakulteter og institutter. Der findes dog enkelte eksempler på større systeminvesteringer. Bl.a. har DTU investeret i et sammenhængende opbevaringssystem, som er i gang med at blive implementeret, så det vil kunne anvendes til langtidsopbevaring og sikker opbevaring af universitetets forskningsdata. SDU og AAU er ligeledes gået i gang med at se på et fælles storage-system.

HVAD ER DE SAMFUNDSØKONOMISKE KONSEKVENSER AF AT INDFØRE FAIR DATA I DANMARK?

En vigtig forudsætning for at prioritere arbejdet med FAIR data i Danmark såvel som i andre lande er, at FAIR data skaber værdi for samfundet. Vi har gennemført en samfundsøkonomisk konsekvensvurdering, som belyser potentielle gevinster og omkostninger ved at indføre FAIR data i Danmark. Udover at give et samlet skøn for den samfundsøkonomiske værdi ved at indføre FAIR data i Danmark giver analysen desuden en forståelse af, hvilke forhold i organiseringen og implementeringen man bør være opmærksom på i forhold til at opnå den størst mulige samfundsøkonomiske værdi af FAIR data.

Analysen tager udgangspunkt i Finansministeriets retningslinjer for udarbejdelse af samfundsøkonomiske konsekvensvurderinger² og bygger på en grundig litteraturgennemgang af den eksisterende viden samt analyser af FAIR forskningsdata og andre åbne data-løsninger i andre lande.

Ud fra et forsigtighedshensyn har vi valgt at basere vores grundscenarie på en række af de mest konservative beregningsforudsætninger, som er fundet i litteraturen. Desuden ser vi bort fra de gevinster, der tilfalder andre lande, hvis FAIR data indføres Danmark, ligesom eventuelle gevinster ved at samarbejde om FAIR data med andre lande heller ikke er medtaget i beregningerne. Øvrige forventede gevinster i form af ny forskning, øget udnyttelse af data, der ellers ikke ville være tilgængelige, og øget tværvideenskabeligt samarbejde inden for forskning og innovation er heller ikke medtaget i beregningerne. Derfor er der formentlig tale om et underkantsskøn for værdien af at indføre FAIR data i Danmark.

Overordnet peger analysen på, at der er en positiv samfundsøkonomisk værdi ved at indføre FAIR data i Danmark. I vores grundscenarie finder vi, at den samfundsøkonomiske nutidsværdi er på ca. 2 mia. kr. over en 40-årig periode, hvis 50 pct. af alle forskningsdata i Danmark gøres FAIR. Det svarer i gennemsnit til en årlig samfundsøkonomisk gevinst på 50 mio. kr. Hvis alle forskningsdata i Danmark gøres FAIR, vil nutidsværdien ligge på knap 4 mia. kr. Analysen peger ligeledes på, at gevinstens størrelse i høj grad afhænger af, hvor succesfuldt og i hvor stort et omfang FAIR implementeres. Tages eksempelvis den andel af forskningen, som ikke ville blive eller kunne skabes uden FAIR forskningsdata, med som ekstra forskning, der skabes som følge af den nye dataløsning, vil den samfundsøkonomiske gevinst ved at indføre FAIR data i Danmark blive væsentligt større.

Der er stor usikkerhed med hensyn til flere af de forudsætninger, som beregningerne bygger på. Flere forhold vil givetvis være anderledes i Danmark, end det er forudsat i beregningerne. Beregningerne skal derfor heller ikke betragtes som præcise estimater, men snarere som en vurdering af værdien ved at indføre FAIR data i Danmark under de valgte forudsætninger. Hvis der ønskes mere præcise skøn for værdien af at indføre FAIR data i Danmark, vil der være behov for at indsamle danske data om en række af de forudsætninger, som beregningerne bygger på.

HVAD ER DE KRITISKE FAKTORER FOR AT IMPLEMENTERE FAIR DATA I DK?

Som vist tyder analysen på, at der er store økonomiske gevinster ved at indføre en FAIR-dataløsning i Danmark. Der er dog en lang række udfordringer og kritiske faktorer, som vil skulle adresseres, hvis det i praksis skal lykkes at implementere FAIR data i Danmark og høste de gevinster ved den øgede datadeling og genanvendelse af data, som en FAIR-løsning potentielt set indeholder.

Implementeringen af FAIR forskningsdata kræver helt overordnet, at de enkelte elementer i økosystemet hænger sammen og understøtter hinanden. For at få et velfungerende økosystem til understøttelse af FAIR forskningsdata skal de enkelte elementer håndteres og afbalanceres, bl.a. er det en forudsætning, at der er de rette kompetencer til at kunne håndtere og dele forskningsdata, at der er en grundlæggende infrastruktur til lagring og opbevaring af data, og at forskerne har en nem og simpel adgang til data. Eftersom det er forskerne, der skal implementere og anvende FAIR data i praksis, er det afgørende at stimulere forskernes efterspørgsel og tage

² Finansministeriet (2017)

udgangspunkt i deres behov, hvad angår genanvendelse og deling af data. Det handler derfor ikke kun om uddannelse og kultur, men også om at støtte forskerne tilstrækkeligt, herunder at de har de rette søgeværktøjer, at de nemt kan tilgå data, og at der er en politik, der specificerer deres praksis.

En national koordination på tværs af de centrale aktører

For at skabe en sammenhæng i systemet er det vigtigt med national koordinering og et samarbejde på tværs af forskningsaktører, biblioteker og forskningsfinansierende aktører. En tværgående koordinering skal samle systemet og sætte de overordnede rammer og mål med øje for det lokale niveau (universiteterne og forskningsbibliotekerne), det nationale niveau (den samlede nationale infrastruktur på tværs af universiteter og fagfelter) samt det internationale niveau (adgang til den internationale infrastruktur). P.t. er der som beskrevet i Danmark ingen national politik eller samlende, koordinerende aktør til at sikre fremdriften på området, fx kan indførelsen af krav til DMP'er (Data Management Planer) understøttet fra de offentlige bevillingsgivere være med til at skubbe til indførelsen af FAIR data.

Forskerne skal have nye kompetencer og incitamenter for at dele data

Kompetencer og kultur er et andet nøgleområde, der må understøttes for at give forskerne de nødvendige kompetencer og evner til at håndtere deling af forskningsdata. I Danmark eksisterer kompetenceudvikling lokalt på nogle universiteter, men der er ikke systematisk fokus på at understøtte en fælles kompetenceudvikling, og flere steder skal forskerne selv stå for deling og håndtering af forskningsdata uden at få nogen kvalificeret, lokal datastøtte. Herudover er der behov for en kulturændring og anden forståelse blandt forskerne af, at ikke kun publikationer, men også forskningsdata, har forskningsmæssig værdi. Det er en grundlæggende udfordring, at det nuværende meriteringssystem ikke tilgodeser forskere, som bruger ressourcer på at indsamle og vedligeholde data, og at deres tid og omkostninger ikke anerkendes. At lægge større vægt på deling og brug af forskningsdata i meriteringen vil givetvis være den mest effektive måde at ændre ved forskernes adfærd. Eftersom den eksisterende meriteringspraksis er internationalt vedtaget, kræver en ændring af denne en international løsning.

Sikre adgang til forskningsdata på tværs af forskningsfelter

Adgangen til forskningsdata er et tredje, helt afgørende element. Der skal en teknisk løsning til at give forskerne adgangen til de samlede data på tværs af forskningsfelter og institutioner. Pointen er her, at man langt hen ad vejen kender til de tekniske krav og løsninger, og i Danmark er der allerede et velfungerende forskningsnet, som forbinder de danske universiteter og forskningsinstitutioner. Hovedudfordringen i Danmark er dog at få samlet eller integreret data og skabt forbindelser på tværs af institutter og fakulteter på de enkelte universiteter, hvor dataopbevaringen og -delingen p.t. foregår på institutniveau og ikke er samlet centralt ét sted. Herudover er der behov for at tage stilling til en lang række af tekniske, juridiske barrierer – ikke bare på nationalt niveau, men også internationalt. Det drejer sig bl.a. om at sikre en international standardisering og tilgængelighed samt at håndtere juridiske gråzoner omkring opbevaring og deling af data, fx i forhold til ejerskab af data og persondatalovgivning.

Investeringer i en specialiseret infrastruktur til at understøtte FAIR data

Grundlaget for økosystemet er en specialiseret infrastruktur til lagring, håndtering, behandling og formidling af forskningsdata. En velfungerende infrastruktur er en forudsætning for, at der overhovedet kan etableres et dansk system for FAIR forskningsdata, og i Danmark peger flere på, at der stadig er behov for investeringer til

etablering af datalagring og datakapacitet. Endvidere er det i den sammenhæng centralt at få en langsigtet og bæredygtig finansiering på plads omkring driften af lagringen og de relaterede datatjenester og supportfunktioner.

LÆRING OG ERFARINGER FRA TYSKLAND

Tyskland er sammen med Holland blandt de lande, som er længst fremme med FAIR forskningsdata. Set i lyset af EOSC og anbefalingerne fra det nedsatte tyske råd for videnskabelig informationsinfrastruktur (Rat für Informationsinfrastrukturen) er GO-FAIR-initiativet omdrejningspunktet for implementeringen af FAIR data i Tyskland. Idéen bag GO-FAIR-initiativet er at få implementeret anbefalingerne fra den nedsatte EOSC-ekspertgruppe i samspil med de initiativer og den infrastruktur, der allerede eksisterer. Derfor er det også intentionen med GO-FAIR, at det er bredt forankret og konsensusdrevet ud fra en erkendelse af, at implementeringen af FAIR ikke bare er tekniske standarder, men i høj grad er betinget af kultur, kompetenceudvikling, finansiering og ansvarsfordeling.

Indtil videre har de tyske initiativer inden for 'Open Science' været ukoordinerede. Det har i høj grad været 'bottom-up'-drevet, men uden en national koordinering. Derfor er situationen i Tyskland i dag, at universiteter og forskningsbiblioteker i de forskellige delstater har igangsat egne initiativer og projekter, og der er derfor ikke i dag et samlet overblik. GO-FAIR-initiativet er tænkt som et brud med den hidtidige praksis og er en politisk bestræbelse på at få koordineret mere på tværs for at undgå løsninger, der ikke er indbyrdes kompatible. Bundesministerium für Bildung und Forschung (det tyske uddannelses- og forskningsministerium) har finansieret GO-FAIR-initiativet, og der er således national og central politisk opbakning til fremme af FAIR forskningsdata.

Vi vurderer, at Danmark har mulighed for at lære af og bygge videre på de tyske erfaringer i implementeringen af en FAIR dataløsning:

For det første er det helt centralt at holde fokus på forskernes efterspørgsel som driver for indsatsen. Det betyder konkret, at en FAIR dataløsning skal udvikles med øje for forskernes brug og behov, fx om de støttes tilstrækkeligt i omstillingen til at arbejde med FAIR data, og hvorvidt de udviklede løsninger tilgodeser forskernes praksisser. I Tyskland har man haft fokus på dialogen med forskningsmiljøerne vedrørende potentialerne ved FAIR og 'Open Science', bl.a. har GO-FAIR-initiativet fokus på at 'modne' forskningsmiljøerne og gå i dialog med forskerne om potentialerne ved FAIR.

For det andet er det en udfordring i Tyskland at få sikret en bæredygtig finansiering af FAIR. Mange af de støttede projekter er finansieret for en begrænset tidsperiode, og risikoen ved projektf finansieringen er, at projekterne strander, så snart projektf finansieringen udløber. I Tyskland bliver det diskuteret, hvilke finansieringsmodeller man kan forestille sig at implementere for at sikre en langsigtet og bæredygtig finansiering. Vi anser det for at være vigtigt, at der også i Danmark tages hul på en diskussion om finansieringsmodellerne, herunder hvordan de kan bidrage til en langsigtet, bæredygtig finansiering til infrastruktur, uddannelse, datastøtte etc.

For det tredje viser erfaringerne fra Tyskland, at der er lang vej, fra man tager en beslutning om at blive FAIR til at realisere visionen. Det tyske forskningssystem er mere fragmenteret end det danske og består af ca. 1.000 offentligt støttede forskningsinstitutioner. Samtidig er det tyske system føderalt, hvilket betyder, at de enkelte

delstater har politisk beslutningskompetence omkring forskningsprioriteter. Det må også forventes at tage lang tid i Danmark at realisere visionen FAIR data, om end det trods alt er et mindre komplekst forskningssystem og politisk er mere centraliseret. Derfor kan en løsning principielt hurtigere rulles ud.

ANBEFALINGER TIL EN DANSK FAIR DATALØSNING?

Som anbefalet i EU-Kommissionens nedsatte EOSC-ekspertgruppes rapport³ er det vigtigt at bygge videre på den eksisterende kapacitet og ekspertise. Mange af elementerne, som skal til for at realisere FAIR data, eksisterer allerede, men udfordringen er, at de er fragmenterede og spredte. Derfor lægger EOSC-ekspertgruppen vægt på koordination og samarbejde på tværs. Denne anbefaling understøttes også i denne analyse, hvis hovedanbefaling er, at en dansk FAIR dataløsning skal følge det princip at sikre balance og skabe sammenhænge i det danske økosystem, så lokale indsats er koordinerede og ikke står alene.

En efterspørgselsdrevet løsning, der tager udgangspunkt i forskerne

Vi vurderer på grundlag af interviews med de forskellige danske interessenter samt de spæde erfaringer fra Tyskland, at det vil være helt afgørende, at en dansk FAIR dataløsning er efterspørgselsdrevet med afsæt i forskernes praksis og den internationale kontekst, forskningen indgår i. Forskeren skal for det første understøttes lokalt på sin institution og have de tilstrækkelige værktøjer og viden til at dele og håndtere forskningsdata. Rammerne for forskerne, herunder udvikling af politikker, incitament, kompetencer og infrastruktur, bør så vidt muligt bygge på internationale erfaringer og samarbejdet i de internationale fora, men med et afsæt i forskningsmiljøernes specifikke behov.

Konkret er der behov for at se på forskernes meriteringssystem, der p.t. ikke tilgodeser, at forskere bruger ressourcer og tid på at indsamle og vedligeholde data. At lægge større vægt på deling og brug af forskningsdata i meriteringen vil givetvis være den mest effektive måde at ændre ved forskernes adfærd. Der ligger derfor et arbejde i gennem deltagelse i internationale fora og samarbejder at finde frem til en måde at håndtere de grundlæggende incitamentsudfordringer.

Lokale initiativer med en sammenbindende national overbygning

En dansk FAIR dataløsning bør derfor bygge videre på de lokale løsninger med en sammenbindende national overbygning. Der er således behov for en national koordination og retning, som dog tager sit udgangspunkt i de igangværende projekter og investeringer, der allerede er foretaget. Analysen har vist, at nogle forskningsinstitutioner allerede har fokus på FAIR og har igangsat projekter, der understøtter FAIR. Eksempelvis er AAU og SDU gået sammen om at investere i et fælles data storage-system, der kan understøtte FAIR-principperne. Samtidig er der også en vis grad af udveksling af viden og inspiration mellem institutionerne, hvor SDU i sin 'Open Science-politik' ser til DTU for inspiration. Disse spæde initiativer er vigtige at få i spil, så erfaringer og viden ikke går tabt, men kan bruges i en bredere, national løsning.

Pas på med en central 'one size fits all'-model, men lad det gro nedefra

Vi vurderer desuden, at det er vigtigt, at en dansk FAIR dataløsning ikke er en 'one size fits all'-løsning, men en løsning, som får lov til at vokse frem med udgangspunkt i de lokale, fag- og forskernære miljøer. Ligesom i Tyskland skal der i Danmark ikke dikteres en national norm for hvert fagområde. I stedet er fremgangsmåden,

³ European Commission (2016): Realising the European Open Science Cloud – First report on recommendations of the Commission High Level Expert Group on the European Open Science Cloud

at det skal komme nedefra, og at institutionerne skal tage FAIR-principperne til sig og herudfra diskutere, hvordan principperne kan omsættes til praksis. De udviklede standarder for deling af forskningsdata skal imødekomme forskernes behov, herunder forskningsdisciplinernes forskellige karakter og specifikationer. Derfor er det helt afgørende, at det er forskerne, der med tilstrækkelig institutionel støtte fra deres eget universitet og forskningsbibliotekerne samt via en fælles, national koordination er dem, som er 'drivere' i at modne og udvikle standarder og løsninger.

1. Indledning

Inden for de senere år er der såvel internationalt som i Danmark kommet et øget fokus på potentialerne for større åbenhed i den videnskabelige proces (Open Science). Open Science kan ses som en bevægelse fra formidling af forskning alene via artikler til videnskabelige tidsskrifter til deling af og udnyttelse af al tilgængelig viden tidligere i forskningsprocessen. Open Science medfører, at viden i langt højere grad gøres direkte tilgængelig og åben for en bred offentlighed, herunder forskere, borgere, institutioner og virksomheder. Det gælder både videnskabelige publikationer (Open Access), som typisk er forskningsprocessens færdige resultat, og data (Open Data) som forskningen bearbejder. Den øgede tilgængelighed af publikationer og data gør forskningsprocessen langt mere gennemsigtig. Det har en positiv effekt på integriteten i forskningen, idet mulighederne for at gentage og afprøve forskningsresultater forbedres. I Danmark er der fokus på især tre elementer af Open Science:

- Open Access (åben adgang) til videnskabelige publikationer
- Open data (åbne forskningsdata)
- Forskningsintegritet.

Denne analyse vedrører åbne forskningsdata. Den øgede forskningsaktivitet og digitaliseringen af forskningsviden og data betyder, at der i dag genereres massive datamængder inden for alle forskningsområder. Dermed er der behov for en indsats på området, dels for at udnytte det store potentiale i data, dels for at effektivisere og øge kvaliteten af forskningen ved at øge mulighederne for samarbejde og reducere omkostningerne til indsamling af data som følge af en udnyttelse af data, der ellers ikke ville være tilgængelige for genbrug.

Som led i EU-Kommissionens planer om at etablere European Open Science Cloud (EOSC), hvor tanken er at skabe åben adgang for 1,7 mio. europæiske forskere til forskningsdata i Europa, er FAIR (Findable, Accessible, Interoperable, Reusable) en konkretisering af principper, der kan danne grundlag for en ny måde at bruge digitale forskningsdata på. FAIR-principperne specificerer ikke tekniske krav, men er retningslinjer for øget genbrug af digitale forskningsdata. Herudover er det en helt central pointe, at principperne ikke understøtter én standard, men at de er vejledende og åbner for forskellige operationelle tilgange.

Denne analyse har til formål at undersøge potentialet for at implementere en FAIR forskningsdataløsning i Danmark. Umiddelbart er potentialet for at indføre FAIR stort, idet en fuld implementering af FAIR forskningsdata ikke alene vil forbedre betingelserne for den individuelle forsker, men i høj grad også vil kunne skabe en samfundsøkonomisk gevinst, idet effektiviteten af både offentlige og private investeringer i videnproduktion forøges. På den anden side skal en FAIR dataløsning i Danmark implementeres i en kompleks kontekst med forskellige interesser og behov. Endvidere vil det kræve massive investeringer i infrastruktur, kompetenceopbygning og datastøttefunktioner.

For at opnå et videnbaseret grundlag har Styrelsen for Forskning og Uddannelse derfor bedt Oxford Research og Højbjerg Brauer Schultz om at gennemføre en foranalyse af indførelsen af FAIR forskningsdata i Danmark med fokus på: 1) at estimere gevinster og omkostninger ved indførelse af FAIR, 2) at undersøge de barrierer

og muligheder, der er ved at implementere FAIR forskningsdata på forskningsinstitutionerne i Danmark. Konkret vil analysen besvare følgende spørgsmål:

- **Hvilke omkostninger og hvilke gevinster kan forventes ved indførelse af FAIR data** – er det samfundsøkonomisk en god forretning?
- **Hvordan arbejder man med implementeringen af FAIR forskningsdata i Tyskland**, herunder hvordan organiseres det, hvad er barriererne, hvilke investeringer er der foretaget/skal der foretages, og hvilke gevinster er opnået/forudses?
- **Hvordan forholder det danske landskab for produktion, lagring og opbevaring af forskningsdata sig til FAIR data-konceptet?** Hvilke muligheder og barrierer ser de for sig, og hvad skal der til – i form af politik, organisation, teknik, nye vaner/kultur, tid (for forskere) og serviceudvikling – for at implementere FAIR data i Danmark?

1.1 LÆSEVEJLEDNING

Som optakt til de to tværgående analyser vil rapporten i det følgende give et kortfattet overblik over baggrunden for FAIR samt over, om og i hvilken grad Danmark allerede arbejder med FAIR data. I kapitel 2 præsenteres den første tværgående analyse – en analyse af omkostninger og gevinster ved at indføre FAIR data i Danmark. I kapitel 3 præsenteres den anden tværgående analyse – en analyse af barrierer og muligheder for at indføre FAIR data i Danmark. I kapitel 4 gives et overblik over erfaringer med FAIR forskningsdata i Tyskland. Analysens metodiske overvejelser og datagrundlag er beskrevet i kapitel 5, og rapportens referencer fremgår af kapitel 6.

1.2 BAGGRUNDEN FOR FAIR

Helt overordnet skitserer FAIR-principperne betingelserne for 'good data management'. Principperne udgør et forsøg på at definere et sæt alment gyldige kriterier for hensigtsmæssig datadeling. Kriterierne skal sikre genomsigtighed gennem tydeligt specificerede krav til (meta)data, som udgør en forudsætning for den praktiske genanvendelighed⁴. Data skal her forstås som digital information, der bruges til produktionen af viden på forskellige forskningsfelter, og kan deraf antage et utal af former. Den eneste praktiske afgrænsning er det fælles digitale format, som udgør en forudsætning for deling på online-platforme⁵.

FAIR dækker over fire principper, der oversat til dansk er: genfindelig, tilgængelig, kompatibel og genanvendelig. Princippernes konceptualisering og operationalisering er beskrevet i tabellen nedenfor.

⁴ Wilkinson et al. (2016): The Fair Guiding Principles for Scientific Data Management and Stewardship.

⁵ Force11 (2017): Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing

Tabel 1.1 Overblik over FAIR-principperne

FAIR-principperne	Konceptualisering	Operationalisering (krav)
Findable <i>(genfindelige)</i>	Det skal være muligt, at data kan findes af maskiner, da datamængderne er for store til, at mennesker kan håndtere dem.	<ul style="list-style-type: none"> • Data er tildelt en unik og konsistent identifikator (O-ID). • Data beskrives med fyldestgørende metadata og er registrerede eller indekserede i en søgbar database.
Accessible <i>(tilgængelige)</i>	Data skal kunne tilgås – dog ikke nødvendigvis af alle.	<ul style="list-style-type: none"> • Data kan genfindes på baggrund af O-ID'er, som beskrives ud fra en standardiseret kommunikationsprotokol. • Metadata er tilgængelige – selv efter at data ikke længere er til rådighed.
Interoperable <i>(kompatible)</i>	Data skal være kompatible, dvs. være forberedte på at blive kombineret med andre datasæt.	<ul style="list-style-type: none"> • Data gør brug af et formelt, tilgængeligt, udbredt og bredt anvendeligt sprog. • Data inkluderer relevante referencer til andre datasæt.
Reusable <i>(genanvendelige)</i>	Data skal kunne genbruges i fremtidige forskningsprojekter og viderebehandles.	<ul style="list-style-type: none"> • Data beskriver et flertal af præcise og relevante egenskaber. • Data frigøres med en klar og tilgængelig datalicens. • Data forbindes med deres oprindelse. • Data imødekommer områderrelevante standarder.

FAIR-principperne udspringer af et frivilligt netværk af interessenter, som på ganske kort tid har formået at præge den politiske 'Open Science'-agenda. Principperne har hurtigt fået international opmærksomhed og vundet indpas som etablerede standarder i flere EU-institutioner. FAIR-principperne er bl.a. blevet en integreret del af EU-Kommissionens arbejde med en European Open Science Cloud (EOSC). EOSC er et initiativ, der følger op på EU-Kommissionens strategi for EU's digitale indre marked, og som sigter mod optimal genanvendelse af forskningsdata. EOSC skal fungere som et bindeled mellem nationale initiativer og skabe rammerne for skabelsen af en fælles europæisk infrastruktur. Der er således tale om en europæisk overbygning på eksisterende nationale systemer i form af ensartede standarder for professionel datahåndtering, ekspertrådgivning, økonomisk støtte, best practice-eksempler og teknisk understøttelse.

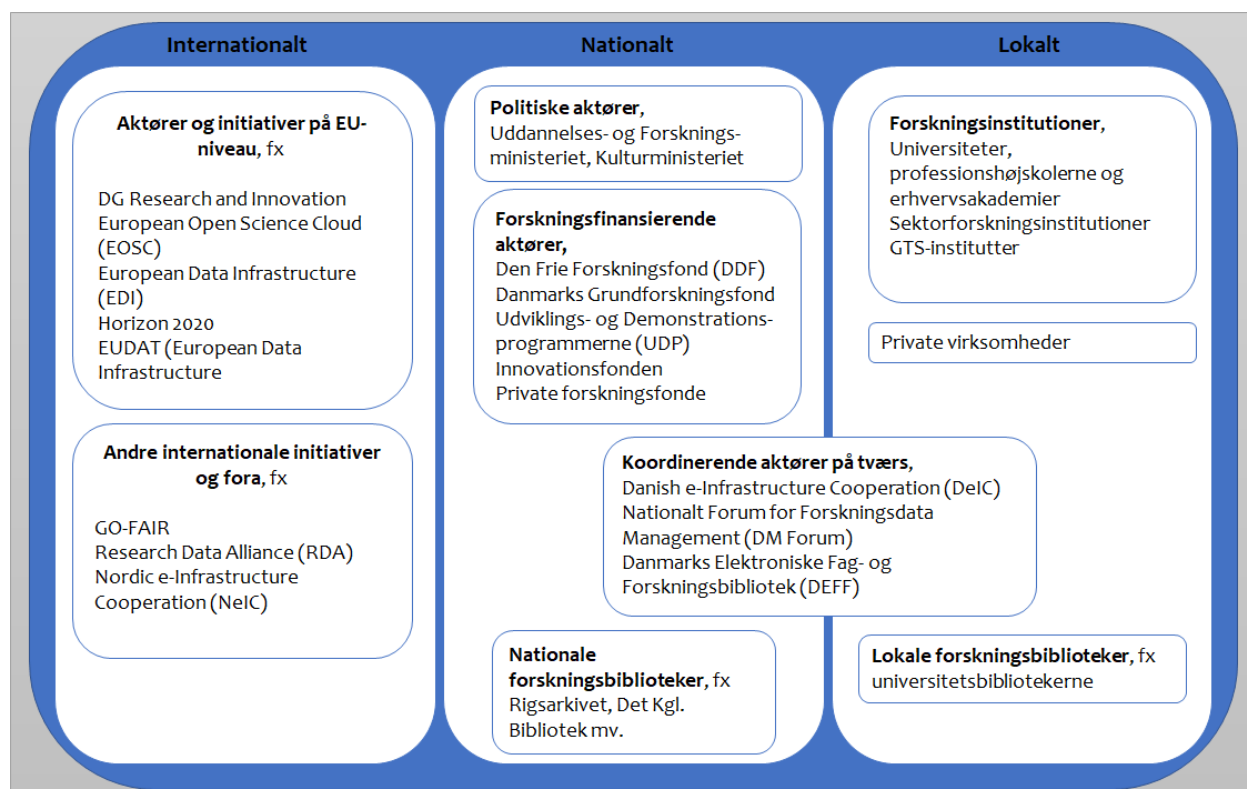
Samtidig har G20 tilsluttet sig principperne, og både Science Europe og EU's forsknings- og innovationsprogram Horizon2020 har fået øje på potentialet i en mere effektiv brug af forskningsdata og stiller nu eksplicite krav om, at modtagere skal følge FAIR-princippernes logik.

1.3 HVOR LANGT ER DANMARK MED FAIR?

Analysen viser, at Danmark p.t. ikke arbejder systematisk med FAIR-principperne. Det var også forventet, da Danmark endnu ikke har hverken en national politik på området eller fælles forpligtende retningslinjer og praksisser for FAIR data. Dog er der en række initiativer – nationalt og lokalt – der kan bidrage til at understøtte FAIR principperne.

Det danske og internationale interessentlandskab i relation til FAIR forskningsdata består af en række interessenter med forskellige roller og interesser. Som det fremgår af figur 1.2 nedenfor, er der i relation til FAIR forskningsdata et lokalt, nationalt og internationalt niveau.

Figur 1.2 Centrale interessenter i relation til FAIR data



FAIR-principperne udspringer af en international dagsorden og er som beskrevet blevet en integreret del af European Science Cloud (EOSC). DG Research and Innovation under EU-Kommissionen har finansieret EOSC-pilotprojektet, som skal understøtte bevægelsen mod EOSC. Med udgangspunkt i EOSC-ekspertgruppens anbefalinger i 'Realising the European Open Science Cloud' er GO-FAIR-initiativet blevet initieret med Holland og Tyskland i spidsen. Formålet er at konkretisere EOSC-politikken inden for tre søjler, hhv. GO-

CHANGE (omhandler kulturen), GO-TRAIN (omhandler kompetenceudvikling) samt GO-BUILD (omhandler den tekniske implementering). Der findes yderligere en række internationale netværk og initiativer, som er relevante ift. FAIR, og hvori Danmark deltager aktivt, bl.a. EUDAT og RDA.

På nationalt niveau er der en lang række forskellige interessenter på området. På det politiske niveau er både Uddannelses- og Forskningsministeriet og Kulturministeriet vigtige aktører med henblik på at sikre en koordineret forsknings-e-infrastruktur. De forskningsfinansierende aktører, herunder Den Frie Forskningsfond, Danmarks Grundforskningsfond og Innovationsfonden, er tillige vigtige aktører på området, da de finansierer en stor del af den danske forskning og således også forskningsdata. Der er samtidig en række nationale forskningsbiblioteker, herunder Det Kongelige Bibliotek, Rigsarkivet mv., som i kraft af deres rolle i at indsamle, bearbejde, lagre og stille forskningsdata til rådighed spiller en helt central rolle på området. Til at fremme koordineringen af de elektroniske forskningsbiblioteker har Uddannelses- og Forskningsministeriet, Undervisningsministeriet og Kulturministeriet samfinansieret Danmarks Elektroniske Fag- og Forskningsbibliotek (DEFF). Ligeledes er DeiC (Danish e-Infrastructure Cooperation) oprettet for at understøtte Danmark som e-Science-nation gennem levering af e-infrastruktur (computing, datalagring og netværk) til forskning og udgør dermed en helt central aktør på FAIR data-området.

På lokalt niveau er der de udførende aktører, dvs. forskere, der selv driver forskning eller anvender forskningsresultater til egne formål. Det er forskerne, der i sidste ende gennem deres praksis skal operationalisere og bruge FAIR-principperne. Virksomheder har ligeledes en væsentlig interesse i at få adgang til forskning, da det bidrager til deres udvikling og vækst. Forskerne bedriver forskning på de danske universiteter samt GTS- og sektor-forskningsinstitutionerne. I tilknytning til de danske forskningsinstitutioner udgør forskningsbibliotekerne også en vigtig aktør på området, idet bibliotekerne understøtter dokumentation og lagring af forskningsdata.

1.3.1 Det danske fokus på Open Access og Open Science

Danmark har siden 2012 taget aktiv del i udviklingen mod Open Science⁶ inden for tre indsatsområder⁷:

- Open Access
- Open Data
- Forskningsintegritet.

I 2012 offentliggjorde de offentlige forskningsråd og fonde (Danmarks Grundforskningsfond, Den Frie Forskningsfond og Innovationsfonden) en fælles Open Access-politik, der medførte, at forskere og andre bevillingsmodtagere fra forskningsrådene skal offentliggøre deres forskningsresultater i Open Access. Politikken er en såkaldt grøn Open Access-politik, som betyder, at forskerne skal parallelarkivere – hvis tidsskrifterne tillader

⁶ Open Science refererer til forskere, regeringer, forskningsfinansierende organer og de videnskabelige miljøers bestræbelser på at gøre de primære resultater af offentligt finansieret forskning – artikler og forskningsdata – offentligt tilgængelige i digitalt format med få eller ingen restriktioner med henblik på at accelerere forskningen. Formålet er at fremme gennemsigthed, samarbejde og innovation” (OECD 2015).

⁷ Regeringen (2016), Forsknings- og Innovationspolitisk redegørelse

det – forskningsartiklerne i et institutionelt eller emnespecifikt repository, dvs. digitalt arkiv, hvortil der vil være åben adgang for alle interesserede via internettet

I 2014 fulgte en National Strategi for Open Access, der bakker op om den ”grønne” vej til Open Access. Til at koordinere og understøtte implementeringen af den nationale strategi er der nedsat en National Styregruppe for Open Access med repræsentanter fra alle universiteterne, Sektorforskningens Direktørkollegium, Danske Professionshøjskoler, Den Frie Forskningsfond, Lundbeckfonden, Novo Nordisk Fonden og DEFF. Styregruppen besluttede bl.a. at etablere en Open Access Indikator, der løbende skal monitorere udviklingen i Open Access i Danmark. I 2017 viser indikatoren, at 23 pct.⁸ af de videnskabelige artikler fra forskningsinstitutionerne var Open Access.

Open Data betyder, at offentligt finansierede forskningsdata gøres tilgængelige i digitalt format med få eller ingen restriktioner. Open Data er bl.a. en central del af EU-Kommissionens planer om at etablere den europæiske åbne forskningscloud (EOSC), der bl.a. sigter på at udvikle et åbent miljø for de 1,7 mio. europæiske forskere, hvor de kan lagre, dele og genbruge videnskabelige data. FAIR-principperne er således tæt forbundet med Open Data-intentionen, dog med undtagelser ud fra princippet: *As Open as Possible, as Closed as Necessary*. I Danmark er der som beskrevet ikke en national politik på området, men flere lokale initiativer er igangsat af enkelte universiteter og forskningsbiblioteker, der understøtter udviklingen mod åbne data.

En del af logikken bag Open Science er, at den øgede tilgængelighed af publikationer og data gør forskningsprocessen mere gennemsigtig, hvilket har en positiv effekt på integriteten i forskningen, idet mulighederne for at gentage og afprøve forskningsresultater forbedres. Uddannelses- og Forskningsministeriet har sammen med Danske Universiteter i den forbindelse lanceret Danish Code of Conduct for Research Integrity. Kodeksen indeholder grundlæggende principper for integritet (ærlighed, gennemsigtighed og ansvarlighed) og standarder for ansvarlig forskningspraksis, som bl.a. omhandler data management. Her står der, ”at data skal gemmes, opbevares og administreres i en klar og præcis form, som gør det muligt at evaluere resultaterne, følge procedurerne og – når det er relevant og muligt - reproducere forskningen”⁹. Alle danske universiteter og en lang række forskningsråd og private forskningsfonde samt GTS- og sektorforskningsinstitutionerne har tilsluttet sig kodeksen.

1.3.2 En spirende dansk interesse for FAIR data

Vores interview med institutledere på de danske forskningsinstitutioner viser, at datahåndteringspolitikker er ved at komme på dagsordenen, og de fleste forskningsinstitutioner har i dag defineret retningslinjer og praksisser for forskernes datahåndtering. Interviewene viser, at FAIR-principperne i enkelte tilfælde er konkretiseret i politikken, men det er stadig uklart, hvordan forskernes praksis skal leve op til principperne, idet de ofte ikke har adgang til en fælles data storage, ikke har ressourcerne eller ikke får tilstrækkelig datastøtte til at realisere principperne.

På nationalt niveau viser interviewene med DeIC og DEFF, at de er begyndt at se nærmere på FAIR, og hvilken rolle de kan spille i den danske FAIR dataløsning. Eksempelvis har DEFF i sit arbejde med at fremme udviklingen af et netværk af elektroniske biblioteker konkretiseret FAIR i sine vurderingskriterier til uddeling af

⁸ Den danske forskningsdatabase, link: <http://www.forskningsdatabasen.dk/>

⁹ Danish Code of Conduct for Research Integrity, p. 9

mindre og større projekttilskud til danske forskningsbiblioteker. DEFF har sammen med DeIC i 2014 tillige formuleret en national strategi for forskningsdata management, hvori formålet er ”at sikre Danmark bedre og mere konkurrencedygtig forskning gennem effektiv opsamling, sikring, formidling og genbrug af relevante forskningsdata”¹⁰. Strategien har siden 2015 understøttet en samlet indsats på data management-området, herunder et samarbejde i et nationalt forum for forskningsdata management (DM Forum) samt lokale initiativer såsom etablering eller udbygning af støttefunktioner på universiteter og andre forskningsinstitutioner. DM Forum forsøger p.t. at indarbejde elementer af den internationale udvikling, som er sket siden strategiens formulering, på europæisk niveau især FAIR-principperne og EOSC. Dette sigter ifølge formand for DM Forum, Anders Sparre Conrad, på at skabe den fornødne sammenhæng mellem lokale, nationale og internationale initiativer.

En gennemgang af interviewene med institutledere fra forskningsinstitutionerne viser overordnet, at der er en spirende interesse for FAIR, set i lyset af at mere forskning er datadrevet, og at forskningsfinansierende aktører, herunder særligt i Horizon 2020, stiller krav om datahåndtering. Det er dog ikke alle institutledere, der har kendskab til FAIR-principperne. Nogle fik først kendskab til FAIR, da vi tog kontakt til dem i forbindelse med analysen. Andre er i gang med at integrere FAIR-principperne i deres politikker, og enkelte har indskrevet FAIR-principperne i deres politikker. ’Open Access’ og ’Open Science’ har været på dagsordenen på flere universiteter i en længere periode, hvorfor bl.a. DTU og KU har investeret i fælles systemer og infrastruktur, der kan understøtte implementeringen af FAIR.

Nedenstående tabel 1.2 giver et overblik over de danske forskningsinstitutioners fokus på FAIR samt eksempler på fundne initiativer i relation til FAIR. Tabellen er udarbejdet med udgangspunkt i 30 interviews med dekaner, institutledere, medarbejdere og forskere på de danske forskningsinstitutioner og øvrige centrale aktører på området. Tabellen er ikke en systematisk gennemgang af alle initiativer, men giver et overblik over de eksempler på politikker og initiativer, der er angivet af de interviewede informanter.

Tabel 1.2. Danske forskningsinstitutioners kendskab til FAIR og eksempler på initiativer, der understøtter FAIR

Institution	Kendskab til FAIR	Eksempler på initiativer i relation til FAIR
KU	Kendskab på ledelsesniveau, men de fleste forskere har ikke kendskab til FAIR-principperne. Ansøgningskrav om data management-planer i H2020 gør bl.a., at kendskabet til FAIR er stigende.	<ul style="list-style-type: none"> ERDA – KU Science har implementeret et repositorium, hvor forskere kan lagre data og gøre dem tilgængelige for andre KU Science-forskere. Dataverse – et samarbejde med Det Kgl. Bibliotek og KU HUM om formidling og tilgængelighed af forskningsdata.
AAU	Har på ledelsesniveau et relativt stort fokus på FAIR-principperne – FAIR skal integreres som en del af forskningen.	<ul style="list-style-type: none"> Samarbejder med SDU om at opbygge data-infrastruktur/lagringskapacitet, der kan understøtte FAIR-principperne.
AU	Kendte ikke på forhånd til FAIR-principperne specifikt, men der er et stigende fokus på Open	<ul style="list-style-type: none"> ’Open Science platform’ – AU har etableret en platform, hvor industrien og forskningen kan dele data og viden.

¹⁰ National strategi for forskningsdata management 2015-2018, p. 3

	Science/data. Bl.a. som følge af krav om data-tilgængelighed fra flere tidsskrifter.	
DTU	FAIR er et prioriteret område på ledelsesniveau, og der er et stigende kendskab til FAIR-principperne blandt forskerne.	<ul style="list-style-type: none"> • 'DTU Research Data Management Forum' – siden 2015 har DTU etableret et centralt kontor for bl.a. at gøre data FAIR. • Data-repositorium – har i nogle år investeret i et sammenhængende repositorium, der forventes at være implementeret inden for de næste 2-3 år.
SDU	Open Access har længe været et stort fokusområde. Der er et stigende fokus på at udvide det til Open Science, hvor FAIR-principperne indgår.	<ul style="list-style-type: none"> • 'SDU Open Science Policy' – har udviklet en open science-politik, som indeholder FAIR-principperne. • Samarbejder med AAU om at opbygge datainfrastruktur/lagringskapacitet, der kan understøtte FAIR-principperne.
CBS	Kendskab til FAIR på ledelsesniveau, men forskerne har generelt ikke kendskab til FAIR. Nogle forskere har erfaring med at bruge eksterne (udenlandske) repositorier til deling af data.	<ul style="list-style-type: none"> • 'CBS Research Data Management Policy' – henviser til FAIR-principperne i datahåndteringspolitikken.
RUC	Kendskab til FAIR på ledelsesniveau, dog et generelt begrænset kendskab til FAIR blandt forskere.	<ul style="list-style-type: none"> • Har udviklet en politik for forskningsdata management, hvori dele af FAIR-principperne indgår implicit.
ITU	Kendskab til FAIR. ITU er i gang med at udvikle en forskningsdatamanagement politik, der bl.a. indeholder FAIR-principperne. Den forventes klar i slutningen af 2018.	
Det Kongelige Danske Kunstakademis Skoler (KADK)	Kendte ikke på forhånd til FAIR-principperne, men oplever en stigende interesse for deling af data.	
Rigsarkivet	Har kendskab til FAIR – oplever, at principperne generelt er et stort fokusområde, da FAIR er en "klassisk arkivtænkning".	<ul style="list-style-type: none"> • Udarbejder p.t. en digitaliseringsstrategi med særligt fokus på forskningsdata.
Det Kgl. Bibliotek	Har kendskab til FAIR, men arbejder ikke aktivt med at integrere principperne.	<ul style="list-style-type: none"> • 'Data management i praksis' – har netop afsluttet et pilotprojekt, hvor der er etableret et data-repositorium, der er stillet til rådighed for alle danske forskere.

Denne gennemgang viser, at der er et relativt stort fokus på den åbne forskningsdagsorden og i nogle tilfælde specifikt ift. FAIR. De forskellige initiativer, hvor nogle arbejder eksplicit med FAIR, mens andre implicit arbejder med initiativer, der understøtter deling af data, viser, at indsatsen er relativt ukoordineret og i høj grad lokalt forankret. Det betyder, at håndtering og deling af data varierer meget på tværs af institutter og fakulteter.

2. De samfundsøkonomiske konsekvenser ved at indføre FAIR forskningsdata i Danmark

En vigtig forudsætning for at prioritere arbejdet med FAIR forskningsdata i Danmark såvel som i andre lande er, at FAIR data skaber værdi for samfundet. I dette kapitel gennemfører vi en samfundsøkonomisk konsekvensvurdering, som belyser potentielle gevinster og omkostninger ved at indføre FAIR data i Danmark. Udover at give et samlet skøn for den samfundsøkonomiske værdi ved at indføre FAIR data i Danmark giver analysen desuden en forståelse af, hvilke forhold i organiseringen og implementeringen man bør være opmærksom på i forhold til at opnå den størst mulige samfundsøkonomiske værdi af FAIR data.

Analysen tager udgangspunkt i Finansministeriets retningslinjer for udarbejdelse af samfundsøkonomiske konsekvensvurderinger¹¹ og bygger på en grundig litteraturgennemgang af den eksisterende viden og analyser af FAIR data og andre åbne dataløsninger i andre lande.

Ud fra et forsigtighedshensyn har vi valgt at basere vores grundscenarie på en række af de mest konservative beregningsforudsætninger, som er fundet i litteraturen. Desuden ser vi bort fra de gevinster, der tilfalder andre lande, hvis FAIR data indføres i Danmark, ligesom eventuelle gevinster ved at samarbejde om FAIR data med andre lande heller ikke er medtaget i beregningerne. Øvrige forventede gevinster i form af ny forskning, øget udnyttelse af data, der ellers ikke ville være tilgængelige, og øget tværvideenskabeligt samarbejde inden for forskning og innovation er heller ikke medtaget i beregningerne. Derfor er der formentlig tale om et underkantsskøn for værdien af at indføre FAIR data i Danmark. Tages eksempelvis den andel af forskningen, som ikke ville eller kunne skabes uden FAIR forskningsdata, med som ekstra forskning, der skabes som følge af den nye dataløsning, vil den samfundsøkonomiske gevinst ved at indføre FAIR data i Danmark blive væsentligt større.

Overordnet peger analysen på, at der er en positiv samfundsøkonomisk værdi ved at indføre FAIR data i Danmark. I vores grundscenarie finder vi, at den samfundsøkonomiske nutidsværdi er på ca. 2 mia. kr. over en 40-årig periode, hvis 50 pct. af alle forskningsdata i Danmark gøres FAIR. Det svarer i gennemsnit til en årlig samfundsøkonomisk gevinst på 50 mio. kr. Hvis alle forskningsdata i Danmark gøres FAIR, vil nutidsværdien ligge på knap 4 mia. kr. Analysen peger ligeledes på, at gevinstens størrelse i høj grad afhænger af, hvor succesfuldt og i hvor stort et omfang FAIR implementeres.

Der er stor usikkerhed med hensyn til flere af de forudsætninger, som beregningerne bygger på. Flere forhold vil givetvis være anderledes i Danmark, end det er forudsat i beregningerne. Beregningerne skal derfor heller ikke betragtes som præcise estimater, men snarere som en vurdering af værdien af at indføre FAIR data i Danmark under de valgte forudsætninger. Hvis der ønskes mere præcise skøn for værdien af at indføre FAIR data i Danmark, vil der være behov for at indsamle danske data om en række af de forudsætninger, som beregningerne bygger på. I kapitel 5 er der en nærmere beskrivelse af metode- og datagrundlaget for beregningerne.

¹¹ Finansministeriet (2017)

I det følgende præsenterer vi først den overordnede model, som vores analyse bygger på. Dernæst præsenterer vi resultater fra vores grundscenarie, og til sidst præsenterer vi en række illustrative beregninger af konsekvenserne af at ændre nogle af de mest kritiske forudsætninger i analysen.

2.1 ANALYSEMODEL

Analysen er en samfundsøkonomisk konsekvensvurdering (også kaldet en cost-benefit-analyse), som følger Finansministeriets metodemæssige retningslinjer. Gevinsterne i analysen er i udgangspunktet modelleret som i Houghton & Gruen (2014), mens opgørelsen af omkostninger ved at indføre FAIR data følger Beagrie et al. (2010), jf. figur.

Figur 2.1. Analysemodel

Opsætning af cost-benefit-analysen

Kilde: Finansministeriet (2017)

- Diskonteringsrente
- Skatteforvridding



Gevinster

Kilde: Houghton & Gruen (2014)

- Sparet dataarbejde for forskere, herunder afkastet af yderligere forskning
- Ny forskning skabt som følge af FAIR data (medtages af forsigtighedshensyn ikke i grundscenarie).



Omkostninger

Kilde: Beagrie et al. (2010)

- Opstartsomkostninger, fx opbyggelse af metadata, workflow-værktøjer, policy mm.
- Driftsomkostninger, køb og drift af servere og dataadgang.

Analysen har i grundscenariet en analyseperiode på 40 år. I de første 20 år implementeres FAIR data gradvist i Danmark, og herefter er der 20 år med fuldt implementerede FAIR data. Ifølge Finansministeriets vejledning bør analyseperioden svare til den forventede levetid for investeringen. I vores analyse består investeringen af omkostninger til at udvikle dataløsninger, retningslinjer, metadata mv. Levetiden af investeringen er forbundet med stor usikkerhed. Når løsningen er udviklet, forventes den at kunne eksistere i mindst 20 år med mindre opdateringer og justeringer som en del af den almindelige drift. En barriere, der påvirker analyseperioden, er forskernes mindset og kulturen omkring beskyttelse af forskningsdata, der bliver utroligt svær at ændre. I afsnit 2.3 er der en følsomhedsanalyse af antagelsen om analyseperioden. Barrierer i forhold til at ændre kulturen kan desuden have betydning for andelen af data, der kan konverteres til FAIR forskningsdata, hvorfor vi i grundscenariet både beregner værdien, når 50 pct. af alle forskningsdata i Danmark gøres FAIR, og når 100 pct. af alle forskningsdata gøres FAIR.

I den samfundsøkonomiske konsekvensvurdering ønsker vi at vurdere den samlede samfundsøkonomiske værdi for Danmark over hele analyseperioden. Det er derfor vigtigt at kunne sammenligne gevinster og omkostninger, som realiseres i forskellige år i analyseperioden. Der findes en række teoretiske og empiriske argumenter for, at gevinster og omkostninger, der realiseres fremover, skal vægtes lavere end gevinster og omkostninger, der realiseres i dag. Derfor kan man ikke bare lægge de umiddelbare værdier af gevinster og omkostninger sammen over tid. Finansministeriet anbefaler, at fremtidige værdier skaleres, således at den typiske præference for nutidigt forbrug frem for fremtidigt forbrug afspejles. Dette gøres ved at tilbageskrive fremtidige værdier med den såkaldte diskonteringsrente.

Vi har anvendt en række centrale parametre, som anbefales i Finansministeriets retningslinjer, herunder diskonteringsrenten og skatteforvridningsfaktoren. Vores resultater er opgjort i en række nøgletal, som ligeledes er anbefalet af Finansministeriet. I figur 1.2. er der en nærmere beskrivelse af de centrale parametre og nøgletal, som anvendes i vores beregninger.

Figur 1.2. Centrale parametre og nøgletal i samfundsøkonomiske konsekvensvurderinger

Diskonteringsrente: Den samfundsøkonomiske diskonteringsrente bruges til at vægte fremtidige fordele og ulemper, sådan at de direkte kan sammenlignes med nutidige værdier. Renten er sammensat af den risikofri realrente og et tillæg for systematisk, ikke-diversificerbar risiko. Diskonteringsrenten følger over meget lange perioder en faldende profil.

Skatteforvridningsfaktor: Skatteforvridningsfaktoren anvendes ved beregning af samfundsøkonomiske konsekvenser af initiativer, hvor det offentlige bærer udgifter (eller indtægter). Faktoren bygger på en antagelse af, at offentlige konsekvenser leder til, at skatten stiger eller falder. Skatteforvridningsfaktoren repræsenterer den marginale samfundsmæssige omkostning, som opstår på grund af behov for generel skattefinansiering.

Nettonutidsværdi: Den samfundsøkonomiske nettonutidsværdi er summen af de tilbagediskonterede gevinster og omkostninger. Nettonutidsværdien udtrykker dermed den samlede værdi af projektet for samfundet. Et projekt er rentabelt, hvis nettonutidsværdien er positiv.

Benefit/cost-ratio: De samlede tilbagediskonterede gevinster divideret med de samlede tilbagediskonterede omkostninger. Udtrykker robustheden af en enten positiv eller negativ nettonutidsværdi.

Afkastratio: Nettonutidsværdien divideret med den tilbagediskonterede nettoinvestering. Udtrykker det samlede samfundsøkonomiske afkast af den samlede investering.

Intern rente: Den interne rente angiver det årlige samfundsøkonomiske afkast af investeringen og repræsenterer den rente, hvor nettonutidsværdien præcis bliver nul.

2.1.1 Samfundsøkonomiske gevinster

De samfundsøkonomiske gevinster i vores analyse dækker over sparet tid for forskere. Derudover er der lavet følsomhedsanalyser af betydningen af ny forskning, som bliver skabt, når data gøres FAIR.

Erfaringer fra datacentre med FAIR data i andre lande indikerer, at det kan lede til store tidsbesparelser for forskere, der arbejder med data. Tidsbesparelserne skyldes, at adgangen til data bliver lettere, og/eller at forskerne ikke selv behøver at generere data. I vores analyse medtages dermed værdien af den tid, som forskerne sparer. Denne beregnes ud fra internationale studier af, hvor meget tid forskerne kan spare, når forskningsdata gøres FAIR og værdisættes ud fra forskernes lønninger. Desuden antager vi, at den tid, forskerne sparer, vil blive brugt på at forske mere, hvilket medfører et yderligere samfundsøkonomisk afkast. Afkastet beregnes ud fra studier af afkast på forskning og den tidligere beregnede direkte værdi af den sparede tid. Værdien af den sparede tid estimeres i vores grundscenarie til 10,5 mia. kr. over en periode på 40 år (for mere om beregningsmetoden, se kapitel 5).

Erfaringerne fra datacentrene indikerer endvidere, at en række af brugerne af centrene ikke ville have kunnet gennemføre deres forskning, hvis datacentrene ikke eksisterede. Med andre ord er FAIR data med til at skabe ny forskning. Det kan hænge sammen med, at forskere først får kendskab til data, når de gøres FAIR, at forskningsprojekter med mindre budgetter nu også får råd til at anvende data, eller at tværfaglige samarbejder bliver gjort nemmere med FAIR data. I vores grundscenarie har vi set bort fra værdien af ny forskning. Det skyldes, at der i den eksisterende litteratur ikke er solid viden om, i hvor høj grad den nye forskning erstatter anden forskning eller giver anledning til mindre forskning på andre områder. Det kan være, at der indsamles mindre datamængder, da der nu er et større udbud af eksisterende data, eller simpelthen at forskernes begrænsede tid flyttes fra én type forskning til en anden. En konservativ antagelse er derfor, at ny forskning blot fortrænger eksisterende forskning. Det svarer til at antage, at værdien af den nye forskning er lige så stor som værdien af den forskning, som fortrænges eller forsvinder, når data gøres FAIR. Man kan dog forvente, at der vil være en tendens til, at forskere vil bruge deres tid på den forskning, som giver det største afkast. Ved at se helt bort fra værdien af ny forskning er der således tale om et konservativt skøn for den samfundsøkonomiske værdi af FAIR data.

På grund af for stor usikkerhed eller manglende data er der i analysen endvidere set bort fra en række andre potentielle gevinster ved at indføre FAIR data i Danmark, se kapitel 5.

2.1.2 Samfundsøkonomiske omkostninger

I vores analyse består omkostninger af opstartsomkostninger og driftsomkostninger. De eksisterende undersøgelser af FAIR data eller andre åbne dataløsninger har typisk fokuseret mindre på omkostninger end gevinster. Omkostningerne i vores analyse er derfor baseret på en enkelt undersøgelse¹² og er behæftet med stor usikkerhed. Vi vurderer dog, at omkostningerne udtrykker et øvre estimat for de potentielle omkostninger, jf. kapitel 5.

Opstartsomkostningerne omfatter analytikerarbejde i forbindelse med opstart, opbyggelse af metadata og rammer for metadata, udvikling af værktøjer og software samt udvikling af retningslinjer og evt. love i forbindelse

¹² Beagrie et al. (2010)

med indførelsen af FAIR data. De er estimeret til at udgøre i gennemsnit 24 pct. af de samlede lønomkostninger til data over implementeringsperioden og antages at stige i perioden. Det er antaget, at opstartsomkostningerne er stigende i perioden, da man kan forvente, at de data, som er nemmest at gøre FAIR, vil blive gjort FAIR først. Samlet set udgør opstartsomkostningerne i hele implementeringsperioden en netto nutidsværdi på 6,1 mia. kr. i grundscenariet.

Driftsomkostningerne dækker over lokale data management-omkostninger såsom køb og drift af servere samt online-adgang og backup og langsigtet opbevaring af data. De er estimeret til at udgøre 3 pct. af de samlede lønomkostninger til data over hele perioden, dog forventes de i implementeringsperioden at være proportionale med den andel af data, som er gjort FAIR. I vores grundscenarie udgør driftsomkostningerne over perioden på 40 år en netto nutidsværdi på 0,8 mia. kr.

2.2 RESULTATER I GRUNDSCENARIET

Som beskrevet ovenfor vurderer vi overordnet set, at resultaterne i grundscenariet repræsenterer et konservativt skøn for værdien af at indføre FAIR data i Danmark. Det skyldes, at vi ser bort fra en række mulige gevinster, potentielt overvurderer omkostningerne og i øvrigt har valgt at basere beregningerne på de mest konservative beregningsforudsætninger, vi har identificeret i litteraturen. Der er dog enkelte af de anvendte beregningsforudsætninger, som muligvis er optimistiske i en dansk kontekst, men hvor vi ikke har været i stand til at finde andre estimater i litteraturen. I afsnit 2.3 er der en række følsomhedsberegninger, som illustrerer konsekvenserne ved at ændre på disse beregningsforudsætninger.

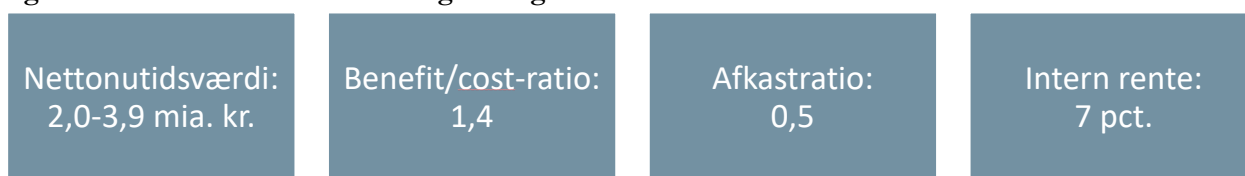
Overordnet peger vores beregninger på, at der er en positiv samfundsøkonomisk værdi ved at indføre FAIR data i Danmark. Det forudsætter, at FAIR data i et nævneværdigt omfang reducerer den tid, forskere bruger på dataarbejde. Hvis ikke det sker, kan indførelsen af FAIR data føre til et samfundsøkonomisk tab. Den samfundsøkonomiske værdi af FAIR data er således meget følsom over for dette parameter, og det er derfor vigtigt at have fokus på at opnå en så stor effektiviseringsgevinst som muligt, hvis man vælger at indføre FAIR data i Danmark.

I vores grundscenarie finder vi, at indførelsen af FAIR data har en samfundsøkonomisk netto nutidsværdi på mellem 2 og 4 mia. kr. over en periode på 40 år. Det svarer til en gennemsnitlig årlig gevinst på 50-100 mio. kr. Værdien afhænger af, hvor stor en andel af alle data, der gøres FAIR. Hvis eksempelvis 50 pct. af alle forskningsdata i Danmark gøres FAIR, er nutidsværdien 2 mia. kr., mens den er knap 4 mia. kr., hvis 100 pct. af alle data gøres FAIR. Dermed stiger den samfundsøkonomiske gevinst, i takt med at flere data gøres FAIR. Hvor stor en andel af de danske forskningsdata, som kan gøres FAIR, afhænger af en række faktorer, som bl.a. er beskrevet i kapitel 3. Der findes allerede flere FAIR data-initiativer i Danmark, ligesom det næppe er realistisk at gøre alle forskningsdata FAIR. Derfor er det ikke muligt at opnå gevinster, der svarer til, at 100 pct. af alle data gøres FAIR.

Resultatet i vores grundscenarie er relativt robust (uanset om det er 50 pct. eller 100 pct. af alle data, der kan gøres FAIR), fordi gevinsterne er ca. 40 pct. større end omkostningerne. Den interne rente indikerer, at det årlige afkast af investeringen i FAIR data er på omkring 7 pct., hvilket er væsentligt over diskonteringsrenten på 4 pct. Afkastratioen viser, at investeringen i FAIR data (her målt ved opstartsomkostningerne) giver en samfundsøkonomisk værdi, der er ca. halvt så stor som investeringen, hvilket er relativt lavt sammenlignet med

estimerer for return on investment (RoI) for investeringer i data-, biblioteks- og informationsservices. Her indikerer estimerne, at de skaber værdi, der er mellem ca. 3 og 5 gange så stor som investeringen.¹³ Dette skyldes sandsynligvis de meget store omkostninger i vores analyse. De resterende nøgletal afhænger ikke af, hvor stor en andel af alle data, der kan gøres FAIR. Det skyldes, at både gevinster og omkostninger antages at stige eksponentielt, således at både de klart største gevinster og de klart største omkostninger realiseres i slutningen af analyseperioden.¹⁴ I de efterfølgende figurer er det antaget, at 75 pct. af alle data gøres FAIR.

Figur 2.2. Samfundsøkonomiske nøgletal - grundscenarie



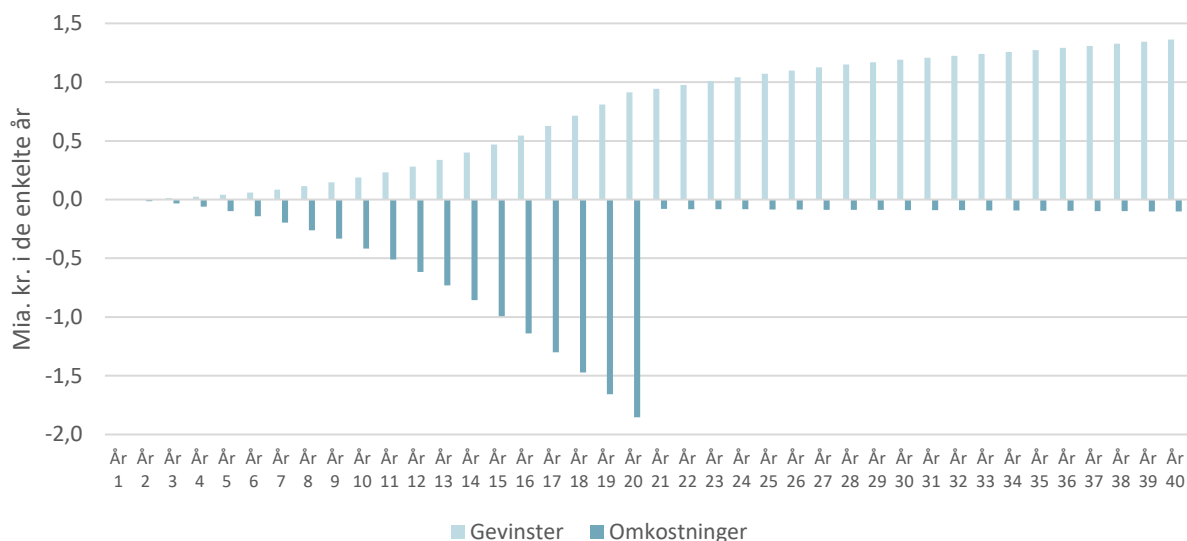
Note: Den lave nettonutidsværdi er, hvis 50 pct. af alle forskningsdata i Danmark gøres FAIR, den høje værdi er for 100 pct. Afkastratioen er beregnet som nettonutidsværdien divideret med den tilbagediskonterede værdi af opstartsomkostningerne. Tallene er afrundet. Kilde: Egne beregninger.

Givet en implementeringsperiode på 20 år og en levetid på FAIR data-løsningen på 20 år vil der være større omkostninger end gevinster ved FAIR data i de første 20 år. Derefter overstiger gevinsterne dog omkostningerne markant, jf. figur 2.3.

¹³ Beagrie (2012)

¹⁴ Antagelsen bygger på, at man i litteraturen fremhæver, at gevinsterne ved åbne data stiger eksponentielt, grundet de netværks- eller systemeffekter, der er forbundet med data – data bliver mere værd, jo flere data, der er tilgængelige. På den anden side for omkostningerne kan man antage, at man ved implementering af FAIR data først gør de data FAIR, som er nemmest at gøre FAIR.

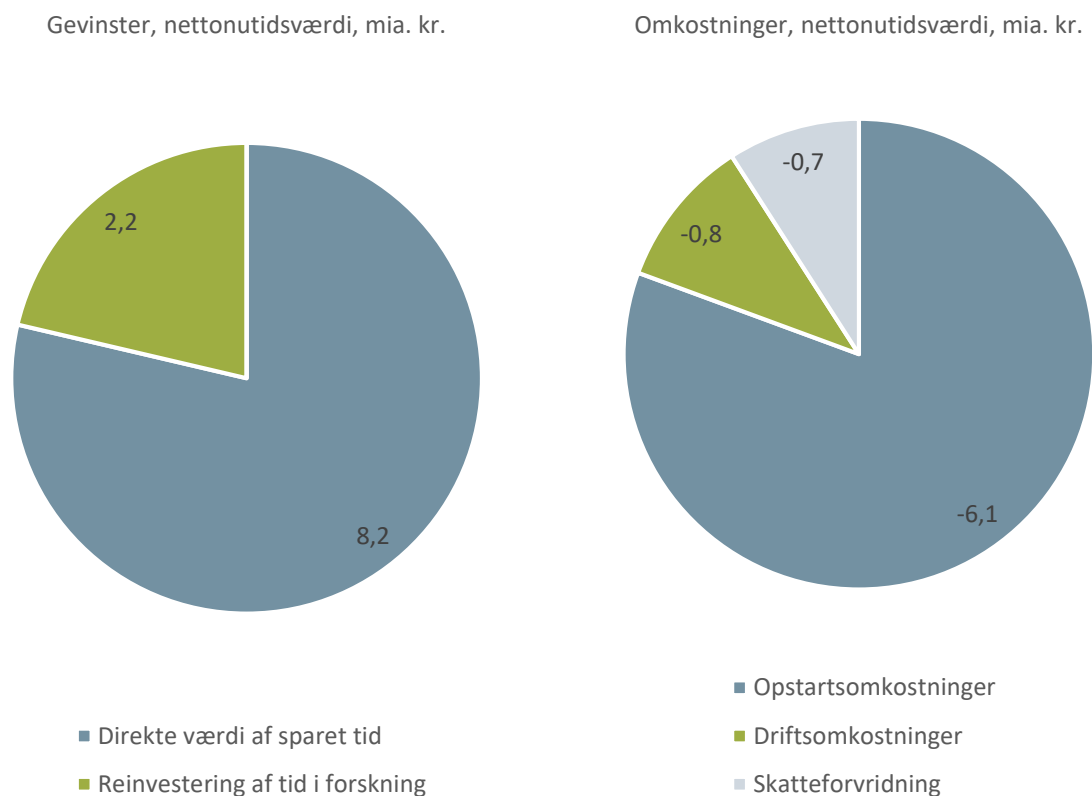
Figur 2.3. Gevinster og omkostninger over tid



Note: Det er antaget, at 75 pct. af alle forskningsdata i Danmark gøres FAIR. Bemærk, at værdierne ikke er tilbagediskonteret.
 Kilde: Egne beregninger.

Den klart største gevinst ved at indføre FAIR data i Danmark er den direkte værdi af den tid, forskere sparer, jf. Figur 2.4. I vores grundscenarie antager vi, at forskerne kan spare ca. 1/3 af deres dataarbejdstid. Det yderligere afkast, der opnås ved at reinvestere den sparede tid i ny forskning, udgør en mindre andel af gevinsterne. De klart største omkostninger er opstartsomkostningerne, som udgør ca. 4/5 af omkostningerne.

Figur 2.4. Fordeling af gevinster og omkostninger



Note: Det er antaget, at 75 pct. af alle forskningsdata i Danmark gøres FAIR. Tallene er afrundet.
Kilde: Egne beregninger.

2.3 FØLSOMHEDSBEREGNINGER

Vores beregninger er som nævnt baseret på en række forudsætninger, som er forbundet med nogen usikkerhed. Vi vurderer dog, at resultatet – at en indførelse af FAIR data i Danmark giver anledning til en samfundsøkonomisk gevinst – er ganske robust. Det skyldes, at dette resultat ikke ændrer sig, hvis der foretages rimelige ændringer i de mest usikre beregningsforudsætninger.

Det er især FAIR data-løsningens levetid og den tid, forskere sparer i dataarbejde, som er afgørende for, at FAIR data kan skabe værdi i Danmark. Hvis løsningen har en kortere levetid end 10 år, vil det være forbundet med et samfundsøkonomisk tab at indføre FAIR i Danmark. Det er således vigtigt at skabe en langtidsholdbar løsning. Hvis tidsbesparelsen er mindre end 23 pct. af den tid, forskere bruger på dataarbejde, så vil indførelsen af FAIR data ligeledes medføre et samfundsøkonomisk tab. Det er således vigtigt at etablere en løsning, som kan effektivisere forskeres arbejdstid mest muligt.

I det følgende præsenterer vi en række illustrative beregninger af konsekvenserne af at ændre nogle af de mest kritiske forudsætninger i analysen. Overordnet vedrører forudsætningerne fem områder: 1) tidshorisont, 2)

værdi af forskningsdata, 3) sparet tid for forskere, 4) reinvestering af tid og ny forskning skabt og 5) omkostninger. Der præsenteres følsomhedsberegninger for hvert område.

2.3.1 Tidshorisont

Figur 2.5. Tidshorisontparametre

- FAIR data implementeres over en periode på 20 år (kilde: interviews).
- FAIR data-infrastruktur har en levetid på 20 år (kilde: interviews).
- Data har en levetid på 10 år (kilde: Houghton & Gruen (2014) m.fl.).
- Data deprecieres med 10 pct. p.a. (kilde: Houghton & Gruen (2014) m.fl.).
- Gevinster og omkostninger realiseres efter en konveks funktion i implementeringsperioden, dvs. at de største gevinster og omkostninger realiseres til sidst (kilde: Houghton & Gruen (2014), m.fl.).

Note: ● Høj validitet ● Middel validitet ● Lav validitet. For en fuld kildeoversigt se kapitel 6.

Der er teoretisk gode argumenter for, at gevinsterne ved at indføre FAIR data vil blive realiseret efter en konveks funktion. Gevinsterne akkumuleres og stiger mere og mere, jo flere data, der er FAIR. Ligeledes kan man argumentere for, at man først vil gøre de data FAIR, som er nemmest at gøre FAIR. Dermed vil omkostningerne først være små, men vil stige, eftersom det bliver mere besværligt at gøre de sidste data FAIR.

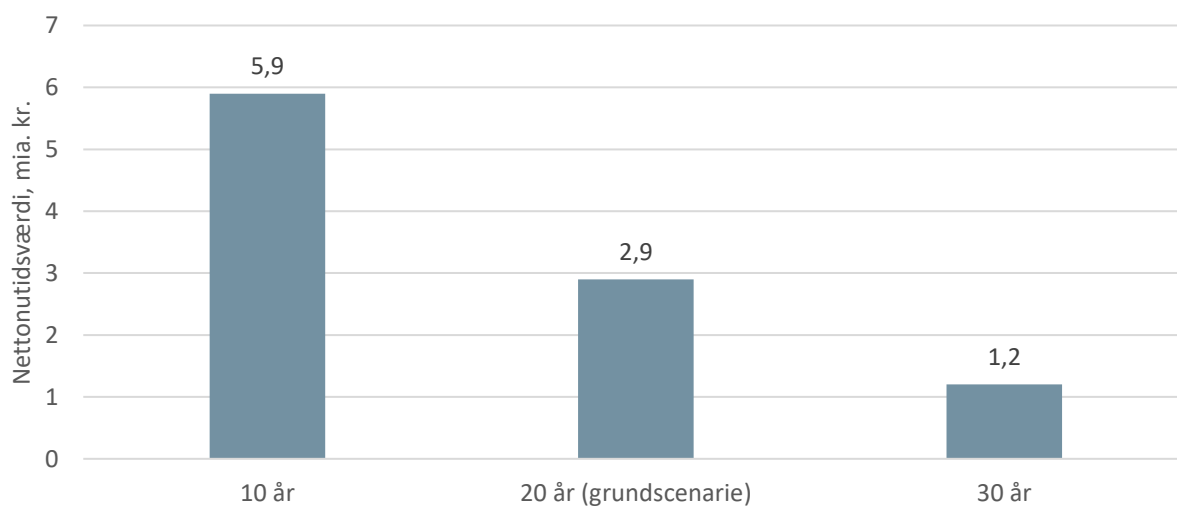
Der findes desuden litteratur, som analyserer offentlige forskningsdatas levetid og depreciering. I forhold til litteraturen har vi lagt os på et relativt højt deprecieringsniveau. Der er lidt større usikkerhed omkring datas levetid, da data i dag bliver genereret meget hurtigere og i meget større mængder end de tidsperioder, der er set på i litteraturen.

De eksisterende analyser af FAIR forskningsdata har kun set på gevinster og omkostninger i et enkelt år og typisk efter, at FAIR data er fuldt implementeret. I vores analyser ser vi derimod på gevinster og omkostninger over tid og før og efter, at FAIR data er fuldt implementeret. Det betyder, at der ikke er noget grundlag fra litteraturen, vi kan basere vores forudsætninger om implementeringsperioden og infrastrukturens levetid på. Vores antagelse om en implementeringsperiode på 20 år er baseret på de tyske nøgleinteressenters tidlige erfaringer og vurderinger af den tyske implementeringsproces samt de umiddelbare vurderinger fra de danske nøgleinteressenter. Konklusionen fra vores interview i Tyskland er, at en fuld implementering af FAIR vil være en lang proces der foregår i flere forskellige tempi, da der er forskel på delstaternes forskningsprioriteter. Derfor er det vanskeligt at estimere et antal år i Tyskland, men som flere peger på, er det en implementeringsproces hvor forskerne og forskningsmiljøerne langsomt skal modnes til at vænne sig til at følge principperne. Herudover er der store generationsforskelle i forskernes holdning til håndtering og deling af forskningsdata, hvorfor

det vurderes at implementeringen først for alvor vil tage fart når de yngre forskere erstatter de ældre. Derfor er en implementeringsperiode på 20-30 år ikke urealistisk.

Vores overordnede vurdering er, at man vil kunne implementere FAIR hurtigere i Danmark end i Tyskland. Derfor har vi valgt at sætte den på 20 år. Det hænger bl.a. sammen med, at det danske forskningsmiljø og politiske system overordnet set er mindre komplekst end det tyske, hvilket trækker i retning af, at implementeringsperioden kan blive kortere i Danmark end i Tyskland. Vores følsomhedsanalyser viser dog, at selv med en implementeringsperiode på 30 år vil der fortsat være en positiv samfundsøkonomisk værdi ved at indføre FAIR data i Danmark, jf. figur 2.6.

Figur 2.6: Betydning af implementeringsperioden



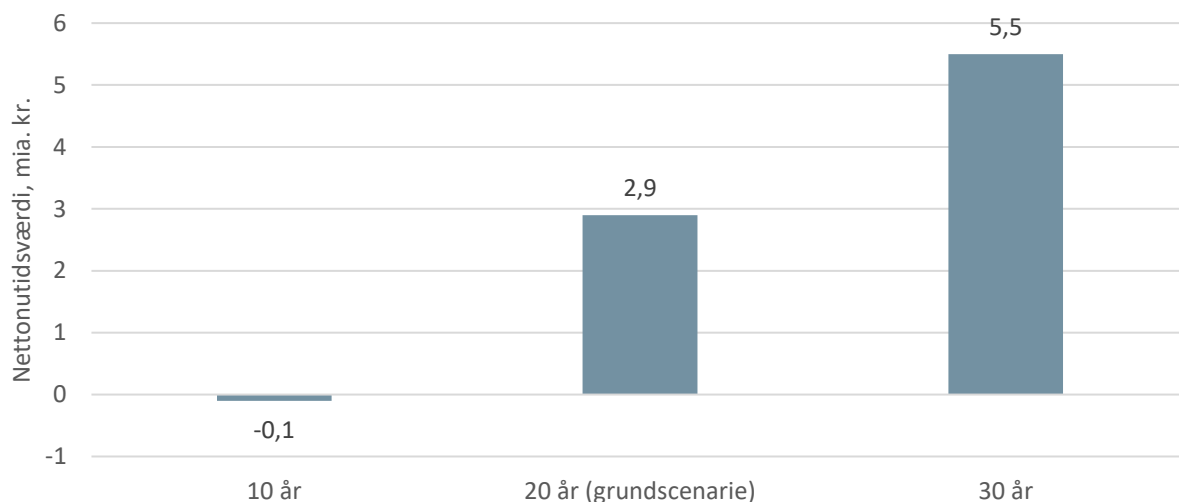
Note: Det er antaget, at 75 pct. af alle forskningsdata i Danmark gøres FAIR.

Kilde: Egne beregninger.

I beregningerne er det antaget, at infrastrukturen har en levetid, som er mindst lige så lang som implementeringsperioden på 20 år. Der er væsentlig usikkerhed forbundet med denne antagelse. Det er sandsynligt, at dele af infrastrukturen, fx metadata eller brugerflade, skal fornyes hurtigere, mens fx retningslinjer, love og knowhow kan have en meget længere levetid. Da vores analyse inkluderer relativt høje driftsomkostninger, anses det for realistisk at antage, at driftsomkostningerne også inkluderer vedligehold af infrastrukturen, herunder løbende opdateringer og evt. udskiftning af fx metadata og brugerflade. Derfor synes en levetid på 20 år at være realistisk. Vores følsomhedsberegninger viser, at den samfundsøkonomiske værdi af FAIR data er positiv, til og med en levetid for infrastrukturen på 11 år, jf.

figur 2.7.

Figur 2.7. Betydning af infrastrukturens levetid



Note: Det er antaget, at 75 pct. af alle forskningsdata i Danmark gøres FAIR.

Kilde: Egne beregninger.

2.3.2 Værdien af forskningsdata

Figur 2.8. Parametre til estimering af værdien af forskningsdata i Danmark

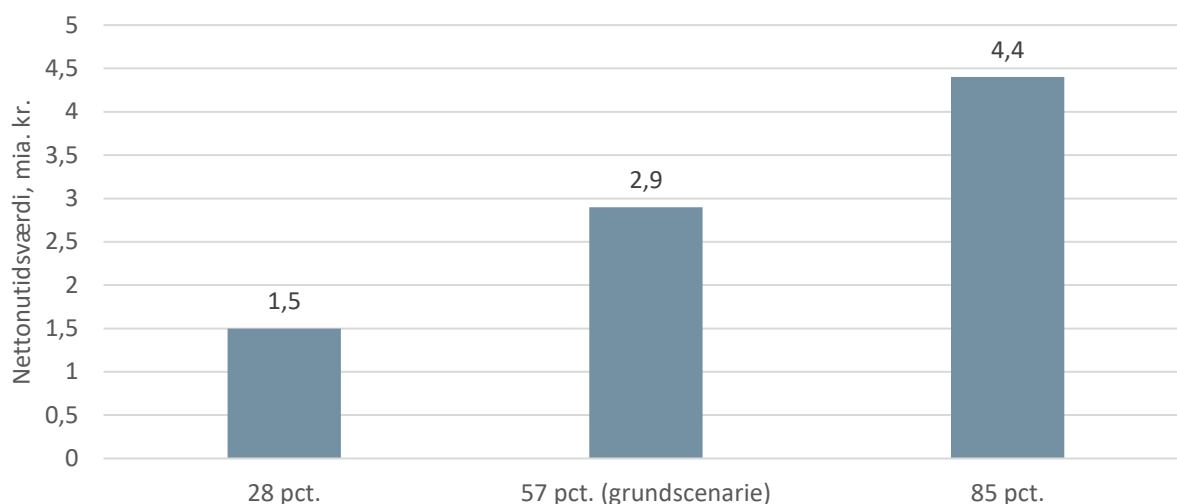
- Andel af forskeres tid anvendt på forskning og viden (kilde: DEA (2015)).
- Andel af forskningstid anvendt på data (kilde: Beagrie & Houghton (2016), m.fl.).

Note: ● Høj validitet ● Middel validitet ● Lav validitet. For en fuld kildeoversigt se kapitel 6.

Forskere anvender deres tid på en række andre aktiviteter end forskning og videnskabelse. DEA lavede i 2015 et studie blandt danske forskere, som fandt, at de brugte ca. 48 pct. af deres tid på forskning. Da dette studie er relativt nyt og for danske forskere, anser vi validiteten for høj. Der findes ingen empiri for danske forskeres tidsforbrug på dataarbejde. Vi har derfor baseret vores forudsætninger om forskeres tidsforbrug på dataarbejde på den internationale litteratur. I litteraturen er oplysningerne typisk indhentet gennem spørgeskemaer, hvor man har spurgt forskere, der bruger datacentre. Derfor vurderes estimerne fra den eksisterende litteratur at være overkantsskøn i forhold til den tid, den gennemsnitlige forsker bruger på dataarbejde. I vores beregninger

varierer både gevinster og omkostninger med den tid, forskere bruger på dataarbejde. Når forskerne bruger mere tid på dataarbejde, skaber de flere data. Dette vil både lede til større gevinster ved at indføre FAIR data og til større omkostninger. I vores beregninger antages det, at både gevinster og omkostninger udgør en fast andel af værdien af forskningsdata i et givent år. Således vil gevinster og omkostninger stige relativt set lige meget, hvis forskere bruger mere eller mindre tid på dataarbejde. Med andre ord vil omkostningerne aldrig komme til at overstige gevinsterne, uanset hvor meget tid forskerne bruger på dataarbejde. Antagelsen påvirker dog størrelsen af den samfundsøkonomiske værdi ved at gøre data FAIR, jf. figur 2.9.

Figur 2.9. Betydning af andelen af forskningstid anvendt på data



Note: Det er antaget, at 75 pct. af alle forskningsdata i Danmark gøres FAIR. Figuren viser nettonutidsværdien ved 50 pct. lavere hhv. højere andel af tid, forskerne bruger på dataarbejde, end der er anvendt i grundscenariet.
Kilde: Egne beregninger.

2.3.3 Sparede tid for forskere

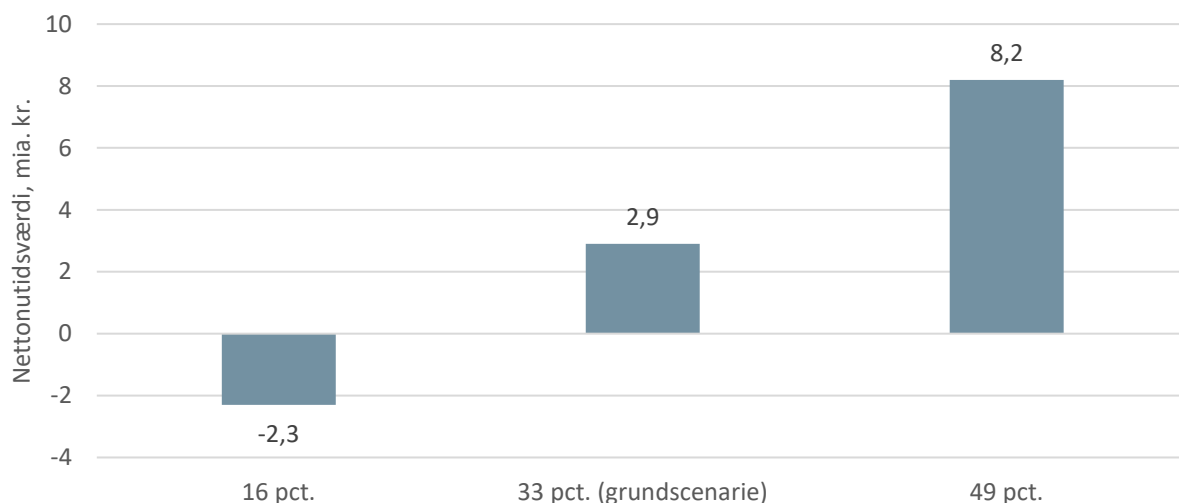
Figur 2.10. Parametre til estimering af værdien af sparet tid for forskere

- Effektivitetsgevinst, andel af dataarbejdstid, der kan spares (kilde: Sullivan et al., m.fl.).

Note: ● Høj validitet ● Middel validitet ● Lav validitet. For en fuld kildeoversigt se kapitel 6.

Andelen af dataarbejdstiden, der kan spares, når data gøres FAIR, er opgjort i flere internationale studier. Studierne bygger på spørgeskemaer blandt brugerne af FAIR datacentre. Derfor betragtes estimerne som overkantsskøn i forhold til den tid, den gennemsnitlige forsker kan spare, når data gøres FAIR. Vi har derfor valgt at anvende de laveste estimater for tidsbesparelsen, som er fundet i litteraturen, og i vores grundscenarie anvender vi dermed en antagelse om, at forskere kan spare ca. 1/3 af deres tid, når data gøres FAIR. I litteraturen findes estimater på op til 46 pct. Vores følsomhedsberegninger viser, at der er en positiv samfundsøkonomisk værdi ved at indføre FAIR i Danmark, hvis tidsbesparelsen/effektivitetsgevinsten er 23 pct. eller større. Med andre ord er resultatet relativt følsomt over for usikkerheden forbundet med effektivitetsgevinsten.

Figur 2.11. Betydning af andel af dataarbejdstid, der kan spares



Note: Det antages, at 75 pct. af alle forskningsdata i Danmark gøres FAIR. Figuren viser nettonutidsværdien ved 50 pct. lavere hhv. højere tidsbesparelse, end der er anvendt i grundscenariet.
 Kilde: Egne beregninger.

2.3.4 Reinvestering af tid og ny forskning skabt

Figur 2.12. Parametre til estimering af værdien af ny forskning

- Afkast af ny offentlig forskning (kilde: DEA (2017), Danske Universiteter (2016)).
- Kun en andel af afkastet realiseres i Danmark (kilde: Houghton & Gruen (2014) m.fl.).
- Der skabes ingen ny forskning (konservativ antagelse).

Note: ● Høj validitet ● Middel validitet ● Lav validitet. For en fuld kildeoversigt se kapitel 6.

Der er en omfattende litteratur om afkastet af forskning og udvikling. I Danmark er der de seneste år lavet flere litteraturstudier, som vurderer, at afkastet ligger i størrelsesordenen 20-40 pct.¹⁵ En stor del af litteraturen fokuserer dog på private investeringer i forskning og udvikling. Vi har valgt at lægge os på 20 pct., som er i den nedre ende af intervallet. Da der er en vis usikkerhed omkring afkastet af offentlig forskning og udvikling, har estimatet dog ikke en høj validitet. Der findes ligeledes flere studier, der peger på, at afkastet af forskning realiseres både lokalt og internationalt. Der findes ingen studier af dette for danske data, og vi har derfor valgt at basere vores beregninger på et estimat i den lave ende af det interval, som de internationale studier har fundet.

Værdien af reinvestering af tid udgør en relativt lille andel af gevinsterne i grundscenariet. Selv uden noget ekstra afkast af den tid, forskerne sparer ved, at data bliver FAIR, vil der være en positiv samfundsøkonomisk værdi ved at indføre FAIR i Danmark.

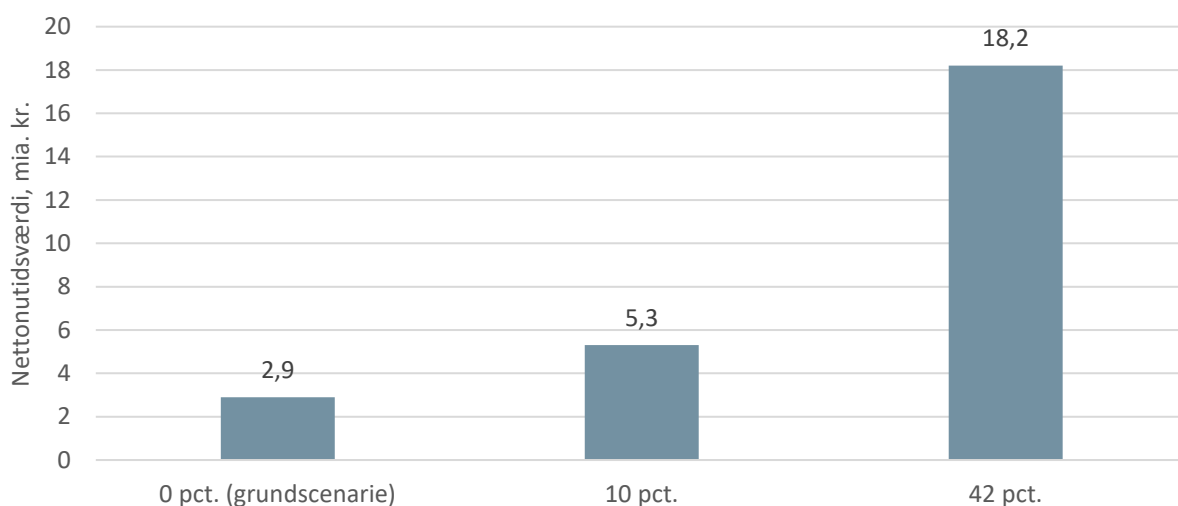
I vores grundscenarie har vi som nævnt set bort fra værdien af den nye forskning, som skabes, når data gøres FAIR. Dette skyldes primært usikkerhed om, i hvor høj grad ny forskning blot vil fortrænge eksisterende forskning. Der kan dog være grund til at tro, at den nye forskning vil have en positiv samfundsøkonomisk værdi. Det kan bl.a. skyldes, at forskere fortsat vil udføre meget af den forskning, de også udførte, før data blev FAIR (svarende til, at deres samlede forskningsaktivitet stiger), men det kan også være, fordi afkastet af den nye forskning, som bliver skabt af FAIR, er højere end afkastet af den forskning, forskerne tidligere udførte. Hvis man antager, at forskere er rationelle, vil de således vælge den forskning, der giver det største afkast.

I Houghton & Gruen (2014) antages det, at ny forskning ikke fortrænger eksisterende forskning. Således anvender de den andel af forskningen, som ikke ville eller kunne skabes uden FAIR forskningsdata, til at estimere værdien af den nye forskning. Andelen er baseret på international litteratur og spørgeskemaer blandt brugerne af FAIR datacentre. Denne tilgang medfører, at indførelsen af FAIR forskningsdata leder til en markant stigning i de samlede forskningsaktiviteter. Hvis vi anvender den samme tilgang som i Houghton & Gruen (2014) og baserer beregninger på litteraturens laveste estimater over, hvor stor en andel af forskningen, der ikke ville eller kunne skabes uden FAIR data (42 pct.), vil den samfundsøkonomiske gevinst ved at indføre FAIR data i Danmark være på ca. 18 mia. kr. Dette eksempel er et klart overkantsskøn og illustrerer først og fremmest det mulige potentiale ved at indføre FAIR data i Danmark.

Hvis der anvendes en mere realistisk antagelse om, at der skabes 10 pct. ny forskning, eller at den nye forskning giver ca. 10 pct. større afkast, bliver den samfundsøkonomiske værdi 5,3 mia. kr., jf. figur 2.13.

Figur 2.13. Betydning af andelen af ny forskning skabt som følge af FAIR data

¹⁵ Kilder: DEA (2017), Danske Universiteter (2016)



Note: Det er antaget, at 75 pct. af alle forskningsdata i Danmark gøres FAIR.
 Kilde: Egne beregninger.

2.3.5 Omkostninger

Figur 2.14: Omkostningsparametre

- Opstartsomkostninger udgør i gennemsnit 24 pct. af den årlige dataværdi i implementeringsperioden (kilde: Beagrie et al. (2010)).
- Driftsomkostninger udgør 3 pct. af den årlige dataværdi (kilde: Beagrie et al. (2010) m.fl.).

Note: ● Høj validitet ● Middel validitet ● Lav validitet. For en fuld kildeoversigt se kapitel 6.

Der findes få studier i litteraturen, som beskæftiger sig med de omkostninger, der er forbundet med at indføre FAIR data eller andre åbne dataløsninger. De fokuserer i høj grad på enkeltstående initiativer og antager typisk, at beslutningstagerne i vid udstrækning selv kan bestemme omfanget af deres investering. Der er dermed stor usikkerhed forbundet med at opgøre omkostningerne forbundet med at indføre FAIR data i Danmark. Driftsomkostningerne er mindre usikre, fordi der er en række eksisterende initiativer, som opgør disse. I vores beregninger har vi valgt at sætte driftsomkostninger til 3 pct. af de samlede lønomkostninger til data i de enkelte år,¹⁶ hvilket er markant højere end i Houghton & Gruen (2014), hvor der kun anvendes 1,5 pct.

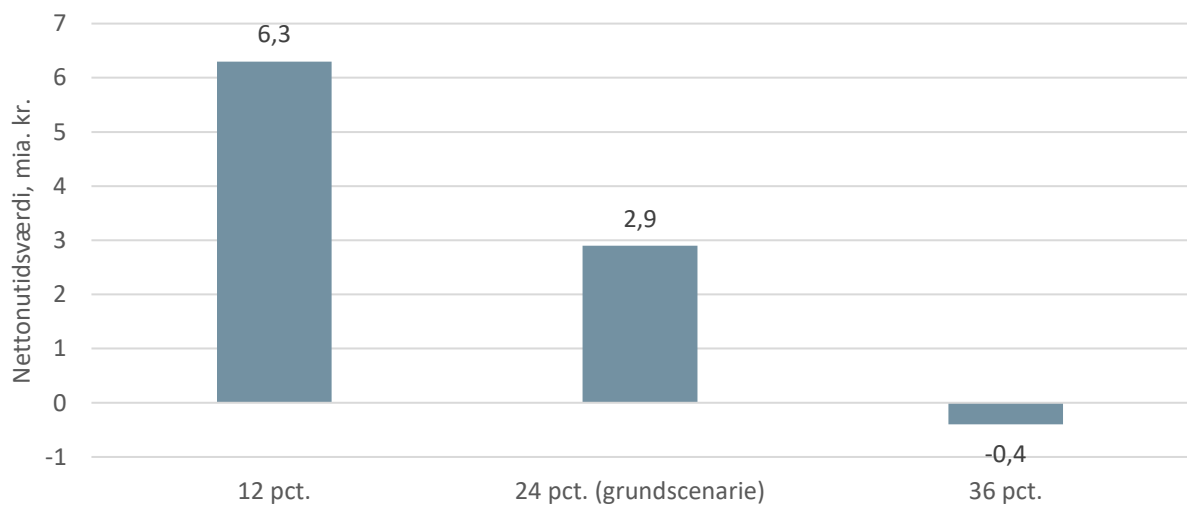
Opstartsomkostningerne er modelleret på baggrund af en enkelt case fra Oxford University, hvor de udgjorde 24 pct. af den samlede dataværdi. Eftersom perioden og projektet adskiller sig fra en eventuel indførelse af FAIR data i Danmark, er der stor usikkerhed omkring validiteten af dette estimat. Det vurderes dog, at estimatet er et øvre estimat, jf. kapitel 5, på baggrund af de afholdte interviews og grundet de potentielle besparelser som

¹⁶ Disse omkostninger estimeres ud fra den samme metode som værdien af data. Det vil sige den tid, forskere bruger på at generere data, og lønomkostningerne til dette.

ny teknologi fx i form af Artificial Intelligence kan give sammenlignet med projektet i Oxford. Vores resultater er ganske følsomme over for opstartsomkostninger, jf.

figur 2.15. Dog er der stadig en positiv samfundsøkonomisk værdi for opstartsomkostninger; op til 35 pct. i gennemsnit over perioden.

Figur 2.15. Betydning af opstartsomkostninger



Note: Det er antaget, at 75 pct. af alle forskningsdata i Danmark gøres FAIR. Figuren viser nettonutidsværdien ved 50 pct. lavere hhv. højere andel, end der er anvendt i grundscenariet.

Kilde: Egne beregninger.

3. Muligheder og barrierer for at indføre FAIR data i Danmark

Formålet med dette kapitel er at analysere de barrierer og muligheder, der er for at implementere FAIR data i en dansk kontekst, samt at undersøge, hvilke politiske, organisatoriske, tekniske og kulturelle tiltag der kan igangsættes for at udbrede FAIR data i Danmark.

3.1 VURDERING AF FAIR-PRINCIPPERNE

Som beskrevet bygger FAIR på følgende fire principper – data skal være Findable, Accessible, Interoperable og Reusable. De fire principper er tæt relaterede, men teknisk uafhængige af hinanden. De kan implementeres i forskellige kombinationer og i forskellige grader eller trin. En fuld implementering af de fire principper i FAIR er nærmere et ideal i et kontinuum, hvor data kan gøres mere eller mindre FAIR¹⁷. At der er tale om principper – ikke fastlagte krav – som kan efterleves i forskellige grader, understreges også af formanden for DM Forum.

”Der er altså flere grader af FAIRness, og institutionen DANS i Holland arbejder p.t. på at udvikle en ranking til de enkelte principper” (Anders Sparre Conrad, formand for DM Forum).

Vores interviews med de danske nøgleinteressenter, herunder universiteterne, forskningsbibliotekerne og de politiske myndigheder, viser, at der i Danmark er en generel opbakning til FAIR-principperne. Flere fremhæver, at der er tale om fornuftige principper, som det er svært at erklære sig uenig i. Dog er det vanskeligt for de fleste repræsentanter fra forskningsinstitutionerne at finde ud af, hvad der helt konkret ligger under de enkelte principper. For at kunne tage reel stilling til de enkelte principper er det nødvendigt at gøre dem mere praksisnære og formulere nogle praktiske implikationer af disse, hvilket også er den bagvedliggende idé med FAIR.

Selvom det er svært at tage konkret stilling til de enkelte principper, vil de følgende afsnit alligevel opsummere nogle af de centrale vurderingsmæssige pointer – dels ud fra de gennemførte interviews i Danmark, dels fra erfaringerne i Tyskland.

3.1.1 Findability

For at opfylde det første FAIR-princip skal data kunne findes. Det kræver gode metadata, der beskriver data. Det er samtidig vigtigt, at maskiner skal kunne finde data, da data typisk er spredt på tværs af grænser – både geografisk og fagligt – og i kraft af at datamængderne i nogle tilfælde er for store til, at mennesker kan håndtere

¹⁷ Mons, Barens, Cameron Neylon, Michel Dumontier, Luiz Olavo Bonino Sila Santos & Mark D. Wilkinson (2017) Cloudy, Increasingly FAIR – Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud.

dem. For at indfri dette princip skal data tildeles en unik og konsistent identifikator, også kaldet PID'er (Persistent Identifiers), og lagres i et repositorium med et ekstra lag af informationer om organisering, strukturering og metadatering mv.

På tværs af de danske interessenter på området er vurderingen, at data i et vist omfang allerede er genfindelige, og princippet om findability er relativt let, teknisk set, at indfri. Dog viser vores analyse også, at 14 af de 20¹⁸ internationale og nationale data-repositorier med dansk deltagelse ikke bruger en PID. Det viser, at størstedelen af den datalagring, danske forskere bruger, er langt fra at opfylde kriteriet om at være Findable (genfindelig) og derfor langt fra at kvalificere som FAIR data.

Der er allerede eksempler på danske lokale repositorier, herunder ERDA¹⁹, Dataverse²⁰, samt nationale repositorier, såsom LOAR²¹ ved det Kgl. Bibliotek, og de danske knudepunkter i internationale repositorier som CLARIN-DK²². Globalt er der også flere generiske og disciplinspecifikke repositorier, fx ELIXIR²³ og CESSDA²⁴. Dog er udfordringen at have tilknyttet en effektiv søgemaskine, som løbende skal vedligeholdes, og at sikre, at de rette kompetencer og incitamenter til registrering, herunder metadatering og deling af store mængder digitale forskningsdatasæt, er til stede hos forskerne. Dette er helt centrale udfordringer for indførelsen af FAIR data, hvilket uddybes senere i nærværende kapitel.

3.1.2 Accessibility

For at opfylde det andet FAIR-princip skal data kunne tilgås, idet de er lagret med henblik på langtidsoptbevaring. Der skal tillige være veldefinerede licensregler, og det skal være let at hente data. Vigtigt er det desuden at pointere, at *accessible*-kriteriet i FAIR ikke er ensbetydende med, at alle former for data skal være åbne og tilgængelige for offentligheden. Data, som af forskellige årsager skønnes at kompromittere personfølsomme oplysninger, national sikkerhed, konkurrencevilkår eller andre tilsvarende vigtige hensyn, er omfattet af særlig lovgivning, som sætter særlige rammer for datas tilgængelighed, og skal kun kunne tilgås efter nærmere aftale/licens. Dette kommer helt konkret til udtryk i Horizon 2020's underliggende guideline om FAIR data management "*as open as possible, as closed as necessary*".²⁵ Det er dog vigtigt at understrege, at selvom data ikke er accessible, må de gerne være findable, forstået på den måde, at man kan finde metadata om materialet, men af forskellige årsager ikke altid selve grundmaterialet.

¹⁸ Re3data.org

¹⁹ ERDA er et lager-, delings- og arkiveringssystem leveret af KU til ansatte og studerende. I første omgang er det begrænset til Det Natur- og Biovidenskabelige Fakultet (Science), men på længere sigt vil det muligvis blive udbredt til hele KU.

²⁰ Det Kongelige Bibliotek og Københavns Universitetsbibliotek og Informationsservice har oprettet KUBIS Dataverse Network, der er et digitalt arkiv for primære forskningsdata.

²¹ LOAR (Library Open Access Repository) er et åbent data-repositorium, som blev etableret i 2016 som en service til at opbevare og få adgang til danske forskningsdata.

²² CLARIN-DK udgør det danske bidrag til en europæisk forskningsinfrastruktur for humaniora. CLARIN-DK råder over en platform, hvor forskere kan få adgang til de sprogbase materialer og værktøjer, som efterhånden deponeres af forskere fra forskellige universiteter og institutioner.

²³ ELIXIR er den europæiske forskningsinfrastruktur for bioinformatik.

²⁴ CESSDA står for Consortium for Social Science Data Archives og giver adgang til samfundsvidenskabelige data på tværs af lande.

²⁵ European Commission (2016), H2020 Programme – Guidelines on FAIR Data Management in Horizon 2020.

Universiteterne og de øvrige forskningsinstitutioner er generelt meget bevidste om princippet *accessibility*. Det forventes, at krav i blandt andet den nye persondataforordning vil betyde, at flere datasæt i fremtiden skal sikkerheds-certificeres for at kunne indgå i forskningssamarbejdet og dermed også vil kræve flere ressourcer til at indfri princippet om, at data skal kunne tilgås. Ligeledes vil mulighederne og barriererne for *adgangen* til relevante forskningsdata som et element i et dansk økosystem for digitale forskningsdata uddybes senere i nærværende kapitel.

3.1.3 Interoperability

For at opfylde det tredje FAIR-princip om *interoperability* er det afgørende, at der er fælles guidelines til at beskrive metadata, og at der anvendes et åbent og veldefineret vokabularium. Det muliggør, at data kan bruges på tværs af forskningsfelter, og at data kan udveksles på tværs af systemer. Full interoperabilitet betyder, at data kan findes, tilgås og udnyttes på tværs af separate fysiske data-repositorier og er uafhængige af en bestemt fysisk og teknisk platform.

I Danmark er der p.t. ikke initiativer, der indfrier det tredje FAIR-princip om interoperability, men i Tyskland har man bl.a. med GO-FAIR-initiativet brudt den hidtidige praksis om en række uafhængige repositorier og forsøger at koordinere mere på tværs for at undgå løsninger, der ikke kan kombineres. GO-BUILD inden for GO-FAIR arbejder fx med behovet for datasæt, der bruger harmoniserede standarder, og som gør det muligt for alle forskere at få adgang til data på tværs af fagspecifikke repositorier.

Interoperabilitet anses af flere interviewpersoner for at være en af de største udfordringer ved at indfri FAIR. I Tyskland har man valgt ikke at satse på en samlet dataløsning, men i stedet på at fremme koordineringen af de allerede etablerede tyske data-repositorier samt et øget internationalt samarbejde, således at de anvendte tekniske standarder for opbevaring af forskningsdata ikke kun gælder i Tyskland, men internationalt, bl.a. igennem RDA (Research Data Alliance) og EUDAT (Research Data Services, Expertise & Technology Solutions).

Ligesom Tyskland deltager Danmark i bl.a. RDA og EUDAT. Den danske repræsentant i RDA forklarer, at princippet om interoperable forskningsdata er et af de sværeste principper at indfri, fordi det kræver, at data kan tale sammen på tværs af forskningsfelter og kan finde frem til et fælles sprog. Dog foregår der nogle internationale projekter i regi af EUDAT, der tester interoperabilitet på metadata-niveau.

”Der er mange format-udfordringer med at gøre data kompatible. Lige nu tester vi det på metadata-niveau i regi af EUDAT. Langt ude i fremtiden kan man måske gøre dem kompatible på rådata-niveau, men det er meget vanskeligt, fordi det er lag på lag-informationer” (Lene Krøl Andersen, Head of eScience Center, DeIC).

3.1.4 Reusability

For at opfylde det fjerde og sidste FAIR-princip skal data kunne genbruges i fremtidige forskningsprojekter og viderebehandles. En af de centrale forudsætninger for, at data kan genbruges i andre forskningsprojekter, er, at data frigøres med en klar og tilgængelig datalicens.

For at kunne genbruge data er der en række juridiske udfordringer, herunder bl.a. at definere brugsrettigheder omkring forskningsdata. Disse rettigheder skal klart beskrives, da tvetydigheder i ophavsret kan begrænse genbrugen af data. Jo mere automatisk processen bliver med at genfinde lagrede data, desto vigtigere er det samtidig at have en klar og tilgængelig datalicens.

På tværs af de gennemførte interviews anses reusability-princippet således også for at være et af de sværeste principper at indfri. Som formanden for DEFF eksempelvis forklarer, skyldes det, at forskeres datadeling ikke meriteres i dag.

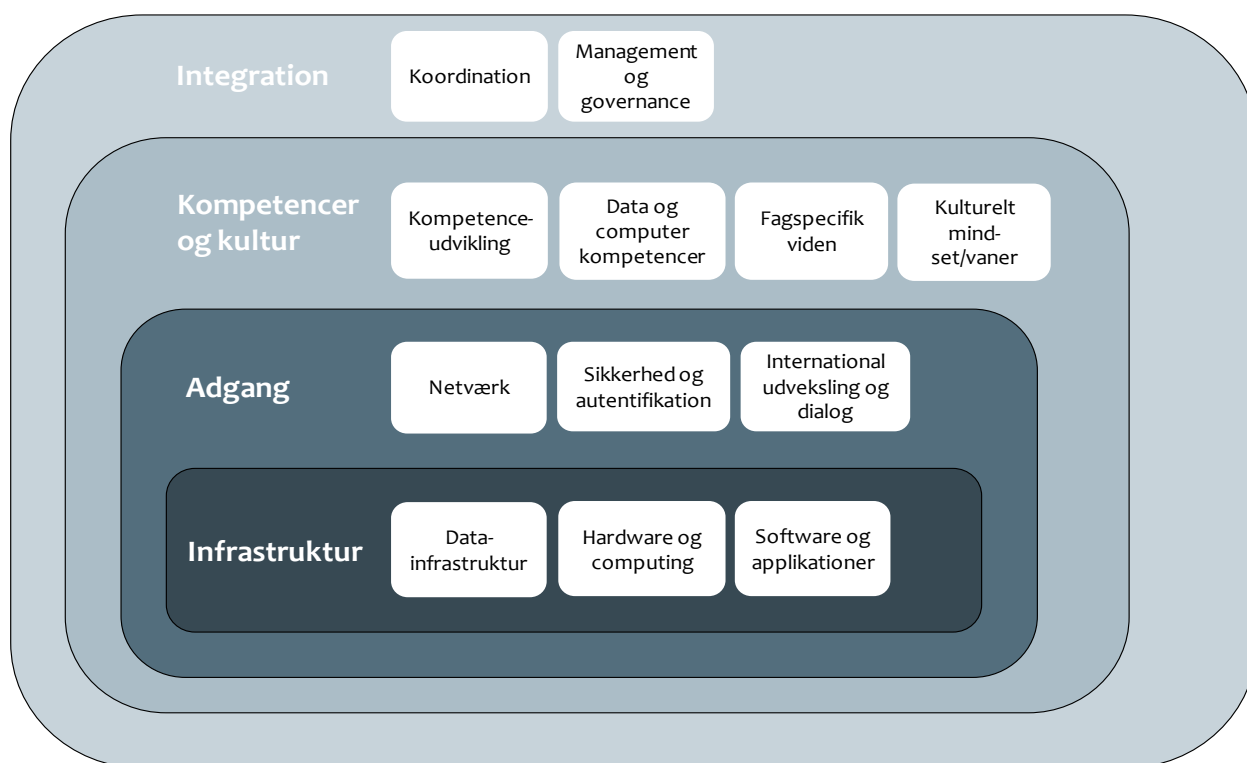
”Det sidste princip om genanvendelighed er vi lidt tvivlsomme overfor. Det kan blive ganske vanskeligt at indfri, da forskerne ikke får merit for deling af data. Det er nødvendigt at kigge på incitamentsstrukturerne” (Børge Obel, professor og centerdirektør, Aarhus Universitet, formand for DEFF).

3.2 ELEMENTERNE I ØKOSYSTEMET

Det danske økosystem for digitale forskningsdata er komplekst og består af flere forskellige elementer, der skal spille sammen. Jf. nedenstående figur 3.1 er hovedelementerne i økosystemet *integration, kompetence og kultur, adgang* samt *infrastruktur*.

Under hvert element er der interessenter og aktører med forskellige krav, interesser og behov. En succesfuld implementering af FAIR data i Danmark afhænger af mange forhold. En helt central pointe er, at den internationale kobling og integration er afgørende, da en isoleret, national FAIR dataløsning ikke giver mening. Ydermere er det en central pointe, at mange af interessenterne kan se fordelene ved FAIR-principperne. Eksempelvis kan mange af forskerne se, at de på længere sigt kan styrke kvaliteten af deres forskning og spare ressourcer i dataindsamling ved at dele og genbruge hinandens data og fortolkninger. At forskningsbibliotekarerne kan være med i forskningsprocesserne og systematisere de digitale data. At de professionelle dataudbydere kan tilbyde tjenester og levere dataanalyse og behandlingstjenester til at understøtte god digital forskningspraksis, og at den offentlige forskningsfinansiering, som i stigende grad berøres af en international dagsorden om 'Open Science', gennem krav til langsigtet datastyring og deling med FAIR-principperne kan være med til at understøtte bedre forskning og nye opdagelser.

Figur 3.1 Elementerne i det danske økosystem for digitale forskningsdata



Kilde: Oxford Research med inspiration fra RCUK e-Infrastructure (2014): E-Infrastructure Roadmap

Grundlæggende er opgaven i økosystemet at afbalancere elementerne og skabe sammenhænge, så indsatsene er koordinerede og ikke står alene. Konkret betyder det eksempelvis, at et fokus på videreuddannelse af forskere til at håndtere og dele digitale forskningsdata skal ses i sammenhæng med, om der etableres tilstrækkelige data-støttefunktioner på universiteterne, og om der er en tilstrækkelig datakapacitet og lagringsplads til at opbevare data.

Integrationen af elementerne i økosystemet skal netop sikre denne sammenhæng i systemet og kræver bl.a. en central, national koordination, der kan skabe et ensartet og integreret system på tværs af miljøer og aktører med øje for de aktiviteter, der sker lokalt, nationalt og internationalt.

Kompetencer og kultur er et andet hovedelement i økosystemet. Pointen er, at selvom de tekniske løsninger og infrastrukturen er på plads, vil potentialerne ved FAIR data ikke kunne høstes uden de rette kompetencer og uden et kulturelt mindset, der understøtter brugen af FAIR data. For det første kræver en korrekt håndtering og deling af digitale forskningsdata, at forskerne har de nødvendige eScience-kompetencer. Disse skal suppleres og understøttes af data stewards, eScience-eksperter og forskningsbibliotekarer, der i højere grad skal indgå som 'medskabere' i forskningsprocessen. For det andet er der den kulturelle dimension, der kræver, at forskerne

og institutionerne omkring dem skal kunne se værdien af forskningsdata og ikke kun i de publicerede forskningsartikler.

Adgang er et tredje element. Det kræver et velfungerende forskningsnet, samt at forskerne har adgang til relevante forskningsdata via en velkendt og simpel proces. Det handler først og fremmest om at få styr på de rent tekniske netværksløsninger og på at forbinde databaser og forskningsnet. Endvidere ligger der et stort koordineringsarbejde i, at der inden for de enkelte forskningsfelter koordineres og videnedveksles på tværs af lande omkring udvikling af metastandarder, dokumentationskrav og datahåndtering.

Infrastrukturen er det sidste element og kan ses som rygraden i økosystemet, idet det muliggør lagring, håndtering, behandling og formidling af digitale forskningsdata. Infrastruktur kan både ligge lokalt på de enkelte institutioner, nationalt fx i nationale dataarkiver og globalt i internationale datacentre. Igen er pointen, at infrastruktur ikke kan stå alene, idet investeringer i og etablering af de nødvendige specialiserede infrastrukturer blot muliggør lagring, håndtering, behandling og formidling af forskningsdata, men at der kræves adfærdsændringer blandt forskerne, for at forskningsdata rent faktisk deles.

3.2.1 Integration

For at integrere elementerne i det digitale økosystem kræves en klar politisk prioritering og central indsats med fokus på FAIR-principperne. Derfor er det vigtigt med en samlende aktør, der kan sætte de overordnede mål og rammer, fx ved at udvikle en strategi på området, og også stå for den tværgående koordination og være med til at skabe enighed i systemet blandt aktørerne. Det er helt centralt, at forskere, universiteter og øvrige forskningsinstitutioner, forskningsråd, fonde og biblioteker arbejder mod samme mål i forhold til eksempelvis data management-politikker, kurser i data management, uddannelse af data stewards og incitament- og meriteringsstrukturer. Integration skal dermed sikre, at FAIR data implementeres på den mest hensigtsmæssige og ressourcebesparende måde.

3.2.1.1 En fælles, koordineret og national indsats

I Danmark er der p.t. ingen national strategi, politik eller fælles retningslinjer, der understøtter FAIR-principperne og omsætter principperne i praksis. I Danmark sker koordineringen med afsæt i de aktiviteter, DeIC, DEFF og DM FORUM igangsætter. DeIC og DEFF har sammen udarbejdet en National Strategi for Forskningsdata Management 2015-18. Strategien er bredt forankret og har været i høring blandt hovedinteressenterne i Danmark. Formålet med strategien er at understøtte en fælles dansk retning for opsamling, sikring, formidling og genbrug af forskningsdata. En anbefaling i strategien er at etablere et nationalt forum for samarbejde om forskningsdata management, der med økonomisk støtte fra DEFF skal løfte de centrale strategiske anbefalinger med fokus på politikker, incitamentsskabelse, infrastruktur, kompetenceudvikling og forskerstøtte samt styringsstruktur. På baggrund af anbefalingerne i strategien er DM Forum teknisk set etableret som et treårigt pilotprojekt og skal være det danske omdrejningspunkt for den praktiske erfaringsudveksling og samarbejde om udvikling af kompetencer og infrastrukturer, ligesom den skal koordinere den danske indsats i forbindelse med internationale aktiviteter. DM Forum understøtter p.t. forskningsdata management generelt og ikke specifikt FAIR. Dog understreger DM Forum, at man er åbne over for FAIR-principperne og gerne vil være med til at koordinere og understøtte implementeringen af dem.

Arbejdet med åbne data er dog stadig relativt ukoordineret, idet de enkelte institutioner, herunder universiteter og forskningsbiblioteker, hver især formulerer politikker og nedsætter grupper, der arbejder konkret med forskningsdata management. Eksempelvis har DTU, som er længst fremme med FAIR data i Danmark, i 2015 oprettet en data management-gruppe (DM-gruppen), som forskerne kan konsultere om datahåndtering. Samtidig er DM-gruppen i gang med at udvikle politikker og praksisser, der kan indgå i DTU's data management-arbejde. Indtrykket fra interviewene er, at universiteterne ikke har et samlet overblik over, hvad der sker i Danmark på området. Dog er der eksempler på samarbejder mellem universiteterne og på, at de udveksler viden og inspiration. SDU's 'Open Science'-politik, der er under udvikling, er bl.a. inspireret af DTU's politik på området, og på SDU bygger man i øjeblikket videre på en nrastruktur, man er gået i udbrud med på AAU. Aftalen er, at det, man bygger på SDU, bygger man også på AAU.

Tyskland, som officielt har valgt at satse på FAIR-principperne, har igangsat GO-FAIR-initiativet som omdrejningspunkt for at implementere FAIR i en tysk kontekst. De foreløbige tyske erfaringer viser, at den ministerielle opbakning og øremærkede finansiering til GO-FAIR-initiativet har været en væsentlig driver for at sætte FAIR data på dagsordenen i Tyskland. Formålet med GO-FAIR-initiativet er at skabe en bred forankring blandt de centrale interessenter samt sætte fokus på potentialerne ved FAIR gennem konkrete best practice-eksempler, der illustrerer, hvordan FAIR-principperne kan operationaliseres, bl.a. gennem projekter som GeRDI²⁶ og RADAR²⁷.

3.2.1.2 Koblingen mellem det lokale, nationale og internationale

En central pointe er, at en isoleret, national løsning for FAIR data ikke giver mening. For at forskerne og institutionerne kan se en interesse i at bruge FAIR-principperne og ændre deres praksis, er den internationale kobling og integration afgørende. Det er der flere grunde til. Først og fremmest kan det begrundes med, at forskningens natur i sit udgangspunkt er international. Det betyder, at forskerne skal kunne dele og genbruge data med 'peers' internationalt, og at det er utilstrækkeligt, hvis forskere kun kan dele og genbruge data nationalt i Danmark.

²⁶ GeRDI (Generic Research Data Infrastructure) har som sit langsigtede formål at gøre det muligt for alle tyske forskere at opbevare, dele og genbruge forskningsdata på tværs af forskningsområder. I den første fase af projektet vil tre datacentre blive sat sammen. Det skal danne pilotcasen for, at det kan rulles ud i Tyskland og være en model for den fremtidige tyske forskningsdatainfrastruktur.

²⁷ RADAR (Research Data Repository) handler om at samle data i et datacenter, som de forskellige medlemmer så har adgang til. Projektet er tværfagligt, og projektets deltagere er fra universiteter i forskellige tyske delstater. Projektet er organiseret således, at der på hver institution er en lokal koordinator, som så er ansvarlig for at understøtte de lokale forskere på deres institution i datadeling og -håndtering. Projektet understøtter 'the long tail' inden for forskningsdata, som er de data, som falder uden for big data-domænet – altså først og fremmest små datasæt. I relation til FAIR er vurderingen, at projektet har opnået kriterierne Findability og Accessibility, men er langt fra at indfri Interoperability og Reusability.

”Udgangspunktet for en løsning er international. Implementeringen af FAIR må tage sit udgangspunkt i forskerne, som arbejder internationalt. Forskerne skal inden for deres egne forskningsdiscipliner udveksle viden, praksisser og erfaringer, som så kan understøttes nationalt” (Henrik Pedersen, dekan, AAU).

For det andet er det vigtigt at følge med internationalt og koordinere den danske indsats i forbindelse med internationale aktiviteter som garanti for, at der ikke udvikles lokale og nationale løsninger, hvis der fx allerede er internationale løsninger, der kan bringes i anvendelse. I øjeblikket arbejdes der på, at den danske indsats er i overensstemmelse med arbejdet i centrale internationale samarbejdsfora som Research Data Alliance (RDA), Knowledge Exchange, e-Infrastructure Reflection Group (e-IRG), European Data Infrastructure (EUDAT) og Nordic e-Infrastructure Collaboration (NeIC). Som supplement for at sikre en kobling mellem de lokale/nationale og internationale initiativer kan GO-FAIR-initiativet ses som en mulig platform for internationalt samarbejde. Ifølge en af initiativtagerne til GO-FAIR, Barend Mons, er de ved at etablere arbejdsgrupper med fokus på implementering af FAIR-principperne.

”I would recommend that Denmark join the GO-FAIR implementation networks which are multi-topical and user case driven – we will for example jointly develop requirement for good data management. Both institutions, funding agencies and the Ministry can join the implementation networks” (Barend Mons, prof., Leiden University).

3.2.1.3 Bæredygtig finansiering

Implementering af FAIR-principperne kræver investeringer i etablering af nye strukturer, processer og systemer. Herudover vil der være en lang transitions- og driftsfase for at få operationaliseret og integreret principperne i forskningspraksis. Der skal investeres i en lang række områder, herunder at skabe de institutionelle rammer for FAIR, infrastrukturinvesteringer, kommunikations- og koordineringsprocesser i de videnskabelige miljøer, reorganisering af eksisterende services, træning og videreuddannelse af forskere.

Generelt er de danske interessenter positive over for de potentialer, der kan høstes ved at implementere FAIR data i Danmark, men flere er usikre på finansieringsmodellen, og hvorvidt omkostningerne skal lægges hos forskeren eller i forskningsinstitutionen.

”Finansieringen er uklar. Omkostningerne forbundet med at vedligeholde data er store. Min erfaring fra fondsverdenen er, at vi modtager rigtig mange ansøgninger om midler til datavedligeholdelse. Men hvornår er det rent faktisk værd at vedligeholde data? Og skaber dette barrierer for generering af nye data?” (Jens Oddershede, professor og formand, Danmarks Forsknings- og Innovationspolitiske Råd (DFiR)).

I Tyskland peger flere af interessenterne, herunder Dr. Stefan Winkler-Nees fra Deutsche Forschungsgemeinschaft, på, at en hovedudfordring i Tyskland er den bæredygtige finansiering. Mange af Open Science-projekterne finansieres med projektmidler, og derfor er der en risiko for, at de strander, når projektfinsieringen udløber. Der er derfor behov for mere langsigtet institutionel finansiering, men også i at tænke i nye forretningsmodeller. Bl.a. har man i det tyske RADAR-projekt udviklet en forretningsmodel, hvor forskere og forskningsinstitutioner skal betale for at deltage.

3.2.1.4 Krav til politikker og praksisser omkring forskningsdata

Data Management Planer (DMP'er) er et centralt værktøj til at fremme og styrke ensretningen af datahåndtering og -deling. En DMP er et dokument, der struktureret beskriver, hvordan forskeren har tænkt sig at håndtere sine forskningsdata med hensyn til genfindning, deling, etiske overvejelser og ikke mindst lagring af data mm. Der er tale om et simpelt tekstdokument, typisk i PDF-format, og det er forskeren selv, der bestemmer, hvad der skal stå.

De tyske erfaringer samt kravene i EU's Horizon 2020 om, at forskere, der modtager støtte fra programmet, skal aflevere en DMP sammen med deres ansøgning, viser, at der er kommet et større fokus blandt forskerne og i institutionernes politikker og praksisser. Deutsche Forschungsgemeinschaft (DFG) i Tyskland stiller krav om, at der udarbejdes en DMP som led i ansøgningsprocessen, og på flere af de danske universiteter vurderer ledelsen, at EU's forskningsprogram Horizon 2020's krav om, at forskere, der modtager støtte fra programmet, skal aflevere en DMP sammen med deres ansøgning, har betydet, at der er kommet et større fokus blandt forskerne og i institutionernes politikker og praksisser. Som det påpeges i flere af de gennemførte interviews, er kravene til DMP'er dog ikke ensbetydende med, at data fra finansierede forskningsprojekter rent faktisk lagres i repositorier, der gør data tilgængelige for andre forskningsprojekter.

Flere understreger, at det er en konkret udfordring i Danmark, at der ikke stilles krav til udarbejdelsen af DMP'er på institutionerne. DM Forum har en overordnet vision om, at Danmark i juni 2018 har et nationalt sammenhængende sæt af DM-politikker hos universiteterne, som 1) bakkes op af forskningsråd, ministerier og bevaringsinstitutioner og er åben for andre forskningsinstitutioner, 2) sikrer, at forskeren får hjælp med at planlægge og effektuere data management for hele datas livscyklus, 3) er forankret såvel lokalt som internationalt og 4) sikrer modeller for finansiering og fordeling af udgifterne til DM mellem interessenterne. Endvidere blev der i 2014 udviklet en Danish Code of Conduct for Research Integrity, som er en national adfærdskodeks for integritet i dansk forskning. Tilbagemeldingerne fra universiteterne er, at den har været med til at sætte gang i overvejelser og tanker omkring forskningsdata management, og at deres politikker er udviklet med inspiration fra denne adfærdskodeks. Samtidig viser interviewene med flere universiteter og forskere, at den endnu ikke har ansporet til konkrete ændringer i deres forskningspraksis, da det vil kræve en bredere indsats på flere områder, herunder en ændring i forskernes incitamentsstruktur, kompetenceudvikling af forskerne og flere investeringer i infrastruktur etc.

3.2.2 Kompetencer og kultur

Kompetencer og kultur er nødvendige forudsætninger for at realisere FAIR data i Danmark. Da det er vigtigt, at forskerne bruger principperne i deres arbejde og er medspillere i en operationalisering, må indførelsen af FAIR tage udgangspunkt i forskerne selv.

For det første spiller kompetencer en helt væsentlig rolle, fordi en korrekt håndtering og deling af digitale forskningsdata kræver, at forskerne har de nødvendige eScience-kompetencer. Disse skal suppleres og understøttes af yderligere specialiserede data stewards, eScience-eksperter og forskningsbibliotekarer, der i højere grad skal indgå som 'medskabere' i forskningsprocessen og have viden om data management med indsigt i forskningsdomænet. Herudover kræves yderligere supplerende datakompetencer til at understøtte udviklingen og vedligeholdelsen af datainfrastruktur.

Det andet er den kulturelle dimension, idet både forskerne og institutionerne omkring dem må ændre deres mindset og kunne se værdien i forskningsdata og ikke kun i publikationerne efterfølgende. Som Barend Mons forudser, udgiver forskere i dag artikler med et appendiks med data, men i fremtiden bliver data den primære udgivelse, hvortil man så kan udgive en supplerende artikel²⁸. For at understøtte den kulturændring kræves der flere indsatser, hvoraf den nok mest centrale er et opgør med meriterings- og incitamentsstrukturen, som i dag er centreret omkring publicering af videnskabelige artikler. En ændring af den eksisterende meriteringspraksis bør i så fald ske i en international kontekst i kraft af forskningens internationale natur.

3.2.2.1 Udvikling af forskernes datakompetencer og datastøtte

Den nuværende situation i Danmark viser, at der ikke er en systematisk tilgang til kompetenceudvikling og træning af forskere med henblik på at give dem de nødvendige værktøjer til at håndtere og arbejde med FAIR data. Forskernes datakompetencer har ikke været på dagsordenen på de danske universiteter, og gennemgående i interviewene vurderes det, at forskernes kompetencer inden for data management ikke er tilstrækkelige. Da flere af universiteterne ikke har haft en koordineret indsats på området, har ansvaret ligget på fakultets- og institutniveau og ofte hos den enkelte forsker. Det har været problematisk, idet forskerne sjældent har den fornødne viden og indsigt. Derfor bliver data ofte gemt lokalt og ikke delt og håndteret på en ensartet måde.

"Forskerne er over én kam stort set uuddannede i data management. Der er dog enkelte universiteter, som eksempelvis KU og AU, der er begyndt at udbyde kurser i det" (Bjarne Andersen, vicedirektør for IT-udvikling og Infrastruktur, Det Kgl. Bibliotek).

Et andet indsatsområde, der er vigtigt at understøtte, er etablering og udbygning af datastøttefunktioner på universiteterne. Formålet med oprettelsen af disse er, at forskerne kan modtage rådgivning i brug af infrastrukturer og informationsteknologiske værktøjer, metadatering og informationsorganisatoriske problematikker samt lovgivning og juridiske problematikker.

²⁸ DeIC (20/12/2016): Data vil afløse den videnskabelige artikel, link: https://www.deic.dk/da/news-2016-12-20_Barend_Mons

”Jeg frygter, at når forskeren oplever, hvor meget tid og hvor mange ressourcer det tager at rense og dokumentere data, så stopper det. De vil spørge sig selv: What’s in it for me? Derfor skal de have data stewards, der kan hjælpe dem” (Nikolaj Helm-Petersen, souschef, Danske Universiteter).

P.t. er der på de danske universiteter en opmærksomhed rettet mod videreudviklingen af støttefunktioner, men ingen større, konkrete investeringer. Ifølge Barend Mons, der er formand for EOSC High Level Expert Group, har de i ekspertgruppen estimeret, at der er behov for ca. én data steward per 20 forskere, hvis potentialet for EOSC skal udnyttes fuldt ud.

Endvidere er der brug for en tættere kobling mellem forskningsbiblioteker, dataarkiver og universiteterne, og i den sammenhæng kan en ny bibliotekaruddannelse, der kombinerer den traditionelle bibliotekaruddannelse med datalogi, være en løsning. John Renner, dekan på KU Science, mener, at der er brug for en ny bibliotekaruddannelse eller en data steward-uddannelse, som ligger i krydsfeltet mellem en traditionel bibliotekar og en datalog. En analyse igangsat af DEFF omkring forskningsdata og Open Access konkluderer, at selv en meget erfaren bibliotekar, der kender til samarbejde med forskere, systemer, forskningsregistrering, metadata, it osv., har yderst svært ved at sætte sig ind i et forskningsdataområde, fordi det er et fremmed stof, og hver case har sit helt eget indhold og formål og sin helt egen baggrund og dermed håndtering²⁹.

3.2.2.2 Tid og ressourcer

Grundlæggende tager det tid og ressourcer for forskerne at skulle omstille deres hidtidige forskningspraksis til et større fokus på datadeling og -håndtering. Det skyldes, at datadeling medfører høje krav til formatering, dokumentation og vedligeholdelse, hvilket så tager tid og ressourcer fra noget andet.

”Jeg tror, at udgiften er størst hos de menneskelige ressourcer for at understøtte hele datas livscyklus og dokumentationsdelen. Det er driftsomkostninger, der er relativt usynlige, fordi de kan blive en del af ens normale arbejde” (Børge Obel, professor og centerdirektør, DEFF).

Det er en grundlæggende udfordring, at det nuværende meriteringssystem ikke tilgodeser forskere, som bruger ressourcer på at indsamle og vedligeholde data, og at deres tid og omkostninger ikke anerkendes. At lægge større vægt på deling og brug af forskningsdata i meriteringen vil givetvis være den mest effektive måde at ændre ved forskernes adfærd. I alle interviews med de danske universiteter er det en gennemgående pointe, at dette system kræver en justering, hvis det skal ændre i forskernes tilgang til forskningsdata. Det samme fremgår af de tyske erfaringer, hvor interviewpersonerne også påpeger den hæmmende effekt af dette system. Det er en international udfordring, som vil kræve internationalt koordinerede løsninger og samarbejde på tværs og først vil blive løst på langt sigt.

²⁹ DTU på vegne af DEFF (2011): Forskningsdata og Open Access

3.2.2.3 Forskernes mindset

Forskere erfarer med datadeling, deres datadelingspraksisser og deres motivation for at dele data varierer afhængigt af forskningsdisciplin. Det konkluderer Knowledge Exchange i sin rapport om forskeres incitamenter og motivation for datadeling³⁰, og der er forskningsdiscipliner såsom genomik og krystallografi, hvor datadeling er så integreret, at det er blevet en helt standardiseret forskningspraksis. Af interviewene fremgår det, at der i mindre grad er en tradition for datadeling inden for humanistiske og samfundsvidenskabelige discipliner, mens det generelt er en mere udbredt del af de faglige miljøer under den naturvidenskabelige og tekniske forskning. På DTU opleves det eksempelvis, at forskerne generelt er positivt indstillede over for potentialerne ved Open Science.

"When we talk to the scientists they say that it is a promising idea, but it involves a lot of changes in culture and requires a strong commitment to use extra time on documentation of data" (fokusgruppe-interview med Forskningsdatagruppen, DTU's bibliotek).

Flere peger også på, at forskernes alder har en betydning for deres tilbøjelighed til at dele data, idet mange af de yngre forskere er mere vant til at dele information og data digitalt og derfor også er med til at skubbe grænserne.

"Der er særligt blandt de yngre forskere en stor interesse. Den er vigtig at bygge videre på" (Peter Munk Christiansen, professor og institutleder for Statskundskab, AU).

Nu og her kan et konkret tiltag, der tilskynder forskerne til at være mere aktive omkring deling og genbrug af forskningsdata være at etablere mere on-site-støtte. Som beskrevet ovenfor er der i Danmark behov for at etablere og udbygge datastøttefunktionerne på universiteterne, men også udviklingen af DMP'er kan udstikke klare retningslinjer for forskerne i deres arbejde med forskningsdata. Endvidere kan en forbedret dataadgang, herunder at få skabt nemme og brugervenlige løsninger til søgning i databaser, tilskynde forskerne til at blive mere aktive med hensyn til deling og genbrug af forskningsdata. For at videndele og sikre en international sammenhæng i de krav, der stilles til datahåndtering, kan udvikling af DMP'erne ske i regi af GO-CHANGE, der omhandler kulturen i de videnskabelige miljøer samt forskernes meriterings- og incitamentssystem.

3.2.3 Adgang

Adgang til digitale forskningsdata er naturligvis en forudsætning for at imødekomme FAIR-principperne. Adgang til data kan gradbøjes, og som Barend Mons et al. understreger, er der grader af FAIRness. FAIR er et

³⁰ Knowledge Exchange (2014): Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective

kontinuum, hvor data som minimum skal være Findable (genfindelige)³¹. Herudover dækker Accessible (tilgængelige) over, at det for hvert datasæt kan vurderes, hvem der kan tilgå data. Det kan eksempelvis være relevant i forhold til medicinsk forskning, hvor der er et hensyn til persondata. På den anden side kan det besluttes, at data er tilgængelige for alle og altså er helt åbne.

Erfaringerne fra Tyskland viser, at der er brug for at præcisere principperne mere konkret under implementeringen, og at der i praksis er grader af FAIR. Eksempelvis har man i RADAR vurderet projektet i henhold til FAIR-principperne, og her er det tydeligt, at det er nemmere at imødekomme Findable (genfindelige) og Accessible (tilgængelige), mens det har vist sig svært at opnå Interoperable (kompatible) og Reusable (genanvendelige).

3.2.3.1 Krav og overvejelser om de tekniske løsninger

En forudsætning for, at forskerne har adgang til forskningsdata, er et velfungerende forskningsnet. I Danmark er der Forskningsnettet, der forbinder de danske universiteter og forskningsinstitutioner. Nettet har omkring 100 institutioner tilsluttet og er desuden internationalt linket op på NORDUnet og GÉANT. Forbindelsen til de internationale forskningsnet er nødvendig, for at de danske forskere kan have mulighed for at deltage i internationale forskningssamarbejder. DeIC er ansvarlig for Forskningsnettet i Danmark og har i sine anbefalinger til en sammenhængende e-infrastruktur til fremme af dansk forskning en ambition om, at det danske forskningsnet skal være 'state-of-the-art' på området.

Herudover er der i forhold til adgang udviklingen af de konkrete politikker og procedurer med henblik på datasikkerhed, brugerauthentifikation (at det er den rigtige bruger) og autorisation (hvad en bruger kan tilgå af data). Dette kompliceres yderligere af persondataforordningen, som skal implementeres i maj 2018.

"Der kan være forskellige typer af omkostninger. Fx glemmer man ofte sikkerhedskostningerne. Kan man ændre data med cyber crime? Hvordan håndterer man persondata? Hvis data ikke er open, hvordan sikrer man så, at en lukket gruppe af forskere kan få fat i dem og ikke alle?" (Børge Obel, professor og centerleder, DEFF).

Overvejelserne om adgangen til data er en central del i FAIR-diskussionen, fordi de hænger sammen med datas troværdighed og kvaliteten af data. Derfor er det for det første vigtigt, at de tekniske løsninger understøtter, at det er de rette kompetente forskere og fageksperter, der har adgang til data, og for det andet vigtigt at tage stilling til, hvorvidt data har en kvalitet og forskningsmæssig interesse, der legitimerer, at de opbevares og dermed også kan blive genanvendt.

³¹ Mons, Barend et al. (2017): Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Science Cloud

”Det skal ikke altid være nemt at gøre data tilgængelige. Man skal ikke bare kunne gemme data, hvis ikke de er godkendt til at blive genanvendt. En af de store udfordringer er derfor at kvalificere genanvendelsen” (Lene Krøl Andersen, Head of eScience Center, DeIC).

3.2.3.2 Forskningspraksis og ansvarsfordeling mellem forsker og institution

Det er den enkelte forskers forskningspraksis og forvaltning af data, der afgør, om andre overhovedet får adgang til de data, forskeren genererer. Det påpeges i flere interviews, at en hæmsko er, at forskerne ikke er oplyste om gældende juridiske regler om fortrolighed, privatliv eller IPR og derfor ofte vælger ikke at dele deres forskningsdata.

Som en kortlægning af forvaltning af forskningsdata i Danmark også har vist, sidder forskerne ofte alene med ansvaret, fordi der ikke er de institutionelle rammer til at understøtte dem³². Forskeren skal således selv sikre, at forskningsdata håndteres i overensstemmelse med bedste praksis inden for det faglige felt, herunder gøre sig sine egne overvejelser om planlægning af håndteringen af de primære materialer og data samt om opbevaring af forskningsdata, datas lagringstid, bortskaffelse af data, adgang til og deling af data.

”Man skal være sikker på, at data er korrekte. Hvem skal validere de data, der ligger tilgængelige? Al forskning er international. Derfor skal vi følge internationale tekniske standarder, for at man kan dele forskningsdata i et større internationalt netværk” (Børge Obel, professor og centerleder, DEFF).

Forskerne kan så understøttes mere eller mindre af institutionernes rammer og politikker, og efterhånden har flere danske universiteter fokus på det. Eksempelvis har DTU udformet en data management-politik, hvor DTU skal stå for at tilvejebringe et lagringssystem, der giver forskere mulighed for at forvalte deres data ansvarligt. Det indbefatter opbevaring, adgangskontrol og sikkerhedskopiering. Endvidere står det beskrevet, at DTU 1) skal give mulighed for uddannelse og efteruddannelse, 2) rådgive om praksis, juridiske problemstillinger og infrastrukturer og 3) udvikle og vedligeholde almindelige systemer og infrastrukturer til forskningsdata management.

3.2.4 Infrastruktur

Infrastrukturen er grundlaget for, at digitale forskningsdata kan lagres, håndteres og behandles. Infrastruktur består af de lokale løsninger på universiteterne, nationalt i fx Rigsarkivet eller Dansk Data Arkiv samt globalt i internationale data-repositorier. Herudover dækker infrastrukturen over den nødvendige software, som fx kan understøtte samarbejdsværktøjer og avancerede beregningsprogrammer til databehandling, samt beregningshardware, der bruges til modellering, simulering, dataanalyse og visualisering.

³² Thestrup B., Jesper et al. (2013): Forvaltning af forskningsdata i Danmark – En undersøgelse af danske universiteters og forskningsråds praksisser, politik og visioner for lagring, langtidbevaring, tilgængeliggørelse og deling af forskningsdata

3.2.4.1 Behov for flere investeringer i etableringen af mere datalagring og kapacitet

Interviewene med de danske universiteter tegner et klart billede af, at der på universiteterne ikke er et fuldt overblik over den danske lagringskapacitet. Heller ikke internt på universiteterne er der nødvendigvis et overblik, fordi datalagring og -håndtering behandles decentralt på fakultets- og institutniveau. Nogle institutter har lagerplads, de stiller til rådighed, andre har ikke. Det netop afsluttede projekt 'Data management i praksis' har vist, at datakapaciteten mangler i Danmark, og at der langtfra er en systematisk tilgang til at lagre data.

"Vi oplevede nogle gange, at de havde opbevaret data på USB-harddiske. Andre steder oplevede vi, at der slet ikke var styr på det. Vi oplevede andre steder, at der var lagring, men at de langtfra understøttede FAIR-principperne. Det hjalp vi dem så med i projektet 'Data management i praksis'. En hovedkonklusion fra det projekt er, at man skal finde måder at gøre forskere mere selvhjulpne ift. metadata-ring" (Bjarne Andersen, vicedirektør for it-udvikling, Det Kgl. Bibliotek).

Selvom der ikke er et fuldt overblik, vurderer en række af de interviewede eksperter, at der i Danmark p.t. ikke er tilstrækkelig datainfrastruktur til at kunne realisere FAIR. Databasen re3data.org viser, at danske forskere er med både nationalt og internationalt i 20 forskellige data-repositorier. 14 ud af 20 bruger ikke en PID (Persistent Identifier System), hvilket i udgangspunktet viser, at en meget stor del af den datalagring, danske forskere bruger, er langt fra at opfylde kriteriet om at være Findable (genfindelig) og derfor langt fra at kvalificere som FAIR data.

"Vi har i varierende grad byggekodserne på plads i Danmark. Vi har i høj grad 'identifiers' på plads, som støtter op omkring findability, men vi mangler den rette infrastruktur – der halter vi bagefter" (Anders Sparre Conrad, formand, DM Forum).

På grund af det manglende overblik er der flere, som påpeger, at løsningen også må findes i samarbejdet mellem universiteterne. DM Forum er en platform, der i den sammenhæng bruges til at sikre koordination på tværs og bidrage til, at der udvikles løsninger i samarbejde.

I forhold til GO-FAIR-initiativet, som bygger på GO-CHANGE (omhandler kulturen i de videnskabelige miljøer samt forskernes meriterings- og incitamentssystem), GO-TRAIN (omhandler kompetenceudvikling) og GO-BUILD (omhandler den tekniske implementering), tyder det på, at der i Danmark også bør fokuseres på GO-BUILD-delen. På det punkt er man længere fremme i Tyskland, idet de interviewede interessenter vurderer, at infrastrukturen til lagring og opbevaring af forskningsdata er opbygget i Tyskland.

3.2.4.2 Fokus på langsigtet og bæredygtig finansiering til drift af datacentre og opbevaring af data

I DeiC's anbefalinger til at sikre en sammenhængende dansk e-infrastruktur for perioden 2018-2028 vurderes de nuværende pilotprojekter på nationalt niveau som utilstrækkelige til at sikre en bæredygtig infrastruktur. Derfor anbefales det, at der afsættes et beløb i en størrelsesorden, som matcher landene omkring os, og som skal dække udvikling, investering og drift af de nuværende anlæg³³. Der er med andre ord behov for en mere langsigtet finansiering til udvikling og drift af infrastrukturen.

I Tyskland er et fokus på en langsigtet finansieringsmodel en central udfordring. En stor del af de tyske investeringer er finansieret med projektmidler, hvor drifts- og vedligeholdelsesudgifterne ikke er indregnet. Derfor vurderer de tyske interessenter, at der er behov for en mere langsigtet, institutionel finansiering og/eller tænkning i nye forretningsmodeller, herunder bl.a. RADAR-projektet, hvor forskere og forskningsinstitutioner skal betale for at deltage.

Generelt er de danske interessenter enige i, at der er behov for at indtænke en langsigtet og bæredygtig finansiering til drift af datacentre og opbevaring af data mv., hvis det skal lykkes at implementere FAIR. Nogle af de interviewede institutledere er i den sammenhæng bekymrede for en model, hvor forskere og forskningsinstitutioner i et for stort omfang skal betale for at deltage, da det vil svække incitamentet til at tage del i udviklingen af FAIR data. I den forbindelse peges der på manglende lovgivning på området. I dag er der en række offentligt finansierede nationale dataarkiver, såsom Det Kgl. Bibliotek og Rigsarkivet mv., og en række lokale universitetsbiblioteker i tilknytning til forskningsinstitutionerne, der afholder udgifterne til langtidsopbevaring. Derudover er det ofte forskerens eget ansvar at langtidslagre forskningsresultaterne, og udgifter til at vedligeholde dataarkivet afholdes af forskeren selv. De forskellige måder at sikre langtidsopbevaring af data i dag viser, at der i en videre konkretisering af FAIR bør være fokus på at sikre en langsigtet og bæredygtig finansiering ved at anerkende, at langtidsopbevaring af data er en ressourcekrævende service, hvis finansiering skal medregnes i de samlede udgifter.

³³ DeiC (2017) Anbefalinger til en sammenhængende e-infrastruktur til fremme af dansk forskning

4. De tyske erfaringer med at indføre FAIR data

Overordnet er Tyskland i starten af en proces, hvor FAIR-principperne skal til at integreres og accepteres af de centrale interessenter, herunder universiteterne, forskningsbibliotekerne, datacentre etc. I 2017 udgav Tyskland sammen med Holland en fælles bekendtgørelse om, at der er behov for at handle nu og her på European Open Science Cloud (EOSC)³⁴. I papiret giver Tyskland og Holland udtryk for behovet for en 'fast track'-implementering for at fremskynde udviklingen af EOSC ved at gøre data FAIR. Forslaget er, at det skal ske ved en decentral tilgang, hvor processen er drevet af forskningsmiljøerne selv og bygger videre på eksisterende infrastruktur og initiativer. Samtidig er det et hovedfokus i initiativet, at det nationale koordineres og hænger sammen med det europæiske for at undgå duplikering af indsatser og investeringer.

GO-FAIR-initiativet (støttet af det tyske forskningsministerium og ledet af Leibniz-Informationszentrum Wirtschaft) er omdrejningspunktet for implementeringen af FAIR data i Tyskland. GO-FAIR-initiativet bygger på tre søjler, hhv. GO-CHANGE (omhandler kulturen i de videnskabelige miljøer samt forskernes meriterings- og incitamentssystem), GO-TRAIN (omhandler kompetenceudvikling) samt GO-BUILD (omhandler den tekniske implementering).

4.1 DET TYSKE INTERESSENTLANDSKAB

Det tyske forskningslandskab indeholder flere forskellige typer af FoU-aktører, herunder universiteter, ikke-universitetsinstitutter, virksomheder samt forskningsinstitutioner drevet af enten føderale eller delstatsmyndigheder. Tyskland har ca. 1.000 offentligt støttede forskningsinstitutioner, hvoraf ca. 400 er højere uddannelsesinstitutioner, herunder universiteter, professionshøjskoler og erhvervsakademier (under fællesbetegnelsen Fachhochschulen (FHHS)). De tyske FHHS er centrale og integrerede aktører i det tyske forsknings- og innovationslandskab.

Den tyske forskningsfinansiering på føderalt niveau varetages af fagministerierne, hvoraf Bundesministerium für Bildung und Forschung (BMBF) er den største finansieringsaktør. De føderale midler går til forskning på én af de cirka 750 forskningsinstitutioner i Tyskland eller til tematisk forskning inden for et af de centrale forskningsområder. I 2014 blev der i alt uddelt ca. 14 mia. EUR på FoU på føderalt niveau, hvoraf Bundesministerium für Bildung und Forschung stod for ca. 60 pct. De tyske delstater står for de resterende ca. 40 pct. af forskningsfinansieringen og uddelte i 2013 ca. 10 mia. EUR.

En stor del af forskningsfinansieringen bliver kanaliseret gennem en række store forskningsfinansierende organisationer. Deutsche Forschungsgemeinschaft (DFG) er den største uafhængige forskningsfinansierende aktør i Tyskland og havde i 2015 et budget på ca. 3 mia. EUR. 2/3 af DFG's finansiering kommer fra føderalt niveau, og den sidste 1/3 fra delstatsniveau. De fire store og internationalt anerkendte forskningsorganisationer Fraunhofer-Gesellschaft, Max-Planck-Gesellschaft, Helmholtz-Gemeinschaft og Leibniz Gemeinschaft tildeler ligeledes midler gennem særskilte støtteprogrammer.

³⁴ Joint Position Paper on the European Open Science Cloud (2017): Germany and The Netherlands

4.2 ET PARADIGMESKIFTE I DEN TYSKE TILGANG

I Tyskland kom forskningsdata management for alvor på dagsordenen i 2010, da Allianz der Wissenschaftsorganisationen (en forening af de mest centrale tyske forskningsaktører) formulerede en række grundregler for håndtering af forskningsdata³⁵. Sidenhen har andre centrale aktører i det tyske forskningssystem sat fokus på emnet, bl.a. Hochschulrektorenkonferenz (en forening for både private og offentlige tyske universiteter), som i 2014 formulerede en række anbefalinger til en mere strategisk tilgang til forskningsdata management³⁶. Endvidere har en række tyske delstater igangsat flere relevante initiativer; Helmholtz-Gemeinschaft (en forening af 18 forskellige forskningscentre) har støttet projekter på området, ligesom Deutsche Forschungsgemeinschaft (den centrale tyske forskningsfinansierende aktør) i sin forskningsstøttepraksis har udviklet retningslinjer.

I Tyskland har området for 'Open Science' været præget af et stort antal ukoordinerede initiativer. Det har i høj grad været en 'bottom-up'-proces, men uden en national koordinering. Der er et stort antal traditionsrige og stolte universiteter, som nødt vil samarbejde og dele deres forskningsdata med andre, konkurrerende universiteter. Inden for delstaterne (Baden-Württemberg nævnes som eksempel) kan det være en udfordring at få universiteterne til at tale sammen.

Idéen bag GO-FAIR-initiativet er at gøre op med den hidtidige praksis i Tyskland og er en politisk vision om at få koordineret mere på tværs for at undgå løsninger, der ikke er indbyrdes kompatible. Dr. Klaus Tochtermann, direktør for ZBW og ansvarlig for GO-FAIR-initiativet, vurderer, at FAIR-principperne er et paradigmeskifte og en ny måde at håndtere forskningsdata på. Som han ser det, er FAIR data noget andet end 'Open Science'.

"To meet the FAIR data and service vision, you need to have a global infrastructure in place, and build up the necessary skills among data stewards and researchers. One major difference in FAIR is the access part, as data does need to be intrinsically open" (Klaus Tochtermann, direktør, ZBW).

For at demonstrere værdien af FAIR data og få forankret og operationaliseret principperne har den tyske tilgang været at igangsætte projekter, der skal demonstrere værdi og være 'best practice' for, hvordan en model for den fremtidige tyske forskningsdatainfrastruktur kan se ud. Især nedenstående projekter er centrale:

- **The Helmholtz Data Federation (HDF)** er et konsortium med seks partnere, bestående af high-performance computercentre samt et community med fokus på big data, som pga. omfanget ikke er overførbare via internettet.
- **GeRDI** (Generic Research Data Infrastructure)-projektet (finansiering fra november 2016 til november 2019), hvis langsigtede formål er at gøre det muligt for alle tyske forskere at opbevare, dele and genbruge forskningsdata på tværs af forskningsområder. I den første fase af projektet vil tre datacentre blive sat sammen. Det skal danne pilotcasen for, at det kan rulles ud i Tyskland og være en model for den fremtidige tyske forskningsdatainfrastruktur.

³⁵ Allianz der Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten

³⁶ HRK (2014): Management of research data – a key strategic challenge for university management

- **RADAR** (Research Data Repository)-projektet (finansiering fra 2013 til 2016) handler om at samle data i et datacenter, som de forskellige medlemmer så har adgang til. Projektet er tværfagligt, og projektets deltagere er fra universiteter i forskellige tyske delstater. Projektet er organiseret således, at der på hver institution er en lokal koordinator, som så er ansvarlig for at understøtte de lokale forskere på deres institution i datadeling og -håndtering. Projektet understøtter 'the long tail' inden for forskningsdata, som er de data, som falder uden for big data-domænet – altså først og fremmest små datasæt. I relation til FAIR er vurderingen, at projektet har opnået kriterierne Findability og Accessibility, men er langt fra at indfri Interoperability og Reusability.

4.3 DEN TYSKE FAIR DATALØSNING

Det er intentionen med GO-FAIR-initiativet, at det er forankret blandt interessenterne og konsensusudrevet ud fra en erkendelse af, at implementeringen af FAIR ikke bare er tekniske standarder, men i høj grad er betinget af kulturen. Som det påpeges af de tyske informanter, er der nu en proces i gang med det formål at 'omvende' de tyske interessenter, forskere og bibliotekarer til at forstå og bruge FAIR-principperne. I den sammenhæng peges der på, at det nationale politiske fokus og den nationale opbakning har spillet en afgørende rolle for at få sat FAIR på dagsordenen.

I modsætning til Danmark har Tyskland en føderal struktur, der giver delstaterne politisk kompetence til at varetage forskningspolitikken, og de sætter også selv initiativer såsom eScience Initiative i Baden-Württemberg, HeFIS (Hessische Forschungsdaten InfraStruktur) i Hessen og Cooperation in Digital Science i Hamburg og Schleswig-Holstein. Endvidere har de store forskningsorganisationer sat gang i store satsninger på området, bl.a. Helmholtz-Data-Federation ved Helmholtz-Gemeinschaft, FORDATIS ved Fraunhofer Gesellschaft og Leibniz Data ved Leibniz Gemeinschaft. Dette øger behovet for koordination og dialog på tværs.

4.3.1 Infrastrukturen er på plads, men der er behov for mere

Grundlæggende vurderer informanterne, at infrastrukturen til lagring og opbevaring af forskningsdata er opbygget i Tyskland. Der er datacentre på 110 tyske universiteter, 233 tekniske uddannelsesinstitutioner, 280 andre forskningsinstitutioner og herudover flere tusinde videnskabelige biblioteker³⁷. Det næste skridt er at bygge videre på denne infrastruktur, linke dem sammen og skabe en ensartet adgang for forskerne. Dr. Stefan Winkler-Nees fra DFG vurderer imidlertid, at det vil kræve en større, national koordination og strategi, fordi det lige nu eksempelvis er en udfordring, at de initiativer, Helmholtz-Gemeinschaft har støttet, prioriterer at give adgang til deres forskningsinstitutioner.

Yderligere er der i Tyskland et stort behov for investeringer i datastøttefunktioner på universiteterne. P.t. er der ikke et samlet overblik over det i Tyskland, men samtlige interessenter understreger i interviewene, at der skal investeres massivt i det område, samtidig med at der ligger en udfordring i at få universiteter og forskningsbiblioteker til at samarbejde endnu mere.

"For research data management we need the libraries on board. In Germany, today about 20 percent of the university libraries are already active, offering research data services to researchers and supporting the

³⁷ Jf. hjemmesiden re3data.org

implementation of data management policies. Regional infrastructure programmes accelerate this rate (e.g. in Baden-Württemberg, North Rhine-Westphalia) (Birgit Schmidt, Scientific Manager, Göttingen State and University Library).

Delft Tekniske Universitet i Holland fremhæves som best practice på et universitet, der har investeret i og arbejdet systematisk med oprettelsen af datastøttefunktioner til sine forskere.

4.3.2 Behov for afklaring af finansieringsmodellen

Der er i Tyskland debat om finansieringen af datacentre og finansieringsmodellerne. Det vil givetvis ikke være én finansieringsmodel, der vil blive implementeret, men en kombination af flere. Nogle data-repositorier vil blive finansieret statsligt eller delstatsligt, men der vil kunne være en betaling for at få særlige services.

En anden model er medlemskabsmodellen (som afprøves i RADAR-projektet), hvor der er flere, som deltager i et konsortium, og hver betaler et medlemsgebyr for at få adgang til data og de relaterede services. Dr. Stefan Winkler-Nees fra DFG understreger, at den model er udfordret af, at datacentre har svært ved at prisfastsætte omkostningerne og ikke har et overblik over prisen på data.

En tredje model er, at forskningsfondene kræver, at projekterne lægger deres data i et data-repositorium, hvor der i forskningsbevillingen afsættes et beløb til dette. Dette er en praksis, som de britiske forskningsfinansierende aktører bruger.

4.3.3 Tekniske standarder udvikles 'bottom-up'

Som beskrevet er den tyske tilgang at modne interessenterne og ikke lægge sig fast på en stor forkromet FAIR dataløsning, som så skal rulles ud over hele landet. De anvendte tekniske standarder skal ikke kun gælde i Tyskland, men internationalt. Derfor deltager Tyskland (ligesom Danmark) bl.a. i Research Data Alliance (RDA), hvis formål er at udvikle praksisser, politikker og værktøjer inden for en række internationalt nedsatte arbejds- og interessegrupper.

DFG stiller i sine retningslinjer krav til standarder, PID'er og metadata-skemaer, men specificerer ikke, hvad ansøgerne skal bruge specifikt. Det afhænger af tradition og praksis inden for de specifikke forskningsfelter og subfelter.

"There was some discussion, if DFG should demand specific standards. However, we are convinced that it is up to the disciplines themselves, what type of standards they will use. Standards can be very complex and specific, and they may need long time to develop. But it is important that there is progress, driven by a bottom-up approach" (Stefan Winkler-Nees, programdirektør, DFG).

4.4 HVAD KAN DANMARK LÆRE AF DE TYSKE ERFARINGER?

Vi vurderer, at Danmark har mulighed for at bygge videre på de tyske erfaringer i implementeringen af en FAIR dataløsning:

For det første er det helt centralt at holde fokus på forskernes efterspørgsel som driver for indsatsen. Det betyder konkret, at en FAIR dataløsning skal udvikles med øje for forskernes brug og behov, fx om de støttes tilstrækkeligt i omstillingen til at arbejde med FAIR data, og om de udviklede løsninger tilgodeser forskernes praksisser. I Tyskland er der fokus på dialogen med forskningsmiljøerne om potentialerne ved FAIR og 'Open Science', bl.a. har GO-FAIR-initiativet fokus på at 'modne' forskningsmiljøerne og gå i dialog med forskerne om potentialerne ved FAIR.

For det andet er det en udfordring i Tyskland at få sikret en bæredygtig finansiering af FAIR. Mange af de støttede projekter er finansieret for en begrænset tidsperiode, og risikoen ved projektf finansieringen er, at projekterne strander, så snart projektf finansieringen udløber. I Tyskland bliver det diskuteret, hvilke finansieringsmodeller man kan forestille sig at implementere for at sikre en langsigtet og bæredygtig finansiering. Vi anser det for vigtigt, at der også i Danmark tages hul på en diskussion om finansieringsmodellerne, herunder hvordan de kan bidrage til en langsigtet, bæredygtig finansiering af infrastruktur, uddannelse, datastøtte etc.

For det tredje viser erfaringerne fra Tyskland, at der er lang vej, fra man tager en beslutning om at blive FAIR til at realisere visionen. Det tyske forskningssystem er mere fragmenteret end det danske og består af ca. 1.000 offentligt støttede forskningsinstitutioner. Samtidig er det tyske system føderalt, hvilket betyder, at de enkelte delstater har politisk beslutningskompetence angående forskningsprioriteter. Det må også forventes at tage lang tid i Danmark at opnå visionen FAIR data, om end det trods alt er et mindre omfattende politisk struktur- og forskningssystem, hvorfor en løsning hurtigere kan rulles ud.

5. Metode og dataindsamling

Analysen bygger på en række forskellige metoder. Estimeringen af omkostninger og gevinster ved indførelsen af FAIR data i Danmark bygger primært på et grundigt litteraturstudie af gevinster og omkostninger ved indførelsen af FAIR data, men vi har forholdt os kritisk til litteraturen ift. at perspektivere i en dansk kontekst. Indhentningen af tyske erfaringer og vurderingen af de danske muligheder og barrierer bygger på en kombination af en grundig desk research af studier, rapporter og artikler samt en bred interviewundersøgelse, der kommer rundt om de centrale interessenters erfaringer og kendskab til FAIR og deres vurderinger af potentialet for at indføre FAIR.

Den metodiske tilgang til hver af de tværgående analyser er i det følgende uddybet, hhv. estimering af omkostning og gevinster (afsnit 5.1) samt muligheder og barrierer for at indføre FAIR data (afsnit 5.2).

5.1 ESTIMERING AF OMKOSTNINGER OG GEVINSTER

Dette afsnit redegør for de forskellige metoder, der kan anvendes til at opgøre værdien af at indføre FAIR data i Danmark, og for den valgte metode i den samfundsøkonomiske konsekvensvurdering i kapitel 2.

Metoderedegørelsen er baseret på en grundig litteraturgennemgang. Litteraturgennemgangen har primært fokuseret på metoder, som tidligere er anvendt til at opgøre værdien af FAIR data, men inkluderer også litteratur om værdien af andre typer af åbne data (fx åbne offentlige data). Metoderne fra denne litteratur er inddraget, hvor de er relevante i forhold til FAIR data. Endelig har vi, hvor det var relevant, inddraget viden fra interviews i Danmark og Tyskland til udviklingen af vores metode.

5.1.1 Overordnede tilgange til samfundsøkonomiske konsekvensanalyser

Vores overordnede tilgang til samfundsøkonomiske konsekvensanalyser følger Finansministeriets vejledning, som er standarden i Danmark. Der er meget få eksempler på samfundsøkonomiske konsekvensvurderinger af FAIR data i litteraturen og ingen eksempler, hvor opstarts- eller driftsomkostninger er inkluderet. Det betyder, at vi har været nødt til at trække på en række forskellige kilder og metoder i vores arbejde med at estimere den samfundsøkonomiske værdi af at indføre FAIR data i Danmark.

Dette gælder fx for timingen af hhv. gevinster og omkostninger. I Houghton & Gruen (2014) modelleres afkastet af forskning, og dermed gevinsterne ved FAIR data, over en 10-årig periode. I adskillige andre studier i litteraturen anvendes en 30-årig periode. Vi anvender den samme tidshorison for gevinsterne som i Houghton & Gruen (2014). Litteraturen, der beskæftiger sig med timingen af omkostningerne, herunder opstarts- og driftsomkostningerne, er meget begrænset. Vi har derfor valgt ganske konservative antagelser.

En anden udfordring er, at en stor del af litteraturen på området består af ex post-analyser eller evalueringer af eksisterende FAIR data-løsninger. En væsentlig begrænsning ved de eksisterende ex post-studier er, at de måler værdien af data, når de *er* FAIR, og ikke – som det er formålet med denne analyse – merværdien af at *indføre* FAIR data. Der er dog enkelte eksempler på ex ante-analyser, som opgør værdien af indførelse af nye FAIR dataordninger eller udbredelsen af eksisterende ordninger. I ex ante-studier estimeres værdien af data, når de ikke er FAIR, og derefter anvendes forskellige metoder til at vurdere den potentielle merværdi, som data får,

når de bliver FAIR. Vores overordnede tilgang er derfor i høj grad baseret på disse ex ante-studier. Særligt er studiet Houghton & Gruen (2014) hovedkilden i forhold til estimeringen af gevinster ved FAIR data.

5.1.2 Estimering af gevinster ved FAIR data

I litteraturen findes der overordnet to forskellige metoder, som kan anvendes til at estimere værdien af data: den velfærdsteoretiske metode og den afkastbaserede metode.

Den velfærdsteoretiske metode estimerer forbrugeroverskuddet (brugerens betalingsvillighed – den faktiske betaling) med udgangspunkt i priselasticiteter. Denne tilgang egner sig dog ikke særligt godt, når prisen på data sættes ned til 0, dvs. at data bliver gratis. Desuden er de nødvendige data til at estimere priselasticiteter ikke fuldt tilgængelige for forskningsdata i Danmark.

Den afkastbaserede metode tager derimod udgangspunkt i litteraturen om afkast af forskning og udvikling. Gode data om forskningsaktiviteten i Danmark og en eksisterende litteratur om afkastet af forskning og udvikling i Danmark gør, at denne metode er mere velegnet til at opgøre værdien af at indføre FAIR data i Danmark. Størstedelen af den eksisterende litteratur er også baseret på denne tilgang.

Figur 5.1. Estimering af afkastet på offentlig forskning

Afkastet estimeres efter samme metode, som er anvendt i Houghton & Gruen (2014). Afkastet opnås på den totale beholdning af forskningsdata, som stiger hvert år, når nye data kommer til. Data **afskrives med 10 pct. p.a.**, baseret på litteratur, som indikerer en afskrivningsprocent på 5-10 (Evans et al., 2008). Vi antager, at data har en **levetid på 10 år**. Flere studier finder dog en levetid på op til 30 år (Sveikauskas, 2007). Afkastet forventes realiseret over 5 år og normalfordeles over år 1 til 5 (Mansfield, 1991; Mansfield, 1998).

I vores estimering af gevinsterne ved FAIR data tager vi som nævnt udgangspunkt i Houghton & Gruen (2014). Det er det eneste studie, som er lavet ex ante og samtidig er baseret på den afkastbaserede metode. Vi vurderer, at vi med udgangspunkt i denne metode har bedst mulighed for at tage højde for, at der i Danmark allerede eksisterer en stor mængde data, og at værdien heraf ikke skal tælles med, når man opgør den samfundsøkonomiske værdi af at indføre FAIR data i Danmark.

Udgangspunktet for estimeringen af gevinsterne er værdien af forskningsdata i Danmark. Dertil kommer en række afledte gevinster ved, at data bliver FAIR, herunder værdien af sparet tid og skabelse af ny forskning. Dette er de gevinster, der ifølge vores interviews er størst potentiale for i Danmark. En mulig gevinst ved FAIR data er, at afkastet på forskning kan blive større, fordi forskningen gøres mere målrettet eller tværfaglig. Denne gevinst er ikke medtaget i vores hovedanalyse, da vi ønsker at estimere gevinsterne konservativt, og da gevinsten ikke fremstår som væsentlig i vores interviews med aktører i Tyskland og Danmark.

I litteraturen findes studier af FAIR data-løsninger for forskellige videnskabsgræne. FAIR data er klart mest udbredte for naturvidenskab, og der findes således også mest litteratur på dette område. Studier peger dog på, at der er forskel på effekterne for naturvidenskab, samfundsvidenskab og humaniora. Vores beregninger af

gevinsterne er derfor baseret på en opdeling i disse tre faggrene. Eftersom der ikke findes separate studier for teknisk videnskab, sundhedsvidenskab og jordbrugs- og veterinærvidenskab, beregnes disse ud fra samme estimater som naturvidenskab.

5.1.2.1 Værdien af forskningsdata

For at kunne estimere gevinsterne ved at gøre data FAIR er det nødvendigt først at estimere værdien af beholdningen af forskningsdata i Danmark. I litteraturen findes en række forskellige metoder til at estimere værdien af forskningsdata. Flere af disse metoder bygger på data, som ikke er tilgængelige for Danmark. Nedenfor opridses kort de vigtigste metoder til at opgøre værdien af forskningsdata:

Investeringsværdien: Udtrykker den samlede værdi af investeringen i FAIR data, herunder drift af infrastruktur, vejledning og service. Indeholder ikke 'sunk costs', dvs. éngangsinvesteringer i forhold til opsætning af infrastruktur og lignende. Baseres i litteraturen på driftsregnskaber fra datacentre samt spørgeskemaer om omkostninger forbundet med upload af data for eksisterende datacentre. Denne metode egner sig bedst til ex post-analyser.³⁸

Betalingsvillighedsværdien: Baseret på stated preferences værdisættes data ud fra brugernes betalingsvillighed. Stated preferences er en metode, hvor brugerne af data bliver bedt om at vurdere værdien af data i et spørgeskema.³⁹

Brugsværdien: Værdien af data estimeres som de omkostninger, der er forbundet med brugen af data, herunder særligt værdien af tid anvendt på dataarbejde, men også omkostninger til fx dataanalyseværktøjer og lignende. I ex post-analyser estimeres denne ud fra spørgeskemaanalyser blandt brugerne af de analyserede datacentre. Houghton & Gruen (2014) estimerer brugsværdien ud fra makrodata om de nationale forsknings- og udviklingsudgifter samt spørgeskemaanalyser fra eksisterende datacentre om andelen af forskningstid, der anvendes på dataarbejde. Makrodata for forsknings- og udviklingsudgifter er tilgængelige for Danmark og af høj kvalitet; derfor anvender vi denne metode i vores analyse.⁴⁰

Houghton & Gruen (2014) arbejder med to estimater for brugsværdien af data. Et øvre estimat, som er baseret på de samlede forsknings- og udviklingsudgifter. Ved at medtage de samlede forsknings- og udviklingsudgifter medtages også udgifter forbundet med dataanalyseværktøjer eller andre kapitalomkostninger forbundet med brugen af data. Estimatet kan dog være et overkantsskøn, da der ikke findes empiri for, hvor stor en andel af kapitalomkostningerne til forskning og udvikling, der relaterer sig til dataarbejde. Det findes der til gengæld for lønomkostningerne eller tidsomkostningerne, baseret på en række forskellige spørgeskemaanalyser i forbindelse med eksisterende datacentre. Dermed anvender de som nedre estimat den andel af lønomkostningerne, som

³⁸ Anvendes i Sullivan et al. (2017), Houghton & Gruen (2014), Beagrie & Houghton (2014), Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012)

³⁹ Anvendes i Sullivan et al. (2017), Beagrie & Houghton (2016), Beagrie & Houghton (2014), Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012)

⁴⁰ Anvendes i Sullivan et al. (2017), Beagrie & Houghton (2016), Houghton & Gruen (2014), Beagrie & Houghton (2014), Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012)

relaterer sig til data. I vores analyse tager vi udgangspunkt i lønomkostningerne, svarende til det nedre estimat i Houghton & Gruen (2014).

5.1.2.2 Værdien af sparet tid

Indførelsen af FAIR data forventes at medføre tidsbesparelser, fordi forskere skal bruge mindre tid på at anskaffe og strukturere data. Dette har i sig selv en værdi, idet forskernes tid er værdifuld og kan bruges på andre aktiviteter. Hurtigere adgang til data gør forsknings- og udviklingsprocessen hurtigere og øger afkastet af forskning, bl.a. ved at forskere ”investerer” den sparede tid i anden forskning, som giver et yderligere afkast.

Tilgangen til at estimere den tidsbesparelse, som er forbundet med FAIR data, er i Houghton & Gruen (2014) baseret på erfaringer fra Storbritannien, hvor man har implementeret forskellige FAIR data-løsninger. Vi anvender den samme metode, men anvender flere kilder til estimatet for tidsbesparelse.

I nogle studier er det alene værdien af tidsbesparelsen (målt i lønkroner), som medtages som effektivitetsgevinst. I Houghton & Gruen (2014) medtages yderligere afkastet af den ekstra forskning, som forskerne investerer deres tid i. Vi anvender samme tilgang i vores analyse.

I enkelte studier kigger man på effektivitetsgevinsterne mere bredt, dvs. ikke udelukkende for forskere, men også for undervisere og studerende, som også anvender disse data.⁴¹ Disse effektivitetsgevinster er ikke medtaget i vores analyse.

5.1.2.3 Værdien af ny forskning skabt som følge af FAIR data

Bredere adgang til data giver bedre muligheder for tværfaglig forskning og samarbejde på tværs af institutioner og sektorer. Dermed kan der opstå ny forskning, som ikke ville opstå uden FAIR data. Den bedre adgang kan også facilitere, at forskere med mindre budgetter kan gøre brug af data, idet de sparer tid og omkostninger forbundet med produktionen af data. Interviews indikerer, at potentialet for tværfaglig forskning kan være begrænset, men der skønnes stadig at være et væsentligt potentiale for skabelse af ny forskning som følge af FAIR data i Danmark.

Houghton & Gruen (2014) estimerer mængden af ny forskning på baggrund af spørgeskemaundersøgelser for eksisterende projekter i Storbritannien. Disse angiver, hvor stor en andel af forskningen der ikke ville være gennemført uden FAIR data. I Houghton & Gruen (2014) medregnes den fulde værdi af denne nye forskning som en gevinst ved FAIR data. Dette svarer til at antage, at forskningsaktiviteten stiger markant, idet man antager, at den forskning, som vil finde sted, når data ikke er FAIR, fortsat gennemføres, samtidig med at der laves ny forskning på baggrund af FAIR. Med andre ord antager man, at den nye forskning ikke fortrænger eksisterende forskning. Vi anser dette for at være en ganske optimistisk antagelse. Da der ikke findes litteratur, som empirisk understøtter graden af fortrængning af eksisterende data, og da vi ønsker at anvende en konservativ tilgang, vælger vi derfor at se bort fra værdien af ny forskning, som er skabt som følge af FAIR data, i vores grundscenarie. Der laves dog følsomhedsanalyser af betydningen af ny forskning.

⁴¹ Se Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)

5.1.2.4 Større afkast af forskning

Nemmere adgang til data kan medføre bedre og mere informeret forskning, herunder en reduktion i den tid, som investeres i blindgyder og dupliserende forskning. Dermed kan man undgå spild af ressourcer og dobbelt-udgifter til den samme forskning. Dette vil gøre forskning mere effektiv og dermed øge det gennemsnitlige afkast af forskningen. I det hele taget kan FAIR data lede til mere effektiv forskning gennem en række kanaler: bedre undervisning, større udbredelse af forskningen, mere tværfagligt samarbejde etc.

I litteraturen er denne gevinst beskrevet, men det er svært at estimere det præcise omfang og dermed værdien af gevinsten. En makrotilgang gør det muligt at illustrere potentialet ved mere effektiv forskning. I Houghton & Sheehan (2009) udføres en potentialeberegning, som tager udgangspunkt i en modificeret Solow-model. Studiet viser, hvordan tilgængelighed og effektivitet (andel af ny viden, som er nyttig) er vigtige parametre, når afkastet af forskning estimeres. Dermed viser de teoretisk, at en stigning i tilgængelighed og effektivitet øger afkastet af forskning. Eftersom studiet ikke indeholder forslag til proxyer for parametrene tilgængelighed og effektivitet, er det dog ikke muligt empirisk at vurdere, i hvor høj grad disse parametre påvirkes af indførelse af FAIR data. Grundet disse usikkerheder er gevinsten ikke medtaget i vores analyse.

5.1.3 Estimering af omkostninger ved FAIR data

Litteraturen har mindre fokus på estimering af omkostninger til FAIR data. Det kan hænge sammen med, at langt de fleste studier er ex post-studier, hvorfor omkostningerne nemt kan opgøres ud fra regnskaberne for de enkelte projekter, og man kan se bort fra opstarts- og investeringsomkostninger, der kan betragtes som 'sunk costs'⁴². Houghton & Gruen (2014) estimerer driftsomkostningerne ud fra deres andel af værdien af forskningsdata. Studiet medtager dog ikke opstarts- eller investeringsomkostninger og modellerer ikke omkostningerne over tid, som det er nødvendigt at gøre i vores analyse.

Beagrie et al. (2010) kortlægger alle de potentielle typer af omkostninger forbundet med opbevaring af forskningsdata ud fra en række forskellige projekter og datacentre. Studiet har fokus på at inddrage og analysere typerne af omkostninger. I de fleste af casene er den samlede dataværdi for datacentrene ikke oplyst, hvilket gør det svært at ekstrapolere omkostningerne til Danmark. Studiet indeholder dog én case, som analyserer fordelingen af de samlede dataomkostninger. I forbindelse med The Embedding Institutional Data Curation Services in Research (EIDCSR)-projektet ved Oxford Universitet er der indsamlet information om alle omkostninger forbundet med dataarbejdet, herunder omkostninger til skabelse af data, omkostninger til opstart af datacenteret og omkostninger til drift af datacenteret. Vi anvender denne fordeling til at estimere henholdsvis opstarts- og driftsomkostninger som en andel af den samlede dataværdi ud fra samme tilgang, som er anvendt i Houghton & Gruen (2014).

Vores estimerer tager dermed hensyn til, at der vil være store opstartsomkostninger forbundet med indførelse af FAIR forskningsdata i Danmark. Estimererne vil være konservative i den forstand, at omkostningerne sandsynligvis vil blive overestimeret. Dette skyldes, at studiet kun omfatter omkostninger forbundet med skabelse af data, mens den samlede dataværdi i vores analyse indeholder alle omkostninger forbundet med brug af data, herunder omkostninger til analyse af data. Dermed vil andelen af opstarts- og driftsomkostninger i forhold til de samlede dataomkostninger være overestimeret. Derudover er andelen baseret på en kortere projektperiode,

⁴² Sunk costs er et begreb anvendt af økonomer til at beskrive omkostninger, som allerede er realiseret, og som dermed ikke bør påvirke beslutningstagningen.

sammenlignet med vores analyseperiode. Det kan derfor være tilfældet, at vi medtager opstartsomkostninger i en for lang periode, og/eller at opstartsomkostningerne vil være faldende over tid. Andelene anvendes dog af forsigtighedshensyn.

5.1.4 Alternativer til cost-benefit-analyse

Litteraturen indeholder adskillige alternative studier af værdien af FAIR forskningsdata, som dog ikke umiddelbart kan bruges i en cost-benefit-analyse. Her opsummeres kort de forskellige metoder og tilgange anvendt i litteraturen.

5.1.4.1 Ekstrapolering ud fra BNP

I Europa-Kommissionen (2015) ekstrapoleres værdien af data ud fra makroøkonomiske tal for værdien af data i pct. af BNP, hvor man dog også tager hensyn til modenhedsniveauet af åbne data i de forskellige EU-lande. Studiet kigger på både den direkte og den indirekte værdi af åbne data. I forhold til vores analyse er en udfordring ved denne tilgang, at det er svært at skelne mellem FAIR forskningsdata og andre typer af åbne data. Studiet er baseret på en ex post-tilgang. Da vi ønsker at isolere den samfundsøkonomiske værdi af at indføre FAIR-principperne i Danmark, er det desuden vigtigt at korrigere for værdien af de eksisterende data. Dette gøres ikke i studiet.

En tilsvarende metode anvendes i OECD (2015) for offentlige data (PSI) og i Gruen (2014) for åbne data mere bredt.

5.1.4.2 Estimering ud fra cases

Gruen (2014) anvender en række forskellige cases, med eksempler på åbne data, til at estimere værdien af åbne data. Estimerne bygger på en række meget usikre antagelser og bør dermed betragtes som potentialeberegninger snarere end egentlige samfundsøkonomiske konsekvensvurderinger. Gruen (2014) anvender hovedsagelig metoden til at argumentere for, at resultatet, som er baseret på en ekstrapolering ud fra BNP, ikke er urealistisk højt.

5.1.4.3 Afkast af investeringen (RoI) fra analyser af værdien af biblioteker

I Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012) fremhæves det, at der findes en relativt veletableret litteratur om afkastet af investeringer i data-, biblioteks- og informationsservices. Denne litteratur kan anvendes som et yderligere tjek af vores beregninger, hvis det forventes, at afkastet af investeringer i FAIR data er af samme størrelsesorden. Litteraturen finder, at investeringer i data-, biblioteks- og informationsservices skaber en værdi, der er mellem ca. 3 og 5 gange så stor som investeringen.⁴³

⁴³ Kilde: Beagrie (2012)

5.1.5 Kilder anvendt i vores beregninger

Parametre og inputs	Kilde(r)
Diskonteringsrente	Finansministeriet (2017)
Skatteforvridning	Finansministeriet (2017)
FoU-udgifter i pct. af BNP efter pct. af BNP og tid	Danmarks Statistik, statistikbanken.dk
FoU-årsværk (videnskabeligt personale) i den offentlige sektor efter køn, personalekategori, fag og tid	Danmarks Statistik, statistikbanken.dk
Standardberegnet månedsførtjeneste, offentlig forvaltning og service	Danmarks Statistik, statistikbanken.dk
BNP i løbende priser	Danmarks Statistik, statistikbanken.dk
Realvækst i BNP	Finansministeriet (2017) – opdateret 2025-forløb
Andel af forskeres tid, som anvendes på forskning og videnskabelse	DEA (2015)
Andel af forskningstid, som anvendes til dataarbejde	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)
Effektivitetsgevinst, andel af dataarbejdstid, der kan spares	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b), Sullivan et al. (2017)
Afkast af offentlig forskning	DEA (2017), Danske Universiteter (2016), Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b), Sullivan et al. (2017)
Andel af afkast, der realiseres lokalt	Houghton & Gruen (2014), Jaffe (1989), Coe & Helpman (1993, 1995), Verspagen (2004), Arundel & Guena (2004)
Deprecieringsrate for forskning	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)
Datas levetid	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)
Andel af forskning, som ikke ville eller kunne skabes uden FAIR data	Houghton & Gruen (2014), Beagrie & Houghton (2016), Beagrie & The Centre for Strategic Economic Studies University of Victoria (2012), Beagrie & Houghton (2013a), Beagrie & Houghton (2013b)
Driftsomkostninger	Houghton & Gruen (2014), Beagrie et al. (2010)
Opstartsomkostninger	Beagrie et al. (2010)

5.2 ANALYSE AF MULIGHEDER OG BARRIERER

Dette afsnit redegør for de valgte metoder og de indsamlede data i den tværgående analyse af barrierer og muligheder (kapitel 3) samt for de tyske erfaringer med at indføre FAIR data (kapitel 4).

5.2.1 Den tyske case

Tyskland er valgt til at indgå i foranalysen som et casestudie, da Tyskland er blandt de 'early movers'-lande, som er gået i gang med at gennemføre GO-FAIR-initiativet og har sat sig for at implementere FAIR-principperne med afsæt i EOSC-ekspertgruppens anbefalinger.

Vi har brugt casestudiet som metode til at indfange de tyske erfaringer i dybden ud fra forskellige vinkler og perspektiver. Indsamling af viden om de tyske erfaringer har været et fokuspunkt i den første fase af foranalysen. Baggrunden for at kortlægge de væsentligste tyske erfaringer først er, at det har givet os mulighed for at analysere mulighederne for at overføre de tyske erfaringer til en dansk kontekst og vurdere, hvorvidt gevinster, omkostninger, muligheder og barrierer også vil gælde i Danmark.

Casestudiet er afdækket via en grundig og systematisk desk research af relevante rapporter, artikler etc. samt seks dybdegående, kvalitative interviews med centrale tyske interessenter. Desk research-fasen består af rapporter, strategier og artikler på området, herunder:

- Bundesministerium für Bildung und Forschung (2016): Open Access in Deutschland – Die Strategie des Bundesministeriums für Bildung und Forschung
- Rat für Informationsinfrastrukturen (2016): Enhancing Research Data Management: Performance Through Diversity
- Allianz der deutschen Wissenschaftsorganisationen (2015): Positionspapier „Research data at your fingertips“
- Germany and The Netherlands (2017): Joint Position Paper on the European Science Cloud
- Hochschulrektorenkonferenz (2014): Management of research data – a key strategic challenge for university management

Med udgangspunkt i ovenstående rapporter på området samt SFU's kontakter til relevante organisationer, som er involveret i den tyske indsats, er der udvalgt seks interviewpersoner, herunder:

Tabel 5.1 Oversigt over interviewpersoner i Tyskland

Organisation	Interviewperson	Personligt/telefonisk
Deutsche Forschungsgemeinschaft (DFG)	Dr. Stefan Winkler-Nees, Direktor	To interview gennemført (Personligt og telefonisk)
FIZ Karlsruhe	Matthias Razum	Personligt
ZBW (Leibniz-Informationszentrum Wirtschaft)	Prof. Dr. Klaus Tochtermann,	Personligt
Leibniz Universität Hannover	Prof. Dr.-Ing. Gabriele von Voigt	Telefonisk
Goettingen State and University Library	Dr. Birgit Schmidt	Telefonisk

Nogle af de tyske aktører, fx det tyske Ministerium for Uddannelse og Forskning, er det ikke lykkedes at gennemføre et interview med. Vi har forsøgt adskillige gange telefonisk og per e-mail at få kontakt til dem, men det er ikke lykkedes. Dog har vi afdækket ministeriets tilgang til FAIR data via de øvrige interviews samt en grundig desk research af relevante strategipapirer og rapporter fra ministeriet.

Interviewene er gennemført ud fra en semistruktureret interviewguide, der er tilpasset den enkelte informants rolle og viden om Tysklands indsats. For at indfange nøgleaktørernes erfaring i dybden er interviewene gennemført personligt i Tyskland.

5.2.2 Danske interessenters holdning til FAIR data

Den tværgående analyse af barrierer og muligheder for FAIR data i Danmark bygger indledningsvis på en grundig desk research af relevante rapporter, analyser og anden litteratur på området, herunder:

- DeIC, DEFF (2015), *National strategi for forskningsdata management 2015-2018*, København, den 30. januar 2015
- Horizon 2020 Programme (2016), *Guidelines on FAIR Data Management in Horizon 2020*
- Force11 (2017), *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing*
- Kommissionen (2017), *European Open Science Cloud*
- Max Planck-Gesellschaft (2002), *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*
- Mons, Barends, Cameron Neylon, Michel Dumontier, Luiz Olavo Bonino Sila Santos & Mark D. Wilkinson (2017), *Cloudy, Increasingly FAIR – Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud*
- OECD (2007), *Principles and Guidelines for Access to Research Data from Public Funding*
- Regeringen (2016), *Forsknings- og innovationspolitisk redegørelse*
- Wessels, Biddette, Rachel L. Finn, Kush Wadhwa, Thordis Sceinsdottir, Lorenzo Bigagli, Stefano Nativi & Merel Noorman. (2017), *Open Data and Knowledge Society*, Amsterdam University Press
- Wilkinson et al. (2016), *The Fair Guiding Principles for Scientific Data Management and Stewardship*.

Denne desk research på området har indledningsvis i foranalysen ført til et baggrundsnotat, der har beskrevet forskellige definitioner af FAIR data samt baggrunden for ønsket om at indføre FAIR data på europæisk plan. Dele af dette baggrundsnotat er indarbejdet i nærværende rapport.

Som en central del i analysen af barrierer og muligheder for implementering af FAIR data i dansk kontekst er der gennemført en lang række interviews med relevante danske interessenter. Der er i alt gennemført 30 dybdegående, kvalitative interviews med aktører på politisk niveau (inkl. EU), systemniveau (herunder koordinerende aktører og nationale forskningsbiblioteker) samt forskningsfinansierende aktører, ledelse og administrative medarbejdere fra de danske universiteter samt forskere. Analysen har bevidst forsøgt at komme bredt omkring de danske universiteter, bl.a. for at skabe ejerskab og opbakning til analysen og for at sikre, at alle føler sig inddraget.

Tabel 5.2 Oversigt over interviewpersoner i Danmark

Organisation	Afdeling	Interviewperson	Titel	Type
EOSC	High Level Expert Group	Barend Mons	Formand	EU
Digitaliseringsstyrelsen	Kontor for data og arkitektur	Jens Krieger Røyen	Kontorchef i Kontor for data og arkitektur	Politisk niveau
Sundhedsdatastyrelsen	Direktionen	Lisbeth Nielsen	Direktør	Politisk niveau
Danmarks Elektroniske Fag- og Forskningsbibliotek (DEFF)		Børge Obel	Centerdirektør	Systemniveau
Det Kgl. Bibliotek	IT-udvikling og infrastruktur	Bjarne Andersen	Vicedirektør for IT-udvikling og Infrastruktur	Systemniveau
Det Koordinerende Organ for Registerforskning (KOR)		Jeppe Klok Due	Specialkonsulent	Systemniveau
Nationalt Forum for Forskningsdata Management (DM Forum)		Anders Sparre Conrad	Formand	Systemniveau
Research Data Alliance (RDA)/DeiC		Lene Krøl Andersen	Kompetencecenterchef	Systemniveau
Danish e-Infrastructure Co-operation (DeIC)		Steen Pedersen	CEO	Systemniveau
Rigsarkivet	Direktionen	Kirsten Villadsen	Vicedirektør	Systemniveau
Danmarks Forsknings- og Innovationspolitiske Råd (DFiR)		Jens Oddershede	Formand	Forskningsfinansierende aktør
Innovationsfonden	Direktionen	Peter Høngaard Andersen	Direktør	Forskningsfinansierende aktør
Novo Nordisk Fonden		Thomas Alslev Christensen	Head of Operations	Forskningsfinansierende aktør
Copenhagen Business School (CBS)	Afdeling for Forskerservice	Lars Nondal	Chefkonsulent	Forskningsinstitut
Danmarks Tekniske Universitet (DTU)	Afdeling for Innovation og Sektorudvikling	Katrine Flindt Holmstrand	Projektleder	Forskningsinstitut
Danmarks Tekniske Universitet (DTU)	Afdeling for Innovation og Sektorudvikling	Paula Maria Martinez Lavanchy	Projektmedarbejder	Forskningsinstitut
Danmarks Tekniske Universitet (DTU)	Afdeling for Innovation og Sektorudvikling	Falco Jonas Hüser	Projektmedarbejder	Forskningsinstitut
Danske Universiteter	Rektorkollegiet	Nikolaj Helm-Petersen	Souschef	Forskningsinstitut

Det Kongelige Danske Kunstakademis Skoler (KADK)	Forskning og kunstnerisk udviklingsvirksomhed	Henrik Oxvig	Forskningsleder	Forskningsinstitut
Københavns Universitet (KU)	Science	John Renner	Dekan	Forskningsinstitut
Roskilde Universitetscenter (RUC)	Administrationen	Peter Kjær	Prorektor	Forskningsinstitut
IT-Universitetet	Forskningsadministration	Jens Christian Godskesen	Prorektor	Forskningsinstitut
Statens Serum Institut		Mads Melbye	Sektordirektør	Forskningsinstitut
Syddansk Universitet (SDU)	Universitetsbibliotek	Bertil F. Dorch	Bibliotekschef	Forskningsinstitut
Aalborg Universitet (AAU)	Det Tekniske Fakultet for IT og Design	Henrik Pedersen	Dekan	Forskningsinstitut
Aarhus Universitet (AU)	SAMF	Peter Munk Christiansen	Instituttleder	Forskningsinstitut
Københavns Universitet (KU)	SUND	Merete Osler	Professor	Forsker
Københavns Universitet (KU)	HUM – Nordisk Forskningsinstitut	Bente Maegaard	Chefkonsulent	Forsker
Københavns Universitet (KU)	NAT/KUNST	Christina Margariti	Ph.d.	Forsker
Syddansk Universitet (SDU)	SUND	Kirsten Ohm Kyvik	Instituttleder	Forsker

Interviewene er hovedsagelig gennemført over telefon ud fra en semistruktureret interviewguide, der er tilpasset til den enkelte aktørtype. Formålet med interviewene har først været at belyse informanternes vurdering og brug af åbne forskningsdata generelt, herunder hvordan de arbejder med åbne forskningsdata. Derefter er der spurgt til informantens vurdering af FAIR-principperne såvel som potentialet for at indføre FAIR-principperne i Danmark, herunder mulige gevinster og udgifter/udfordringer. I hvert interview har konsulent skitseret den tyske case vedrørende implementeringen af FAIR data med fokus på 1) modellen for FAIR data, 2) muligheder/barrierer og 3) gevinster og efterfølgende bedt informanten om at vurdere, om og hvordan de tyske erfaringer kan overføres til en dansk kontekst.

6. Referencer

- Mansfield, E. (1991), 'Academic research and industrial innovation,' *Research Policy* 20(1), 1991, pp. 1-12.
- Mansfield, E. (1998), 'Academic research and industrial innovation: an update of empirical findings,' *Research Policy* 26(7/8), 1998, pp. 773-776.
- Evans, P., Hatcher, M. and Whittard, D. (2008), 'The preliminary satellite account for the UK: a sensitivity analysis,' *Economic & Labour Market Review* 2(9), September 2008, pp. 37-43.
- Sveikauskas, L. (2007), *R&D and Productivity Growth: A Review of the Literature*, US Bureau of Labor Statistics, Washington DC, Working Paper 408.
- Finansministeriet (2017), *Vejledning i samfundsøkonomiske konsekvensvurderinger*.
- Sullivan, K. P., Brennan-Tonette P. & Marxen L. J. (2017), *Economic Impacts of the Research Collaboratory for Structural Bioinformatics*.
- Beagrie N. & Houghton J. (2016), *The Value and Impact of the European Bioinformatics Institute*.
- Europa-Kommissionen (2015), *Creating Value through Open Data*.
- OECD (2015), *Assessing government initiatives on public sector information*.
- Houghton, J. W. & Gruen, N. (2014), *Open Research Data*.
- Beagrie N. & Houghton J. (2014), *The Value and Impact of Data Sharing and Curation. A synthesis of three recent studies of UK research data centres*.
- Lateral Economics (2014), *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*.
- Schrier, B. (2014), *Government Open Data: Benefits, Strategies and Use*.
- Beagrie, N. and Houghton J. W. (2013a), *The Value and Impact of the Archaeology Data Service: a study and methods for enhancing sustainability*.
- Beagrie, N. and Houghton J.W. (2013b), *The Value and Impact of the British Atmospheric Data Centre*.
- McKinsey (2013), *Open Data: Unlocking innovation and performance with liquid information*.

De Vries (2012), Re-use of Public Sector Information - Catalogue and highlights of studies, cases and key figures on economic effects of changing policies.

Beagrie, N. & The Centre for Strategic Economic Studies University of Victoria (2012), Economic Impact Evaluation of the Economic and Social Data Service.

Houghton, J. W., Swan, A., & Brown, S. (2011), Access to Research and Technical Information in Denmark, Forsknings- og Innovationsstyrelsen and Denmark's Electronic Research Library.

Houghton, J. W. (2011), Costs and Benefits of Data Provision.

Vickery (2011), Review of recent studies on PSI re-use and related market developments.

Zangenberg og Co. (2011), Kvantificering af værdien af åbne offentlige data.

Beagrie, N., Lavoie, B. & Woollard, M. (2010), Keeping Research Data Safe 2.

Houghton, J. W. & Sheehan P. (2009), Estimating the Potential Impacts of Open Access to Research Findings.

Beagrie, N. (2009), Draft Guide to Cost/Benefit Analyses for Research Data Services.

Fry, J., Lockyer, S., Oppenheim, C., Houghton, J. W., & Rasmussen, B. (2008), Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes, London and Bristol: JISC.

Dekkers, M., Polman, F., te Velde, R., & de Vries, M. (2006), MEPSIR: Measuring European Public Sector Information Resources, Brussels: European Commission.

PIRA (2000), Commercial exploitation of Europe's public sector information.

Jaffe, A. (1989), 'Real effects of academic research,' American Economic Review 79, pp. 957-970.

Coe, D. T. and Helpman, E. (1993), International R&D Spillovers, NBER Working Paper 4444, National Bureau of Economic Research, Cambridge MA.

Coe, D. T. and Helpman, E. (1995), 'International R&D Spillovers', European Economic Review 39, pp. 859-887.

Arundel, A. and Geuna, A. (2004), 'Proximity and the use of public science by innovative European firms,' Economics of Innovation and New Technology 3(6), pp. 559-580.

- Verspagen, B. (2004), The Impacts of Academic Knowledge on Macroeconomic Productivity Growth: An Exploratory Study, Eindhoven Centre for Innovation Studies, Eindhoven.
- DeIC, DEFF (2015), National strategi for forskningsdata management 2015-2018, København, den 30. januar 2015
- Horizon 2020 Programme (2016), Guidelines on FAIR Data Management in Horizon 2020
- Force11 (2017), Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing
- Kommissionen (2017), European Open Science Cloud
- Max Planck-Gesellschaft (2002), Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities
- Mons, Barends, Cameron Neylon, Michel Dumontier, Luiz Olavo Bonino Sila Santos & Mark D. Wilkinson (2017), Cloudy, Increasingly FAIR – Revisiting the FAIR Data Guiding Principles for the European Open Science Cloud
- OECD (2007), Principles and Guidelines for Access to Research Data from Public Funding
- Regeringen (2016), Forsknings- og innovationspolitisk redegørelse
- Wessels, Bidgette, Rachel L. Finn, Kush Wadhwa, Thordis Scainsdottir, Lorenzo Bigagli, Stefano Nativi & Merel Noorman. (2017), Open Data and Knowledge Society, Amsterdam University Press
- Wilkinson et al. (2016), The Fair Guiding Principles for Scientific Data Management and Stewardship.
- Graham Pryor (2012): Why manage research data?
- Bundesministerium für Bildung und Forschung (2016): Open Access in Deutschland – Die Strategie des Bundesministeriums für Bildung und Forschung
- Rat für Informationsinfrastrukturen (2016): Enhancing Research Data Management: Performance Through Diversity
- Allianz der deutschen Wissenschaftsorganisationen (2015): Positionspapier „Research data at your fingertips“
- Germany and The Netherlands (2017): Joint Position Paper on the European Science Cloud
- Hochschulrektorenkonferenz (2014): Management of research data – a key strategic challenge for university management



DANMARK

Oxford Research A/S
 Falkoner Allé 20
 2000 Frederiksberg
 Danmark
 Tel: (+45) 3369 1369
 office@oxfordresearch.dk

NORGE

Oxford Research AS
 Østre Strandgate 1
 4610 Kristiansand
 Norge
 Tel: (+47) 4000 5793
 post@oxford.no

SVERIGE

Oxford Research AB
 Norrlandsgatan 11
 103 93 Stockholm
 Sverige
 Tel: (+46) 08 240 700
 office@oxfordresearch.se

FINLAND

Oxford Research Oy
 Fredrikinkatu 61a, 6krs.
 00100 Helsinki, Suomi
 www.oxfordresearch.fi
 office@oxfordresearch.fi

BRUXELLES

Oxford Research
 C/o ENSR
 5. Rue Archimède
 Box 4, 1000 Brussels
 www.oxfordresearch.eu
 office@oxfordresearch.eu