

National strategi for forskningsdata management 2015-2018

København, den 30. januar 2015

Indholdsfortegnelse

0. Forord	3
1. Executive summary	5
2. Om styregruppen	7
A. <i>Baggrund, nedsættelse og kommissorium</i>	7
B. <i>Styregruppens arbejdsproces</i>	8
C. <i>Høring og tilpasning af strategien</i>	9
3. Baggrund for strategien	10
A. <i>Digitalisering og eksponentiel vækst af forskningsdata</i>	10
B. <i>Voksende forventninger og fremspirende betingelser</i>	10
C. <i>De faglige miljøers behov og præferencer</i>	12
4. Strategiske indsatsområder	14
A. <i>Politikker</i>	15
B. <i>Incitamentsskabelse</i>	15
C. <i>Infrastruktur</i>	16
D. <i>Kompetenceudvikling og forskerstøtte</i>	17
E. <i>Fremtidig styringsstruktur</i>	18
F. <i>Pilotprojekter</i>	20
5. Handlingsplaner og finansiering	24
A. <i>Nationalt forum for samarbejde om forskningsdata management</i>	24
B. <i>Train the trainer kursus</i>	26
C. <i>Fagspecifikke pilotprojekter</i>	26
D. <i>Akademisk "dropbox"</i>	26
E. <i>Projektleder</i>	27
F. <i>"Data management i praksis" projekt</i>	27
6. Executive summaries af fem afdækninger	28
A. <i>Bevillingsgiveres krav til forskningsdata management</i>	29
B. <i>Danske juridiske rammer og udfordringer</i>	30
C. <i>International best practice for forskningsdata management politikker</i>	32
D. <i>Udenlandsk inspiration vedr. Incitamenter</i>	35
E. <i>Danske faglige miljøers behov og præferencer</i>	37
7. Udvalgte referencer	39
A. <i>Web dokumenter</i>	39
B. <i>Bøger</i>	39
8. Bilagsliste	40

0. Forord

Forskningsdata er vigtige aktiver for såvel de danske universiteter som for nationen – både forskningsmæssigt og økonomisk. De lever ofte en isoleret og ikke-fremtidssikret tilværelse – privat hos forskeren, hos IT afdelinger eller på kommercielle skytjenester. Sjældent entydigt linket til forskere, projekter og publikationer, og derfor svære at dele, genfinde for verifikation eller genbruge til nye forskningsformål.

Formålet med nærværende strategi er at sikre Danmark bedre og mere konkurrencedygtig forskning gennem effektiv opsamling, sikring, formidling og genbrug af relevante forskningsdata.

Strategien er vedtaget af DeiC og DEFF. Forud herfor har et udkast udfærdiget af Styregruppen for National Data Management været i høring blandt primærinteressenterne – Rektorkollegiet og de otte universiteter samt bevaringsinstitutionerne, Det Kongelige Bibliotek, Statsbiblioteket og Rigsarkivet. Udkastet blev efterfølgende revideret på baggrund af høringssvarene, og den endelige strategi vedtaget i december 2014 af DeiC bestyrelsen og DEFF styregruppen.

Sektorforskningen har ikke været repræsenteret i styregruppen, og strategiens eventuelle implementering på dette område bør derfor overvejes særskilt. Der har heller ikke været repræsentation af Danske Regioner, registre, forskningsråd og fonde. I det videre forløb bør dialogen om forskningsdata management med disse parter fremmes.

Arbejdet har omfattet en række afdækninger, der har været centrale for styregruppens udarbejdelse af strategiske anbefalinger med tilhørende handlingsplaner og økonomiske estimater. Executive summaries findes bagerst i dokumentet. Afdækningerne i deres helhed og yderligere bilag, der fremgår af bilagsoversigten, findes på deic.dk/datamanagement.

DeiC og DEFF, 30. januar 2015.

Styregruppen for National Data Management

Formand

- Chef for Bibliometri & Data Management, **Mogens Sandfær**, Danmarks Tekniske Universitet

Medlemmer

- Professor **Brian Vinter**, Niels Bohr Institutet, Københavns Universitet
- Director of Center for Massive Data Algorithmics, professor **Lars Arge**, Institut for Datalogi, Aarhus Universitet
- Institutleder, professor **Kirsten Ohm Kyvik**, Institut for Regional Sundhedsforskning, Syddansk Universitet
- Bibliotekschef **Bertil F. Dorch**, Syddansk Universitet
- Director of Center for Data-intensive Systems, professor **Christian Søndergaard Jensen**, Institut for Datalogi, Aalborg Universitet
- Chefkonsulent **Lars Nondal**, CBS Bibliotek, Copenhagen Business School
- Institutleder **Jens Christian Godskesen**, IT-universitetet
- Udviklingschef og vicedirektør, **Birte Christensen-Dalsgaard**, Det Kongelige Bibliotek
- Områdedirektør, **Bjarne Andersen**, Statsbiblioteket
- Vicedirektør **Kirsten Villadsen Kristmar**, Indsamling og Bevaring, Rigsarkivet

Sekretariat

- Projektleder **Diba Markus**, DeiC Sekretariatet
- Chef for Internationale Relationer **Rene Belsø**, DeiC Sekretariatet

(→ Bilag 01 for kontaktinformation)

1. Executive summary

Digitaliseringen indebærer en eksponentiel stigning i mængden af forskningsdata, nye udfordringer og nye potentielle fordele:

Eksponentielt stigende mængder af data er en realitet, og inden for forskningen udgør dette både udfordringer og væsentlige muligheder for at effektivisere forskningen. Udfordringerne består ikke mindst i at opbevare og organisere de enorme mængder forskningsdata uden at gå på kompromis med problemstillinger om oprindelse, kontekst, internationalisering og sikkerhed. Fordelene er mange og betydelige. Med den rette infrastruktur kan forskere i forskellige forskningsmiljøer samarbejde og dele datasæt og dermed styrke den tværdisciplinære forskning. / Uddannelses- og Forskningsministeriets webside¹

Styregruppen for National Data Management har gennem en række afdækninger søgt at danne sig bedst muligt overblik over såvel relevante udenlandske erfaringer som de danske forskningsmiljøers overordnede udfordringer, behov og præferencer.

Nationalt er der tale om en meget stor variation i hovedfagområdernes betingelser, løsninger og behov på dette område. Variationen fortsætter desuden inden for hovedområderne, hvorfor man overvejende må adressere udfordringerne på et væsentlig mere fagspecifikt niveau for sikre det nødvendige match med forskningsmiljøers konkrete udfordringer, betingelser, traditioner, behov og præferencer på forskningsdataområdet.

Internationalt viser f.eks. de svenske erfaringer, at en national top-down politik uden udbud af infrastruktur og støttefunktioner og uden tilpasning til og engagering af forskningsmiljøerne reelt er virkningsløs. Til gengæld viser de australske erfaringer, at en parallel og gradvis opbygning af lokale og nationale politikker, infrastrukturer og støttefunktioner koblet med en stærk samarbejdsorganisation og en betydelig national økonomisk støtte – er særdeles virkningsfuld.

På denne baggrund er strategiens hovedpunkter:

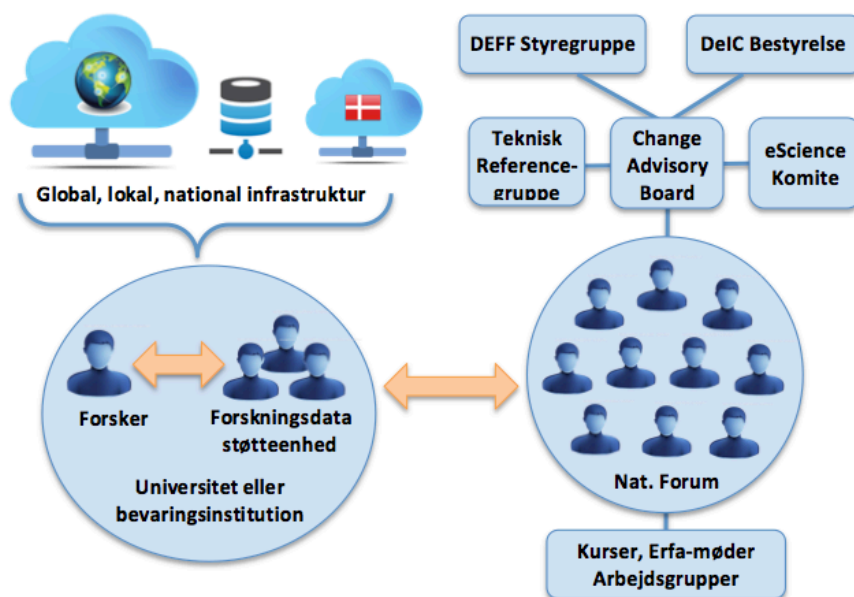
- At lokale, fag- og forskernære initiativer i universitetsregi kobles med et nationalt samarbejde med fokus på tværgående videndeling - samt tværgående projekter, hvor disse kan lette og kvalificere arbejdet - og tværgående infrastrukturer, hvor disse efterspørges og kan etableres på bæredygtig vis.
- At forskningens internationale natur tages i ed og udnyttes i denne sammenhæng, således at forskningsdata management initiativer – det være sig indenfor politikker, incitamentsskabelse, kompetenceudvikling eller infrastruktur – så vidt muligt bygger på international erfaring og samarbejde og sigter på en sammenhæng mellem aktiviteter på lokalt, nationalt og internationalt niveau; eksempelvis at internationale infrastrukturer udnyttes, når det er muligt, og nationale og lokale infrastrukturer kun etableres, når det er påkrævet.
- At DEFF og DeiC faciliterer og økonomisk støtter denne proces i samarbejde med danske universiteter og de kulturministerielle bevaringsinstitutioner.

¹ <http://ufm.dk/forskning-og-innovation/samspil-mellem-viden-og-innovation/open-science/fri-adgang-til-forskningsdata> den 31.7.2014

Hovedindsatsen for 2015-2018 år er således mere konkret:

- Etablering eller udbygning støttefunktioner på universiteterne indenfor forskningsdataområdet - der på baggrund af et globalt overblik kan rådgive forskerne omkring brug af infrastrukturer og informationsteknologiske værktøjer, metadatering og informationsorganisatoriske problematikker samt lovgivning og juridiske problematikker
- Etablering af et samlende forum for disse støttefunktioners nøglepersoner med økonomisk støtte fra DEFF med henblik på videndeling og erfaringsudveksling på tværs af institutioner samt på identifikation og specifikation af fælles infrastrukturer og supportfunktioner, hvor de måtte være nødvendige
- Etablering eller udbygning af tilsvarende støttefunktioner på forskningsdataområdet hos Det Kongelige Bibliotek, Statsbiblioteket og Rigsarkivet, hvorved nøglepersoner herfra kan indgå i det nationale forum for samarbejde om forskningsdata management
- Igangsættelse af et antal pilotprojekter støttet af DEFF og/eller DeiC - det drejer sig dels om en række fagspecifikke projekter til støtte for afklaring af best practice på fagområdet, dels om etablering af en dansk akademisk "dropbox".

Organiseringen af indsatserne er illustreret nedenfor og uddybes i afsnit 4 i overensstemmelse med kommissoriets seks indsatsområder og feedback fra høringen.



2. Om styregruppen

A. Baggrund, nedsættelse og kommissorium

Rektorkollegiet, Danske Universiteter behandlede i november 2013 et oplæg fra DeiC og DEFF vedrørende *National strategi og proces for data management*. Oplægget var baseret på en møderække i en interessentgruppe med repræsentanter fra universiteterne, universitetsbiblioteker, Rigsarkivet, Dansk Data Arkiv, det Kongelige Bibliotek, Statsbiblioteket, Det Frie Forskningsråd, DEFF og DigHumLab samt DeiC. Oplægget indstillede, at der udarbejdes handlingsplaner og økonomiske estimater for en række strategiske indsatsområder for data management, som samlet set ville tegne en national strategi.

I opfølgning herpå nedsatte Rektorkollegiet, DeiC og DEFF i februar 2014 en styregruppe for strategiarbejdet, "Styregruppen for National Data Management" med 11 medlemmer. 8 udpeget af Rektorkollegiet og 1 udpeget af hhv. Det Kongelige Bibliotek, Statsbiblioteket og Rigsarkivet. Rektorkollegiet, DeiC og DEFF udpegede formanden, og DeiC Sekretariatet har forestået sekretariatsbetjeningen (→ Bilag 01).

Kommissoriet (→ Bilag 02) fastsætter formålet som at sikre, at Danmark, i et samarbejde mellem primærinteressenterne, får en national strategi for forskningsdata management. Missionen er at sikre at værdien af danske forskningsdata bevares og realiseres nu og for fremtiden. Denne udfordring tager afsæt i visionen om at tilvejebringe bedre og mere konkurrencedygtig forskning gennem effektiv opsamling, sikring, formidling og genbrug af relevante danske forskningsdata.

Kommissoriet henviser endvidere til DeiC's og DEFF's oplæg af 12.12.2013 vedr. National Strategi og Proces for Datamanagement (→ Bilag 03). Heri opregnes seks strategiske indsatsområder:

- Politikker
- Incitamentsskabelse
- Infrastrukturer
- Kompetenceudvikling og forskerstøtte
- Styringsstruktur
- Pilotprojekter

Af oplægget fremgår:

Såfremt man nationalt opnår enighed om strategiens implementering forventes det, at universiteter, arkiver og biblioteker medfinansierer såvel en konkret implementering som en levedygtig model for data management fremover indenfor de lovgivningsmæssige og økonomiske rammer som de er underlagt.

I tillæg hertil lister oplægget ni styrende principper som ramme for såvel udformningen af strategien som implementeringen af den:

- **Datas hele livscyklus** (alle stadier fra planlægning til bevaring og genbrug)
- **Incitamenter til data management** (f.eks. BFI-point, øget citations-impact)
- **Et bredt spektrum af forskningsdata** (kræver bredt spektrum af løsninger og politikker)
- **Nationale løsninger og tiltag, hvor det giver mening** (klarlægge niveau for løsning: lokalt, nationalt, internationalt)
- **Flere infrastrukturelle løsninger** (spektret af stadier og fagområder peger på en flerhed af løsninger)

- **Samarbejde på tværs af organisatoriske skel** (på tværs af institutioner og på tværs af kompetencer)
- **Interoperabilitet og standarder** (flerhed af løsninger med åbne grænseflader, Open Source, leverandøruafhængighed)
- **Prototyper og pilotforsøg** (3 trin med indlagte stop/go beslutninger: prototype → pilotforsøg → drift)
- **Udnyttelse af internationale erfaringer og samarbejdsmuligheder** (andre lande er langt på området) med henblik på effektiv ressourceudnyttelse

På ovenstående grundlag indledte styregruppen arbejdet.

B. Styregruppens arbejdsproces

Styregruppen afholdt i alt seks heldagsmøder i perioden marts-august: 5. marts 2014, 10. april 2014, 19. maj 2014, 6. juni 2014, 26. juni 2014 og 21. august 2014.

Det første møde havde fokus på tre forhold: (1) En gennemgang af baggrunden og grundlaget for arbejdet (2) En præsentation af styregruppens institutioner og medlemmer, hvor alle bl.a. forholdt sig til spørgsmålet: Hvad er det vigtigste for din institution hhv. at opnå og at undgå ifm. styregruppens arbejde? (3) En første diskussion af styregruppens holdning til de seks strategiske indsatsområder – baseret på et diskussionsoplæg med en række åbne spørgsmål til hvert af områderne (→ Bilag 04).

Mødet gav et godt overblik over styregruppens store spændvidde i udgangspositioner, hvad angår de strategiske indsatsområder. Ligeledes blev det klart, at der var stort behov for at gennemføre en række afdækninger af centrale forhold på området for at etablere et tilstrækkeligt grundlag for styregruppens endelige konklusioner og anbefalinger. Disse afdækninger blev umiddelbart sat i værk (→ Bilag 05) og kom således til at strukturere de følgende tre møder.

Det drejer sig om følgende afdækninger, der alle indgår i denne afrapportering:

- A. Bevillingsgiveres krav
- B. Juridiske rammer og udfordringer
- C. Udenlandsk inspiration vedr. politikker
- D. Udenlandsk inspiration vedr. incitament
- E. Danske faglige miljøers behov og præferencer

De fem afdækninger er kort opsummeret i afsnit 6 og indgår i deres fuldstændighed som bilag tilgængelige på deic.dk/datamanagement.

Afdækningerne har således haft stor betydning for styregruppens arbejde og anbefalinger, men det er vigtigt at fremhæve, at mange af afdækningerne er foreløbige, subjektive og/eller bygger på et begrænset datagrundlag. Derfor tages der forbehold for repræsentativiteten af konklusionerne i afdækningerne, ligesom det pointeres, at konklusionerne ikke nødvendigvis afspejler styregruppens holdning.

I løbet af møderækken har styregruppen endvidere holdt sig orienteret om de parallelt løbende nationale arbejder med henholdsvis "The Danish Code of Conduct for Research Integrity"², hvor en

² <http://ufm.dk/forskning-og-innovation/rad-og-udvalg/andre-udvalg-og-fonde/arbejdsgruppe-faelles-retningslinjer-for-forskningsintegritet/arbejdsgruppe-faelles-retningslinjer-for-forskningsintegritet>

repræsentant for ministeriet har deltaget i et af styregruppens møder, og "Danmarks Nationale Strategi for Open Access"³, hvor flere af styregruppens medlemmer sidder i "Den Nationale Styregruppe for Open Access" og således løbende har kunnet orientere om arbejdet.

De to sidste møder i styregruppen havde atter fokus på den samlede strategiske udfordring. Med udgangspunkt i den fælles læring i forbindelse med de gennemførte afdækninger udvikledes de strategiske anbefalinger mv.

C. Høring og tilpasning af strategien

Primo september 2014 sendte DeiC og DEFF udkast til strategi med tilhørende handlingsplaner og økonomiske estimater i høring hos universiteterne og bevaringsinstitutionerne med frist 10. oktober 2014.

De indkomne høringssvar viste, at der blandt interessenterne er enighed om, at forskningsdata management er vigtigt, og der var således overordnet set tilfredshed med strategiens tilblivelse og generelle indhold. Dog var en uddybning af styringsstrukturen og den skitserede økonomi – specifikt fordeling af omkostninger og finansieringsmodel – et gennemgående ønske. Høringssvar inklusiv sammenfatning er bragt i bilag (→ Bilag 11).

På baggrund af DeiC bestyrelsens og DEFF styregruppens behandling af høringssvarene er der foretaget en revision af anbefalingerne for såvel indsatsområdet 'Fremtidig styringsstruktur' (afsnit 4. E) som handlingsplaner og finansiering (afsnit 5).

Som konsekvens af denne revidering er anbefalingerne i de øvrige dele af dokument justeret let – primært med hensyn til hvilke aktiviteter der anbefales støttet økonomisk af hhv. DeiC og DEFF, og hvorledes evaluering anbefales foretaget.

I tillæg hertil har DeiC bestyrelsen og DEFF styregruppen ønsket en tydeliggørelse af den forudsatte medfinansiering fra institutionerne og af mulighederne i det internationale samarbejde.

DEFF's styregruppe ønskede endvidere beskrevet, at DEFF allerede har bevilget et større beløb til et data management projekt ("Data management i praksis" projektet), der således er indført i afsnit 5. Handlingsplaner og finansiering.

³ <http://ufm.dk/forskning-og-innovation/samspil-mellem-viden-og-innovation/open-science>

3. Baggrund for strategien

A. Digitalisering og eksponentiel vækst af forskningsdata

Uddannelses- og Forskningsministeriets hjemmeside: Fri adgang til forskningsdata⁴:

Den digitale udvikling giver betydelige potentialer på forskningsområdet. Opbevaring, tilgængeliggørelse og videre brug af forskningsdata har afgørende betydning for effektivisering af forskningen, herunder muligheder for at undgå dobbeltfinansieret forskning samt at styrke videnskabens legitimitet.

Eksponentielt stigende mængder af data er en realitet, og inden for forskningen udgør dette både udfordringer og væsentlige muligheder for at effektivisere forskningen. Udfordringerne består ikke mindst i at opbevare og organisere de enorme mængder forskningsdata uden at gå på kompromis med problemstillinger om oprindelse, kontekst, internationalisering og sikkerhed.

Fordelene er mange og betydelige. Med den rette infrastruktur kan forskere i forskellige forskningsmiljøer samarbejde og dele datasæt og dermed styrke den tværdisciplinære forskning. Forskere kan bruge, genbruge og kombinere data og dermed øge produktiviteten i forskningen. Med andre ord giver mere åbenhed omkring forskningsdata forskere mulighed for at indgå i en ny og forbedret form for videnskabelig tilgang, der har betydelige samfundsmæssige fordele.

Der er store perspektiver i, at forskere kan anvende data, der ikke nødvendigvis er produceret inden for eget forskningsfelt. Men for at dette kan realiseres er der et behov for, at information og data bliver struktureret i et langtrækkende perspektiv i et globalt perspektiv.

Her afspejles mange af de vurderinger og forventninger, som ikke bare den danske, men en stor del af den vestlige verdens regeringer og forskningsfinansierende instanser har kommunikeret i de senere år.

B. Voksende forventninger og fremspirende betingelser

De ovennævnte forventninger og ambitioner omsætter sig gradvist i krav fra myndigheder og forskningsfinansierende instanser.

Herhjemme kendes krav til håndtering, arkivering og deling af forskningsdata primært fra DFF Sundhed & Sygdom samt DFF Samfund & Erhverv, hvor man henviser til infrastrukturen Dansk Data Arkiv, der drives som en del af Statens Arkiver (pr. 1.10.2014 Rigsarkivet) og som endvidere bistår historiske/arkæologiske forskningsprojekter. For øvrige råd og fonde gælder generelt, at dette er et forhold, der løbende overvejes, men som ikke umiddelbart står foran udmøntning som egentlige krav til bevillingstager.

Udenlands ses en væsentlig udvikling. I forbindelse med EU's Horizon 2020 program er der introduceret en Open Research Data Pilot, ganske som var tilfældet med Open Access under det forrige rammeprogram FP7. Syv faglige indsatsområder er valgt ud til piloten med krav om planlægning og effektivering af data management med henblik på videst mulig deling med andre forskere, erhvervslivet og offentligheden.

⁴ <http://ufm.dk/forskning-og-innovation/samspil-mellem-viden-og-innovation/open-science/fri-adgang-til-forskningsdata> den 31.7.2014

Tiden vil vise om også denne pilot omveksles til et alment krav for alle indsatsområder i næste rammeprogram, således som var tilfældet med Open Access piloten. De væsentlige forskelle mellem fagområdernes betingelser og udfordringer på dataområdet betyder formentlig, at en uniform og tværgående politik i givet fald vil være ganske overordnet, og at den i givet fald vil blive suppleret med mere konkrete politikker på de enkelte fagområder.

Internationalt er der ligeledes en væsentlig udvikling i gang i forbindelse med tidsskriftpublicering. Mange tidsskrifter giver nu mulighed for at beskrive og linke til datasæt som en del af (eller bilag til) artiklen, der beskriver forskningsresultaterne. Dette er ofte et tilbud til forfatterne, men i stigende grad eskaleres dette til et krav – en forudsætning for artiklens peer-review og eventuelle publicering. Fremtrædende eksempler på dette er det prestigefyldte tidsskrift Nature's "Policy on availability of data and materials":

"An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Therefore, a condition of publication in a Nature journal is that authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript, including details of how readers can obtain materials and information."

Samt verdens største tidsskrift PLOS One, der har følgende "Policy on sharing of data, materials, and software":

"PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception. When submitting a manuscript online, authors must provide a Data Availability Statement describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the final article. Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal editors encourage researchers to contact them if they encounter difficulties in obtaining data from articles published in PLOS journals. If restrictions on access to data come to light after publication, we reserve the right to post a correction, to contact the authors' institutions and funders, or in extreme cases to retract the publication."

Dette medfører, at danske forskere mødes med krav om management og deling af deres forskningsdata, når de ønsker at publicere i internationale tidsskrifter som de ovennævnte.

Forskningsdata management er et vigtigt element i "The Danish Code of Conduct for Research Integrity"⁵. Her nævnes data management som det andet af seks bærende elementer. Det pointeres, at:

"Responsible conduct of research includes proper management of primary materials and data. The key purpose of data management is to guarantee credible and transparent research. ... Data and primary materials should be retained, stored and managed in a clear and accurate form that allows the result to be assessed, the procedures to be retraced and – when applicable – the research to be reproduced."

Ansvar for dette fordeles her mellem institutionerne, der gives ansvar for dels at udarbejde, vedligeholde og kommunikere forskningsdatapolitikker og dels at etablere infrastrukturer og støttefunktioner på forskningsdataområdet, og forskerne, der gives ansvar for at følge disse politikker, at benytte disse infrastrukturer samt - med mindre andre regler gælder - konkret at vurdere, hvor meget der skal gemmes og hvor længe.

⁵ <http://ufm.dk/publikationer/2014/the-danish-code-of-conduct-for-research-integrity>

På den baggrund må det forventes, at alle universiteter vil starte (i nogle tilfælde fortsætte) indsatsen til afklaring af den nærmere udformning af politikker, infrastrukturer og støttefunktioner på forskningsdataområdet.

I betragtning af det betydelige forskningsmæssige samarbejde internationalt og på tværs af universiteterne vil det være til fordel for såvel forskerne som for institutionerne, hvis der videndeles og samvirkes intensivt i denne udviklingsproces – med blik for såvel den danske som den internationale sammenhæng.

Herved kan det undgås, at forskningsprojekter på tværs af universiteterne bliver udfordret af modstridende eller inkompatible politikker og procedurer på dataområdet. Et sådant samarbejde vil endvidere kunne afdække, hvor der måtte være gevinst ved fælles eller nationale infrastrukturer og støttefunktioner til supplement af de lokale og globale.

Strategiens indsatsområder sigter mod at facilitere denne videndeling og samvirken – lokalt, nationalt og internationalt.

C. De faglige miljøers behov og præferencer

Hele styregruppen var engageret i en undersøgelse af de danske faglige miljøers behov og præferencer på området (→ Afsnit 6.E og Bilag 10). Undersøgelsen gennemførtes parallelt på fem hovedområder, med to styregruppemedlemmer/institutioner som ansvarlige for afdækningen af hvert hovedområde:

- Humaniora
- Samfundsvidenskab
- Sundhedsvidenskab
- Naturvidenskab
- Teknisk videnskab

Denne afdækning blev central for styregruppens forståelse af områdets udfordringer og dermed for strategien.

Overordnet bekræftede undersøgelsen, at der er reelle og betydelige udfordringer med at håndtere forskningsdata på tværs af landskabet. Dette gælder indenfor alle hovedområder, men til gengæld findes der adskillige eksempler på (snævrere) fagområder, hvor man har løst udfordringerne på forbilledlig vis – typisk i internationalt samarbejde. Et eksempel herpå er den europæiske forskningsinfrastruktur for bioinformatik, ELIXIR, som Danmark i 2014 blev medlem af på basis af en bevilling på 25 mio. kr. fra den nationale pulje for forskningsinfrastruktur⁶.

Afdækningen viste endvidere, at der er en meget stor variation i behov og præferencer mellem hovedområderne såvel som inden for de enkelte hovedområder. Der er således behov for at forstå behov og præferencer på et meget mere fagspecifikt niveau. Kun herved kan løsninger designes, der reelt adresserer forskernes behov og som reelt vil støtte deres forskning og fremme den danske management af forskningsdata. Denne observation motiverer strategiens fagspecifikke indsats, som løftes bedst i nationalt og/eller internationalt samarbejde.

Variationerne mellem de faglige hovedområder afspejler sig også i opfattelsen af hvad forskningsdata er for noget. I oplægget til styregruppen (→ Bilag 03) defineres termen forskningsdata således:

⁶ <http://ufm.dk/aktuelt/nyheder/2014/danmark-far-adgang-til-biovidenskabelige-data-fra-hele-europa>

Forskningsdata afgrænses til digitale data, der indsamles eller skabes mhp. på forskning - i modsætning til offentlige registre og andre kilder, der også kan gøres til genstand for forskning og videnskabelig analyse.

I denne definition pointeres administrative og juridiske skel mellem forskningsrelevante datasæt. Afdækningen af forskernes behov og præferencer viser dog, at forskerne generelt ikke opfatter denne skelnen som central. I praksis forstås forskningsdata oftest som alle de data, der er relevante for et givet forskningsprojekt. Men den endelige afgrænsning af, hvad der er forskningsdata, og dermed hvad der skal og kan være genstand for data management, må foretages på fagspecifikt niveau.

Afdækningen viste dog også, at to behov er repræsenteret på alle hovedområder:

- En Dropbox-lignende infrastruktur til nem lagring og deling af data med kolleger i ind- og udland – sikkert, lovligt, kraftfuldt, brugervenligt og økonomisk attraktivt.
- Et tilbud om rådgivning og undervisning, der adresserer alle faser af datas livscyklus og er tilpasset fagområdets betingelser, regler og best practices.

To af strategiens indsatsområder sigter således mod at løse disse behov.

4. Strategiske indsatsområder

Styregruppen har dannet sig et overblik over såvel de udenlandske erfaringer som de danske forskningsmiljøers overordnede udfordringer, behov og præferencer.

- Nationalt er der tale om en meget stor variation i hovedfagområdernes betingelser, løsninger og behov på dette område. Variationen fortsætter desuden inden for hovedområderne, hvorfor man overvejende må adressere udfordringerne på et væsentlig mere fagspecifikt niveau for sikre det nødvendige match med forskningsmiljøers konkrete udfordringer, betingelser, traditioner, behov og præferencer på forskningsdataområdet.
- Internationalt viser f.eks. de svenske erfaringer, at en national top-down politik uden udbud af infrastruktur og støttefunktioner og uden tilpasning til og engagering af forskningsmiljøerne reelt er virkningsløs. Til gengæld viser de australske erfaringer, at en parallel og gradvis opbygning af lokale og nationale politikker, infrastrukturer og støttefunktioner koblet med en stærk samarbejdsorganisation og en betydelig national økonomisk støtte – er særdeles virkningsfuld.

På denne baggrund er strategiens hovedpunkter:

- At lokale, fag- og forskernære initiativer i universitetsregi kobles med et nationalt samarbejde med fokus på tværgående videndeling - samt tværgående projekter, hvor disse kan lette og kvalificere arbejdet - og tværgående infrastrukturer, hvor disse efterspørges og kan etableres på bæredygtig vis.
- At forskningens internationale natur tages i ed og udnyttes i denne sammenhæng, således at forskningsdata management initiativer – det være sig indenfor politikker, incitamentsskabelse, kompetenceudvikling eller infrastruktur – så vidt muligt bygger på international erfaring og samarbejde og sigter på en sammenhæng mellem aktiviteter på lokalt, nationalt og internationalt niveau; eksempelvis at internationale infrastrukturer udnyttes, når det er muligt, og nationale og lokale infrastrukturer kun etableres, når det er påkrævet.
- At DEFF og DeiC faciliterer og økonomisk støtter denne proces i samarbejde med danske universiteter og de kulturministerielle bevaringsinstitutioner.

Hovedindsatsen for de næste tre år er mere konkret:

- Etablering eller udbygning støttefunktioner på universiteterne indenfor forskningsdataområdet - der på baggrund af et globalt overblik kan rådgive forskerne omkring brug af infrastrukturer og informationsteknologiske værktøjer, metadatering og informationsorganisatoriske problematikker samt lovgivning og juridiske problematikker
- Etablering af et samlende forum for disse støttefunktioners nøglepersoner med økonomisk støtte fra DEFF med henblik på videndeling og erfaringsudveksling på tværs af institutioner samt på identifikation og specifikation af fælles infrastrukturer og supportfunktioner, hvor de måtte være nødvendige
- Etablering eller udbygning af tilsvarende støttefunktioner på forskningsdataområdet hos bevaringsinstitutionerne, hvorved nøglepersoner herfra kan indgå i det nationale forum for

samarbejde om forskningsdata management

- Igangsættelse af et antal pilotprojekter støttet af DEFF og/eller DeiC - det drejer sig dels om en række fagspecifikke projekter til støtte for afklaring af best practice på fagområdet, dels om etablering af en dansk akademisk "dropbox".

Strategien har seks indsatsområder, der er konkretiseret i det følgende. Finansiering og handlingsplaner er beskrevet i hovedafsnit 5.

A. Politikker

Det må forventes, at alle universiteter i den nærmeste fremtid igangsætter processer for at udvikle politikker for forskningsdata management. Med henblik på at lette arbejdet for forskere og forskningsprojekter på tværs af institutioner, bør universiteterne videndele og samarbejde for at harmonisere politikkerne i videst muligt omfang. Med henblik på at lette det internationale samarbejde, bør disse processer løbende informeres om udviklingen af politikker i de lande, hvormed der i særlig grad samarbejdes forskningsmæssigt, således at harmonisering kan overvejes, når det er muligt og ønskeligt.

Dette arbejde vil kunne drage nytte af afdækning C: "International best practice for forskningsdata management politikker" – særligt konsulentrapporten: "Current Best Practice for Research Data Management Policies" by Simon Hodson and Laura Molloy – Codata, May 2014 (→ Bilag 08).

Tilsvarende vil arbejdet med udviklingen af politikker kunne drage nytte af styregruppens afdækning B: "Juridiske rammer og udfordringer" (→ Bilag 07).

→ Den praktiske implementering af dette indsatsområde kan ske gennem etablering af et nationalt forum for samarbejde om forskningsdata management som beskrevet i afsnit 4.F.

B. Incitamentsskabelse

Der bør arbejdes konsekvent (og i nationale og internationale fællesskaber) mod etablering af incitamenter for forskerne i forbindelse med forskningsdata management. Der bør her fokuseres på:

- Klare og let forståelige politikker på universiteterne med bedst mulig kompatibilitet med politikker hos samarbejdspartnere og forskningsfinansierende instanser.
- Kompetente og let tilgængelige støttefunktioner på universiteterne, der kan bistå forskerne ved udfordringer på alle relevante områder: de informationsteknologiske (IT-infrastruktur), de informationsorganisatoriske (metadatering) og de juridiske.
- Effektive og attraktive tilbud om IT-infrastruktur i overensstemmelse med fagområdernes behov og præferencer.
- Udbredelse af kendskabet til forskningsområder, hvor oparbejdning, dokumentation og tilrådighedsstillelse af datasæt er med til at fremme den enkelte forskers status og karriere - for eksempel gennem meriterende citering af forskningsdata.

- Sikring af den nødvendige infrastruktur til meriterende citering af forskningsdata i form af unikke og globalt anerkendte identifikatorer for forskere og datasæt.
 - Bistå forskerne med at synliggøre deres forskningsdataindsatser gennem eksponering af datasæt og forskere på tværs af institutioner, nationalt og internationalt i overensstemmelse med fagområdets best practice.
- Flere af disse indsatser kan i praksis implementeres gennem etablering af et nationalt forum for samarbejde om forskningsdata management som beskrevet i afsnit 4.F. De øvrige knytter sig til indsatser på infrastrukturuområdet, se afsnit 4.C nedenfor.

C. Infrastruktur

Universiteterne og bevaringsinstitutionerne bør i samarbejde fortsætte og udbygge det arbejde, styregruppen kun har påbegyndt, med at skabe klarhed over:

- De enkelte forskningsmiljøers og –specialers konkrete behov og præferencer – med blik for såvel lokale, nationale som globale løsninger og deres samspil.
 - En uddybning kun på overordnet hovedområdeniveau er ikke tilstrækkeligt grundet den store diversitet hvad angår forskningsdata-udfordringer og –traditioner inden for de enkelte hovedområder.
 - Det eksisterende udbud af systemer og services – lokalt, nationalt og globalt – med angivelse af omfanget af deres opfyldelse af danske behov – med redegørelse for deres faciliteter, kapaciteter, brugervenlighed, omkostninger, erfaringer, juridiske forhold, forretningsmodeller, bæredygtighed mv.
 - Alle relevante systemer og services bør identificeres og beskrives – og de mest lovende bør aftestes og evalueres i forhold til de relevante fagområders behov.
 - De systemer og services, der pt. mangler i infrastrukturen for at give danske forskere effektive og konkurrencedygtige betingelser for management og deling af forskningsdata og opfyldelse af det fælles kodeks for videnskabelig redelighed.
 - En specifikation af manglende systemer eller services og deres samspil med den øvrige, eksisterende infrastruktur.
- Disse indsatser kan i praksis implementeres ved
- etablering af et nationalt forum for samarbejde om forskningsdata management (se afsnit 4.F nedenfor)
 - at indkalde og støtte projekter, der på velafgrænsede fagspecialer klarlægger områdets forskningsdatamæssige udfordringer, behov, præferencer, eksisterende udbud af systemer og services, manglende systemer og services – gerne koblet med praktisk implementering og aftestning – samt bred erfaringsudveksling.

Dernæst er der identificeret et infrastrukturtiltag, der er klar til umiddelbar igangsættelse:

- En nem og sikker måde forskere at dele filer og forskningsdata på nettet. En akademisk "Dropbox", hvor danske forskere sikkert, lovligt og brugervenligt kan lagre og dele aktive forskningsdata, metadata, manuskripter mv. med relevante og specificerede kolleger i ind- og udland.

- Dette behov er i afdækningerne konstateret på tværs af hovedområder, og det illustreres endvidere klart af den udbredte, men juridisk uholdbare anvendelse af services som www.dropbox.com
- Da forskningsverdenen kan have væsentligt andre behov, end dem et kommercielt produkt som f.eks. Dropbox er målrettet, bør der gennemføres en uddybende behovsanalyse inden etablering af en sådan infrastruktur.

Endelig bør to allerede igangsatte initiativer omkring globalt anerkendte identifikatorer følges og evt. forstærkes – bl.a. til støtte for meriterende citering af forskningsdata.

- En dansk udbredelse af det globale forsker-ID ORCID, således at alle forskere bliver unikt identificerbare på nettet (jf. underafsnit 4.B ovenfor)
 - Dette er allerede sat i gang med økonomisk støtte fra DEFF og opfølgning og opbakning fra Styrelsen for Forskning og Innovation og den Nationale styregruppe for Open Access, hvis formål ligeledes fremmes af udbredelsen af en sådan identifikator.
 - Projektet bør følges nøje af det nationale forum for forskningsdata management for at sikre, at behovene på dette område bliver dækket, og at evt. behov for supplerende indsatser bliver identificeret og adresseret.
- En dansk service til støtte for udtagelse af globalt anerkendte DOI-identifikatorer til datasæt, således at alle datasæt bliver unikt identificerbare og citerbare (jf. underafsnit 4.B ovenfor)
 - DTU har som en del af et afsluttet DEFF-projekt etableret et dansk medlemskab af og dansk kontor for DataCite, der er et globalt DOI-registreringssamarbejde for datasæt, etableret i et direkte samarbejde mellem universiteter, forskningsinfrastrukturenheder og forskningsbiblioteker.
 - Det anbefales, at denne opgave og service fremover varetages af DeiC, der vil være et naturligt regi for en sådan national service. Endvidere anbefales, at DeiC overvejer en anden måde for dækning af de årlige omkostninger på ca. 150.000 kr. end løbende fakturering af brugerne.

D. Kompetenceudvikling og forskerstøtte

Alle universiteter og bevaringsinstitutioner bør etablere lokale tilbud om forskerstøtte på området forskningsdata management og samarbejde om udvikling af disse støttefunktioner og kompetencer. Der vil givet være stor variation mellem udformningen på de enkelte universiteter - mellem større og mindre, mellem mono- og multifakultære, etc.

Dette har fire aspekter:

- Støttefunktioner på universiteterne og bevaringsinstitutionerne, der kan bistå forskerne ved udfordringer på alle relevante områder: De informationsteknologiske (IT-infrastruktur), de informationsorganisatoriske (metadatering) og de juridiske.
- En effektiv ramme om nationalt samarbejde, videndeling og erfaringsudveksling på området kompetenceudvikling og forskerstøtte.
- Et tæt samarbejde med det fremvoksende eScience kompetencecenter i DeiC-regi.

- En kickstart af den lokale kompetenceudvikling såvel som det nationale samarbejde gennem etablering af et fælles uddannelsesforløb for universiteternes nøglepersoner. Dette forløb bør trække på de bedste danske og udenlandske kræfter og have en klar "train the trainer" orientering.
- Disse indsatser kan i praksis implementeres ved etablering af et nationalt forum for samarbejde om forskningsdata management (se afsnit 4.F nedenfor).

E. Fremtidig styringsstruktur

Strategien omfatter organisatoriske tiltag såvel lokalt som nationalt samt et udstrakt internationalt samarbejde. Styringsstrukturen er således vigtig og er revideret på baggrund af DeiC og DEFF's behandling af høringsvarene. Den vil blive evalueret efter det første år i funktion.

- **Nationalt Forum for Forskningsdata Management** etableres med henblik på tværgående erfaringsudveksling, videndeling og samarbejde, således som beskrevet ovenfor og uddybet nedenfor i afsnit 4.F.

Dette forum samler nøglepersoner fra universiteternes forskerstøtteenheder vedrørende forskningsdata sammen med tilsvarende aktører fra bevaringsinstitutionerne.

Forummet er omdrejningspunkt for den praktiske erfaringsudveksling og det praktiske samarbejde om udvikling af kompetencer og infrastrukturer, ligesom det koordinerer den danske indsats i forbindelse med internationale aktiviteter og deltager i eller følger relevante organer, følger projekter over landegrænser som ELIXIR og følger EU på området åbne data. Forum står dermed som garant for, at der ikke opbygges lokale løsninger, hvis udenlandske løsninger kan anvendes.

Forummet etableres som et 3-årigt pilotprojekt, hvor aktiviteter og leverancer afrapporteres til DeiC bestyrelsen og DEFF styregruppen, som årligt evaluerer det og tager stilling til, hvorvidt og hvordan det skal fortsættes.

Formmets aktiviteter ledes af en formand, der udpeges af DeiC bestyrelsen og DEFF Styregruppen på baggrund af indstilling fra interessenterne.

- **Den Nationale eScience Komite** under DeiC har ansvar for de videnskabelige aspekter i arbejdet i det nationale forum. eScience Komiteen rummer udpegede forskere fra alle 8 universiteter og alle fagområder og udvider sit virkefelt til også at omfatte de videnskabelige aspekter af forskningsdata management.
- **Teknisk Referencegruppe for Forskningsdata Management (DM TekRef)** etableres som en åben referencegruppe med fri deltagelse fra alle interessenter indenfor forskningsdata managementområdet.

Referencegruppens aktiviteter ledes af en formand, der udpeges af DeiC bestyrelsen og DEFF Styregruppen efter indstilling fra referencegruppen.

- **Ledelses Change Advisory Board for Forskningsdata Management (DM LedelsesCAB)** etableres som et ledelsesmæssigt knudepunkt for aktiviteterne.

DM LedelsesCAB behandler indstillinger fra eScience komiteen, DM TekRef og Nationalt

Forum for Forskningsdata Management og sender sine anbefalinger til beslutning hos henholdsvis DeiC bestyrelsen og DEFF Styregruppen.

DM LedelsesCABen har ansvar for at føre tilsyn med den nationale indsats og rapportere til DeiC bestyrelsen og DEFF Styregruppen.

DM LedelsesCAB'en består af:

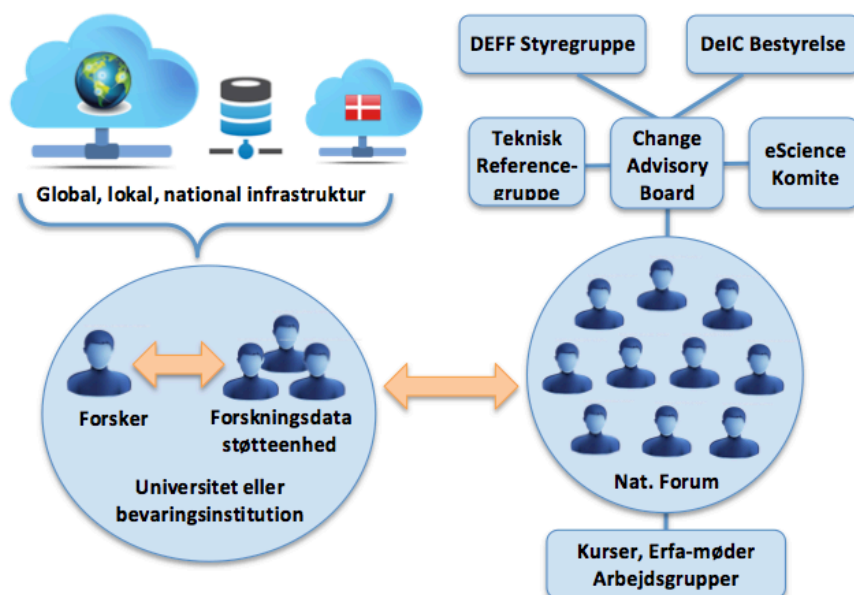
- Formanden for Nationalt Forum for Forskningsdata Management
- Formanden for eScience komiteen
- Formanden for DM TekRef
- En repræsentant udpeget af Rektorkollegiets CIO-forum
- En repræsentant udpeget af bevaringsinstitutionerne
- En repræsentant udpeget af registre (KOR)
- En repræsentant udpeget af operatøren af Nationalt Forum
- En repræsentant udpeget af DEFF Sekretariatet
- En repræsentant udpeget af DeiC Sekretariatet

DM LedelsesCAB ledes af en formand, der udpeges af DeiC bestyrelsen og DEFF Styregruppen på baggrund af indstilling fra interessenterne.

- **DeiC bestyrelsen og DEFF Styregruppen** behandler indstillinger fra Data Management LedelsesCAB. Beslutninger skal træffes i enighed.

DeiC og DEFF evaluerer løbende indsats, resultater og medfinansiering, og tager på den baggrund beslutning om fortsat finansiering af aktiviteterne. Formanden for forummet skal således to måneder før udløbet af bevillingsperioden indsende evaluering samt plan for næste bevillingsperiode gennem DM LedelsesCAP til DeiC bestyrelsen og DEFF Styregruppen, førend midler for næste bevillingsperiode kan frigives.

- **Rektorkollegiet's CIO-forum** indgår i arbejdet og har bl.a. ansvar for at undersøge og koordinere valg af relevante infrastrukturløsninger universiteterne imellem og for at etablere fælles løsninger sammen med DeiC, hvor det giver mening.



F. Pilotprojekter

Dette afsnit opsummerer de pilotprojekter, der er nævnt ovenfor under afsnit A.-E.

Strategien peger overordnet på:

1. At etablere et nationalt forum for samarbejde om forskningsdata management som et konsoliderende pilotprojekt, der med økonomisk støtte fra DEFF løfter hovedparten af de strategiske anbefalinger bekræftet ovenfor under:
 - A. Politikker
 - B. Incitamentsskabelse
 - C. Infrastruktur
 - D. Kompetenceudvikling og forskerstøtte
 - E. Fremtidig styringsstruktur
2. At supplere dette store konsoliderende pilotprojekt med nogle mindre, specifikke pilotprojekter, der med økonomisk støtte fra DEFF og/eller DeiC sigter mod at løse problemer, der allerede nu er velforståede og derfor kandidater til en umiddelbar indsats.

Ad. 1. Nationalt forum for samarbejde om forskningsdata management

Hensigten er dels at fremme fag- og forskernære initiativer i universitetsregi og dels at koble disse sammen i et udstrakt nationalt samarbejde med fokus på tværgående videndeling - samt tværgående projekter, hvor disse kan lette og kvalificere arbejdet - og tværgående infrastrukturer, hvor disse i praksis efterspørges og kan etableres på bæredygtig vis.

Dette forudsætter, at universiteterne og bevaringsinstitutionerne etablerer forskerrettede støttefunktioner vedr. forskningsdata – tilpasset det enkle universitets størrelse og faglige diversitet mv. Der vil givet være stor variation mellem udformningen på de enkelte universiteter - mellem de større og de mindre, mellem de mono- og de multifakultære, etc.

Personalet ved disse enheder er basis for såvel den lokale indsats og service som det praktiske nationale samarbejde på området.

- Der etableres et nationalt forum, der samler universiteternes og bevaringsinstitutionernes praktiske nøglepersoner på området - med kompetencestøtte, projektlederstøtte og sekretariatsbetjening fra DeiC Sekretariatet.
- Det nationale forum udgør omdrejningspunktet for den praktiske erfaringsudveksling og det praktiske samarbejde om udvikling af kompetencer og infrastrukturer.
- Det nationale forum støttes økonomisk i form af frikøb af
 - en formand på halv tid til at lede arbejdet – udpeget af DeiC og DEFF
 - institutionernes nøglepersoner (på op til halv tid) til at indgå det nationale arbejde med dets afklarings- og analyseprojekter, arbejdsgrupper og erfa-grupperHerigennem sikres forum de nødvendige ressourcer samt den nødvendige fremdrift og ledelse.
- Det nationale forum spiller det første år en særlig rolle som kickstarter af såvel den lokale kompetenceudvikling som det nationale samarbejde gennem etablering af et fælles

uddannelsesforløb for universiteternes nøglepersoner på området.

- Det nationale forum beskæftiger sig typisk med områder som:
 - Uddannelse og opkvalificering på data management området
 - Fagspecifikke behovsafklaringer og identifikation af fagspecifikke best practices
 - Rådgivning og skabeloner ifm. udarbejdelse af data management planer
 - Afklaring og rådgivning ifm. metadatering og identifikatorer
 - Afklaring og rådgivning omkring juridiske aspekter
 - Afklaring og rådgivning på det tekniske område – herunder erfaringer med eksisterede IT-værktøjer og identifikation og specifikation af manglende værktøjer
 - Optimering af samspillet mellem lokale, nationale og globale infrastrukturer
 - Kommunikation og markedsføring af de lokale støtteenheders service
- Det nationale forum identificerer potentielle behov for nye fælles infrastrukturer og supportfunktioner og danner ramme om ideudvikling af disse.
- Det nationale forum koordinerer den danske indsats i forbindelse med internationale udviklings- og driftstiltag på det operative, implementerende niveau, og deltager i eller følger relevante internationale samarbejdsfora/projekter, som har en vedvarende natur og/eller er af aktuell strategisk betydning, f. eks.: Research Data Alliance, Knowledge Exchange, e-Infrastructure Reflection Group e-IRG, European Data Infrastructure (EUDAT), Nordic e-Infrastructure Collaboration (NeIC).
- Det nationale forum fremlægger hvert halve år til godkendelse i DM LedelsesCAB en rapport over arbejdet og de opnåede resultater samt en plan for det kommende halve års indsatser. I denne forbindelse behandles ligeledes eventuelle nye udfordringer og muligheder.
- Dette kan i tur føre til nye samarbejdsprojekter/projektansøgninger på forskningsdata området – for alle eller for særligt interesserede universiteter o.a.. Sådanne projekter vil forventeligt kunne ansøge om økonomisk støtte fra DEFF, DeiC og/eller forskningsrådssystemet. De enkelte universiteter og institutioner afgør selv deres behov for deltagelse i sådanne fælles projekttiltag.

Eksempler på det nationale forums leverancer på de seks strategiske indsatsområder gives nedenfor. Sådanne indsatser og leverancer beskrives nærmere i forums halvårsplan inklusive forslag til prioritering og forelægges styregruppen til godkendelse:

A. Politikker

- En kogebog/inspirationskatalog over elementer i forskningsdata politikker

B. Incitamentsskabelse

- En guide til meriterende management og meriterende citering og af forskningsdata
- Fagspecifikke afklaringer af forskernes behov og præferencer

C. Infrastruktur

- Fagspecifikke best practice guides til valg/brug af forskningsdata-infrastruktur
- Fagspecifikke guides til metadatering af datasæt – formater, ontologier, systemer
- Aftestning og evaluering af udbudte IT-løsninger på forskningsdataområdet
- En løbende opdateret oversigt over relevante udbudte IT-løsninger/systemer – generiske og fagspecifikke – med danske erfaringer og evalueringer
- Identifikation og specifikation af evt. manglende IT-infrastruktur

- Eventuelle roller i forbindelse med etablering af manglende IT-infrastruktur
- En guide til tildeling af unikke identifikatorer til forskere og datasæt

D. Kompetenceudvikling og forskerstøtte

- En effektiv national ramme om nationalt samarbejde, videndeling og erfaringsudveksling på området kompetenceudvikling og forskerstøtte
- Et indledende "train the trainer" kursus for medlemmerne af forum og øvrigt personale ved universiteternes støtteenheder på forskningsdata området
- Et samarbejde om udvikling af fagspecifikke kurser i forskningsdata management
- Et fælles repository for kursusmaterialer på forskningsdataområdet (tekster, præsentationer/slides, videoer, quiz'er etc.)
- Fælles afklaring af juridiske spørgsmål
- Rådgivning og skabeloner ifm. udarbejdelse af data management planer
- Et fælles inspirationskatalog til markedsføring af støttefunktioner og -tilbud

E. Fremtidig styringsstruktur

- Halvårlige rapporter om indsatser og resultater – til godkendelse i DM LedelsesCAB
- Halvårlige aktivitetsplaner til godkendelse i DM LedelsesCAB
- Ad-hoc forslag om nye tiltag, projektforslag mv. til godkendelse og videre opfølgning i DM LedelsesCAB.

Ad. 2. Andre pilotprojekter

Hensigten er at supplere det ovennævnte store konsoliderende pilotprojekt med mindre, specifikke pilotprojekter, der sigter mod at løse problemer, der allerede nu er velforståede og derfor kandidater til en umiddelbar indsats. Disse pilotprojekter vil således kunne indkaldes inden det ovennævnte brede forum er på plads.

- **Fagspecifikke pilotprojekter** til støtte for udvikling af best practice anbefalinger for området. På velafgrænsede fagområder og i tæt samarbejde med forskningsprojekter og deres forskere. Praktisk afprøvning af koncepter og systemer i forbindelse med (potentielt alle) dele af forskningsdatas livscyklus – fra data management planlægning til publicering og overførsel til arkiv med henblik på langtidsbevaring. Løbende afrapportering til/videndeling med det nationale forum.

Dernæst anbefales et konkret infrastrukturtiltag, for hvilke behovet allerede er klart konstateret, og som således er kandidat til umiddelbar igangsættelse:

- En nem og sikker måde forskere at dele filer og forskningsdata på nettet. En **akademisk "Dropbox"**, hvor danske forskere sikkert, lovligt og brugervenligt kan lagre og dele aktive forskningsdata, metadata, manuskripter mv. med relevante og specificerede kolleger i ind- og udland.
 - Dette behov er i afdækningerne konstateret på tværs af hovedområder, og det illustreres endvidere klart af den udbredte, men juridisk uholdbare anvendelse af services som www.dropbox.com
 - Da forskningsverdenen kan have væsentligt andre behov, end dem et kommercielt produkt som f.eks. Dropbox er målrettet, bør der gennemføres en uddybende behovsanalyse inden etablering af en sådan infrastruktur.

Endelig anbefales det at følge og evt. forstærke to allerede igangsatte indsatser omkring udbredelse af globalt anerkendte indikatorer – bl.a. til støtte for meriterende citering af forskningsdata.

- En dansk udbredelse af det **globale forsker-ID ORCID**, således at alle forskere bliver unikt identificerbare på nettet (jf. underafsnit 4.B ovenfor)
 - Dette er allerede sat i gang med økonomisk støtte fra DEFF og opfølgning og opbakning fra Styrelsen for Forskning og Innovation og den Nationale styregruppe for Open Access, hvis formål ligeledes fremmes af udbredelsen af en sådan identifikator.
 - Projektet bør følges nøje af det nationale forum for forskningsdata management for at sikre, at behovene på dette område bliver dækket, og at evt. behov for supplerende indsatser bliver identificeret og adresseret.
- En dansk service til støtte for udtagelse af globalt anerkendte **DOI-identifikatorer** til datasæt, således at alle datasæt bliver unikt identificerbare og citerbare (jf. underafsnit 4.B ovenfor)
 - DTU har som en del af et afsluttet DEFF-projekt etableret et dansk medlemskab af og dansk kontor for DataCite, der er et globalt DOI-registrerings-samarbejde for datasæt, etableret i et direkte samarbejde mellem universiteter, forskningsinfrastrukturenheder og forskningsbiblioteker.
 - Det anbefales, at denne opgave og service fremover varetages af DeiC, der vil være et naturligt regi for en sådan national service. Endvidere anbefales, at DeiC overvejer en anden måde for dækning af de årlige omkostninger på ca. 150.000 kr. end løbende fakturering af brugerne.

Samtidigt som det forudsættes, at de nationale tiltag er forankret i universiteternes behov og evt. eksisterende tiltag, og gennemføres i et samarbejde mellem lokale universitetskræfter og nationale enheder, så skal de så vidt muligt bygge på international erfaring og samarbejde. Dobbeltarbejde skal undgås og de sparsomme ressourcer fokuseres, sådan at tiltagene får størst mulig gennemslagskraft og nytte for dansk forskning.

Herudover gælder følgende generelle kriterier for pilotprojekter:

- Bred anvendelighed, dvs. resultatet er brugbart på tværs af flere institutioner.
- Tværinstitutionelt samarbejde med sigte på at styrke det nationale og internationale samarbejde om infrastruktur og støttefunktioner, således at dobbeltarbejde og redundant infrastruktur undgås.
- Den skitserede løsning skal demonstrere merværdi ud over det, som i forvejen ligger i institutionens opgaveportefølje.
- Der skal være brugerforankret udvikling og aftestning af det nyttige i infrastrukturen, servicen eller kompetenceudviklingstiltaget, der påtænkes.
- Plan for økonomisk bæredygtighed over længere tid skal indgå såvel i projektansøgning som i afrapportering.

5. Handlingsplaner og finansiering

Den samlede økonomi for 2015-18 fremgår af nedenstående oversigt:

Aktivitet	Aktivitet indhold	Finansieres af DEFF	Finansieres af DeiC	Medfinansiering (in-kind) fra universiteter	Medfinansiering (in-kind) fra bevaringsinst.
A	Nationalt Forum	9.500 Kkr.		6.000 Kkr.	2.250 Kkr.
B	Train the trainer kursus	500 Kkr.			
C	Fagspecifikke pilotprojekter		5.000 Kkr.	5.000 Kkr.	
D	Akademisk "dropbox"		1.000 Kkr.		
E	Projektleder		2.250 Kkr.		
F	"Data management i praksis" projekt	4.522 Kkr.		3.708 Kkr.	
I alt		14.522 Kkr.	8.250 Kkr.	16.958 Kkr.	

Medfinansieringen fra universiteterne og bevaringsinstitutionerne er en forudsætning for, at midlerne fra DeiC og DEFF kan udløses.

Forslag til finansiering og organisering af de enkelte aktiviteter A-F fremgår nedenfor.

A. Nationalt forum for samarbejde om forskningsdata management

DEFF finansierer denne aktivitet, hvis opgaver og leverancer er detaljeret beskrevet i afsnit 4.F. DEFF indgår aftale med et universitet om at være operatør og økonomisk og juridisk ansvarlig for det 3-årige projekt, der realiserer det nationale forum og dets leverancer. Aktiviteten evalueres årligt af DeiC bestyrelsen og DEFF styregruppen, der tager stilling til, hvorvidt og hvordan der skal fortsættes.

Samlet økonomi 2015-18

	DEFF tilskud	Medfinansiering
A.1. Frikøb af formand	750 Kkr.	
A.2. Deltids frikøb af nøglepersoner på universiteter	6.000 Kkr.	6.000 Kkr.
A.3. Deltids frikøb af nøglepersoner på bevaringsinstitutioner	2.250 Kkr.	2.250 Kkr.
A.4. Driftsomkostninger	500 Kkr.	
I alt	9.500 Kkr.	8.250 Kkr.

Ad. A.1:

Der udpeges en formand med såvel projektlederkompetencer som indsigt i forskningsinfrastruktur og forskningens vilkår og organisering. Omkostningen er beregnet som 250 Kkr. årligt til frikøb i 3 år.

Ad. A.2:

En række nøglepersoner fra universiteternes støttefunktioner på forskningsdata management området tilknyttedes det nationale forum. Disse nøglepersoners løn dækkes delvist af DEFF på basis af årlige ansøgninger fra universiteterne.

For den enkelte nøgleperson kan frikøbet maksimalt udgøre 50% af vedkommendes allokering til forskningsdata management opgaver. Omfanget af frikøb vil endvidere afhænge af nøglepersonernes aktivitetsniveau i det nationale samarbejde.

Omkostningen er beregnet på basis af en gennemsnitsløøn på 500 Kkr./år i 3 år for 8 universiteter. Dette er en gennemsnitsbetragtning, idet nogle universiteter antagelig kun vil have behov for at tilknytte f.eks. et halvt årsværk til forskningsdatastøtte og dermed kan ansøge om frikøb af en kvart stilling, mens andre vil have behov for at tilknytte flere personer og således vil have muligheden for at ansøge om tilsvarende større omfang af frikøb i den udstrækning, de er indstillede på at stille deres nøglepersoner til rådighed for forums aktiviteter.

Ad. A.3:

Tilknytning af nøglepersoner ved bevaringsinstitutionerne følger de samme retningslinjer og omkostningsberegninger som ved universiteterne.

Ad. A.4:

Beløbet er afsat til dækning af driftsomkostninger.

Møderne forventes afholdt på medlemsinstitutionerne, hvor hvert mødes vært betaler lokaler og forplejning.

Stipuleret handlingsplan for etablering og første år 2015-16

2015	Q1	DeiC og DEFF udsender oplæg til universiteterne og bevaringsinstitutionerne med opfordring til <ol style="list-style-type: none"> at indstille medlemmer og formandskandidater til det nationale forum og DM LedelsesCAB, og at allokere lokale ressourcer til støttefunktioner vedr. forskningsdata management (medfinansiering som betingelse for tilskud)
2015	Q1-2	Universiteterne og bevaringsinstitutionerne forbereder deres støttefunktioner vedr. forskningsdata management
2015	Q1	Projektleder og operatør udarbejder forslag til "Train the trainer" kursus i andet halvår (overgår til formanden, når denne er tiltrådt)
2015	Q2	1. evaluering af lokale forberedelser og potentiale for medfinansiering efter universiteternes og bevaringsinstitutionernes tilbagemeldinger
2015	Q2	Nedsættelse af det nationale forum og DM LedelsesCAB
2015	Q2	Formand og sekretariat udarbejder forslag til arbejdsplan
2015	Q3	Universiteterne og bevaringsinstitutionerne afslutter bemanding af deres støttefunktioner
2015	Q3	Første møde i forum - behandling af forslag til "Train the trainer" kursus samt forslag til arbejdsplan for etableringsåret
2015	Q4	Første møde i DM LedelsesCAB - behandling af forslag til arbejdsplan for etableringsåret og evt. indsendte forslag til projekter
2015	Q4	2. evaluering af medfinansiering og involvering foretages af DeiC bestyrelsen og DEFF Styregruppen
2015	Q4	"Train the trainer" kursus gennemføres
2016	Q1	To måneder før udløbet af bevillingsperioden indsender forum evaluering samt plan for næste bevillingsperiode gennem DM LedelsesCAB til DeiC bestyrelsen og DEFF Styregruppen
2016	Q2	3. evaluering af resultater og organisering, samt beslutning vedr. fortsat finansiering foretages af DeiC bestyrelsen og DEFF styregruppen

B. Train the trainer kursus

DEFF finansierer denne aktivitet.

Samlet økonomi 2015

	DEFF tilskud
Konsulent- og rejseudgifter til udenlandske eksperter	500 Kkr.

Stipuleret handlingsplan 2015

Q1-2	Kurset planlægges af projektleder og opeartør (overgår til formanden, når denne er rekrutteret)
Q3	Kursusplanen forelægges forum
Q4	Kurset gennemføres

C. Fagspecifikke pilotprojekter

DeIC finansierer fem fagspecifikke pilotprojekter.

Samlet økonomi 2015-18

	DeIC tilskud	Medfinansiering
C.1. Pilotprojekt 1	1.000 Kkr.	1.000 Kkr.
C.2. Pilotprojekt 2	1.000 Kkr.	1.000 Kkr.
C.3. Pilotprojekt 3	1.000 Kkr.	1.000 Kkr.
C.4. Pilotprojekt 4	1.000 Kkr.	1.000 Kkr.
C.5. Pilotprojekt 5	1.000 Kkr.	1.000 Kkr.
I alt	5.000 Kkr.	5.000 Kkr.

Stipuleret handlingsplan 2015-18

2015	Q2	Forum identificerer i samarbejde med eScience Komite og DM TekRef kriterier og områder
2015	Q2	Projektkald i DeIC regi
2015	Q3	Sekretariatet evaluerer projekter og medfinansiering, indstiller til DM LedelsesCAB
2015	Q4	DM LedelsesCAB beslutter hvilke pilotprojekter, der gives bevilling
2016	Q1	Projektstart
		Løbende orientering gennem hele projektforløbet til såvel forum som DM Ledelses CAB og resten af styringsstrukturen
2017	Q4	Projekt slut
2018	Q1	Formidling på et fælles event for dansk forskningsdata management

D. Akademisk "dropbox"

DeIC finansierer denne aktivitet, bestående i behovsanalyse, specifikation og etablering af en dropbox-lignende facilitet, der opfylder kravene for sikker opbevaring og udveksling af forskningsdata. Dette skal ske i samarbejde med CIO-forum, og arbejdet kan påbegyndes nu.

Beløbet er afsat til lønning af f.eks. konsulenter til behovsanalyse og specificering.

Samlet økonomi 2015-16

	DeiC tilskud
Afklaring, specifikation, forretningsmodel	1.000 Kkr.

E. Projektleder

DeiC finansierer og dedikerer en projektleder for arbejdet i en periode på tre år.

I den første periode vil det Nationale Forum endnu ikke være forsynet med en formand og DeiC's projektleder vil spille en vigtig rolle med at støtte igangsætningen af arbejdet, herunder planlægningen af det indledende Train the trainer kursus mv. Senere vil antallet af pilotprojekter, arbejdsgrupper, internationale relationer mv. være så stort, at Forum og dets formand vil have stor glæde af at kunne trække på en kyndig og erfaren arbejdskraft.

Samlet økonomi for aktiviteten 2015-18

	DeiC tilskud
Projektleder på fuld tid	2.250 Kkr.

F. "Data management i praksis" projekt

DEFF har i december 2014 bevilget 4.522.000 kr. til "Data management i praksis" projektet, der inklusive medfinansiering andrager 8.230.000 kr. Projektet løber i perioden marts 2015 til februar 2017 og har deltagere fra 6 universiteter og 3 bevaringsinstitutioner.

Projektet tager fat på at løse udfordringen omkring data management af forskningsdata gennem etablering af en række tjenester i praksis. I projektet implementeres tjenester til understøttelse af alle dele af data management livscyklus.

Tjenesterne vil blive etableret i samarbejde med udvalgte igangværende forskningsprojekter og -infrastrukturer fordelt over flere forskningsmæssige hovedområder. Konkret vil udviklingen ske i syv cases, som hver skal levere en løsning i forhold til en bestemt forskergruppe. Projektet vil med det nationale forum som referencegruppe sikre koordinering med øvrige nationale og lokale initiativer på området.

6. Executive summaries af fem afdækninger

Det bemærkes, at mange afdækningerne er foreløbige, subjektive og/eller bygger på et begrænset datagrundlag. Derfor tages der forbehold for repræsentativiteten af konklusionerne i afdækningerne, ligesom det pointeres, at konklusionerne ikke nødvendigvis afspejler styregruppens holdning.

A. Bevillingsgiveres krav til forskningsdata management

Introduktion

Mhp. at informere Styregruppens overvejelser ift. det strategiske element "politikker for data management" besvarer afdækningen spørgsmålet: Hvilke eksplicitte krav stiller bevillingsgivere pt. til forskningsdata management, som danske forskningsprojekter skal leve op til? Afdækningen oplister således krav, henstillinger og anbefalinger fra væsentlige indenlandske og udenlandske offentlige såvel som private bevillingsgivere med relevans for dansk forskning.

Metode

Undersøgte domæner: EU Horizon 2020 - specifikt Open Research Data Pilot (OA til offentligt finansierede forskningsdata), danske offentlige forskningsråd og fonde, danske private fonde og enkelte store US og UK fonde.

Kilder: Internettilgængelige ressourcer, fx overordnede bevillingsgiver datadelingspolitikker, informationsmateriale med bevillingsvilkår og implementeringsguidelines, opslag, generiske tilskudsaftaler (model grant agreements) mv. Herudover redegørelse fra Dansk Data Arkiv og mail korrespondance med repræsentanter for private fonde.

Hovedpunkter

- **EU Open Research Data Pilot:** Detaljerede data management planer er et bevillingskrav. Data nødvendige for at validere forskningsresultater, der ligger til grund for publikationer, o.a. videnskabelige oplysninger, skal deponeres og stilles til rådighed for andre forskere, industrier og borgere. 3. part skal så vidt muligt kunne tilgå, data mine, udnytte og formidle data, som derfor skal følges af en genanvendelseslicens. Et forskningsprojekt kan på et vilkårligt tidspunkt helt eller delvist fravælge piloten (opt out).
- **EU (udenfor pilot):** Alle projektansøgninger under "Research and Innovation actions" og "Innovation actions" skal indeholde en sektion om forskningsdata management, hvoraf bl.a. skal fremgå hvordan data udnyttes og/eller deles/gøres tilgængelig for verifikation og genbrug, og hvordan de kureres og bevares.
- **Danske offentlige forskningsråd og fonde:** Det Frie Forskningsråd (DFF Sundhed og Sygdom og DFF Samfund og Erhverv) kræver, at indsamlede/købte kvantitative eller kvalitative data afleveres i dokumenteret stand til Dansk Data Arkiv. Datatilsynets regler tilskriver, at personhenførbare data destrueres eller anonymiseres, når de ikke længere benyttes til det angivne formål. Senere behandling af sådanne data, der alene sker i historisk, statistik eller videnskabeligt øjemed, er imidlertid tilladt. Fortsat opbevaring af data efter endt brug er mulig, såfremt de overføres til arkiv efter reglerne i arkivloven. Overføres til arkivering i Statens Arkiver, anses lovens krav om sletning som opfyldt.
- **Danske private fonde:** Ingen eksplicitte krav vedr. data management eller deling af data.
- **US og UK fonde:** Her opfordres til/forventes at data deles, men der stilles ikke krav. Data management planer er kun et bevillingskrav fra US National Science Foundation og i visse tilfælde UK Wellcome Trust.

(→ Bilag 06 for den fulde afdækningsrapport)

B. Danske juridiske rammer og udfordringer

Introduktion

Mhp. at informere Styregruppens overvejelser ift. de strategiske elementer "politikker for data management" og "incitamentskabelse" skitserer afdækningen lovgivnings- og retsområderne, aktiviteterne på hvert af trinene i forskningsdata livscyklus sætter i spil. Afdækningen danner grundlag for videre uddybning af de juridiske rammer og udfordringer knyttet til indsamling, håndtering og formidling samt genanvendelse af forskningsdata.

Metode

Juridisk institut, AAU, har udarbejdet oversigt over relevante juridiske og lovgivningsmæssige rammer/ domæner. Styregruppens institutioner har suppleret med uddybende notater om 1) krav til forskningsprojekter ihht. lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter, 2) pligtafleveringslovens forordninger vedr. offentliggjort materiale, herunder forskningsdata, og 3) krav til institutionerne ihht. ISO 27001-sikkerhedsstandard (forvaltning af informationsaktiver), persondataloven og arkivloven.

Hovedpunkter

I eksisterende lovgivning eller uddybende retningslinjer er der i vidt omfang direkte eller indirekte taget stilling til indsamling og håndtering af forskningsdata, ligesom juridisk forskning og litteratur på områder som ophavsret, patentret, persondataret, IT-ret mv. rummer stillingtagen til mange af de juridiske problemstillinger knyttet hertil.

Der er imidlertid behov for en beskrivende afdækning af relevante dele af eksisterende lovgivning, identifikation og analyse af forskningsdata-problemstillinger i lyset af forskning/'state of the art' på diverse relevante retsområder, samt en samlende beskrivelse i et dynamisk data management-perspektiv.

- **Forskningssamarbejder:** Den samlede mængde af dansk lovgivning af betydning herfor er omfattende. Hertil de retningslinjer, aftaleretlige dokumenter og EU retten, der regulerer EUs forskningsprogrammer. Der er behov for en beskrivende overordnet afdækning af de lovgivningsmæssige rammer for forskningssamarbejder mhp. en vurdering af, om der er behov for regler, der fastlægger ejerskab/deling af ejerskab til data og beskyttelse af data og ophavsrettigheder, eller om stillingtagen er bedst på aftalebasis.
- **Indsamling og generering af data:** Der behov for kvalificering af forskellige typer af data i lyset af ophavs- og patentretlige regler og principper mhp. at klarlægge de glidende overgange mellem rettigheders opståen og beskyttelse og benyttelse af data. Forskerforpligtelser ifm. indsamling af data er tæt reguleret, og i særdeleshed indenfor sundhedsvidenskab, men bør generelt kvalificeres yderligere ift. fx indsamling af meget store datamængder fra forskellige kilder (big data).
- **Lagring af aktive data:** Et i vidt omfang ureguleret felt, udover de regler, der følger af persondataloven og regler om databasehåndtering. Specielt må enhver form for information om identificerbare personer ikke opbevares længere tidsrum end det, der er nødvendigt for formålet, hvortil de behandles, og den dataansvarlige skal træffe fornødne tekniske og organisatoriske sikkerhedsforanstaltninger mod at de hændeligt eller ulovligt tilintetgøres, fortabes, forringes, kommer til uvedkommendes kendskab, misbruges eller i øvrigt behandles i strid med loven. Men juraen må ses i sammenhæng med teknologiudviklingen, og fx rejser cloud-infrastrukturer særlige problemstillinger.
- **Bearbejdning af andres data og generering af afledte data:** De juridiske problemstillinger er kun delvist afdækket i den ophavsretlige litteratur og behøver afklaring.

- **Deling/overdragelse af rettigheder til forskningsdata:** Aftaler er fremadrettet et helt afgørende instrument. Der må bygges videre på erfaringerne gjort i Techtrans-sammenhæng og ifm. EU-programmer.
- **Publicering af data:** Output skabt ifm. forskning og udgivet i Danmark eller med specifik dansk målgruppe er omfattet af pligtafleveringsloven. Materiale, som er tilgængeligt for offentligheden fra en dansk server enten gratis eller mod betaling, høstes af netarkivet, som kan anmode om at få udleveret nødvendige logins/passwords. Netarkivet giver efterfølgende adgang til det indhøstede iht. persondataloven og ophavsretsloven.
- **Videreanvendelse af data:** Ophavsberettigede forskningsdata falder ikke ind under direktiverne om videreanvendelse af offentlige informationer, og videreanvendelse kan derfor reguleres på aftalebasis, men hvor forskere stiller data til rådighed via OA, er der en grænseflade, som bør undersøges nærmere.
- **Samlinger af forskningsdata:** Betragtes som informationsaktiver, som institutionen er forpligtet på at behandle i overensstemmelse med ISO 27001-sikkerhedsstandard.
- **IT-systemer:** Indeholder de forskningsdata skabt af statsansatte forskere, skal de iflg. arkivloven anmeldes til Statens Arkiver, som afgør om data skal bevares for eftertiden, og som i givet fald overtager ansvaret for den fremtidige bevaring af data og for at stille data til rådighed efter reglerne i arkivloven.

(→ Bilag 07 for den fulde afdækningsrapport)

C. International best practice for forskningsdata management politikker

Introduktion

Mhp. at informere Styregruppens overvejelser ift. det strategiske element ”politikker for data management” besvarer afdækningen spørgsmålet: Hvad anses internationalt som best practice indenfor politikker for forskningsdata management – generelt og specifikt for hovedområderne HUM/SAM/NAT/TEK/SUND? Notatet, der er resultatet af afdækningen har potentiale til at kunne bruges som framework for at udvikle og evaluere politikker.

Metode

Kommissorie fremsendt til en shortlist af kompetente og erfarne udenlandske data management konsulenter. Kontrakt indgået med Simon Hodson, executive director, CODATA, der har forfattet 26 s notat excl. bilag, som giver overblik over pt. best practise indenfor forskningsdata management politikker internationalt, formuleret som de er af primært diverse nationale bevillingsgivere, men også forskningsinstitutioner og tidsskriftredaktioner i Nord-amerika, Europa og Australien. Notatet

- Oplister de drivere og generelle principper der optræder i forskningsdata management politikker.
- Diskuterer af de vigtigste elementer i de mest indflydelsesrige politikker.
- Fokuser på de væsentligste variationer og distinkte træk ved politikkerne indenfor hovedområderne HUM/SAM/NAT/TEK/SUND.

Hovedpunkter

De fleste institutionelle politikker skrives som respons på bevillingsgiveres ditto, der igen i særlig grad er influeret af OECD's Principles and Guidelines for Access to Research Data from Public Funding (2007). Der er senest kommet stort fokus på Open Data, princippet om 'open by default' og kriterierne for data genbrug (specielt konceptet 'intelligent openness'), og de nødvendige begrænsninger herfor.

Det er god praksis at en forskningsdata management politik indledes med overordnede betragtninger om:

- Drivers og principper for politikken – fx validering af forskningsresultater, øget forskningspotentiale, offentligt finansieret forskning som et offentligt gode, bredere samfundsmæssige og økonomiske fordele.
- Kravene til effektiv data deling – fx at data er ”genfindelige, tilgængelige, vurderbare, forståelige, brugbare og om muligt interoperable ihht. givne kvalitetsstandarder”.
- De nødvendige begrænsninger for 'openess' – specielt kravet om at beskytte personhenførbare informationer og til at udvise kommercielle og sikkerhedsmæssige hensyn.

Nøgleelementer i en politik/specifikke elementer i implementeringen af generelle principper omfatter:

- En definition af forskningsdata
- En klar defineret af, hvilke data er omfattet af politikken – alle data produceret af forskningsprojekter skal manages, men politikken skal være klar omkring:

- Hvilke forskningsdata, der er centrale for politikken og som er vigtigst at tilgængeliggøre.
- Hvilke data der ikke nødvendigvis er underlagt 'open data'/datadelingsperspektivet i politikken.
- Der skal altså være en differentiering af data ud fra et eller flere definerede kriterier, fx nødvendige for validering, sensitivitet, unikhed etc., og - hvor det er relevant - evt. suppleret med en angivelse af tidsrammerne indenfor hvilke de forskellige typer af data forventes tilgængeliggjort.
- Data falder generelt indenfor to overordnede kategorier:
 - Data der underbygger publicerede forskningsresultater, dvs. data nødvendige for validering.
 - Data der er skabt af forskningsprojektet, men som ikke direkte underbygger publicerede forskningsresultater.
- Software, kode, algoritmer, protokoller, værktøjer etc. nødvendige for at generere/replikere data og analyseresultater anses i stigende grad som væsentligt at tilgængeliggøre, og dermed omfattet af en politik.
- En oversigt over ansvarsområder fordelt på stakeholders – bevillingsgiver, forsker, institution, data center/service ift. at sikre langtidsbevaringen, tilgængeligheden og genbrugen af data.
- En indikation af tilgængelig infrastruktur og ansvaret for omkostningerne – dette gælder for forskningsdata management i et projekts levetid ('in-project costs'), men er specielt vigtigt ift. langtidsbevaringen af data. Hvor internationale eller nationale infrastrukturer ikke kan anvendes, skal politikken indikere, hvor ansvaret og omkostningerne ligger.
- En stipulering af data management plan krav og procedurer – hvilke problemstillinger bør en data management plan adressere, hvordan bliver den vurderet, hvem har ansvaret for at kvalitetssikre den og tilse at den eksekveres. For de tilfælde, hvor ansvaret for langtidsbevaringen tilfalder institutionen, kan politikken indikere, hvor lang tid institutionen forventer at opbevare data.
- Krav/anbefalinger ift. at muliggøre opdagelse og genbrug – fx krav om at publiceret forskning anfører, hvordan underliggende data kan tilgås, og anbefalinger vedr. brugen af passende metadatering, permanente identifikatorer (PID) og licenser der muliggør genbrug.
- Bestemmelser, som tilskynder til anerkendelse og belønning af forskere, der gør data frit tilgængelige – krav om at genbrugt data krediteres og eksplicit citeres, hvor det underbygger til nye forskningsresultater, evt. suppleret med en bestemmelse om forskerens eneret til data /'privileged access' i en given periode.
- En oversigt over afrapporteringskrav, monitorering og evt. sanktioner – politikken bør indikere hvordan overholdelse af den monitoreres, hvilken afrapportering/opfølgning der kræves og hvilke sanktioner der kan blive pålagt.

Det er god praksis, at politikken linker til information og vejledninger, der understøtter implementeringen af den.

Der kan ikke drages stærke konklusioner vedr. fagspecifikke variationer i forskningsdata management politikker:

- Variationerne i bevillingsgiverpolitikker er tilfældige og proceduremæssige snarere end udtryk for særlige iboende udfordringer ift. datadeling indenfor et givent forskningsområde.
- Politikkerne præsenterer brede højniveau principper. De bevillingsgivere, der fremhæver disciplinkarakteristiske problemstillinger, adresserer dem oftest kun i supplerende vejledninger – ikke i selve politikken.
- De mest signifikante variationer relaterer sig til:
 - Lovmæssige og etiske krav, der er specifikke for den type forskning der bedrives, fx beskyttelse af persondata indenfor SUND og SAM.
 - Eksistensen af mere etableret data management infrastruktur og praksis indenfor nogle forskningsområder, fx kan appendikser til politikker have "white lists" over infrastruktur.
 - Accepterede tekniske tilgange indenfor et specifikt forskningsområde, fx de specifikke dataformater eller metadatastandarder, der er de facto standard indenfor nogle miljøer under SAM/NAT/TEK kan findes nævnt i nogle politikker eller supplerende vejledninger.

(→ Bilag 08 for den fulde afdækningsrapport)

D. Udenlandsk inspiration vedr. Incitamenter

Introduktion

Mhp. at informere Styregruppens overvejelser ift. det strategiske element "incitamentsskabelse" besvarer afdækningen spørgsmålet: Hvilken inspiration kan hentes udenlands fsva. incitamenter for forskningsdata management?

Metode

Afdækningen tager afsæt i et case study kommissioneret af Knowledge Exchange (KE), "Incentives for sharing research data seen from researcher perspective". Undersøgelsen er gennemført som kvalitative interviews af fem forskningsteams med en etableret praksis for deling af forskningsdata – et fra hvert af KE partnerlandene (UK, DK, Holland, Tyskland og Finland) og fordelt på humaniora, samfundsvidenskab, biomedicin og kemi samt biologi. Undersøgelseslederne præsenterede de foreløbige resultater for Styregruppen primo juni.

Hovedpunkter

- Deling af data er ikke bare deling af data - forskere opererer med forskellige niveauer/forståelser: Deling af data med kolleger i forskningssamarbejder, deling af data via publikationer og frigivelse af rådata/fulde datasæt til ukendte brugere, dvs. til fri afbenyttelse af enhver.
- Incitamenterne for forskere kan opdeles i kategorier:
 - 1. Direkte forskningsmæssige fordele for forskerne**
 - Deling er essentielt for forskning generelt, men ikke alle dataproducenter også data-konsumenter, dvs. der er ubalance mellem udbud og efterspørgsel, og motivationen kan mangle for nogle. Andre strategier er derfor stadig er nødvendige (politikker mv.).
 - Deling har pt. negativ indflydelse på karrieren, idet impact-målinger kun har fokus på artikler. Nødvendigt med alternativer; metrikker for datacitation og social media impact.
 - 2. Minimering af omkostninger og risici**
 - De reelle omkostninger og ekspertise i at producere delbare data skal anerkendes.
 - Tid/omkostninger ved at producere data skal minimeres, fx one-stop entry af metada.
 - Der skal udvikles skabeloner til fx samtykkeerklæringer og data management planer.
 - 3. Normer der støtter op om datadeling**
 - Social media har vænnet unge forskere til at dele information og dermed data, men de replikerer praksissen i forskningsmiljøet; dvs. deling af data er nødt til at være default indenfor forskningsområdet, forskningsgruppen og institutionen.
 - Progressive forskere skubber grænserne ("guerilla data sharing"), men mange frygter for kritik og for at blive overhalet og miste konkurrencemæssige fordele. Tillid er essentiel, og klare aftaler kan derfor være nødvendige.
 - 4. Bevillingsstrategier og politikker**
 - Bevillingsgivere bør støtte datadelingsprojekter, fx databaseudvikling og sekundær analyse programmer mm., hvor delte data er output, forskningsressource eller værktøj.
 - Udgiver- og bevillingsgiverpolitikker/forventninger ændrer langsomt holdninger, praksisser og normer. Regler skal der til.

- Lederskab fra bevillingsgivere, institutioner, samfund og udgivere nødvendig for at skabe de rette rammebetingelser, for at datadeling bliver normen. Rammebetingelser der giver tillid: Data skal citeres, der skal være klare forventninger/aftaler og arbejdet i at dele data skal anerkendes. Incitamentsstrukturerne skal endvidere tage højde for datas livscyklus og forskerens karrierestadier.

(→ Bilag 09 for den fulde afdækningsrapport)

E. Danske faglige miljøers behov og præferencer

Introduktion

Mhp. at informere Styregruppens overvejelser ift. de strategiske elementer "Infrastrukturer" og "Kompetenceudvikling og forskerstøtte" besvarer afdækningen spørgsmålet: Hvad er den eksisterende forskningsdata management relaterede praksis, og hvilke behov for systemer/infrastrukturer og støttfunktioner (fx rådgivning af juridisk, teknisk, informationsorganisatorisk art og kompetenceudvikling mv.) har man indenfor hovedområderne HUM/SAM/SUND/NAT/TEK. Udgangspunktet er en faseopdelt model for forskningsdatas livscyklus.

Metode

Interviews gennemført parallelt på hovedområderne, med 5-9 forskere per område og på tværs af universiteter. Respondenterne repræsenterede både større forskningsgrupper og solo forskningsindsatser. Grundet det begrænsede datagrundlag, må der tages forbehold for repræsentativiteten og dermed entydigheden af konklusionerne.

Hovedpunkter – infrastruktur:

Praksis og organiseringsgrad ift. lagring af aktive data varierer meget hovedområderne imellem. Brug af lokal HD, eksterne lagringsmedier og 3. partssystemer (Dropbox, Amazon etc.) er udbredt. Der efterspørges generelt plads, back up, versionering og fleksibilitet ift. adgang (interface/metoder og devices) og ift. at dele data med andre sikkert. Sikkerhedsaspektet er facetteret; for nogle handler det om adgangskontrol, for andre handler det om, at data sikres til en standard accepteret af evt. samarbejdspartner, og for andre igen er flytning af personfølsomme data ifm. databehandling/HPC udfordringen. Holdningen til, hvorvidt behovene for lagerløsninger til aktive data bedst løses lokalt/nationalt/internationalt varierer hovedområderne imellem og i nogle tilfælde også på det enkelte hovedområde. Nogle forskergrupper ønsker skræddersyede løsninger; andre synes at kunne stilles tilfreds med noget mere generisk. Tilsvarende er både praksis og behov ift. metadata meget forskellig. "One size fits all" er mao. ikke mulig. Dog synes der generelt at være behov for en Dropbox-lignende infrastruktur.

Praksis ift. intensiv databehandling og dataanalyse varierer, men på især SAM, NAT og TEK er der ønske om adgang til (flere) HPC ressourcer til databehandling, og for SAM og TEK tilsvarende ressourcer til dataanalyse.

Ift. deponering af data varierer praksis hovedområderne imellem og i nogle tilfælde også på det enkelte hovedområde. Decentral opbevaring og deponering i de infrastrukturer, hvor data er i forvejen (fælles/eget netværks-drev, egen PC eller ekstern HD) er udbredt, og data er derfor typisk ikke citérbare. Validering og genbrug kan være vanskeligt uden at software compilere, specifikke operativsystemer etc. er gemt samlet (HUM og TEK). Der udtrykkes således interesse for infrastrukturer til deponering af data, hvor også komplette, kørende systemer (image af virtuel maskine) kan deponeres. SAM udtrykker behov for infrastruktur til deponering og deling af data, hvor også studerende har adgang. Hvorvidt infrastrukturer til deponering helst ses som institutionelle eller nationale er det svært at udlede noget om. Konklusionen er, at der på tværs af hovedområder er et behov for infrastrukturer til deponering, men på hvilket niveau er ikke entydigt.

Fsva. publicering og citering af data er der pt. generelt ikke den store tradition eller krav herom, undtagen NAT hvor der indenfor visse fagområder er regler for publicering af data ifm. publikationer. Krav om tilgængelighed af data mødes i nogen grad ved – på baggrund af personlig henvendelse – at formidle data via e-mail. HUM og TEK peger på et behov for at kunne citere data, og dermed infrastrukturer til at tildele persistente identifikatorer.

Også ift. langtidsbevaring og deling af data varierer praksis hovedområderne imellem. Generelt er der ringe eller ingen tradition for langtidsbevaring. Kun SUND og NAT gør det i nogen grad; hos førstnævnte i Statens Arkiver mens det hos sidstnævnte varetages af nogle af forskningsgrupperne - i egne systemer eller i internationalt organiserede infrastrukturer. Fsva. deling er dette lidt mere udbredt; SAM gør det gennem dataportaler, TEK gennem open source fora og websites og på case-by-case basis. Sidstnævnte praksis er udbredt både på de fleste hovedområder, og her bruges bl.a. mail, fildelingstjenester og Dropbox. Hovedparten af hovedområderne udtrykker behov for infrastrukturer til langtidsbevaring; SUND, NAT og TEK mener infrastrukturer bør være lokale (alt. at ansvaret ligger hos den instans, der stiller krav).

Praksis omkring opdagelse af data er relativt ens på tværs af hovedområder; litteraturen er den mest udbredte vej til at opdage data – dernæst det personlige netværk. NAT og SUND anvender derforuden hhv. dataportaler og databaser. Det er svært at udlede meget fsva. behov for infrastrukturer til opdagelse af data; HUM ønsker tværfaglig formidling, og SAM ser et behov for bedre formidling af datas kontekst; specifikt links mellem data og projektbeskrivelser, projektansøgninger, publikationer eller andre formidlede præsentationer af data. De øvrige hovedområder ser ingen behov for infrastrukturer til opdagelse af data.

Hovedpunkter – støttefunktioner:

Der er bred interesse og behov på tværs af hovedområderne for kurser og især rådgivning ift. data management delen af planlægnings- og ansøgningsfasen eller dataindsamlingsfasen, dvs. i starten af forskningsdatas livscyklus. Behovet er så udtalt, at der for to af hovedområderne er villighed til at betale herfor af forskningsmidler. Behovene er næsten enslydende knyttet til enten "data management planning" (DMP) eller juridiske aspekter omkring især persondata. Skræddersyet support/projektspecifik hjælp nævnes også som et behov. Hvorvidt kurserne, rådgivningen eller supporten bedst løses på universitetsniveau eller nationalt niveau er det svært at konkludere noget om. På SUND og NAT synes nogle at have behov for kurser indenfor dataanalyse, hvilket muligvis kunne være relevant at udbyde på nationalt niveau.

(→ Bilag 10 for den fulde afdækningsrapport)

7. Udvalgte referencer

A. Web dokumenter

The Danish Code of Conduct for Research Integrity

Arbejdsgruppen vedr. en fælles national kodeks for integritet i forskning
Uddannelses- og Forskningsministeriet. Styrelsen for Forskning og Innovation
København, 5. november 2014.
<http://ufm.dk/publikationer/2014/the-danish-code-of-conduct-for-research-integrity>

Danmarks Nationale Strategi for Open Access

Uddannelses- og Forskningsministeriet. Styrelsen for Forskning og Innovation
København, 23. juni 2014.
<http://ufm.dk/forskning-og-innovation/samspil-mellem-viden-og-innovation/open-science/danmarks-nationale-strategi-for-open-access.pdf>

FIF - Fælles Infrastruktur for Forskningsdata - Rapport fra fase I

Det Kongelige Bibliotek, Statsbiblioteket, Roskilde Universitetsbibliotek, DTU Bibliotek og DeiC.
København, DEFF, 11. juli 2014
http://www.deff.dk/fileadmin/user_upload/dokumenter/DEFF/aktuelt/Afrapporteringer/FIF_afsluttende_rapport_20140718.pdf

How to Develop Research Data Management Services – a guide for HEIs

Sarah Jones, Graham Pryor and Angus Whyte
Edinburgh: Digital Curation Centre, March 2013
<http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services>

How to Discover Requirements for Research Data Management Services

Angus Whyte and Suzie Allard
Edinburgh: Digital Curation Centre, March 2014
<http://www.dcc.ac.uk/how-discover-requirements>

B. Bøger

Delivering Research Data Management Services - Fundamentals of good practice

Graham Pryor, Sarah Jones and Angus Whyte, editors
London: Facet Publishing, December 2013
ISBN: 978-1-85604-933-7
256 pages
<http://www.amazon.co.uk/Delivering-Research-Data-Management-Services/dp/1856049337>

Research Data Management: Practical Strategies for Information Professionals

Joyce M Ray (Editor)
West Lafayette, Indiana: Purdue University Press, January 2014
ISBN: 978-1-557536648
436 pages
<http://www.amazon.co.uk/Research-Data-Management-Information-Professionals/dp/1557536648>

Managing and Sharing Research Data - A Guide to Good Practice

Louise Corti, Veerle Van den Eynden, Libby Bishop and Matthew Woollard
London: SAGE Publications, March 2014
ISBN: 978-1-44626-726-4
240 pages
<http://www.amazon.co.uk/Managing-Sharing-Research-Data-Practice/dp/1446267261>

8. Bilagsliste

Strategien samt alle bilag er tilgængelige på deic.dk/datamanagement.

Bilag 01. Styregruppens medlemsliste med kontaktinformation

Bilag 02. Styregruppens kommissorium

Bilag 03. DeiC og DEFF oplæg vedrørende national strategi og proces for datamanagement

Bilag 04. Diskussionsoplæg vedrørende strategiske indsatsområder

Bilag 05. Forslag til afdækninger og videre forløb

Bilag 06. Afdækning A: Bevillingsgiveres krav

Bilag 07. Afdækning B: Juridiske rammer og udfordringer

Bilag 08. Afdækning C: Udenlandsk inspiration vedr. politikker

Bilag 09. Afdækning D: Udenlandsk inspiration vedr. incitamenter

Bilag 10. Afdækning E: Faglige miljøers behov og præferencer

Bilag 11. Høringssvar