

Forvaltning af forskningsdata i Danmark

En undersøgelse af danske universiteters og forskningsråds praksisser, politik og visioner for lagring, langtidsbevaring, tilgængeliggørelse og deling af forskningsdata.

Udarbejdet af:

Jesper B. Thestrup, Filip Kruse, Lars Nondal, Bertil F. Dorch, Mikkel Andersen, Niels Jørgen Blaabjerg, Thea Drachen, Jeannette Ekstrøm, Mikael Elbæk, Ole Ellegaard, Karsten Kryger, Asger V. Larsen, Diba Markus, Anna Mette Morthorst, Erik Sonne, Sacha Zurcher og Ellen V. Knudsen

Indhold

1. Executive summary.....	3
2. Indledning	5
3. Forskningsdata er kommet på dagsordenen	7
3.1 De bevilligende myndigheder - forskningsrådene	10
4. Forvaltning af forskningsdata i Danmark	12
4.1 Understøtter universitetet forskningsprocessen med henblik på at opbevare data (storage)?.....	14
4.2 Understøtter universitetet forskningsprocessen ved at sikre forskernes data for fremtiden (data preservation)?	15
4.3 Har universitet på linie med initiativer fra andre universiteter i blandt andet Europa, USA og Australien gjort sig nogle overvejelser mht. at etablere eget dataarkiv eller indgå i samarbejde omkring et?	16
4.4 Har universitetet tiltag i forhold til at sikre forskernes opmærksomhed på spørgsmålet om bevaring, synliggørelse og evt. tilgængeliggørelse af forskningsdata?	17
4.5 Indgår eksponering af forskningsdata som en del af universitetets strategi?	17
4.6 Undersøgelsens konklusioner og den nuværende udvikling i forbindelse med forvaltning af forskningsdata	18
5. Forskningsbibliotekerne og forvaltningen af forskningsdata.....	19
6. Data management i udvalgte lande - fælles og forskelligt	20
7. Hvad er en data management plan	22
7.1 Rammerne for indholdet.....	22
7.2 Gennemgående elementer i data management planer	22
7.3 Eksempel på indhold i en data management plan	23
7.4 Eksempler på data management planer i praksis.....	25
8. Konklusioner og anbefalinger	26
9. Litteratur	28

1. Executive summary

The long range ambition of this project is to improve the management of research data in Danish universities and other research institutions. In order to achieve this, the first step on the road will be to establish a catalogue of the knowledge and experience already existing on the subject of research data management in Denmark. One central aim of this project is to fulfill this.

We present findings from an interview based survey on the practices of Danish universities regarding storage and preservation of research data and the extent to which data are made accessible to the research community and others, by means of policies, open access etc.

To provide the necessary background for reflection on the Danish practice of research data management we analyze selected developments from abroad - both from the field of practice and from the field of policy. To analyze this is another central aim of this project. The analysis underlines the importance of a strong focus on description of research data, content, formats, metadata etc. and of organizational and policy issues.

Among our findings is that as a rule Danish universities support research data storage, although in different ways. This uneven practice does not imply a lack of interest or awareness on behalf of the universities; rather it reflects differences in scientific traditions and cultures and especially differences in the use and organization of IT. Long term preservation of data is not widespread, apart from the data which are preserved in the Danish Data Archive and various international systems. Although the universities seek to enhance the researchers' awareness of the importance of data management by different means, we cannot ascertain the existence of a common overall strategy for the management of research data. An appropriate characterization of the attitude and practice of the universities could be 'wait-and-see'.

The various research councils which fund research seem to adopt an attitude and practice similar to that of the universities. As a rule they welcome data sharing and the establishment of a national solution to the problems of data preservation. Closely related to this is their recognition of the need for legal clarification of issues relevant to the question of access to research data.

Based on our findings we put forward these six recommendations for the preservation and sharing of research data, this being another central aim of the project:

- A more binding awareness on the importance of the issue from the part of universities and researchers.
- Formulation of a national policy on data preservation and sharing.
- An addressing of the legal problems.
- Analysis of the researchers' behavior in relation to data preservation and sharing and their future needs in this field.
- Formulation of a policy regarding data preservation and sharing on the part of the research councils and funds.

- Establishment of one or more technical systems in order to facilitate the activities and work processes.
- A safeguard from the part of the universities and the research libraries to ensure that the relevant members of staff have the required qualifications.

2. Indledning

På langt sigt er målet med dette projekt at forbedre forvaltningen af forskningsdata eller som det også kaldes, 'data management'. Et skridt på vejen til dette mål - og et vigtigt formål med projektet - er etablering af et katalog over eksisterende viden og erfaringer, primært udenlandsk baserede, men også fra Danmark. Nærværende rapport skal ses som et sådant foreløbigt videnskatalog. I udlandet er der generelt en bredere erfaringsbaggrund at trække på, hvorfor denne umiddelbart kan virke nærliggende at tage afsæt i. Men en forudsætning for at reflektere anvendelsesorienteret over disse erfaringer vil være, at eksisterende praksis i Danmark forinden er analyseret således, at der kan foretages koblinger herimellem og relevante dele af den udenlandske praksis. Hvad forvaltning af forskningsdata er og hvilke aktiviteter, der kan indgå heri, introducerer vi i kapitel 3.

En væsentlig del af projektet har bestået i en interviewbaseret undersøgelse - som kan læses i kapitel 4 - af praksis på danske universiteter. Særligt fokuseredes i denne på, om lagring ('storage') af forskningsdata i løbet af forskningsprocessen støttes, om langtidsbevaring ('preservation') af de færdige data og deling og tilgængeliggørelse af data - herunder muligheden for etablering af dataarkiv(er) - tilsvarende støttes.

Undersøgelsen af den udenlandske praksis for forvaltning af forskningsdata viser, at netop disse forhold er af central betydning. De hviler på flere forudsætninger blandt andet, at data er velbeskrevne, for så vidt angår indhold, formater og metadata, at der består fornøden datasikkerhed, at en organisation med ressourcer til at varetage arbejdet er etableret og endelig, at der eksisterer en politik - national eller institutionsbaseret - for adgang til og hermed også for deling og genbrug af forskningsdata.

I Danmark viser undersøgelsen, at universiteterne generelt foretager lagring af forskningsdata, forstået som forskellige back-up aktiviteter i forbindelse med forskningsregistrering og dette især i forbindelse med forskellige projekter. Praksis er dog uensartet. Det antages at skyldes forskellige faglige traditioner og ikke mindst forskellig praksis i forhold til IT, uden at dette dog undersøges nærmere. Men universiteterne er opmærksomme på området og dets betydning. Langtidsbevaring er ikke udbredt - bortset fra de data, der bevares i Dansk Data Arkiv - og universiteterne har fokus på specielt omkostningssiden ved langtidsbevaring. De søger på forskellig vis at skærpe forskernes opmærksomhed på hvordan forskningsdata forvaltes, primært gennem kampagner og forskelligt støttemateriale. Men vi har ikke kunnet konstatere en fælles overordnet strategi for dataforvaltning på universiteterne.

Dette billede af en nærmest afventende praksis på området genfinder vi hos forskningsrådene. Undersøgelsen viser, at der ses klare fordele ved deling af data og ved etablering af en national løsning i form af et fælles dataarkiv. Samtidig peges på vigtigheden af at afklare de juridiske forhold omkring deling og anden tilgængeliggørelse af data.

Ud over de her nævnte aktører, er to andre væsentlige. Dels forskningsbibliotekerne – se kapitel 5 - som på forskellig vis medvirker i forskningsregistreringen og har en væsentlig ekspertise inden for digital formidling og -bevaring. Dels Danish E-infrastructure Cooperation (DEIC), som driver forskningens digitale infrastruktur i form af Forskningsnettet. Disse aktørers nærmere virksomhed undersøges ikke særskilt.

Hovedsigtet med analysen af data management i Danmark er at give et aktuelt overblik, når det gælder praksis i udlandet trækker vi dels sammenlignelige træk frem, dels betoner vi de specifikke omstændigheder, der er af betydning. En kort fremstilling kan læses i kapitel 6, mens bilag 12-19 er mere dybtgående behandlinger af praksis og politik i de enkelte lande. Sigtet er her at vise, hvorledes dataforvaltning kan indgå i selve planlægningen af forskningsprojekter - også kaldet 'data management planning' eller DMP. Dette er i stigende grad ved at blive en betingelse for tildeling af offentlige forskningsmidler og viser sig samtidig som et redskab, der kan lette forskernes arbejde.

Hvad der kan være indholdet i en data management plan, uddyber vi i kapitel 7, hvor forskellige modeller for indhold nærmere diskuteres. I bilag 19-21 kan læses autentiske eksempler på sådanne planer inden for forskellige fagområder. Spørgsmålene om deling og anden tilgængeliggørelse af forskningsdata indgår som oftest i dataforvaltningsplaner, i bilag 1-3 opridser vi eksemplarisk hovedtendenserne inden for tre fagområder.

Det er også et væsentligt formål med projektet at fremkomme med anbefalinger ud fra de samlede undersøgelses- og analyseresultater:

- En større og mere forpligtende opmærksomhed fra universiteternes og forskernes side på bevaring og deling af forskningsdata.
- Formulering af en national politik på området.
- Undersøgelse af hvorledes forskerne aktuelt bevarer og deler forskningsdata og deres fremtidige behov på dette område.
- Håndtering af de juridiske problemer, der er forbundet med bevaring og deling af forskningsdata.
- Formulering af en egen politik på dette område fra forskningsråds og -fondes side.
- Etablering af et eller flere tekniske systemer til at understøtte de forskellige aktiviteter inden for databevaring og datadeling.
- Sikring af de nødvendige kvalifikationer hos de relevante medarbejdere på universiteterne og forskningsbibliotekerne.

3. Forskningsdata er kommet på dagsordenen

Der kan peges på både teknologiske, politiske og økonomiske begrundelser for, at forskningsdata og håndteringen af forskningsdata, er kommet på dagsordenen. Det er sket i en sådan grad at de bevilgende myndigheder og forskningsmiljøerne selv på forskellig måde begynder at formulere ændrede praksisser og politikker på området.

Den følgende korte gennemgang af de bagvedliggende årsager rummer kun udvalgte nedslag og eksempler, primært hentet fra udviklingen i Storbritannien.

Den *teknologiske* baggrund vedrører i høj grad den digitale revolution. Pryor taler om "the data deluge" (Pryor, 2012), og nævner, at den årlige vækst i produktionen af forskningsdata i elektronisk form er i omegnen af 30 %. Og de største datamængder findes indenfor naturvidenskaberne, eksempelvis indenfor bio-videnskaberne (f.eks. genom-projektet og 'high energy physics' (Se hertil bilag 1 til 3 for en nærmere uddybning).

Udover de store digitale datamængder, der genereres i forskningsprojekter, giver den digitale udvikling generelt i samfundet også forskere muligheder for at arbejde med store mængder digitale data eller kilder. Internettet og de sociale medier giver nye muligheder, måske især indenfor humaniora og samfundsvidenskaberne, hvor der er opstået begreber som 'digital humanities' og 'computational social science'.

Den måske seneste trend er 'big data'¹, som både er relevant for planlæggere, virksomheder og egentlig forskning. 'Big data' som begreb dækker over de store datamængder, som mennesker efterlader sig f.eks. på de sociale medier eller i form af transaktionsdata i net-betalingsystemer, andre 'digitale fodspor' og meget andet. Disse data kan eksempelvis analyseres ved hjælp af data-mining, og derved bruges til analyser og evt. også forudsigelser af befolkningsgrupperes adfærd, til brug for byplanlægning, trafikforskning, markedsføring med mere.

Den *politiske baggrund* kan stærkt simplificeret karakteriseres med en række begreber, der alle starter med 'Open'. Lige fra 'Open Society' og 'Open Government' på et overordnet, ideologisk plan, til 'Open Science' og 'Open Research' for forskningens generelle rolle i samfundet. På det helt konkrete plan betegner 'Open Access' og 'Open Data' fri adgang til forskningens resultater, hvad enten det er i form af publikationer eller data.

¹ Se eksempelvis McKinsey (2011); Big data: The Next Frontier for innovation, competition and productivity; McKinsey 2011

Begrebet Open Society forbindes måske oftest med filantropen George Soros' Open Society Foundation², hvis erklærede mål er: "to build vibrant and tolerant democracies whose governments are accountable to their citizens".

Den konkrete anledning var murens fald og de østeuropæiske landes farvel til kommunismen. Dette initiativ og en lang række andre lignende initiativer er udtryk for en politisk trend i retning at skabe et mere åbent og gennemsigtigt samfund, hvor begreber som 'transparency' og 'accountability' er centrale og hvor statsmagt og myndigheder står til regnskab for deres handlinger overfor civilsamfundet og den enkelte borger.

'Open Government' er nu et centralt begreb i den politiske debat og der eksisterer f.eks. et netværk kaldet Open Government Partnership (OGP)³, startet i 2011, med lande som Norge, USA og Storbritannien blandt stifterne og med Danmark som en af første medunderskrivere. Et par citater fra hjemmesiden:

We all want more transparent, effective and accountable governments - with institutions that empower citizens and are responsive to their aspirations (...)....to secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance

Det kan føles som om der er langt fra sådanne overordnede politiske proklamationer til begreber som 'Open Science', 'Open Research' og måske især 'Open Access' og 'Open Data', der er velkendte begreber i forsknings- og biblioteksverdenen. Men der er faktisk en i en række tilfælde en klar sammenhæng:

Soros' Open Society Foundation arbejder meget med medier og forskning og var medinitiativtager til Budapest Open Access Initiative⁴. 'Open data' er et centralt element i den britiske regerings Open Government-initiativ.

Scholarly Publishing and Academic Resources Coalition (SPARC)⁵ har etableret The Alliance for Taxpayers Access⁶ for, som det fremhæves på websiden: "American taxpayers are entitled to the research they've paid for".

Så åben adgang til forskningsdata er ikke en isoleret størrelse, der kun har sin oprindelse i interne forhold i forskningsverdenen. Det indgår i en større diskurs.

² <http://www.opensocietyfoundations.org/>

³ <http://www.opengovpartnership.org/>

⁴ <http://www.opensocietyfoundations.org/openaccess/read>

⁵ <http://www.arl.org/sparc/about/index.shtml>

⁶ <http://www.taxpayeraccess.org/>

Derudover er der selvfølgelig en række *økonomisk bevæggrunde* for interessen for forskningsdata, som ikke kan adskilles fra de politiske og teknologiske trends nævnt ovenfor. Og også her er begreber som 'accountability' og 'transparency' centrale. I en mere snæver økonomisk fortolkning af begrebet 'accountability' end behandlet ovenfor, drejer det sig blandt andet om at håndtere offentlige udgifter - og offentligt genererede værdier - bedst muligt og så ansvarligt som muligt. Dette gælder også forskningsinstitutioner, hvor det forventes, at de på linie med andre offentlige institutioner, kan dokumentere, at bevillinger forvaltes ordentligt og, at skatteyderne får noget for deres penge.

Kravet om ansvarlig omgang med offentlige, skatteyderbetalte, midler kommer blandt andet til udtryk i IT-revisionens henvendelse til en række danske universiteter, blandt andet Copenhagen Business School (CBS) og Aalborg Universitet (AAU), om at redegøre for hvordan de håndterer forskningsdata.

At der er interesse for at holde øje med bevillinger til forskning generelt, og herunder om forskningsdata forvaltes ordentligt og økonomisk ansvarligt, kan yderligere begrundes i den kraftige vækst i forskningsudgifterne. Som eksempel brugte EU i FP7-rammeprogrammet⁷ ca. 50 mia. euro i perioden 2007-2013. Det forventes at stige til ca. 80 mia. euro i næste rammeprogram Horizon2020⁸, for perioden 2014-2020.

Bortset fra de måske lidt bogholder-agtige argumenter for at have styr på forskningsudgifterne, og sikre, at forskerne og institutionerne forvalter de bevilgede midler ordentligt, er der også en række andre økonomiske argumenter eller rationaler for at interessere sig for forskningens økonomiske effekter. Forskning anses i sig selv som en vækstmotor. Væksten i store dele af den industrialiserede verden er gået i stå, og politikere og andre leder efter mekanismer, der kan befordre vækst og innovation i samfundet. De leder efter mekanismer, der kan øge vidensindholdet i produktionen, og dermed produktiviteten og væksten. Og et middel til dette er mere forskning.

Men det må også gerne gå hurtigere. Hurtigere forskningsresultater og navnlig hurtigere formidling af disse til virksomheder og erhvervsliv, anses for centrale elementer i innovationspolitikken. Og for eksempel 'data sharing', i form af øget og nemmere adgang til forskningsdata og andre offentligt genererede data, forventes at have effekter i form af innovation, vækst og teknologisk udvikling i den private sektor. At dele forskningsdata er dog ikke uden juridiske udfordringer (Se bilag 4 for en kort indføring i disse).

Et andet økonomisk argument vedrører det uhensigtsmæssige i den ringe grad af genbrug af producerede forskningsdata i forskningsverdenen selv - at forskere ofte starter forfra og gentager eksperimenter og dataindsamling, som andre forskere har foretaget tidligere. Genbrug ('re-use') og 'data sharing' er både billigere og mere befordrende for hurtigere forskningsresultater end hvis, som det påstås nu, forskerne ikke genbruger og deler data.

⁷ http://ec.europa.eu/research/fp7/index_en.cfm

⁸ http://ec.europa.eu/research/horizon2020/index_en.cfm

3.1 De bevilligende myndigheder - forskningsrådene

Men hvordan kommer den øgede fokus på forskningsdata så til udtryk hos de bevilligende myndigheder, her forstået som forskningsrådene? Der er selvfølgelig forskelle landene imellem, men et af de steder, hvor der tales meget om forskningsdata, er Storbritannien.

De forskellige engelske forskningsråd, der i alt årligt fordeler knap 4 mia. euro til forskningsprojekter, besluttede i 2011 nogle fælles principper for håndteringen af forskningsdata, RCUK Common Principles on Data Policy⁹, hvor første 'bullet point' fastslår:

Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.

På denne baggrund er alle Storbritanniens universiteter og forskningsinstitutioner, der søger om støtte fra Engineering and Physical Science Research Council (EPSRC) blevet bedt om at udarbejde 'road maps' for deres arbejde med at kunne leve op til disse krav. De skulle aflevere foreløbige 'road maps' pr. 1. maj 2012 og har så indtil 1. maj 2015 til at sikre, at de harmonerer med forskningsrådenes krav.

Det vil ikke være uden konsekvenser for institutionerne ikke at kunne opfylde disse krav – som det hedder hos EPSRC:

If, after 1st May 2015, an institution is found to be deliberately obstructing the sharing of research data or otherwise seriously failing to comply with the EPSRC's expectations, then the EPSRC may ultimately withdraw its funding.

I følge en mail fra Martin Donnelly fra Digital Curation Center (DCC)¹⁰ er det pt. lidt uklart, om alle EPSRC-støttede universiteter har fået udarbejdet disse 'road-maps', da der oprindeligt ikke var noget krav om, at de skulle godkendes af eller overhovedet sendes til EPSRC. Ifølge Donnelly er der stor forskel på detaljeringsgraden i de 'road-maps', der er blevet indsendt.

En eksemplarisk udgave (ifølge DCC) af en sådan 'road map' er 'University of Bath Roadmap for EPSRC. Compliance with Research Data Management Expectations'¹¹,

Forskningsrådene forventer, at forskerne indsender Data Management Plans (DMPs), hvori de redegør for hvilke data, de har tænkt sig at arbejde med i deres forskningsprojekt, om det er data, der findes i forvejen eller data, de selv skal producere, hvad de har tænkt sig at gøre ved dem i løbet af forskningsprocessen og endelig, hvordan de har tænkt sig at arkivere dem, publicere dem eller på anden måde stille dem til rådighed for andre.

⁹ <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

¹⁰ modtaget januar 2013

¹¹ <http://www.bath.ac.uk/rds/University-of-Bath-Roadmap-for-EPSRC.pdf>

En anden central, international drivkraft bag den øgede opmærksomhed på forskningsdata, der måske endda var tidligere ude end de britiske forskningsråd, er EU. EU's arbejde med forskningsdata gennemgås nærmere i afsnit 7. Her skal bare som opsummering af baggrunden for interessen for forskningsdata nævnes, at EU i rapporten *Riding the wave: how Europe can gain from the rising tide of scientific data* (EU, 2010) fremhæver, hvor vigtig en ressource forskningsdata er for samfundet, hvorfor der skal investeres i både forskningsinfrastruktur og lagring af forskningsdata:

Our vision is a scientific e-infrastructure that supports seamless access, use, reuse and trust of data. In a sense the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure - a valuable asset, on which science, technology, the economy and society can advance.

II EU's nuværende rammeprogram (FP7) har der været pilotprojekter inden for bestemte områder, hvor der blev stillet krav Open Access-adgang til de publikationer forskningsprojekterne resulterede i. Interessen for og betydningen af forskningsdata, fører nu til, at i det næste rammeprogram, Horizon2020, forventes på tilsvarende vis introduceret pilotprojekter inden for udvalgte områder, hvor kravet om Open Access vil gælde for både publikationer og data. Og krav om Open Access til publikationerne vil komme til at gælde for alle områder.

Lignende tendenser, men med nationale forskelle, går igen i lande som Holland, Australien, USA med flere, som det fremgår af landerapporterne i bilagsdelen (bilag 12-18). De bevilgende myndigheder i form af forskningsrådene begynder at stille krav til institutionerne og forskerne i forhold til håndteringen af data. Hvorvidt det ligeledes er tilfældet i Danmark er temaet for den interviewbaserede undersøgelse, der afrapporteres i afsnit 4. Her vil vi nøjes med at konstatere, at de danske forskningsråd i hvert fald ikke adresserede spørgsmålet om forskningsdata i den fælles Open Access-politik, der blev offentliggjort i sommeren 2012 (Det Frie Forskningsråd et. al., 2012). Den undtager eksplicit forskningsdata fra kravet om Open Access-publicering.

4. Forvaltning af forskningsdata i Danmark

Dette afsnit beskriver de resultater som projektgruppen fik ved en interviewundersøgelse af de danske universiteter og en række råd og fonde. Baggrunden var et ønske om at afdække aktuel status, herunder hvorledes arbejdet omkring forskningsdata er organiseret og hvilke umiddelbare planer, der foreligger for arbejdet med disse på universiteterne. For fondenes vedkommende ønskede vi at undersøge eventuelle krav fra bevillingsgiverne, da vi har set en tendens til at fondene i udlandet stiller sådanne krav.

- Det skal bemærkes, at digitale forskningsdata og forskningsobjekter er ikke entydige størrelser. I forbindelse med interviewundersøgelsen har vi ikke anvendt en bestemt, teoretisk funderet definition. Målet med dette var at få respondenterne til at give så mange oplysninger som muligt, uden at vi styrede dialogen ud fra en bestemt teoretisk forståelse. For en nærmere gennemgang af hvad forskningsdata kan omfatte, se bilag 5.-

Projektgruppen valgte ikke at undersøge private forskningsinstitutioner, da gruppen vurderede, at disse har helt specielle krav, når det drejer sig om deling af data. Det er på ingen måde uvigtigt at afklare dette spørgsmål, men det vil kræve et selvstændigt forskningsprojekt. De private forskningsinstitutioner kunne overvejes inddraget i en proces omkring etablering af en dansk infrastruktur til langtidsopbevaring af forskningsdata.

Universiteternes oplysninger blev indsamlet ved mundtlige interviews ud fra på forhånd fremsendte spørgsmål, fondene havde mulighed for skriftlig besvarelse eller telefonisk interview. Samme interviewer gennemførte alle interviews for at sikre ensartethed i dataindsamlingen. Bortset fra to interviews, deltog d et til to 'lokale' medlemmer fra projektgruppen i interviewene for at sikre, at relevante lokale forhold blev inddraget. Der blev i alt gennemført otte interviews og et universitet besvarede spørgerammen telefonisk.

På forhånd havde gruppen ikke foreskrevet, hvem der ønskedes interviewet, men i stedet stilet brevene til selve institutionen. Det var således op til den enkelte at vælge hvem den ønskede skulle repræsentere organisationen. Det skulle – i en vis grad – sikre, at de enkelte modtagere kunne udvælge de mest egnede respondenter. Projektgruppen ønskede at gennemføre undersøgelsen på så højt et organisatorisk niveau som muligt, hvilket også lykkedes. En samlet oversigt over respondenter kan ses i bilag 6, og spænder over central ledelse på institutionerne, over deres it-afdelinger og til bibliotekerne.

Undersøgelsens fokus på ledelsesniveauet skal ikke ses som en negligering af behovet for viden om forskernes position i forhold til arbejdet med forvaltning af forskningsdata, men som en efterlevning af projektets målsætninger.

Det skal her bemærkes, at en undersøgelse fra 2008 blandt forskerne på Aarhus Universitet (Sørensen et. al, 2009) af forskningskommunikation og implikationerne for bevaring af forskningsdata, peger på flere interessante forhold: forskerne udtrykker et klart behov for bedre bevaringsmuligheder end de eksisterende. Videre ønsker de bevaring af foreløbige forskningsdata og ser gerne, at disse gøres

tilgængelige for andre end de involverede forskere. Ansvar for bevaring ses som i højere grad forskernes eget ansvar end institutionens eller universitetets.

En opfølgning på dele af denne undersøgelse ville have været ønskelig, men har ikke været mulig inden for rammerne af nærværende projekt.

For at sikre sammenhæng mellem besvarelserne fra universiteterne blev en række fonde og forskningsråd bedt om at besvare samme spørgeramme som universiteterne. Sideløbende gennemførtes en omfattende søgning på nettet og der blev optaget en række uformelle direkte telefoniske kontakter til nærmere uddybning af besvarelserne. Undersøgelsen bygger på oplysninger fra tretten forskellige råd, fonde og myndigheder, inklusive Forsknings- og Innovationsstyrelsen (se bilag 7).

Spørgerammen blev opdelt i fem forskellige spørgsmål (se bilag 8). Et, hvor universiteterne blev spurgt hvorvidt det pågældende universitet støttede forskningsprocessen i forhold til at sikre opbevaring (storage) af forskningsdata. Et, hvor de blev spurgt, om de sikrer langtidsbevaring af forskningsdata (preservation). Derefter fulgte to spørgsmål, der på forskellig måde berørte tilgængeliggørelse af forskningsdata. Videre et spørgsmål om det pågældende universitet ville oprette (eller indgå i) et dataarkiv og et spørgsmål om forskningsdata indgår i branding af universitet. Endeligt blev universiteterne spurgt hvorvidt de søger at øge forskernes opmærksomhed omkring lagring, bevaring og tilgængeliggørelse af forskningsdata. Spørgerammen til rådene og fondene afveg en smule herfra, da disse naturligvis ikke kunne svare på spørgsmål om universiteternes interne policies (se bilag 9).

Spørgsmålene blev formuleret på baggrund af projektgruppens research i data management planning. I litteraturen herom understreger Donnelly (2011) og Jensen (2011) at data skal kunne gemmes, langtidsopbevares og deles for at kunne opfylde de krav, data management planning stiller forskeren. For at afklare dette nærmere valgte gruppen at stille de fire spørgsmål om lagring, bevaring og deling af data. Det sidste spørgsmål om at påvirke forskernes opmærksomhed skulle belyse universiteternes forsøg på at udvikle de eksisterende aktiviteter.

Analysen er struktureret ud fra spørgsmålene og sammenfatter svar fra såvel universiteter som fonde. Da der er forespurgt centralt på universiteterne, vil de omtalte initiativer også antageligt være aktiviteter koordineret på tværs af det enkelte universitet. Der kan således være fakulteter, institutter eller andre lokale enheder, hvor der f.eks. er tiltag, der ikke koordineres fra centralt hold og som derfor ikke er med i denne analyse. Der er således ikke nødvendigvis tale om et retvisende billede for universitetets enkelte enheder, men om et overblik. Som det blev fremført på workshoppen den 31/10 i Aarhus (se bilag 10) i forbindelse med rapporten, vil der også være forskningsmiljøer hvor f.eks. traditionen for at dele data vil være stærkere end andre, og inden for de enkelte fagdiscipliner vil der være forskellige kulturer, opbygget på baggrund af traditioner, forskellige incitamentsmodeller mm.

Svarene fra universiteterne viser, at der er stor variation i aktiviteterne. Forskellene virker til at skyldes tilsyneladende universiteternes forskellige institutionelle opbygning, forskelle i størrelse og i faglige profiler. Det ses tydeligt af besvarelserne, at respondenterne forventer, at universiteter med et fakultet, og mindre universiteter, har lettere ved at opbygge en service, der sikrer lagring, opbevaring og deling af

forskningsdata. Dette virker, som det fremgå senere, ikke til at være helt korrekt. Universiteterne står basalt set med den samme udfordring, uanset størrelse og faglig orientering.

4.1 Understøtter universitetet forskningsprocessen med henblik på at opbevare data (storage)?

Den generelle tendens er, at universiteterne i princippet stiller rå lagerplads til rådighed i form af netværksdrev til deres forskere. Ofte er politikken den, at forskerne selv bestemmer hvilke typer af opbevaring, de ønsker at bruge men, at de er forpligtede til at sikre, at data er forsvarligt gemt. Ansvaret er ofte forankret på institutterne, der så praktiserer dette forskelligt.

Der er eksempelvis ikke en fælles løsning på tværs af Aalborg Universitet (AAU). Nogle institutter stiller lagerplads til rådighed, andre gør ikke. Der arbejdes på en politik på området, som skal adressere problemet. På IT-Universitet (ITU) er der fokus på storage, som en service, der varetages af IT-afdelingen. Der er en politik på området, der beskriver hvordan opbevaring skal foretages, men det er frivilligt for forskerne, om de vil benytte sig af servicen. Københavns Universitet (KU) tilbyder alle fakulteterne storage, men der er ikke formuleret en politik på området. Erfaringen er her, at forskerne ofte selv finder deres egen løsning. Danmarks Tekniske Universitet (DTU) har som eneste institution både en infrastruktur og en politik på området. Der er dog frihed for de enkelte institutter at forvalte politikken og implementere den teknologi, på den måde de finder er bedst for dem. Der er i skrivende stund en ny samlet infrastruktur for hele institutionen under implementering (se også næste afsnit nedenfor).

Det fravær af services på området, som (med enkelte undtagelser), samlet set præger universiteterne forklares af respondenterne ved:

- at der ikke er fokus på opbevaring, beskrivelse, deling og bevaring af forskernes data. Der er et ensidigt fokus på formidling af forskningsresultaterne i form af publikationer.

- at der hersker stor usikkerhed om, hvad forskningsdata er og hvordan de skal opbevares. Dette fører til de pragmatiske tilgange, som kan spores på institutionerne, hvor ansvaret er decentralt placeret hos fakulteterne og institutterne og hvor det er op til de enkelte forskere at gøre som de mener er bedst - under skyldig hensyntagen til gældende krav vedrørende opbevaring af personhenførbare data og projektets konkurrenceudsathed. Herudover skal forskernes resultater naturligvis kunne dokumenteres og eventuelt også reproduceres.

Der er dermed en vis skepsis i forhold til centrale politikker og valgte teknologier – de såkaldte 'one-size-fits-all'-løsninger, men der foretages flere steder undersøgelser af, hvordan opbevaringen foregår på de enkelte institutioner med henblik på at udforme anbefalinger til ledelsen.

Fondene og rådene har ingen krav eller politikker på området, men forventer, at institutionerne og forskerne overholder gældende lovgivning og følger faglige standarder.

4.2 Understøtter universitetet forskningsprocessen ved at sikre forskernes data for fremtiden (data preservation)?

Der foregår ikke mange aktiviteter vedrørende arkivering af forskernes data på de danske universiteter, det vil sige aktiviteter, som rækker ud over almindelig dataopbevaring. Der er ved at blive etableret nye centrale IT systemer på universiteterne til backup af forskningsdata, men lagring er ikke nødvendigvis det samme som langtidsbevaring og mulig gørelse af deling af data.

Langtidsbevaring kræver, at data lagres på en sådan måde, at de ikke ødelægges ved at der ændres i den basale opbygning af data helt ned på bit niveau, at data kan åbnes og genbruges også selvom softwareformater og hardware ændrer sig samt, at data er søgbare (jævnfør f.eks. <http://digitalbevaring.dk/viden-om/#Bevaring>; Higgins, 2011 og Klumb, 2011).

Kun et enkelt universitet - DTU - har i 2011 indkøbt et sammenhængende storage system, som, når det er færdigimplementeret, vil kunne anvendes til langtidsbevaring og sikker opbevaring af universitetets egne forskningsdata. Men på de øvrige universiteter findes ingen fælles systemer til databevaring. Roskilde Universitet (RUC) udtaler " at de ikke har nogen lagring, der giver sikring af muligheden for fremtidig fortolkning af data fx migration til nye formater eller emulering". Copenhagen Business school (CBS) fremhæver, at "de teknologiske muligheder ikke har været tilstrækkelige".

Langtidsbevaring ses af respondenterne som en ressourcekrævende service. Der er enkelte eksempler på data som langtidsbevares lokalt som tvillingeregistret fra Syddansk Universitet (SDU), der også arkiveres i Dansk Data arkiv (DDA)¹², men SDU ser det ikke som universitetets opgave at sikre alle typer forskningsdata i en længere tidshorisont, det være sig 10, 20 eller 30 år og nævner projektstørrelse som et parameter. Kun særlige forskningsressourcer som universitetet ejer og driver langtidsbevares. KU peger også på manglende lovgivning på området.

Dette standpunkt understøttes også fra øvrige universiteter og på DTU er holdningen eksempelvis, at det ikke nødvendigvis giver mening at transformere alle former for data i en permanent tilgængelig form. DTU nævner også som forudsætning for langtidsbevaring, at data er dokumenteret og tildelt en identifikator, så de kan citeres.

Der findes ikke nogen politikker eller værktøjer som f.eks. data management planer, der nærmere definerer, hvordan lagring af data skal foregå. AAU fremhæver, at 'data life cycle' er længere end blot forskningsprojektperioden.

Blandt de 10 adspurgte fonde og råd har ingen en formuleret politik for sikring af forskernes data for fremtiden. Det Frie Forskningsråd vurderer, " at opbevaring og sikring af forskningsdata i sig selv næppe generelt vil spille en større rolle for rådets støttetilsagn i fremtiden" men er dog enig i grundtanken om at dele forskningsdata" (se bilag 11). Danmarks Grundforskningsråd omtaler det som værende institutionernes ansvar at sikre lagring af data.

¹² <http://www.sa.dk/dda/>

4.3 Har universitet på linie med initiativer fra andre universiteter i blandt andet Europa, USA og Australien gjort sig nogle overvejelser mht. at etablere eget dataarkiv eller indgå i samarbejde omkring et?

Flere universiteter har i varierende grad lokal lagring af forskningsdata (storage). DTU har som nævnt allerede et arkivsystem, der vil kunne varetage langtidsbevaring af egne forskningsdata. Også ITU har eget arkiv til lagring af data. AAU og KB har overvejelser i retning af etablering af eget arkiv, men afventer resultater af et nærmere analysearbejde.

Statens arkiver, herunder Dansk Dataarkiv, nævnes som en aktør på nationalt plan, men det anses for nødvendigt med deltagelse af flere i det Dansk Dataarkiv kun opbevarer en meget lille del af den samlede mængde forskningsdata, der produceres.

Forhold som store forskelle i faglige traditioner, økonomiske omkostninger og forskningsprojekter, der omfatter flere universiteter, taler ifølge respondenterne for at oprette en national løsning til at supplere de eksisterende løsninger.

DTU forestiller sig eksempelvis en infrastruktur med forskellige typer arkiver. Lokale arkiver ses primært til brug for understøtning af lagring på forskningsområder, hvor der i forvejen *ikke* eksisterer internationale infrastrukturer og datalagre. Hertil kommer, at mindre datasæt og bilag til publikationer formodes at kunne lagres i bibliotekets repository. DTU deltager i forvejen i opbygningen af internationale arkivstrukturer inden for 'big science'. DTU forestiller sig ikke én samlet arkivløsning for arkivering af forskningsdata men løsningsmodeller, der udvikler sig i takt med, at krav differentieres i medfør af de eksisterende forskelligheder. Efter afviklingen af interviewene har RUC tilkendegivet, at man ønsker at indgå i en drøftelse af, hvordan en fælles løsning for sikring af forskningsdata for de danske universiteter kan etableres. Også CBS støtter en national eller international dataarkivløsning.

De danske fonde og råd har ikke egne infrastrukturer med tilhørende dataarkiver. Der er blandt fondene og rådene principiel støtte til deling af forskningsdata. En deling vil sikre de bevilligende myndigheder og fonde en bedre udnyttelse af de tildelte midler, da den vil gøre det muligt, at forskningsdata kan bruges af flere.

Det Frie Forskningsråd nævner flere forhold, der skal belyses, før en national strategi for forvaltning af forskningsdata implementeres. Rådet nævner, at opbevaring og tilgængelighed af forskningsdata bør ses i sammenhæng med udvikling og vedligeholdelse af forskningsinfrastrukturer og, at Statens Arkiver, herunder Dansk Dataarkiv, bør inddrages i drøftelsen af tilgængeliggørelse. Endvidere nævnes, at der savnes en klar definition af forskningsdata, standardisering af metadata, afklaring af juridiske forhold og præmisserne for tilgængeliggørelse af data, herunder tidsrammer. Et nationalt arkiv forventes at sikre ensartede krav til forskerne, afklarede ejendomsforhold og en mere gennemskuelig økonomi omkring denne type service.

4.4 Har universitetet tiltag i forhold til at sikre forskernes opmærksomhed på spørgsmålet om bevaring, synliggørelse og evt. tilgængeliggørelse af forskningsdata?

Nogle universiteter; AU, AAU, DTU, RUC og CBS har alle svaret, at de på nuværende tidspunkt er i gang med – eller på sigt planlægger – kampagner eller andre former for aktiviteter for at gøre forskerne opmærksomme på udfordringer og muligheder i forhold til forskningsdata. Det gælder eksempelvis orientering i form af guides, undervisningstilbud og lignende.

ITU har ikke umiddelbart nogle særlige tiltag fra centralt hold, men afholder sig ved nyansættelse introduktion tilgældende politikker og muligheder. På KU er der for tiden ikke planlagt aktiviteter, men et udvalg er nedsat til at komme med bud på forskellige kommunikationstiltag, dog med det forbehold, at man her næppe skal forvente en 'one-size-fits-all'-løsning som resultat. SDU har ikke planlagt særlige aktiviteter, men henviser dog til, at der på decentralt niveau kan være opmærksomhed på spørgsmålet, særligt hvor man arbejder med store datasæt.

Fondene og forskningsrådene har ikke umiddelbart skulle forholde sig til dette spørgsmål. Deres rolle kan dog, sammen med andre bevillingsgivere, have stor indflydelse på netop dette spørgsmål, i fald de stiller øgede krav, eller der kommer krav herom fra udgivere.

4.5 Indgår eksponering af forskningsdata som en del af universitetets strategi?

Næsten alle universiteter nævner, at eksponeringen af forskningsdata i dag indgår implicit, idet det er selve outputtet af forskningen i form af publiceringen, der er det væsentlige. ITU nævner eksponering af forskningsdata som et 'must have', idet det også indgår som et opmærksomhedspunkt i deres forskningsstrategi. DTU nævner eksponering i forbindelse med myndighedsbetjening og som et opmærksomhedspunkt, uden det dog fremgår eksplicit af strategien og forventer, at området vil få større betydning i årene fremover, ligesom det allerede har hos enkelte bannerførere på universitetet. SDU omtaler det ikke som et specifikt element i strategien, men nævner Tvillingeregistreret og OPEN (bio- og databank finansieret af Odense Universitetshospital) som services, der gøres opmærksom på over for omverdenen. På AAU indgår forskningsdata på nuværende tidspunkt ikke direkte i strategien, men arbejdet med denne kommer i høj grad til at gå via de beslutningstagere der i forvejen på decentralt niveau har ansvaret for forskningen. Det samme gælder for KU's vedkommende. På CBS er det ej heller en del af strategien på nuværende tidspunkt, men det vurderes dog, at der vil være potentiale i at lade rådata indgå som en del af en brandingstrategi.

Både DTU og CBS nævner manglen på citationsmetoder for forskningsdata som noget, der vil kunne påvirke publicering af forskningsdata. DTU nævner helt specifikt h-indexet som kunne tænkes opdelt flere varianter, hvor datasæt kan blive en del af branding for den enkelte.

4.6 Undersøgelsens konklusioner og den nuværende udvikling i forbindelse med forvaltning af forskningsdata

Undersøgelsen viser, at der behov for yderligere aktiviteter i forbindelse med lagring, langtidsbevaring og deling af data. Dette gælder såvel på de enkelte institutioner som på nationalt niveau.

I Danmark er Danish e-Infrastructure Cooperation (DeIC)¹³ blevet etableret i 2012. DeIC's mission er blandt andet at udbrede adgangen til e-Infrastruktur til den danske forskning. Dette tænkes gennemført blandt andet ved "sikring af en stabil operativ leverance af e-infrastruktur" (DeIC, 2012).

Det er således en del af DeIC's strategi, at:

DeIC skal sikre en effektiv national ressourceudnyttelse og sikre adgang til e-Infrastruktur, der lever op til højeste internationale standarder, for derigennem at være med til at facilitere et højt internationalt niveau på e-Science området. (DeIC, 2012)

De igangværende aktiviteter på universiteterne og forskningsbibliotekerne kan således suppleres og styrkes af DeIC's strategiske fokus på området. DeIC har i efteråret 2012 afholdt en række møder med universiteternes ledelser, hvor services i forhold til forskningsdata indgik¹⁴. Der er således udsigt til yderligere initiativer, som på nationalt plan kan sikre bedre adgang for landets forskere til relevante services.

I forhold til at udviklingen på området kan den fælles Open Access-politik som Det Frie Forskningsråd, Det Strategiske Forskningsråd, Danmarks Grundforskningsfond, Højteknologifonden og Rådet for Teknologi og Innovation fælles Open Access politik (Det Frie Forskningsråd et al., 2012) har udformet, anvendes som model for hvordan forskningsrådene kan være med til at præge udviklingen. To af de øvrige adspurgte fonde angiver, at de vil følge en politik for lagring, bevaring og tilgængeliggørelse af data, hvis ovennævnte råd og fonde udvikler en politik på området.

¹³ <http://www.forskningsnettet.dk>

¹⁴ Oplæg af Centerleder Børge Obel, Interdisciplinary Center for Organizational Architecture, Aarhus Universitet og formand for DeIC på workshoppen 31/10

5. Forskningsbibliotekerne og forvaltningen af forskningsdata

Forskningsbibliotekerne har siden 2010 arbejdet målrettet med at analysere udviklingen indenfor håndteringen af forskningsdata. I 2010 startede Danmarks Elektroniske Fag- og Forskningsbiblioteker (DEFF) et mindre projekt 'Forskningsdata og Open Access'¹⁵. Projektet afsluttedes i 2011 og lagde efterfølgende grunden til flere aktiviteter, heriblandt dette projekt.

Blandt andet som led i projektet fra 2011 har en række biblioteksansatte besøgt biblioteker og institutioner i Australien, Holland og Storbritannien. Besøgene omfattede Australian National Data Service (ANDS), University of Melbourne, Monash University, 3TU Datacenter i Delft og Digital Curation Center i Edinburgh.

Forskningsbibliotekerne har allerede services og kompetencer på området. Det drejer sig om registrering, herunder metadatering, langtidsopbevaring og deling af forskningsdata samt deling af store mængder digitale forskningsobjekter. Eksempelvis er forskningsbibliotekerne involveret i en række projekter og services. Det gælder f.eks. registrering af forskningsdata i ORBIT på DTU, RUC's registrering af forskningsdata i PURE og data lagring i RUC's Digitale Arkiv (RUDAR)¹⁶, samarbejde med DDA og Danmarks Statistik¹⁷, systemet DATAVERSE (udbudt af Det Kongelige Bibliotek (KB)¹⁸; LARM¹⁹ projektet; DigHumLab²⁰; Netarkivet²¹.

Disse kompetencer vil der yderligere blive brug for efterhånden som stadig flere data skal langtidsbevares og tilgængeliggøres. I takt med udviklingen af services kan bibliotekarernes rolle ses at skifte fra først at have været aktive efter publiceringen af data til at være involverede i selve forskningsprocessen. Det skyldes, at forskningsbibliotekarerne fungerer som 'data librarians', der hjælper med at skaffe de rådata som forskeren har brug for (Büttner et al, 2011).

Forskningsbibliotekerne kan derfor være med til at udvikle de relevante services sammen med forskerne. Bibliotekerne kan, ud fra deres allerede opbyggede kompetencer, være med til at designe systemer til at forvalte forskningsdata, betjene systemerne og vejlede brugere i at anvende disse (Corall, 2012).

De aktiviteter, som danske forskningsbiblioteker varetager, minder i høj grad om de tilsvarende i udlandet. Heriblandt i Australien, hvor forskningsbibliotekerne er vigtige aktører. Det danske system har dog behov for at blive mere udbygget og mangler i høj grad at arbejde med konkrete dataforvaltningsplaner for at nærme sig det australske (Heller og Nondal, 2011).

¹⁵ <http://projekter.bibliotekogmedier.dk/projekt/forskningsdata-og-open-access-et-pilotprojekt>

¹⁶ <http://rudar.ruc.dk>

¹⁷ eksempelvis <http://www.dst.dk/da/TilSalg/Forskningssservice.aspx>

¹⁸ <https://data.kb.dk/dvn/>

¹⁹ <http://www.larm-archive.org>

²⁰ <http://dighumlab.dk>

²¹ <http://netarkivet.dk>

6. Data management i udvalgte lande - fælles og forskelligt

I det følgende sammenfatter vi udvalgte udviklingstræk i Skandinavien, Holland, Tyskland, Canada og USA. Et særligt fokus er naturligt nok på datamanagement og på eventuelle planer herfor, men herudover ser vi også på nationale dataarkiver og deres virksomhed og på Open Access. En mere uddybende gennemgang af forholdene i de enkelte lande kan ses i bilag 12-18.

Et fælles træk ved de nordiske lande (se bilag 12) er, at der findes nationale dataarkiver, der i en årrække har haft fokus på bevarings- og formidlingsopgaver. Aflevering af data til disse arkiver kan udgøre en betingelse for tildeling af offentlige forskningsmidler, men er det ikke i alle tilfælde. Data management planer indgår ikke som krav til projektbeskrivelser, bortset fra i Finland, hvor en sådan plan er betingelse for offentlig støtte. De nationale dataarkiver varetager bevaring af data fra afsluttede forskningsprojekter og formidling af data til forsknings- og studieformål, idet personfølsomme data selvfølgelig er underlagt særlig beskyttelse. Et andet fælles træk er, at Open Access til forskningspublikationer ses som en parallel til databevaring og -formidling i det nationale arkiv - som vi altså kan se som en slags national vidensbase - således bør også forskningspublikationer være åbent tilgængelige for offentligheden. Også her er der tale om blød styring, om opfordringer og ikke krav.

Vi kan anvende disse træk som udgangspunkt for at iagttage de øvrige lande nærmere, hvilket dog ikke skal opfattes således, at de nordiske lande udgør en form for 'gylden standard' i forhold til de øvrige undersøgte lande.

I Holland (se bilag 13) genfinder vi den samme underliggende politiske målsætning, at forskningsdata bør bevares og gøres tilgængelig for videre forskning. Ligeledes findes nationale dataarkiver for de forskellige fagområder, hvor forskningsinstitutionerne kan være forpligtet til at deponere data fra offentlig finansieret forskning. Til forskel fra de nordiske lande er der en mere etableret praksis for datamanagement med bl.a. fælles retningslinjer for data management planer (DMP'er).

I Tyskland (se 14) anbefales brug af DMP'er, som også i nogle tilfælde er obligatoriske. Et samlet billede vanskeliggøres af, at forskningsområdet er præget af mange aktører på de forskellige niveauer, forbunds-, delstats-, universitets-, forskningsråds- osv. Der foreligger på forbundsniveau et forslag til fælles indhold i en DMP, som indgår i den politiske proces. Også anvendelse af OA anbefales stærkt. Vi ser også her en tydelig bevægelse hen imod data management, mest præget af blød styring.

På overnationalt plan - EU og OECD (se bilag 15) - finder vi formuleringer af politiske målsætninger, der går igen på nationalt plan. Det gælder således OECD's anbefaling (OECD, 2007) til de respektive landes regeringer om at skabe forudsætninger for deling af forskningsdata ved hjælp af teknologisk infrastruktur, øgede forskningsbevillinger, nye samarbejdsformer osv. Samme budskab gentages i flere EU-rapporter (EU, 2010; Van der Graaf, M. og Waaijers, L., 2011), som stadig stærkere understreger betydningen af netværk mellem forskningsmiljøerne og den større rolle den frie udveksling af data kommer til at spille.

USA (se bilag 16) udviser i forhold til Europa flere interessante træk. Flere store forskningsfonde, både private og offentlige, kræver fri adgang til forskningsdata og DMP'er som betingelse for støtte.

Universitetsbibliotekerne spiller en central rolle for den praktiske del af data management, afklaring af rettigheder, databevaring osv. Dette betragtes af universiteterne som væsentlige aktiviteter i forhold til synliggørelse af universiteternes forskning og til arbejdsmæssig aflastning af forskerne. Samtidig erkendes nødvendigheden af at udvikle særlige kompetencer inden for data management for at kunne løfte denne opgave.

I Canada (se bilag 17) går dette sidste igen - at data management kræver særlig viden og færdigheder - og, at disse mest praktisk må udvikles i (universitets-)biblioteksregi. Krav om fri adgang til forskningsdata og, at forskningsresultater publiceres med OA, stilles ofte i bevillingssammenhænge.

Australien (se bilag 18) markerer sig med en tydelig profil inden for data management. Fri adgang til forskningsdata betragtes som en vigtig konkurrenceparameter. Der er oprettet en fælles national dataportal, hvorfra data kan hentes. Der er ikke pligt til gøre data tilgængelige, men data skal være beskrevet med metadata. Ligesom i USA har universiteterne specialister i data management, som underviser og hjælper forskerne med arbejdet. Udarbejdelse af DMP'er anbefales stærkt, men er ikke obligatorisk. Der er udarbejdet flere hjælperedskaber til støtte for planlægning af data management.

Vi ser altså følgende udviklingstræk i de undersøgte lande:

- Aktiviteter i relation til data management tager hyppigt afsæt i eksisterende institutioner og strukturer såsom nationale dataarkiver, universitetsrepositories, retningslinjer for bevillinger, praksis for OA osv.
- Politiske målsætninger har oftere karakter af anbefalinger, snarere end bindende retningslinjer.
- I de lande, hvor aktiviteter i relation til data management er mest udviklet, anses udvikling af særlige kompetencer herindenfor som nødvendige, disse lokaliseres i (universitets-) biblioteksregi.

7. Hvad er en data management plan

7.1 Rammerne for indholdet

Et af dette projekts mål er at afklare hvilke konkrete elementer en plan for forvaltning af forskningsdata bør indeholde. Det følgende er en oversigt over de forhold, der kan behandles i en sådan plan, til inspiration og støtte for forskere og andre ansatte i forskningssektoren.

Kort fortalt er en DMP et dokument, der skitserer, hvordan data vil blive forvaltet, både under et forskningsprojekt og efter projektet er afsluttet. Dokumentets afsender er i princippet en forskningsleder eller en ansøger i forhold til et forskningsprojekt, modtageren er aftagere af data, samt eventuelle bedømmer, i tilfælde af, at planen udarbejdes som led i en ansøgning.

Udover dokumentation af data, er et af hovedformålene med en DMP, at projektejeren skal overveje de forhold, der kan være med til at sikre, at data, der er produceret i forskningsprojektet forvaltes korrekt samt, at de er forberedt til eventuel langtidsbevaring.

På den ene side, kan vi ikke her opstille en endegyldig model for hvad en DMP skal indeholde. Dertil er forskellene fagligt, nationalt og institutionelt i forståelsen af hvad forskningsdata er og hvordan en DMP kan anvendes, de tekniske krav til data og metadata samt forskellige nationale lovgivninger indenfor persondatabeskyttelse mm., alt for store.

På den anden side, er der nogle fælles grundbetingelser, der har betydning for indholdet i en DMP, nemlig de fysiske, økonomiske og etiske perspektiver for forskningsdata. Disse er belyst i kapitel 3 og i bilag 5. Disse grundbetingelser gør, at vi mener, at man opstille visse fællestræk, som det vil fremgå af de to kommende afsnit.

7.2 Gennemgående elementer i data management planer

Der findes flere forskellige beskrivelser af dataforvaltningsplaner. Der er her en række fællestræk for hvad forskerne må overveje i forbindelse med ansøgninger om forskningsstøtte, publicering, arkivering mm.

Et vigtigt forhold, der skal tages højde for i planen, er de forskellige faser i projektets livscyklus, hvad skal der foretages i planlægningsfasen, undervejs i forskningsprocessen - hvor data skal lagres - og efter afslutningen, hvor data måske skal langtidsbevares (Jensen, 2011). De enkelte faser i processen kan således indgå i en DMP. En detaljeret beskrivelse af planens enkelte dele kan findes hos Donnelly (2012), baseret på Digital Curation Centers beskrivelse af en data management plan (Digital Curation Centre 2012a).

En række fælles temaer kan ses i flere af modellerne. Der stilles krav om generelle oplysninger om forhold som: ejendomsret, ophavsret og deling af data, datas beskaffenhed, langtidsbevaring mm. Endelig skal forskeren sikre sig fornøden finansiering af udgifterne til håndtering af data, undervejs i projektet og efter dets afslutning.

Herudaf fra kan vi uddrage disse overordnede indholdskategorier i en DMP:

- Generelle oplysninger: Information om data og dataformater.
- Ejerskab: ejerskab og ophavsret, politik for adgang, deling og genbrug.
- Beskaffenhed: Metadataindhold og formater.
- Betingelser for opbevaring af data: Langtidsopbevaring og datastyring.
- Organisation og finansiering.

Disse overordnede kategorier går i nogenlunde denne form igen, i mange forskellige sammenhænge og der er en direkte kobling fra disse emner til *Responsible Conduct of Research* og dermed til videnskabsteoretiske overvejelser, så som f.eks. i det klassiske CUDOS-system.

7.3 Eksempel på indhold i en data management plan

Med udgangspunkt i ovennævnte kategorier, kan vi skitsere det grundlæggende indhold i en DMP:

Generelle oplysninger

- En beskrivelse af de data, der produceres eller indsamles af forskningsprojektet. Her kan der være forskellige fagspecifikke definitioner af forskningsdata. Er det data, der stammer fra observationer, er skabte ved hjælp af eksperimenter, modeller o.l., er indsamlede ved surveys, er det fysiske/materielle data osv.?
- Hvorledes er data producerede, hvornår og i hvilken sammenhæng?
- Efter indsamlingen af data, hvordan behandles eller reduceres data yderligere, hvilke metoder, anvendes her(f.eks. software, algoritmer)?
- Beskrivelse af dataformater, konventioner mm.
- Beskrivelse af foranstaltninger til kvalitetssikring og kvalitetskontrol.
- Hvis der anvendes allerede eksisterende data, hvad er disses proveniens og vil nyindsamlede data blive kombineret med eksisterende data?
- Hvorledes forvaltes data på kort sigt? Foretages der versionskontrol og sikkerhedskopiering af rådata og behandlede data og hvem er ansvarlig for dataforvaltningen?

Ejerskab

- Beskrivelse af eventuelle forpligtelser, der eksisterer i forhold til deling eller publicering af data.
- Oplysninger om, hvorledes data skal deles, herunder hvornår de vil være tilgængelige, hvor længe de vil være til rådighed, hvordan adgang kan opnås, og eventuelle rettigheder, som dataejeren har til udnyttelse.
- Planer for at håndtering af eventuelle spørgsmål om privatlivsbeskyttelse og andre etiske spørgsmål i forbindelse med tilgængeliggørelse af data.
- Hvem har ophavsretten til data?
- Hvem har ejendomsretten til data?
- Beskrivelse af den påtænkte fremtidige anvendelser af data.
- Angivelse af hvordan data ønskes citeret.

Metadata

- Hvordan skabes metadata?
- Hvilket format og hvilken standard vil blive anvendt til metadata?
- Hvilke minimumskrav skal metadata opfylde?

Langtidsopbevaring og datastyring

- Hvilke data bør eventuelt langtidsbevares?
- Hvordan skal data behandles, for at de kan langtidsbevares?
- Hvor og hvordan vil data blive langtidsbevaret?
- Hvilke filer bør langtidsbevares sammen med data? Hvem er ansvarlig for arkiverede data og hvem sikrer, at kontaktoplysninger altid er up-to-date?

Organisation

- Hvilke formelle roller og funktioner skal være aktive i forhold til data, hvem er ansvarlig for hvad?
- Hvem er ansvarlig for datastyring og hvilke udgifter budgetteres der med?
- Hvor meget tid afsættes i projektet til klargøring af data, dataforvaltning, dokumentation og bevaring?
- Forefindes der eller anskaffes der hardware eller programmel til dataforvaltning, sikkerhedskopiering, anden sikkerhed, dokumentation og bevaring? Er udgifterne hertil medtaget i budgettet?
- Er der omkostninger i forbindelse med fremsendelsen af data til arkiv eller datacenter, er disse medtaget i budgettet?

7.4 Eksempler på data management planer i praksis

I Bilag 19 til 21 vedlægger vi tre eksempler på DMP'er, der alle stammer fra ansøgninger fra UC San Diego. Eksemplerne er projekt- og fagspecifikke og meget forskellige, alle ansøgningerne blev bevilget støtte af UC San Diego, bl.a. på baggrund af den obligatoriske DMP. Eksemplerne er forskellige i forhold til mange af de punkter vi har stillet op ovenfor.

Dermed er eksemplerne med til at illustrere, at det formentlig hverken er hensigtsmæssigt eller muligt at etablere en universel model for en dataforvaltningsplan, der indholdsmæssigt er mere detaljeret end hvad der kan rummes under de generelle kategorier ovenfor.

8. Konklusioner og anbefalinger

Som det er fremgået af det foregående og uddybet i rapportens bilag, er forvaltning af forskningsdata et tema, det bliver stadig vigtigere for forskningssektoren at forholde sig til. Denne bliver i stigende omfang nødt til at sikre en velfungerende digital infrastruktur, der kan langtidsbevare og tilgængeliggøre forskningsdata.

Dette er blandt andet betinget af den teknologiske udvikling, der gør det muligt at skabe flere former for forskningsdata og digitale forskningsobjekter indenfor alle forskningsområder. Denne udvikling foregår på alle samfundsområder og stadig flere typer data kan deles via internettet. Eksempelvis gøres artikler i stigende omfang tilgængelige online via Open Access.

Der kan på internationalt plan iagttages klare tendenser til, at forskningsråd, forskningsfonde, regeringer og internationale organisationer stiller krav om, at forskningsdata skal kunne deles og, at forskerne i dataforvaltningsplaner kan dokumentere, hvorledes forskningsdata skal behandles. Denne udvikling påvirker naturligvis danske forskere og forskningsorganisationer, der skal agere i et internationalt miljø. I det omfang, der stilles krav til langtidsbevaring og deling af forskningsdata fra internationale bevilligende myndigheder, er den danske forskningssektor nødt til at etablere en forskningsinfrastruktur, der lever op til internationale normer og standarder.

Nærværende undersøgelse har vist, at universiteterne i et vist omfang sikrer opbevaring af forskningsdata under selve forskningsprocessen, men ikke altid systematisk og ofte med store faglige og lokale variationer. Samtidigt viste undersøgelsen, at forskningsdata i langt mindre omfang bliver langtidsbevaret, med alt hvad det indebærer af overvågning af datas integritet, metadatering og søgbarhed. Dette bliver set som omkostningstungt og vurderes ofte som værende en national opgave. Der er blandt respondenterne en klart forventning om, at de løsninger, der vælges, skal kunne håndtere mange forskellige typer data og leve op til forskellige krav. 'One-size-fits-all' løsninger er ikke vejen frem.

På denne baggrund er det forståeligt, at universiteterne ikke i nævneværdigt omfang stiller services til rådighed, der kan langtidsbevare og tilgængeliggøre forskningsdata, såsom egne dataarkiver. Det skal dog bemærkes, at der her er store faglige forskelle. Eksempelvis deler fysikerne traditionelt forskningsdata, men her er der tale om løsninger, der er etableret på internationalt plan. På andre områder bevarer og deler forskere data via Dansk Data Arkiv, på atter andre har forskere adgang til langtidsbevarede forskningsobjekter via samarbejde med forskningsbibliotekerne. Men det generelle billede er, at universiteterne ikke stiller muligheder for langtidsbevaring til rådighed. Samtidig er der imidlertid blandt universiteterne en generel interesse for at kunne synliggøre forskningsdata som led i en brandingstrategi. På fem af de undersøgte universiteter er der i øjeblikket aktiviteter, der skal gøre forskerne mere opmærksomme på bevaring og tilgængeliggørelse af forskningsdata.

Forskningsrådene har ikke tidligere stillet krav på dette område. Generelt ønskes der deling af forskningsdata, da rådene og fondene på denne måde kan sikre, at resultaterne af tildelte bevillinger bliver større i og med, at forskningsresultater kan genbruges i nye projekter.

Universiteterne, fondene og forskningsrådene peger på en række problemstillinger, der bør afklares på området:

- En fælles forståelse af, hvad forskningsdata er, bør etableres.
- Standarder for metadata og citationer, så data bliver søgbare og tilgængelige, bør udvikles.
- Juridiske problemstillinger omkring adgang, ophavs- og ejendomsret til data skal afklares.
- Det bør undersøges, om der skal etableres en national infrastruktur for forskningsdata.

En gennemgang af forskellige landes aktiviteter i forhold til forvaltning af forskningsdata viste, ikke uventet, store forskelle mellem de nationale løsninger.

Forskningssektoren i Danmark virker umiddelbart mindre parat til at løfte opgaven end i lande som f.eks. Australien og Storbritannien. Videre inspiration for danske aktører kan med fordel søges i udlandet, men samtidig findes der i de danske dataarkivers virksomhed værdifuld praksis at bygge videre på.

Rapportens gennemgang af hvad forvaltning af forskningsdata i form af dataforvaltningsplaner kræver, understreger interviewpersonernes opfattelse af, at der ikke findes en standard for dataforvaltningsplaner, endsize en universel sådan. Dertil er de faglige, institutionelle og nationale krav for forskellige. De store forskelle stiller derfor krav til forskningssektoren og til infrastrukturen for bevaring og langtidsbevaring af forskningsdata om fleksibilitet og internationalt orientering for at kunne leve op til interessenternes behov.

På baggrund af undersøgelsen vil projektgruppen fremkomme med disse handlingsanbefalinger:

- En større og mere forpligtende opmærksomhed fra universiteternes og forskernes side på bevaring og deling af forskningsdata.
- Formulering af en national politik på området.
- Undersøgelse af hvorledes forskerne aktuelt bevarer og deler forskningsdata og deres fremtidige behov på dette område.
- Håndtering af de juridiske problemer, der er forbundet med bevaring og deling af forskningsdata.
- Formulering af en egen politik på dette område fra forskningsråds og -fondes side.
- Etablering af et eller flere tekniske systemer, der understøtter de forskellige aktiviteter inden for databevaring og datadeling.
- Sikring af de nødvendige kvalifikationer hos de relevante medarbejdere på universiteterne og forskningsbibliotekerne.

9. Litteratur

- Stephan Büttner, Stefanie Rümpel og Hans -Christoph Hobohm (2011); Informationswissenschaftler im Forschungsdatenmanagement; I Stephan Büttner, Hans– Christoph Hobohm og Lars Müller (Red.): Handbuch Forschungsdatenmanagement, Bad Honnef: BOCK + HERCHEN Verlag, 2011, side 203-218.
- Sheila Corral (2012); Roles and responsibilities; I Graham Pryor (Red.): Managing Research Data, London: Facet Publishing, 2012; 106-133.
- Digitalbevaring (tilgået januar 2013); Om bevaring af digitale data; <http://digitalbevaring.dk/viden-om/#Bevaring>.
- DeIC (2012); Strategi 2012-2016; <http://www.forskningsnettet.dk/strategi>.
- Martin Donnelly (2012); Data management plans and planning, I Graham Pryor (Red.): Managing Research Data, London: Facet Publishing, 2012, side 83-104.
- EU (2010); Riding the wave - How Europe can gain from the rising tide of scientific data Final report of the High level Expert Group on Scientific Data; Bruxelles, 2010; <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.
- Det Frie Forskningsråd, Det Strategiske Forskningsråd, Danmarks Grundforskningsfond, Højteknologifonden og Rådet for Teknologi og Innovation (2012); Open Access-politik for offentlige forskningsråd og fonde; København, 2012; <http://www.fi.dk/raad-og-udvalg/det-frie-forskningsraad/open-access-politik>.
- Alfred Heller og Lars Nondal (2011); Forschungsdata Down Under – og måske right here?; DF Revy, 2011, Vol. 34, nr. 1, side 16-18.
- Sarah Higgins (2012); The Life of data management; I Graham Pryor (Red.): Managing Research Data; London: Facet Publishing, 2012; side 22-44).
- Uwe Jensen (2011); Datenmanagementpläne; I Stephan Büttner, Hans–Christoph Hobohm og Lars Müller (Red.): I Handbuch Forschungsdatenmanagement; Bad Honnef: BOCK + HERCHEN Verlag, 2011; side 71-82.
- Jens Klump (2011); Langzeiterhaltung digitaler Forschungsdaten; I Stephan Büttner, Hans–Christoph Hobohm og Lars Müller (Red.): Handbuch Forschungsdatenmanagement; Bad Honnef: BOCK + HERCHEN Verlag, 2011; side 115-122.
- OECD (2007); OECD Principles and Guidelines for Access to Research Data from Public Funding; <http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf>
- Graham Pryor (2012); Why manage research data?; I Graham Pryor (Red.); Managing Research Data; Facet publishing, 2012; side 1-16.
- Annette Balle Sørensen, Filip Kruse, Jørn Thøgersen, Laura Molloy, John W. Pattenden-Fail, Bart Ballaux (2009); Report based on DT/7 questionnaire; Project: IST-2006-033789 Planets <http://ebookbrowse.com/planets-dt7-d4-questionnaire-report-pdf-d384840500>.
- Van der Graaf, M. og Waaijers, L. (2011); A surfboard for riding the wave – Towards a four country action programme on research data; KE Knowledge Exchange, 2011.