

DeiC's og DEFF's oplæg til

National strategi og proces for data management

1. Baggrund – Forskningsdata er et aktiv

Forskningsdata er et vigtigt aktiv for såvel universiteterne som for samfundet, samtidigt med at mængden af digitale forskningsdata er stor og hurtigt voksende. Som situationen er nu, lever mange forskningsdata en isoleret og ikke fremtidssikret tilværelse privat hos forskeren, hos IT afdelinger eller på kommercielle skytjenester. Forskningsdata er sjældent entydigt linket til forskere, projekter og publikationer og er derfor ofte svære at dele, genfinde for verifikationsformål eller genbruge til nye forskningsformål. Der er bred enighed om, at værdien af forskningsdata øges signifikant, når de er "forvaltede, forbundne, søgbare og genbrugbare" og derved muliggør nye typer af forskning med udgangspunkt i genbrug af datasæt – det såkaldte "4. forskningsparadigme". Bedre data management vil medføre et markant løft til dansk forskning i den offentlige og private sektor og til dermed nationens konkurrenceevne.

Internationalt driver udviklingen en stigende faglig bevidsthed om normer for god videnskabelig praksis, som udmøntes i politikker og krav fra institutioner og bevillingsorganer, f.eks. i EU's udkast til Horizon 2020 programⁱ og diverse Open Access politikkerⁱⁱ.

Denne udvikling indebærer fælles udfordringer af økonomisk, teknisk, juridisk og politisk art for institutionerne, der med fordel kan håndteres på nationalt niveau og i tæt samarbejde. Gennem specialisering, arbejdsdeling og samarbejde opnås effektiv ressourceudnyttelse, stordriftsfordele og mere fokuseret indsats med større effekt. Bestyrelsen for DeiCⁱⁱⁱ er i aftalen mellem universiteterne og Forskningsstyrelsen blevet pålagt at varetage udviklingsaktiviteter, der har national og tværinstitutionel karakter, herunder indenfor området forskningsdata^{iv}, og styregruppen for DEFF^v vurderer ligeledes, at området er helt centralt.

DeiC og DEFF har på denne baggrund etableret et samarbejde, og en interessentgruppe med repræsentation fra universiteterne, bibliotekerne og arkiverne har opstillet en række principper for en data management strategi og peget på centrale indsatsområder. Der er behov for de danske rektorers opbakning til dette oplæg til strategi og proces, såfremt udfordringerne skal kunne imødegås med tilfredsstillende økonomisk, politisk og organisatorisk beslutningskompetence.

2. Mission og vision

DeiC's og DEFF's interessentgruppe foreslår følgende mission og vision for strategien:

Mission:

*Sikre at værdien af danske forskningsdata bevares
og realiseres nu og for fremtiden*

Vision:

*Bedre og mere konkurrencedygtig forskning gennem effektiv opsamling, sikring, formidling
og genbrug af relevante danske forskningsdata*

3. Genstandsfelt – Forskningsdata og data management

Forsknings- og revisionsmæssigt er forskningsdata aktiver, som institutionerne er stærkt interesserede i at registrere og håndtere forsvarligt. Der er vigtigt for at undgå tab af et fremtidigt forsknings- og innovations-potentiale af stor betydning.

Forskningsdata:

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

Data management:

Aktiviteter, der sikrer at forskningsdata nemt og sikkert kan tilgås, deles, arkiveres, genfindes og forstås på kort såvel som langt sigt

4. Principper for DeiC's og DEFF's interessentgruppes strategiske overvejelser om data management

De elektroniske forskningsdata, der søges organiseret og bevaret er ikke kun omfattende i mængde, men er også mangeartede i diverse stadier i forskningsprocessen. En national data management strategi må afgrænse og strukturere genstandsfeltet som del af at pege på behov for tiltag. Der foreslås derfor følgende styrende principper som ramme for videre tiltag:

- **Datas hele livscyklus**

Jo tidligere data management tænkes ind i forskningsprocessen, desto større effektivitet og desto mere værdi opnås. Indsatsen bør derfor adressere hele datas livscyklus fra planlægning, indsamling, analyse og publicering til langtidsbevaring og genbrug^{vi}.

- **Incitamenter til data management**

Det er en afgørende forudsætning, at forskerne og institutionerne kan se en fordel i data management og de infrastrukturer, der stilles til rådighed samt de politikker, der formuleres. Incitamenter kan f.eks. være mulighed for verifikation af forskningsresultater, at forskningsdata tæller med ved beregning af den bibliometriske indikator, at vigtige datasæt kan genbruges i fremtiden, og særligt når der er evidens for, at det fører til øget citation^{vii}.

- **Et bredt spektrum af forskningsdata**

Forskningsdata kommer i mange formater, grader af struktur og varierende kompleksitet. Forskningsdata bør ikke defineres snævert. I tillæg til gældende lovgivning er der behov for nationale og institutionelle politikker for, hvilke data der skal bevares på kort, mellemlangt og langt sigt.

- **Nationale løsninger og tiltag, hvor det giver mening**

Ressourcerne, der er til rådighed på området, skal anvendes så hensigtsmæssigt som muligt, så der ikke opbygges parallelle infrastrukturer. En national data management strategi bør derfor klarlægge, hvilket niveau de identificerede behov skal løses på, af hvilke aktører og eventuelle snitflader til internationale strukturer.

- **Flere infrastrukturløsninger**

Hvert trin i datas livscyklus har sine udfordringer og potentialer. Tilsvarende er det med praksis og data management indenfor de forskellige fagområder. Derfor bør der tænkes i en flerhed af løsninger.

- **Samarbejde på tværs af organisatoriske skel**

IT-afdelinger, universitetsbiblioteker og andre aktører på området, såsom arkiver og biblioteker, har hver for sig unikke kompetencer på hver sine felter indenfor data management. Derfor bør der sigtes på samarbejde på tværs af organisatoriske skel.

- **Interoperabilitet og standarder**

De bestående og kommende løsninger vil have forskellige brugs- og systemgrænseflader. En flerhed af løsninger nødvendiggør således et absolut fokus på grænsefladerne mellem systemerne og på eventuelle standarder på området for at opnå størst mulig interoperabilitet. Open source løsninger bør så vidt muligt anvendes for at undgå leverandørfhængighed samt fremme fleksibilitet i interessenternes udviklings-samarbejde.

- **Prototyper og pilotforsøg**

Nationale data management services bør etableres i 3 trin: Prototyper, pilotforsøg og egentlig drift for at kompensere for de mange usikkerheder og uafklarede problemstillinger, der stadig er på området. Udviklingstiltag og prototyper bør nemmere og hurtigere kunne kasseres til fordel for de, der videreføres.

- **Udnyttelse af internationale erfaringer og samarbejdsmuligheder**

I en række andre lande er man langt med strategi- og løsningsopbygning^{viii}. En dansk strategi bør indtænke alle muligheder for at trække på disse erfaringer og evt. samarbejde – igen med henblik på den mest effektive ressourceudnyttelse.

5. Indstilling til Danske Universiteter – Struktur og handlingsplan

DeiC har pt. budgetafsat DKK 10 millioner til implementering af strategien. DEFF råder over ca. DKK 10 millioner årligt til udviklingsprojekter. DEFF har data management som indsatsområde og ser for sig, at dele af disse midler kan anvendes til data management. 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.

Det indstilles,

- A)** at der udarbejdes strategi, handlingsplaner og økonomiske estimater for følgende strategiske indsatsområder:
1. Politikker for data management på områder, hvor det giver mening at etablere dem på hhv. lokalt, nationalt og internationalt niveau i samarbejde med universiteterne, Styrelsen for Forskning og Innovation, Forskningsrådene samt arkiver og biblioteker.
 2. Incitamentsskabelse, således at forskerne, universiteter, forskningsråd mv. motiveres til at efterleve god data management skik gennem forskellige virkemidler.
 3. Data management infrastrukturer for alle stadier af forskningsdatas livcyklus. Infrastrukturen skal sikres optimal synergieffekt med andre initiativer, såsom det planlagte European Spallation Source - Data Management and Software Centre (ESS-DMSC).
 4. Kompetenceudvikling og forskerstøtte herunder identificering af kompetencer, som forskere må besidde for at forvalte forskningsdata.
 5. Fremtidig styringsstruktur for data management området i Danmark.
 6. Forslag til evt. pilotprojekter, som kan fremme fremdriften i udvikling og implementering den nationale strategi.
- B)** DeiC nedsætter en "Styregruppen for National Data Management" bestående af 11 personer. Rektorkollegiet (RK) udpeger 8; Statsbiblioteket (SB) udpeger 1; Det Kongelige Bibliotek (KB) udpeger 1; Statens Arkiver (SA) udpeger 1; Styregruppen får ansvaret for at ovennævnte strategi, handlingsplaner og økonomiske estimater udarbejdes.
- C)** Udkast til strategi, handlingsplaner med økonomiske estimater forelægges RK, SB, KB og SA senest 1. april 2014 til orientering og eventuel kommentering med henblik på eventuel justering før endelig vedtagelse af DeiC og DEFF med sigte på opfølgning.
- D)** Ovennævnte arbejde finansieres af DeiC og DEFF. DeiC varetager sekretariatsfunktionen.

ⁱ http://www.consilium.europa.eu/uedocs/cms_Data/docs/pressdata/en/intm/138118.pdf

ⁱⁱ <http://fivu.dk/publikationer/2011/anbefalinger-til-implementering-af-open-access-i-danmark-afsluttende-rapport-fra-open-access-udvalget>

ⁱⁱⁱ Danish e-Infrastructure Cooperation, under Ministeriet for Forskning, Innovation og Videregående Uddannelser

^{iv} Aftale med universiteterne om reorganisering og styrkelse af dansk e-Science af 30. september 2011:

http://www.deic.dk/sites/default/files/uploads/Aftale_af_30_september_2011_om_reorganisering_og_styrkelse_af_dansk_e-Science_DOK2044078%20%282%29.pdf

^v Danmarks Elektroniske Fag- og Forskningsbibliotek, er et samarbejde mellem Kulturministeriet, Ministeriet for Forskning, Innovation og Videregående Uddannelser samt Ministeriet for Børn og Undervisning

^{vi} <http://www.knowledge-exchange.info/surfboard>

^{vii} <http://arxiv.org/abs/1111.3618>

^{viii} Se eksempelvis rapporten "European Landscape Study of Research Data Management" (SIM4RDM Support infrastructure models for research data management, udgivet maj 2013).