

DATA MANAGEMENT IN PRACTICE

En fortælling om et projekt og en udflugt i the Matrix

Målet var fra starten praktisk: at give danske forskere adgang til tjenester, der kunne understøtte alle dele af forskningsdatas livscyklus. Det vil sige etablering af en infrastruktur, der kunne spænde fra data management planlægning til data bevaring og publicering over praktisk data håndtering og analyse. Ud over dette i sig selv rigeligt ambitiøse mål skulle der også udvikles værktøjer til træning og uddannelse i data management og endelig indgik en undersøgelse af forskningsbibliotekernes aktuelle og fremtidige roller i data management også som delmål.

Hvorfor "in practice"? Fordi udviklingen af tjenesterne skulle ske i samarbejde med forskere og som integreret del af igangværende forskningsprojekter på flere fagområder. Dette for at sikre den praktiske anvendelighed og for at understøtte læring som del af udviklingsarbejdet.

Men hvordan skulle det organiseres? En mulighed ville være en ren projektorganisation med hver institution arbejdende på egne projekter, en anden ville være temabaseret, hvor hver institution arbejder inden for et eller flere temaer, f.eks. databevaring, dataanalyse osv. I stedet valgtes en hybrid mellem disse to, en matrix-organisation, som fik stor betydning for projektets resultater.

"Neo, sooner or later you're going to realize just as I did that there's a difference between knowing the path and walking the path". (Morpheus, The Matrix, filmen fra 1999)

-og forskellen mellem at kende vejen og vandre vejen skulle netop matrix-organisationen gerne kunne opløse.

Form og indhold

Da projektet sigtede mod udvikling af brugbare forskertjenester, måtte de praktisk være forbundet til konkrete forskningsprojekter. Samtidigt skulle hver af projekterne spænde over hele forskningsdatacyklussen.

Derfor kom organisationen til at se sådan ud:

	HUM cases (SB og KB)	SAMF / SUND cases (DDA, SDU, RUC)	NAT cases (KB)	TEK cases (DTU)	Projektstyring & Koordinering
RUC	Data Management Planning				Internt i projektet samt
KB	Data capture, storage and documentation (basic metadata)				
DTIC	Data identification, citation and discovery				DEFF / DelC: Nationalt Forum for Forskningsdata management
SB	Select and deposit for long term preservation				
AUB	Training and marketing toolkits				
SUB	Sustainability				

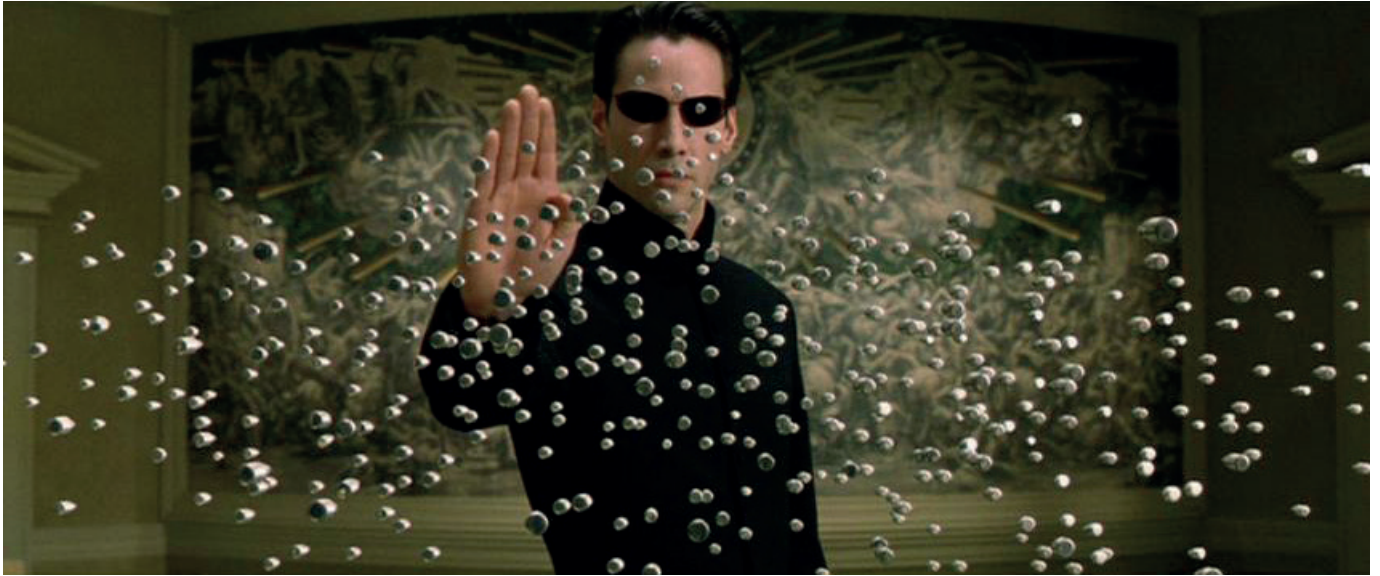
De vandrette striber er temaer, baseret på livscyklussen for forskningsdata. Temaerne dækker processen fra planlægning af et projekt, over den aktive forskningsproces til arkivering. Herudover dækker to af temaerne henholdsvis uddannelse og træning og hvordan den enkelte tjeneste kan sikres overlevelse. Temaerne sikrer en sammenhæng og koordinering på tværs af projekterne og giver mulighed for at homogenisere resultaterne. De lodrette søjler er casene - forskningsprojekterne - inden for hovedfagområderne. Her ligger integrationen i igangværende forskningsprojekter og samarbejdet med forskerne.

Hvilke projekter?

Forskningsprojekterne - casene - spændte bredt, både over de traditionelle fagområder og inden for hver af disse. Fælles var, at hver af dem skulle gennemløbe faserne i forskningsdatas livscyklus, beskrevet i temaerne. Den indholdsmæssige bredde i forskningsprojekternes kan ses ud af de anvendte data:

HUM: Radio- og TV-data fra LARM (LydArkiv for RadioMedier: <https://larm.sites.ku.dk/>), fra Netarkivet (<http://netarkivet.dk/>) og fra Søren Kierkegaards Skrifter (<http://www.sks.dk/forside/indhold.asp>)

SAM/SUND: data fra kommunalvalg 2013 fra SDU og Rigsarkivet og CALPIU's (<http://calpiu.dk/ruc/Home.html>) audio- og videobaserede data om brug af sprog i universitets-sammenhænge, formelle og uformelle, undervisnings- og sociale sammenhænge mm. Endvidere sundhedsdata fra GALAXY (<http://>)



livergalaxy.eu/), der er et projekt om behandlingsmetoder for skrumpelever.

NAT: data om grønlandske vandrefalke til brug for monitorering af biodiversiteten i Arktis. Desuden astronomiske data om stjerner, stammende fra Kepler missionen.

TEK: globale meteorologiske data for vind og data for jordens magnetfelt.

Hvilke temaer og hvilke mål?

Data Management Planning skulle udvikle en dansk DMP Online, som en national installation, med et indhold matchende danske forhold.

Data capture, storage... havde fokus på at skabe forudsætningerne for at data kan genbruges. Det vil sige infrastruktur, herunder repositorer til aktive data, metadata, datadokumentation og redskaber til dataindsamling.

Data identification... handlede om muligheden for at identificere forskere ved hjælp af ORCID og datasæt ved hjælp af f.eks. DOI'er (som er en slags ISBN-nummer til andet end bøger, såsom datasæt, billeder, tekst osv.). Dette for at lette genfindning og citering af data.

Select and deposit... skulle sikre rutiner og workflow for udvælgelse af data til langtidsbevaring og herunder også bevaringsløsninger, der kunne sikre fremtidig adgang og mulighed for genbrug af data.

Opgaven for Training and marketing toolkits var udarbejdelse af træningsmateriale og kursusformater for forskere.

Sustainability var fælles mål for alle cases, idet det skulle undersøges hvordan de udviklede redskaber kunne leve videre efter projektperioden.

Blev målene for projekterne nået?

Ja, de allerfleste. De, der kun nåedes delvist, afdækkede til gengæld nye muligheder og betingelser.

På tværs af de forskellige forskningstraditioner i SAM - HUM - NAT - TEK - SUND og forskellig praksis på universitetsbibliotekerne, skabtes en både fælles og fleksibel DMP Online. Som et spin-off af samarbejdet udvikledes en juridisk modelaftale for datahåndtering.

Der viste sig ikke umiddelbart at være behov for en fælles infrastruktur i form af nationale repositorer, da flere forskellige allerede var i drift. Som en sidegevinst viste det sig, at PURE var velegnet til at håndtere metadata og informationer om data management, men ikke til at håndtere egentlige datasæt.

DataCite og de DOI'er, som den udsteder, viste sig at være velegnet til at gøre data søgbare ud fra standardiserede metadata. For at kunne koble forsker med data - og omvendt - vil det være vigtigt at integrere forsker-ORCID med DataCite. Dette er imidlertid ude af projektets hænder.

LOAR (Library Open Access Repository), der giver danske forskere mulighed for at arkivere data, dækket af Creative Commons licenser, blev udviklet af Det Kgl. Bibliotek, Aarhus, sammen med LOAR's mere restriktive søster LCAR (Library Controlled Access Repository), der er særligt målrettet personfølsomme data. Dette repository afventer dog beslutning om etablering. Det Kgl. Bibliotek, København, har tilsvarende udviklet KUBIS Dataverse Network til forskningsdata fra Københavns Universitet. Herudover opstilledes en række praktiske anbefalinger til bevaring og tilgængeliggørelse af forskningsdata.

Endelig udvikledes DataFlowToolkit (<https://dataflowtoolkit.dk/>), som er et online værktøj, der hjælper med tænke datamanagement ind i forskningsprocessen.

At vandre vejen i matrixen - resultater

Udviklingen af den fleksible DMPOnline, der kunne tilpasses alle forskningsområders behov, blev muliggjort af input fra arbejdet med de konkrete cases.

Et andet resultat af arbejdet med projekterne igennem temaerne var, at personfølsomme eller ophavsretligt beskyttede data udgjorde særskilte problemer. Det førte bl.a. til udvikling af et muligt fremtidigt repository med begrænset adgang og af en datamodelaftale.

De gennemgående temaer muliggjorde feedback fra andre cases og modvirkede isolation og silotænkning inden for den enkelte case (jf. Burton et al., 2015).

Matrixorganisationen sikrede tydeligvis effektive informationsstrømme, organisatorisk fleksibilitet og praktisk samarbejde på tværs af projekterne. En organisation af rene enkeltprojekter blot under en fælles hat, ville ikke nødvendigvis have givet disse resultater.

Ulemperne ved matrixorganisationen har på den ene side været et stort ressourceforbrug til kommunikation og koordinering. På den anden side har den medført tillid og åbenhed deltagerne imellem og herigennem bidraget til fælles vidensopbygning.

-og til Morpheus kan en afslutningsreplik være, at det kan være vanskeligt at skelne mellem at kende vejen og vandre den.

For yderligere oplysninger, kontakt

Filip Kruse, fkr@kb.dk

Jesper Boserup Thestrup, jbt@kb.dk

Bodil Ammitzbøl Jurik, baj@kb.dk

Referencer:

R. M. Burton, B. Obel og D.D. Håkansson: "How to get the Matrix Organization to Work", Journal of Organizational Design, 2015, 37-45

Data Management In Practice. Results and Evaluation. Supplementary Files. Jan. 2018. DEFF. Danmarks Elektroniske Fag- og Forskningsbibliotek. <http://projektbank.dk/data-management-i-praksis>

Mere info:

I projektet deltog: RUC - Roskilde Universitet, KB - Det Kongelige Bibliotek, nu Det Kgl. Bibliotek, DDA - Dansk Data Arkiv - Rigsarkivet, DTIC, DTU Bibliotek, SB - Statsbiblioteket, nu Det Kgl. Bibliotek, AUB - Aalborg Universitetsbibliotek og SUB, Syddansk Universitetsbibliotek. Projektet blev finansieret ligeligt af DEFF, Danmarks Elektroniske Forskningsbibliotek, og de deltagende institutioner. Projektperioden var fra marts 2015 til juni 2017. Endelig afrapportering i januar 2018.

DMP Online er et værktøj, som guider brugeren igennem udarbejdelsen af hendes projekts datamanagement plan. Det er udviklet af Digital Curation Center i UK. Den danske version er tilgængelig hos DeIC: <https://dmponline.deic.dk/>

Datamodelaftale: <http://www.au.dk/samarbejde/erhvervssamarbejde/samarbejde-med-forsker/modelaftale-for-samarbejde-om-forsknings-data/>

DataCite er en non-profitorganisation og infrastruktur, som udsteder DOI'er til forskningsdata. Drives i Danmark af DeIC.