

# Data Management in Practice

Supplementary files  
January 2018

ISBN: 978-87-7507-416-7

DOI: 10.7146/aul.244.175

### **Contributors**

Aalborg University Library: Karsten Kryger Hansen

Danish National Archives: Christina Guldeldt Madsen, Anne Sofie Fink, and Bodil Stenvig

DTU Library: Falco Jonas Hüser and Paula Maria Martinez Lavanchy

Roskilde University Library: Søren Møller, Stine Vejlebo, and Sacha Zürcher

The Royal Danish Library, Aarhus: Tony Brian Albers, Bjarne Andersen, Katrine Hofmann

Gasser, Knud Åge Hansen, Bolette Ammitzbøll Jurik, Filip Kruse, and Jesper Boserup Thestrup

The Royal Danish Library, Copenhagen: Anders Sparre Conrad and Michael Svendsen

University Library of Southern Denmark: Janni Brødbæk and Asger Væring Larsen

# Contents

This collection of supplementary files illustrates the work and findings of the project Data Management in Practice. The collection consists of the original application, a model agreement on cooperation between researchers regarding handling of research data, several data management plans and other files. Some texts are in Danish and other are in English. The files are listed in the order they are mentioned in the report.

<b>APPENDIX</b>	<b>PAGE</b>
1. Data Management in Practice (18 pages) .....	4
2. The contribution of the cases in relation to the themes of the project (4 pages) .....	22
3. Larm Case Data Management Plan (10 pages) .....	26
4. Tjekliste for Probing a Nation's Web Domain (11 pages) .....	36
5. Probes — The Cook Book (7 pages) .....	47
6. Det Kgl. Biblioteks spørgeramme i forbindelse med at udarbejde datamanagement planer (3 pages) .	54
7. Skabelon til aftalemodel DEIC projekt (14 pages) .....	57
8. DM plan: Kommunalvalgsundersøgelsen 2013 (2 pages) .....	71
9. DMP: CALPIU (4 pages) .....	73
10. Galaxy - Project management plan (19 pages) .....	77
11. Galaxy - Initial data management plan (9 pages) .....	96
12. DMP: Teknisk støtte til at operationalisere dele af CAFE's Terrestrial Circumpolar (4 pages) .....	105
13. Type 1: Data set part of a publication - European wind atlas (2 pages) .....	109
14. Type 2: Experimental data not part of a publication - Øresund experiment data bank (2 pages) .....	111
15. Type 3: Modelled data not part of a publication, wind resource map for the Wasa domain (2 pages).	113
16. DTU Space - Geomagnetic Ground Stations (3 pages) .....	115
17. KASOC Data Management Plan (6 pages) .....	118
18. Technical setup for DataFlowToolkit.dk (4 pages) .....	124

## 1. DATA MANAGEMENT IN PRACTICE

### Økonomisk/juridisk ansvarlig

2. Bjarne Andersen
3. områdedirektør, Statsbiblioteket, CVR: 10100682, P-nummer: 1003404662
4. bja@statsbiblioteket.dk
5. Victor Albecks Vej 1, 8000 Aarhus C
6. KB, DTIC (DTU Bibliotek), SUB, RUB, Statens Arkiver (DDA), AUB

### 7. Hvilke(t) indsatsområde(r) hører projektet under?

Indsatsområde nummer 1: Datamanagement af forskningsdata.

### 8. Overordnet beskrivelse af projektet. De vigtigste elementer i projektet beskrives, herunder metodevalg og tilknytning til det valgte indsatsområde: Hvordan løser projektet den problemstilling, som det valgte indsatsområde peger på?

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, og projektet kommer således bredt rundt om indsatsområdet.

Tjenesterne vil omfatte værktøjer til træning af forskere og udarbejdelse af data management planer, samt den praktiske håndtering, indsamling og bevaring af forskningsdata. Dermed skabes mulighed for at opsamle, dele, formidle, videreudnytte, videreudbygge, citere, kontrollere og bevare forskningsdata, hvorved forskere forbedrer synligheden, troværdigheden og værdien af deres forskning.

Projektet bidrager til at skabe bedre adgang til forskningsdata og -resultater - for andre forskere såvel som for erhvervslivet og offentligheden. Herved styrkes forskningen selv såvel som forskningsens omsætning til samfundsmæssig innovation og værditilvækst. Erfaringer fra Dansk Dataarkiv viser evidens for genbrug af forskningsdata, når de beskrives fyldestgørende: ca. 80 pct. af studierne i DDA's samling har været udleveret en eller flere gange, og de mest populære datasæt (valgundersøgelser) har været udleveret mellem 42-88 gange.

Tjenesterne vil blive etableret i samarbejde med udvalgte igangværende forskningsprojekter og – infrastrukturer, fordelt over flere forskningsmæssige hovedområder. Konkret vil udviklingen ske i et antal cases, som hver skal levere en løsning i forhold til en bestemt forskergruppe. Hver case vil blive uddelegeret til en eller flere projektpartnere, og hver case vil have en hovedansvarlig part. De enkelte cases vil alle skulle dække den fulde livscyklus, men der vil samtidig blive organiseret en koordinering af de forskellige cases i forhold til en række temaer, inspireret ud fra livscyklustrin. Dette vil sikre at opgaver der bedst kan løses på tværs af de forskellige cases, vil blive håndteret effektivt.

Overordnet set vil projektet bidrage til at danske forskningsresultater giver et større afkast. Projektet er i fuld overensstemmelse med DeIC og DEFFs udkast til National strategi for data management som er i høring pt. Derudover vil projektet sikre en tæt koordinering med det

nationale forum for forskningsdata management som anbefales etableret i strategien.

Fire spørgsmål rejses i projektkaldet for dette DEFF-indsatsområde:

- (1) **Hvordan kan FFU-bibliotekerne bidrage til etablering og drift af data management services for forskningsdata?** - I projektet vil et bredt udvalg af FFU-biblioteker arbejde praktisk og forskernært med etablering, aftestning, rådgivning af sådanne services, ligesom de vil samarbejde og videndele med relevante øvrige aktører på nationalt niveau.
- (2) **Hvilke problemområder kan identificeres i forbindelse med etablering og drift af forskningsdatabaser?** - Projektet adresserer problemområderne meget bredt, idet alle dele af forskningsdatas livscyklus inddrages i analyser og løsninger. Der henvises her til Bilag 1, der redegør for livscyklus og problemområder- fra data management planlægning til digital bevaring for eftertiden.
- (3) **..., og hvilke metoder og værktøjer kan anvendes til at løse de identificerede problemstillinger?** - I overensstemmelse med anbefalingerne i den nationale strategi for data management anvender projektet en tredobbelt metode: (1) Dels organiseres et antal fagspecifikke og forskernære pilotprojekter, der løser praktiske data management udfordringer i tæt samarbejde med de pågældende forskergrupper. (2) Dels bindes disse fagspecifikke cases sammen af seks tværgående indsatser, der på nøgleområder sikrer videndeling, synergi og fælles løsninger, hvor sådanne kan etableres på tværs af fagområder. (3) Endelig indgår såvel cases som tværgående temaer i det kommende nationale forum for forskningsdata management, hvor erfaringsudveksling og etablering af fælles løsninger faciliteres og koordineres på nationalt niveau.
- (4) **Hvad er effekten af de valgte metoder og værktøjer?** - Grundet det store forarbejde i form af DEFF- afklaringsprojekter og DEFF/DEIC-strategien, som denne projektansøgning udnytter og forlænger, forventes en betydelig og blivende effekt af indsatsen - jf. afsnit 14 nedenfor. Effekten består blandt andet i udviklede kompetencer på bibliotekerne, relevante services etableret, større afkast af forskningen gennem genbrug og bedre adgang til videnskabelige resultater. Overordnet vil projektet i høj grad bidrage til en bedre forskningspraksis og understøtte reproducerbarhed. Projektet afsluttes af en international ekspertevaluering, der nøje vil granske dets resultater og deres effekt.

Nedenstående diagram giver et overblik over projektets organisering med vertikale cases (praktiske data management løsninger og tjenester) og seks horisontale faglige temaer der sikrer sammenhæng og synergi - internt i projektet såvel som som nationalt, repræsenteret ved søjlen til højre, der viser sammenhængen med DEFF/DeIC's forventede nationale initiativer og samarbejdsfora på området.

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

Ud over projektets organisering som figuren illustrerer, vil der blive nedsat en styregruppe som får det overordnede ansvar for at sikre fremdrift og leverancer. Se i øvrigt bilag-4 omkring projektorganiseringen.

De implementerede løsninger vil tage udgangspunkt i såvel den nationale strategi som de evalueringsresultater og anbefalinger der er kommet ud af det nyligt afsluttede DEFF-projekt: *Fælles Infrastruktur for Forskningsdata (FIF)*. Løsningerne vil være af en kvalitet, så de efterfølgende vil kunne benyttes til forvaltning af forskningsdata af de pågældende forskerteams og af beslægtede forskningsprojekter. Løsningerne vil samtidig tjene som grundlag for fortsat evaluering af både systemer, teknologier og af værdien af de tjenester som forskningsbibliotekerne og andre aktører vil kunne tilbyde på området.

I de enkelte cases involveres forskerne og konkrete forskningsprojekter, hvorved deres behov og praktiske udfordringer tages med i den færdige løsning.

### HUM cases

- Statsbibliotekets cases kobler sig til igangværende forskningsprojekter inden for humaniora: LARM (på vej til at blive DigHumLab tema-2b) med forskning i audiovisuelle medier og DigHumLab tema-2a (Netlab) med forskning i internetmateriale. Der er i disse projekter allerede udtrykt behov for håndtering og bevaring af forskningsdata. Der vil være et særligt fokus på forholdet mellem kildedata (fx. kulturarv) og nye forskningsdata produceret i projekterne.
- Det Kongelige Biblioteks case omhandler tekstkorpora som forskningsdata. Casen vil dreje sig om hvordan specielt TEI-opmærkede tekster kan bruges til at strukturere forskningsrelateret information, et arbejde som startede i Clarin.dk og fortsætter i DigHumLab tema 1. Temaer i casen vil være håndtering og forvaltning af berigelse, annotationer og relationer i forhold til standardkorpora som Dansk Referencekorpus og litterære udgivelser som Søren Kierkegaards Skrifter.

**SAMF/SUND cases**

- Syddansk Universitetsbibliotek og DDA's ene case koncentrerer sig om forskningsdata, der kombinerer surveys, fysiske målinger, registerudtræk og evt. andet. Hele datas livscyklus vil være i fokus, og der vil desuden være fokus på kobling og citation mellem forfattere, publikationer og datasæt. Et forskningsprojekt hos Statens Institut for Folkesundhed, SDU vil danne fundament for casen.
- Syddansk Universitetsbibliotek og DDA's anden case vil fokusere på projekter inden for kommunalforskning. Her vil der med casen arbejdes med, hvordan opbygning af forskningsdokumentation i henhold til 'best practices' gennem hele forskningsprocessen både er med til styrke forskerens arbejde, datadeling mellem forskere, bevaring af forskningsdata, citeringsmuligheder mv.
- RUBs case forholder sig til kvalitative data. Det drejer sig om typisk samfundsvidenskabelig/humanistisk empiri i form af videomateriale og transskriberet interviewtekst. RUB har med interessetilkendegivelse fra forskningsprojektet CALPIU<sup>1</sup> specifik fokus på langtidsof bevaring og adgangsstyring.

**NAT cases**

- Det Kongelige Bibliotek vil adressere etableringen af proveniens data som en integreret del af forskningsprocessen og vil have fokus på et stort internationalt netværk omkring Kepler Data. Alle dele, primærdata, forarbejdede data, ansøgninger, anvendte programmer samt publikationer vil blive struktureret og annoteret i forhold til relevante internationale standarder.

**TEK cases**

- DTIC's cases ligger på det teknisk-naturvidenskabelige område. Der vil bl.a. være fokus på cases inden for astrofysik og geomagnetisme med problematikker omkring såvel data, der er indhentet af forskergruppen selv, som data, der opstår ved efterbehandling af data fra ESA ol. internationale eksperimenter og dataleverandører. Der vil også være fokus på problematikker omkring visualisering af datasæt og de bevaringsmæssige udfordringer der rejstes herved. Her tænkes især på eksponering og meritering, f.eks. ifm. citation af datasæt.

Med udgangspunkt i DCCs livscyklusmodel (se også bilag-1) arbejdes indenfor seks tværgående temaer.

---

<sup>1</sup> Research Centre for the Study for Cultural and Linguistic Practices in the International University

1. **Data management planning.** Der bygges på resultaterne af FIF-projektet og etableres adgang til et værktøj til udarbejdelse af data management planer. Der udarbejdes data management planer for hver case, og værktøjet evalueres på baggrund af de indhøstede erfaringer. Projektet vil tilpasse DMPonline i samarbejde med DCC således, at de skabeloner der tilbydes i den danske installation i videst muligt omfang passer til danske forhold og danske forskeres projekter ud fra den viden der opsamles i arbejdet med de enkelte cases.
2. **Data capture, storage and documentation.** Der skabes repositories for de aktive forskningsdata, der understøtter metadatastandarder relevante for projektets cases, mulighed for citation af datasæt og digitale objekter (f.eks. til formidling af data). Optimale data capture teknikker (f.eks. integration med KUs nye tilbud om central lagerservice) efterforskes og implementeres.
3. **Data identification, citation and discovery.** Forskerne forventes identificeret vha. ORCID-identifikatorer. Datasæt forventes identificeret ved tildeling af DataCite DOI'er og/eller DanPID'er. Metadata om datasæt høstes fra såvel lokale repositories som de nationale langtidsarkiver og gøres søgbare i den nationale forskningsdatabase og i internationale databaser. Krydslinking af forskere, forskningspublikationer, datasæt og relaterede digitale ressourcer (f.eks. visualiseringer) udforskes i paradigmatiskke eksempler for hver case.
4. **Select and deposit for long term preservation.** Der udarbejdes strategier og kriterier for udvælgelse og hensigtsmæssig pakning af materiale til langtidsbevaring. Workflows og integration til bevaringsløsninger baseret på Dansk Dataarkiv og softwareudviklingen fra Det Nationale Bitmagasin implementeres for de relevante fagområders cases. Globale løsninger kan gøre sig gældende på andre fagområder. Der etableres adgang til de arkiverede data på passende måde og via passende kanaler.
5. **Training and marketing toolkits.** Der udarbejdes informationsmateriale, vejledninger, træningsmateriale, kursusmateriale. På de deltagende bibliotekers universiteter gennemføres informationskampagner for at udvide anvendelsen af projektets løsninger. Der vil være særligt fokus på at udarbejde skabeloner for kurser i data management for PhD'er.
6. **Sustainability.** Der vil i alle cases blive arbejdet med at sikre at de tjenester, der etableres, lever videre efter projektets afslutning. Der vil på tværs af cases blive udviklet forretningsmodeller, der beskriver hvad en fortsat drift af de etablerede tjenester koster og forslag til frembringelse af den nødvendige finansiering.

Der nedsættes en temaleder til hvert tværgående tema. Opgaven er at uddrage det generelle vedr. temaet og påvirke de enkelte cases til at anvende fælles begreber og værktøjer inden for temaet, hvor det er muligt. Temalederne er angivet til venstre på ovenstående figur.

På tværs af de fire første temaer (1-4) vil der også blive fokuseret på, hvordan forskningsdata kan indgå i de institutionelle repositories (pt baseret på Pure). Herunder om metadata om forskningsdata, eller aktiviteter relateret hertil, med fordel kan blive registreret i Pure. Dette gøres på baggrund af Pures eksisterende rolle i.f.m. at registrere forskere, publikationer og aktiviteter, hvor det derfor vil være oplagt at se på, hvorledes forskningsdata kan kobles hertil.

Endvidere vil der blive undersøgt om Pure (eller andre af de enkelte institutioners systemer) kan være datagrundlag til andre eksterne systemer. Projektet vil analysere hvor og hvordan forskellige typer af metadata skabes således at det sikres, at information om forskningsdata kan identificeres og genbruges på tværs af relevante systemer nationalt og internationalt.



Det kunne eksempelvis være relevant at undersøge, om der kunne skabes kobling mellem Pure og DMPonline således at projekter kun skulle beskrives et sted.

Den blå søjle til højre på figuren illustrerer projektets interne koordinering samt koordineringen med Det Nationale Forum for Forskningsdatamanagement. Her koordineres også de tværgående temaer, og her arrangeres de workshops og andre tema-møder, der afvikles som en del af projektet hen over hele projektperioden. Hvis Det Nationale Forum for Forskningsdatamanagement mod forventning ikke etableres, vil projektet selv nedsætte en referencegruppe med deltagere fra alle relevante institutioner. Denne referencegruppe vil således sikre videndeling, erfaringsudveksling samt national koordinering og samarbejde - skulle implementering af DEFF/DeIC strategien vise sig vanskeligere end antaget.

**9. Er der indgået aftale med samarbejdspartnere, og ønskes der på sigt at indtage andre aktører i projektet?**

Der er indgået aftale med alle medansøgerne. Derudover er der indgået aftaler med nedsatte projektpartnere i de fire cases.

HUM Cases: DigHumLab har behandlet ansøgningen i sin temaledergruppe den 27/10 og der er opbakning til at bidrage til de humanistiske cases i dette projekt.

SAMF/SUND cases; Der er indgået aftale med Statens Institut for Folkesundhed, SDU og Forskningsprojektet CALPIU

NAT cases: Der er indgået aftale med Stellar Astrophysics Centre (SAC), ved Aarhus Universitet om at arbejde med en strukturering af deres data med henblik på at sikre fremtidig adgang.

TEK cases: Der er indgået aftale med DTU Space om et eller flere pilotprojekter. Der er dialog med to andre institutter om supplerende pilotprojekter, skulle der blive kapacitet til dette i projektet.

Derudover vil projektet koordinere og videndele med det forventede nationale Forum for forskningsdata management og det er muligt at udvide projektet med flere eksternt finansierede cases såfremt behovet opstår. Der er dialog med DeIC om mulig fremtidig drift af udvalgte services, hvilket fastlægges i løbet af projektet.

**10. Hvordan løser projektet en problemstilling, der berører hele FFU-bibliotekssektoren? Hvilken værdi skaber projektet for FFU-bibliotekerne og deres brugere?**

Forskningsdata er et central element i DEFF-strategien og er således en problemstilling som berører hele FFU-bibliotekssektoren. Projektet involverer for første gang en række FFU-biblioteker i praktiske løsninger af denne problemstilling på tværs af faglige skel og omfattende alle trin af forskningsdatas livscyklus. Den systematiske tilgang sikrer, dels at der indhøstes erfaringer af relevans for sektoren, dels at områder der egner sig til nationale løsninger på tværs af faglige skel, bliver identificeret og adresseret, sådan at der så vidt muligt skabes generelt brugbare løsninger.

Projektet skaber umiddelbar værdi for de forskere hvis cases der bliver taget hånd om og for de biblioteker der bliver i stand til at tilbyde tjenester på forskningsdataområdet. For både forskeres og bibliotekers vedkommende betyder dette et skub til videreudvikling af metoder, kompetencer og best practice. For brugerne/forskerne gælder desuden at der bliver adgang til data f.eks. gennem Den Danske Forskningsdatabase, hvilket åbner op for at nye målgrupper, herunder almindelige borgere, kan få adgang til og anvende dansk produceret forskningsdata.

Der sikres en bred formidling af projektets erfaringer til FFU-bibliotekerne og skabes nye

kompetencer, som bidrager til at give momentum og troværdighed til bibliotekernes arbejde med forskningsdata. For nationale tjenester som bliver identificeret og udviklet i projektet, gælder at disse vil kunne stilles til rådighed for alle FFU-biblioteker. For de enkelte cases gælder at disse vil kunne tjene som inspiration og eventuelt videreudvikling/genbrug for fremtidige projekter i sektoren. Projektet styrker desuden samarbejdet mellem fag-, forsknings- og uddannelsesbibliotekerne gennem den brede sammensætning af projektorganisationen og gennem koordineringen med den nationale strategiudvikling på området.

**11. Hvordan understøtter projektet implementeringen af DEFFs strategi 2012-2016 (download strategien fra <http://www.deff.dk/om-deff/formaal-og-strategi/>)?**

Projektet er koblet til DEFF strategiens pkt. 3.1.5 Adgang til nye former for viden afsnittet Forskningsmateriale, hvor det blandt andet nævnes, at fremtidig fri adgang til forskningsdata vil være af afgørende betydning for vækst og innovation i alle sektorer. Projektet tager fat på alle de tre initiativer som nævnes i afsnittet. 1) Forskningsdata skal beskrives på et metadataniveau, gøres forståelige og derved genfindelige. Dette sker i data management planning og data capture (se figuren). 2) DEFF skal deltage i nationale tiltag omkring eScience og forskningsdata og arbejde proaktivt for at sikre, at bibliotekernes kompetencer tænkes ind i nationale løsninger. Dette sikres ved, at projektet kobler sig på anbefalingerne i den nationale strategi, via koordinering med det nationale forum og i projektet som helhed. 3) Der skal laves undersøgelser, som fastlægger bibliotekernes optimale rolle. Blandt andet grundet "FIF" projektet er bibliotekerne nu nået dertil, at der i dette projekt kan arbejdes med datamangement i praksis. Der er således en meget klar sammenhæng mellem DEFF strategien, DEFFs mission og projektet.

**12. Har andre nationalt eller internationalt arbejdet med samme emne og i givet fald, hvad er sammenhængen med nærværende projekt?**

Projektet og dets cases skal fungere som pilotprojekter i stærk tilknytning til det Nationale Forum for Forskningsdatamangement, som blev foreslået i udkast til national strategi for forskningsdatamangement. Det nationale forum vil få en løbende evaluerende og koordinerende rolle ift. cases og aktiviteter, der gennemføres i dette projekt.

Projektet kobler sig stærkt til det igangværende DEFF-projekt omkring ORCID og forskningsdatabasen idet de enkelte cases vil anvende disse projekters resultater i det omfang det giver mening ift. den enkelte case.

Projektet tager udgangspunkt i etableret best practice i lande som USA (Stanford), England (Oxford, Edinburgh), Holland (DANS, 3TU.Datacentrum, SURF), Tyskland og Australien (ANDS). I den forbindelse udnyttes etablerede standarder og referencemodeller, herunder data livscyklusmodellen fra JISC som ligger til grund for projektets struktur. På nationalt plan bygges videre på erfaringerne fra tidligere DEFF-projekter på området, især det nyligt afsluttede Fælles Infrastruktur for Forskningsdata. Projektet har planlagt en evaluering af resultaterne af en internal(e) ekspert(er) så det sikres at danske løsninger har en international standard.

**13. Hvilken målgruppe vil få gavn af projektet, og er der evt. en sekundær målgruppe?**

Den primære målgruppe er de forskere, der er inddraget i de forskellige cases. Både dem som producerer og dem som aftager forskningsdata. Af sekundære målgrupper kan nævnes

1. Forskningsbibliotekerne, som skal finde deres rolle i data management livscyklussen
2. Universiteterne og de enkelte forskningsområder, som tilbydes services indenfor data management.
3. De fremtidige forskere som kan få tilbudt løsninger baseret på de løsninger som etableres i projektet.
4. Brugere som kan anvende de datasæt der bliver frit tilgængelige ved projektets afslutning.

#### 14. Hvilke målbare succeskriterier opstilles for projektet, og hvordan skal disse vurderes?

##### Generelle succeskriterier

- a) Der er etableret et dansk setup til håndtering af den samlede cyklus for forskningsdata. Løsningerne er på niveau med international best practice og demonstrerer, at de møder forskernes behov.  
Vurdering: Internationale eksperter gennemfører en evaluering af projektets resultater.
- b) Projektet har demonstreret, at FFU-bibliotekerne har en vigtig rolle i forvaltningen af forskningsdata og har medvirket aktivt til at sikre, at bibliotekernes kompetencer udvikles og tænkes ind i nationale løsninger.  
Vurdering: Projektets bud på FFU-bibliotekernes rolle i forskningsdata indgår i evalueringen. Derudover evalueres det, hvordan det nationale forum for forskningsdata anvender resultaterne og i hvor høj grad bibliotekernes kompetencer er udviklet og tænkes ind i dette forums arbejde.

##### Data management planning

- c) Der etableres dansk DMPonline installation  
Vurdering: Installationen er tilgængelig for danske forskere
- d) DMPonline tilpasset til danske forskere og danske forhold generelt  
Vurdering: Forskerne vurderer den danske DMPonline som relevant i deres kontekst

##### Data capture, storage and documentation

- e) Nye eller eksisterende repositories står til rådighed for lagring af aktive datasæt med tilhørende metadata.  
Vurdering: Aktive forskningsdata fra alle cases bliver håndteret i henhold til de opstillede data management planer
- f) Der er implementeret teknikker til datafangst som integrerer direkte med forskernes arbejdsprocesser og sikrer at aktive data kommer under forvaltning.  
Vurdering: Data kan hentes ind i repositorerne direkte fra det udstyr forskerne benytter, fx via cloud services, delte fildrev, osv.

##### Data identification, citation and discovery

- g) Der benyttes unikke ID'er for de håndterede datasæt, således at de kan citeres og refereres til på en entydig og persistent måde. Dette vil typisk være i form af DataCite DOI og DanPID.  
Vurdering: Alle datasæt er repræsenteret ved metadata indeholdende et unikt ID
- h) Der benyttes unikke ID'er for forskere og andre personer knyttet til datasættet. Herved kan forskere mv. refereres til på en entydig og persistent måde, og det sikres at meritring ifm. data management tilfalder de rette. Disse ID'er vil alle være i form af ORCID.  
Vurdering: Alle personer er repræsenteret ved metadata indeholdende et ORCID

- i) De håndterede datasæt kan (uanset hvor der er arkiverede) fremsøges i en national søgetjeneste for datasæt etableret som en del af/i tilknytning til Den Danske Forskningsdatabase. Dette realiseres ved at høste metadata fra alle de systemer (forskningsdatabaser, repositories, mv.) der indgår i projektet.  
Vurdering: Metadata er høstet og indekseret i en national tværgående søgetjeneste

#### Select and deposit for long term preservation

- j) Der udarbejdes relevant dokumentation og best practice samt kriterier for hvordan data der skal langtidsbevares, udvælges  
Vurdering: Dokumentation forefindes
- k) Data er formatteret og pakket på en relevant måde, der så vidt muligt muliggør såvel langtidsbevaring som genbrug  
Vurdering: Data fra alle cases findes lagret og ekspertevaluering vurderer metoder og formater relevante

#### Training and marketing toolkits

- l) Der udarbejdes relevant træningsmateriale  
Vurdering: Det konstateres at der er blevet udviklet træningsmateriale på nettet til fælles brug på tværs af alle projektets cases.
- m) Markedsføring og outreach gennemført hos udvalgte forskergrupper  
Vurdering: Det konstateres, at der er lagt en plan for markedsføring og outreach i udvalgte forskergrupper.

#### Sustainability

- n) Der er udviklet forretningsmodeller for relevante services  
Vurdering: Forretningsmodeller findes dokumenteret og tilgængelige

#### 15. Beskrivelse af kommunikationsplan: Hvordan skal resultaterne af projektet formidles og til hvem? Og hvordan skaber projektet synlighed omkring DEFFs virke?

Tema 5, *Training and marketing toolkits*, har som mål at etablere træning og information i forhold til projektets primære målgruppe, forskerne. Herudover vil der blive udviklet en kommunikationsplan som fastlægger niveauet af formidling til øvrige interessenter. Formatet og omfanget på den afsluttende resultatformidling vil blive aftalt nærmere i efteråret 2016. Der kunne sigtes efter en national konference om emnet data management.

Projektet skaber synlighed omkring DEFFs virke via forskellige koordineringstiltag såsom det nationale forum for forskningsdata management, via arbejdet med partnerne i de enkelte cases, i de forskellige formidlingstiltag som planlægges ift. FFU-bibliotekerne og internationalt via evalueringen. Dokumentation for eksempelvis projektets cases, services og best practices udarbejdes og offentliggøres på relevante kanaler. Her tænkes eksempelvis på website fra Det Nationale Forum for Forskningsdatamanagement og/eller website fra DeIC's e-Science kompetencecenter. Såfremt ingen af disse fora mod forventning har relevante eller brugbare kanaler udarbejder projektet selv et website til formidling

#### 16. Hvordan har projektet haft en effekt på/for sektoren som helhed efter afslutning?

Ved projektets afslutning har FFU-bibliotekerne leveret en samlet løsning, af international høj standard, til håndtering af de 4 faser som forskningsdata indgår i. FFU-bibliotekerne har for alvor påvist kompetencer ift. håndtering af forskningsdata og der er skabt grobund for FFU-

bibliotekernes fortsatte aktive og centrale rolle på området. Derudover er der ved projektets afslutning så vidt muligt fri adgang til datasæt for alle interesserede (relevant jura taget i betragtning) og potentialet synliggøres for alvor.

17. **Har projektet en blivende effekt? Hvordan sikres projektets leverancer i en driftsfase?** Efter projektets afslutning tager projektpartnerne ansvaret for drift af de løsninger som benyttes, indkøbes eller udvikles i projektperioden. Institutionerne bag projektet garanterer, at publicerede forskningsdata, som projektets partnere tager ansvar for, kan tilgås i minimum 10 år.

Et af projektets tværgående temaer er sustainabilty og der arbejdes således gennem hele projektet struktureret med at indsamle viden om, hvilke omkostninger der er forbundet med de enkelte faser. Projektet bidrager således med erfaringsbaseret input til drøftelser af fremtidige forretningsmodeller for håndtering af forskningsdata samt mulige finansieringskilder.

Det faktum at projektet her baserer sig på konkrete cases skulle gerne resultere i løsninger der kan bruges af forskerne og dermed får disse i sig selv en blivende værdi.

## 18. Tidsplan

Projektet løber fra marts 2015 til februar 2017. Se bilag-2 for mere detaljeret tidsplan

### Budget

Projektet har et samlet budget på kr. 8.230.000 hvoraf der søges finansiering til kr.

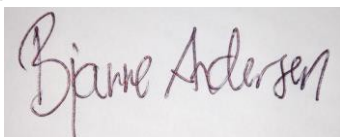
4.522.000 Se bilag-3 for detaljeret budget

Forskerinddragelse i de forskellige cases betales af case-budgettet således at partnerne selv laver aftaler om evt. økonomisk kompensation ift. projektdeltagelsen. DelC er indstillet på efter nærmere aftale at stille yderligere ressourcer til rådighed for projektet i form af fx. server- og lagerkapacitet, såfremt det er hensigtsmæssigt for projektets løsninger.

19. **lønudgifter:** ansøgt: 3.564.000 - total budget: 5.940.000  
 20. **udgifter til møder og rejser:** ansøgt: 100.000 - total budget: 100.000  
 21. **udgifter til projektledelse:** ansøgt: 528.000 - total budget: 660.000  
 22. **udgifter til evaluering:** ansøgt: 100.000 - total budget: 100.000  
 23. **udgifter til konsulentbistand:** ansøgt: 130.000 - total budget 130.000  
 24. **andre udgifter:** ansøgt: 100.000 - total budget: 1.300.000  
 25. **Bemærkninger**  
 26. **Dato og underskrift af den økonomisk ansvarlige**

Århus 31/10 2014

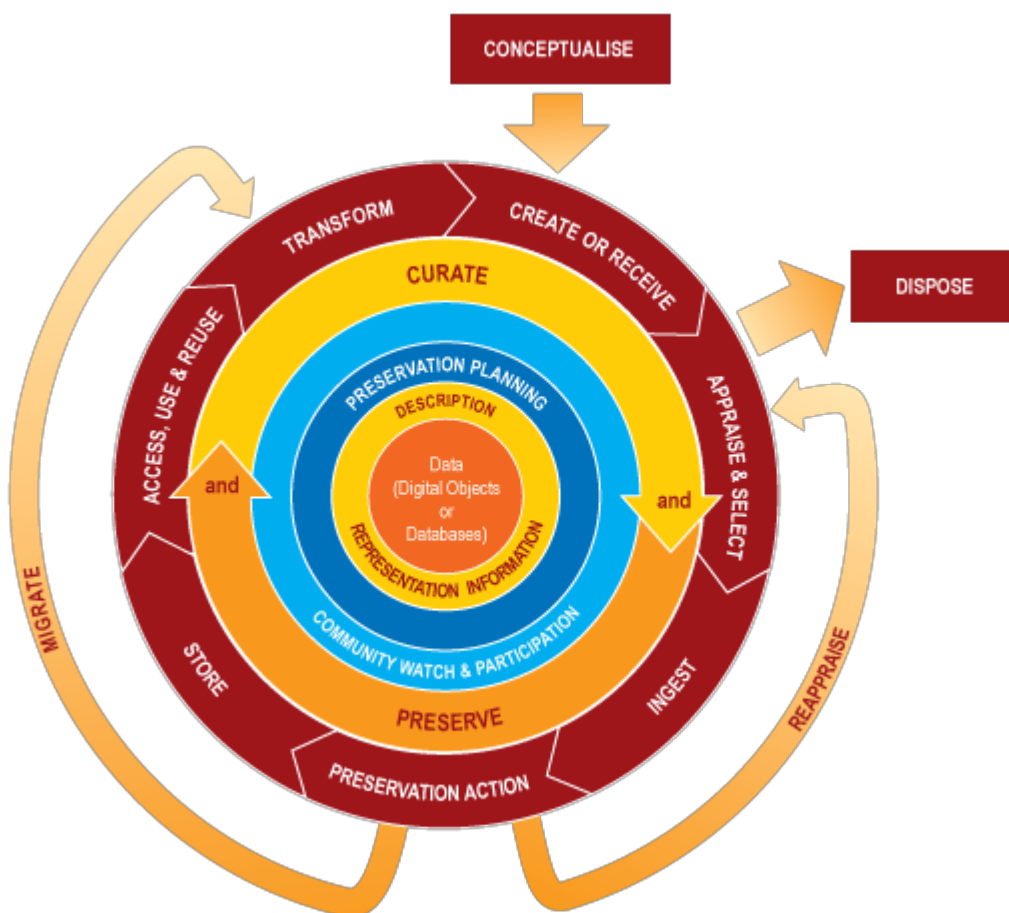
Bjarne Andersen



## Bilag - 1 Livscyklus for forskningsdata

Hvad er baggrunden for projektets metode og temaer?

Udviklingen af effektive tjenester til støtte for forskning forudsætter en god forståelse af forskningsprocessen og forskningsdatas livscyklus. Data skabes på mange tidspunkter i denne proces, og meget information, som er afgørende for effektiv langtidsbevaring og genbrug af data skabes allerede, når forskningsprojektet konceptualiseres og data indsamles. Mange organisationer har derfor fundet det nyttigt at sammenfatte deres forståelse af forskningsdatas livscyklus i en model. En rapport udarbejdet af CEOS (Committee on Earth Observation Satellites) identificerer 55 sådanne modeller.<sup>1</sup>



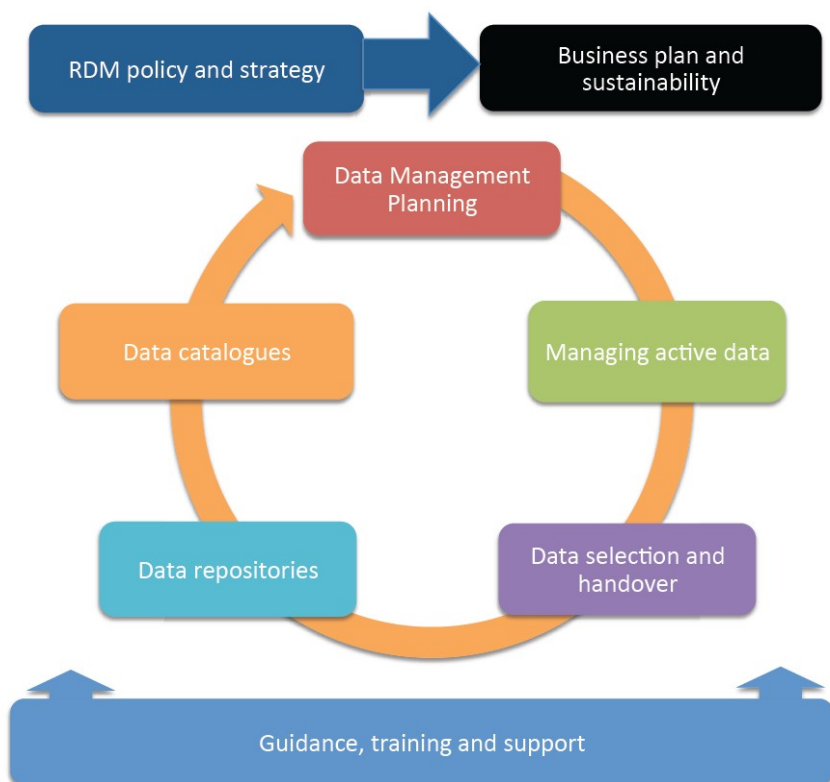
Figur 1: DCCs Curation Lifecycle Model

<sup>1</sup> CEOS Working Group on Information System and Services, Data Lifecycle Models and Concepts, version 1.2, <http://www.ceos.org/images/DSIG/Data%20Lifecycle%20Models%20and%20Concepts%20v13.docx>, 30/10-14

Digital Curation Centre (DCC) er et nationalt kompetencecenter for forskningsdata i Storbritannien, som har sammenfattet betydelige erfaringer med forvaltning af forskningsdata og udviklingen af institutionelle strategier og tjenester til støtte for forskningsdata i både modeller og guides.<sup>2</sup> Disse danner derfor - sammen med de øvrige erfaringer og analyser fra det tidligere DEFF projekt, Fælles Infrastruktur for Forskningsdata<sup>3</sup> - i høj grad grundlag for projektets tilgang til udviklingen af danske løsninger. Figur 1 viser DCCs Curation Lifecycle Model,<sup>4</sup> og Figur 2 viser centrale komponenter i tjenester til støtte for forvaltning af forskningsdata.<sup>5</sup>

Figur 2

### Components of research data management support services



<sup>2</sup> Se navnlig Jones, S., Pryor, G. & Whyte, A. (2013). 'How to Develop Research Data Management Services - a guide for HEIs'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Tilgængelig online: <http://www.dcc.ac.uk/resources/how-guides>

<sup>3</sup> Asger Væring Larsen, Sacha Zurcher, Mikkel Hvidtfeldt Andersen, Jesper Boserup Thestrup, Søren Ærendahl Mikkelsen, Mogens Sandfær, Anders Conrad, *FIF - Fælles Infrastruktur for Forskningsdata. Rapport fra fase 1*. 11/7-2014.

<sup>4</sup> Se <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

<sup>5</sup> Jf. <http://www.dcc.ac.uk/resources/developing-rdm-services>

Bilag 2. Tidsplan								
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
<b>Tværgående</b>								
Styregruppetmøder	1		2		3		4	5
Projektmøder og koordinering med Nationalt Forum								
Identifikation af relevante formidlingskanaler til resultater								
Eventuel etablering af eget website om nødvendigt								
Undersøgelse af Pure's rolle ift. Data Management								
Afsluttende konference med ekspertevaluering								
<b>Cases</b>								
Alle cases udarbejder detaljeret tids- og leverance-plan								
Definition af hvilke data der arbejdes med								
Identifikation af eksisterende / nye services								
Etablering af evt. nye pilot-services								
Dokumentation af services og best practices								
<b>Data Management Planning</b>								
Etablering af DMPOnline installation								
Opsamling af krav fra cases og andre relevante kilder								
Tilpasning af DMPOnline til dansk version								
Udarbejdelse af DM-planer for alle cases								
<b>Data capture, storage and documentation</b>								
Overblik over eksisterende data-flows								
Specifikation af krav til metadata								
Data opsamlet og dokumenteret i alle cases								
<b>Data identification, citation and discovery</b>								
Definition af relevante identifikatorer for alle elementer								
Definition af relevante workflows mellem systemer								
Implementation af workflows mellem systemer								
<b>Select and deposit for long term preservation</b>								
Definition af strategier for udvælgelse af data til bevaring								
Definition af relevante pakkeformater til bevaring								
Udvalgte data bevaret i relevante systemer i alle cases								
Adgang til bevarede data på relevant måde i relevante systemer								
<b>Training and marketing toolkits</b>								
Afklare arbejdsdeling med relevante aktører (fx. DeIC kompetencecenter)								
Udarbejdelse af materiale / skabeloner til vejledning, undervisning, kursus								
Informationskampagner på alle universiteter								
<b>Sustainability</b>								
Etablering af prismodeller for fortsat drift af relevante services								
Forslag til finansieringsmodeller (eg. brugerbetaling)								



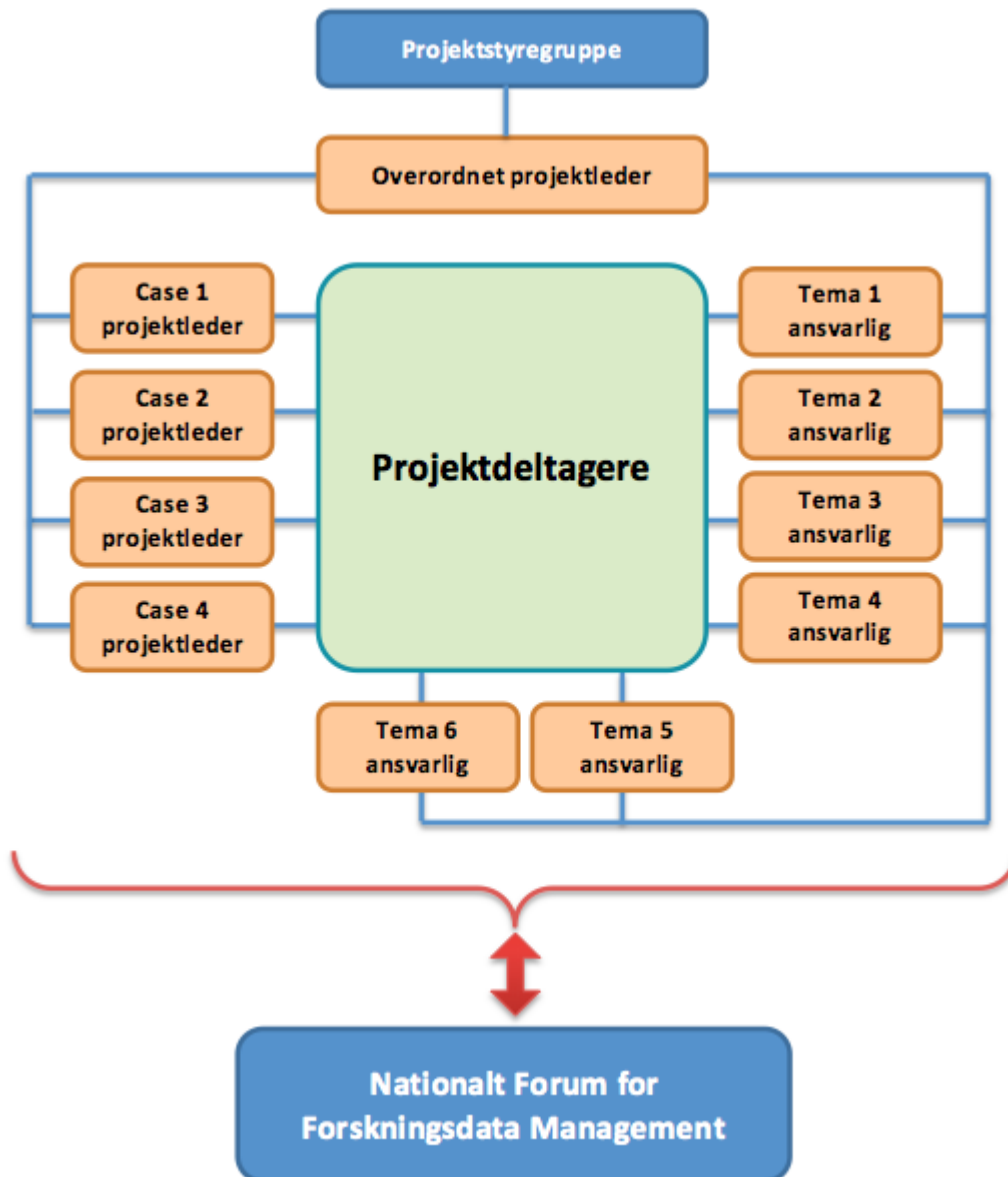
## Bilag 3. Budget

	Person måneder	Pr. måned inkl. OH	Udgift i alt	Egen- finans.	DEFF tilskud
<b>DTU: TEK cases + Tema 3 leder</b>					
Projektmedarbejder til cases (postdoc el.)	24	60,000	1,440,000	576,000	864,000
Projektmedarbejder til tema (specialkons.)	2	66,000	132,000	52,800	79,200
Projektledelse af case (specialkons.)	2	66,000	132,000	132,000	0
Ekstern IT-konsulentbistand			130,000		130,000
Intern IT-hardware & -service (anslået værdi)			400,000	400,000	0
<b>I alt</b>			<b>2,234,000</b>	<b>1,160,800</b>	<b>1,073,200</b>
<b>SB: HUM cases + Tema 4 leder</b>					
Projektmedarbejder til cases	24	60,000	1,440,000	576,000	864,000
Projektmedarbejder til tema	2	66,000	132,000	52,800	79,200
Projektledelse af case	2	66,000	132,000	0	132,000
Ekstern IT-konsulentbistand					0
Intern IT-hardware & -service (anslået værdi)			400,000	400,000	0
<b>I alt</b>			<b>2,104,000</b>	<b>1,028,800</b>	<b>1,075,200</b>
<b>KB: HUM+NAT cases + Tema 2 leder</b>					
Projektmedarbejder til cases	24	60,000	1,440,000	576,000	864,000
Projektmedarbejder til tema	2	66,000	132,000	52,800	79,200
Projektledelse af case	2	66,000	132,000	0	132,000
Ekstern IT-konsulentbistand					0
Intern IT-hardware & -service (anslået værdi)			400,000	400,000	0
<b>I alt</b>			<b>2,104,000</b>	<b>1,028,800</b>	<b>1,075,200</b>
<b>RUC: SAM cases + Tema 1 leder</b>					
Projektmedarbejder til cases	4	60,000	240,000	96,000	144,000
Projektmedarbejder til tema	2	66,000	132,000	52,800	79,200
Intern IT-hardware & -service (anslået værdi)					0
<b>I alt</b>			<b>372,000</b>	<b>148,800</b>	<b>223,200</b>
<b>SDU: SAM cases + Tema 6 leder</b>					
Projektmedarbejder til cases	4	60,000	240,000	96,000	144,000
Projektmedarbejder til tema	2	66,000	132,000	52,800	79,200
Intern IT-hardware & -service (anslået værdi)					0
<b>I alt</b>			<b>372,000</b>	<b>148,800</b>	<b>223,200</b>
<b>AUB: Tema 5 leder+ PURE undersøgelse</b>					
Projektmedarbejder til tema	4	60,000	240,000	96,000	144,000
Intern IT-hardware & -service (anslået værdi)					0
<b>I alt</b>			<b>240,000</b>	<b>96,000</b>	<b>144,000</b>
<b>DDA/SA: SAM cases</b>					
Projektmedarbejder til cases	4	60,000	240,000	96,000	144,000

Intern IT-hardware & -service (anslået værdi)					0
---	--	--	--	--	---

	<b>I alt</b>			<b>240,000</b>	<b>96,000</b>	<b>144,000</b>
<b>Tværgående udgifter</b>						
	Overordnet projektledelse (SB)	4	66,000	264,000	0	264,000
	Møder + Rejser			100,000	0	100,000
	Videndeling + Kommunikation			100,000	0	100,000
	International ekspertevaluering			100,000	0	100,000
	<b>I alt</b>			<b>564,000</b>	<b>0</b>	<b>564,000</b>
<b>Sammentælling</b>						
				<b>Udgift ialt</b>	<b>Egen- finans.</b>	<b>DEFF tilskud</b>
19	Lønudgifter (inkl. OH, eks. projektledelse)			5,940,000	2,508,000	3,564,000
20	Møder + Rejser			100,000	0	100,000
21	Projektledelse			660,000	132,000	528,000
22	Evaluering			100,000	0	100,000
23	Konsulentbistand			130,000	0	130,000
24	Andre udgifter			1,300,000	1,200,000	100,000
	<b>I alt</b>			<b>8,230,000</b>	<b>3,840,000</b>	<b>4,522,000</b>

## Bilag 4. Projektorganisering



Projektet organiseres med en styregruppe der har det overordnede ansvar for projektets fremdrift og leverancer. Styregruppen sammensættes af et medlem fra hver af de deltagende organisationer.

Projektet har en hoved-projektleder som placeres hos Statsbiblioteket og denne danner bindeled mellem styregruppen og projektgruppen.

Projektgruppen består af temaansvarlige for de 6 tværgående temaer, projektlederne for de 4 cases samt Statens Arkiver (DDA). Projektgruppen kan suppleres med relevante medarbejdere fra de deltagende institutioner

Nationalt Forum for DM fungerer som referencegruppe og hvis denne mod forventning ikke etableres nationalt, nedsætter projektet selv en referencegruppe med blandt andet forsker-repræsentation, der kan vejlede og sparre med projektet omkring retning og mål.

Styregruppen forventes at afholde 5 møder. Et møde per halvår samt et ekstra møde i det sidste halvår for at afslutte projektet og planlægge den afsluttende konference og ekspert-evaluering.

Projektgruppen mødes efter behov med jævnlige fysiske møder (fx. 4 gange om året) og virtuelle møder derimellem.

### The contribution of the cases in relation to the themes of the project

This table shows how the different cases can contribute to the themes. In the table, you can also find an estimate of how relevant the case is the specific theme. See Section 2.1 in the report for further information. The estimate has three levels: low, medium and high. The relevance was estimated at an early phase of the project. The Cases are listed in the same order as they are described in section three to seven in the report.

Case	Theme: Data Management Planning	Theme: Data capture, storage and documentation	Theme: Data identification, citation and discovery	Theme: Select and deposit for long term preservation	Theme: Training and marketing tool	Theme: Sustainability
<b>The LARM Case</b> See section 3.1	Evaluation of putting an institutional template into DMPonline.  (high)	Establish code that extract information from the metadata of LARM.  (high)	Establish metadata based on the DataCite model.  (medium)	No research data will be stored in an open repository since the data are sensitive.  (high)	No plans yet  (high)	N/A.  (high)
<b>The NETLAB Case</b> See Section 3.2	Evaluation of the template from Royal Danish Library - Aarhus  (medium)	Develop algorithms that extract the wanted information and document the validity of the algorithms.  (high)	Establish metadata based on the DataCite model.  (medium)	Only the algorithms will be stored in an open repository, but not the research data since the data are sensitive.  (high)	May contribute with methods for working with research data.  (high)	The case may tell how to work with the Netarchive through a big data system.  (high)

<p><b>The Kierkegaard Case</b> See Section 3.5</p>	<p>Use of DMPonline and DMP for planning and organizing data in dialogue w. IT developr</p>	<p>All data extracted from original running web server and documented in the TEI header</p>	<p>Currently no discovery solution</p>	<p>Currently stored at archiving solution ERDA at UCPH, but future long term preservation has to be organized</p>	<p>Usage as DMP example in Ph.d.-courses</p>	<p>Looking for 50 years of accessibility for active research</p>
<p><b>The Local Elections Case</b> See Section 4.1</p>	<p>(low)</p> <p>Evaluation of templates and guidance in DMPonline</p>	<p>(low)</p> <p>N/A</p>	<p>(low)</p> <p>DOI as persistent identifier. Data is made visible, findable and re-usable</p>	<p>(medium)</p> <p>Data is long term preserved by The National Archives together with previous municipal elections surveys</p>	<p>(low)</p> <p>N/A</p>	<p>(medium)</p> <p>Sustainability is secured through The National Archives</p>
	<p>(high)</p>	<p>(high)</p>	<p>(high)</p>	<p>(medium)</p>	<p>(medium)</p>	<p>(medium)</p>

<p><b>The CALPIU Case</b> See Section 4.2</p>	<p>Evaluation of DMPonline</p>	<p>Already managed by the research team</p>	<p>N/A</p>	<p>The case is used to assess how sensitive data can be handled</p>	<p>Not clear yet</p>	<p>The case has been used to assess whether The Danish National Archives support storage of the file formats and software used in the project</p>
<p><b>The GALAXY Case</b> See Section 5.1</p>	<p>(high)</p> <p>Used DMPonline</p>	<p>(low)</p> <p>Uses RedCAP for some data and metadata</p>	<p>N/A</p> <p>Suitable data will be published in Zenodo but not for a long time</p>	<p>(high)</p> <p>Data will be preserved by The National Archives</p>	<p>(low)</p> <p>Not clear yet</p>	<p>(high)</p> <p>Sustainability secured through OPEN and The National Archives</p>
<p><b>The CAFF terrestrial CBMP Case</b> See Section 6.1</p>	<p>(high)</p> <p>Evaluation of templates and guidance in DMPonline</p>	<p>(high)</p> <p>Already managed by the research team</p>	<p>(high)</p> <p>Establish metadata based on the DataCite model</p>	<p>(high)</p> <p>Already managed by the research team at vandrefalk.dk and CAFF website</p>	<p>(high)</p> <p>Not clear yet</p>	<p>(medium)</p> <p>Awaiting acceptance of the Arctic Biodiversity Data Service (ABDS)</p>
	<p>(high)</p>	<p>(low)</p>	<p>(medium)</p>	<p>N/A</p>	<p>(low)</p>	<p>(medium)</p>



<p><b>The DTU Wind Energy Case</b> See Section 7.1</p>	<p>Evaluation of DMPonline tool, DTU template and guidelines.</p>	<p>Focus on historical data already collected. Identification of relevant data documentation that enable re-use of WIND data.</p>	<p>Use of DataCite MDS. Minting DOIs as persistent identifiers. The main goal in this project is making data visible, findable and re-usable.</p>	<p>Test of Zenodo as a data repository. Defining relevant features and requirements for an institutional repository.</p>	<p>Not clear yet.</p>	<p>Defining the role of the institution for long-term preservation of research data.</p>
<p><b>The DTU Space (Geomagnetic data) Case</b> See Section 7.2</p>	<p>(high)</p>	<p>Data Capture: (high) Documentation: (high)</p>	<p>(high)</p>	<p>(high)</p>	<p>N/A</p>	<p>(medium)</p>
<p><b>The Kepler Case</b> See Section 7.3</p>	<p>Evaluation of templates and guidance.  (high)</p>	<p>Handling of real-time data.  (low)</p>	<p>Main focus in this project is making data visible and promote the research.  (medium)</p>	<p>Defining requirements for an institutional repository.  (high)</p>	<p>Not clear yet.  N/A</p>	<p>Defining the role of the institution for long-term preservation of research data.  N/A</p>
	<p>Experience with DMPonline. Usage of DMP as tool for research and dialogue with researchers  (high)</p>	<p>Already organized by research team  (high)</p>	<p>Work w. DataCite MDS and discipline specific search terms  (high)</p>	<p>Prepare and document data bundles to be identified and reused for research  (high)</p>	<p>Usage as DMP example in Ph.d.-courses  (medium)</p>	<p>Looking for 50 years of accessibility for active research  (medium)</p>

# Larm Case Data Management Plan

12. december 2016

<b>1: Basisoplysninger om projektet</b>	
<b>Projektnavn</b>	Arkivering af brugerskabte data i LARM.fm
<b>Projekt-ID</b>	Hum_SB_LARM
<b>Bevillingsgiver</b>	Det Kgl. Bibliotek som led i DMiP (Data Management i Praksis).
<b>Bevillingsnummer</b>	Ikke relevant
<b>Projektbeskrivelse</b>	Arkivering af brugerskabte data i LARM.fm. Data skal høstes fra ny version af LARM software, for evt. at kunne tilbyde service fremover. Der er tale om én høstning. LARM.fm er udviklet og drives af DIgHumLab (dighumlab.com). LARM.fm hører under temaet medieværktøjer: Lyd- og billedmedier.
<b>Varighed</b>	2017 pga. DMiP
<b>Primærforsker(e)</b>	<p><b>Hovedansvarlig</b> Iben Have Institut for Kommunikation og Kultur - Medievidenskab Helsingforsgade 14 bygning 5335, 246 8200 Aarhus N Danmark</p> <p>---</p> <p><b>Andre kontakter</b></p> <p>Institut for Kommunikation og Kultur - Medievidenskab</p> <p>Det Kgl. Bibliotek <u>Bjarne Andersen,</u> Det Kgl. Bibliotek,</p> <p>ChaosInsight</p> <hr/>

<b>Primærforsker(e)s forsker-ID</b>	Iben Have orcid.org/0000-0003-1724-0898
<b>Primær kontaktperson</b>	Iben Have, se ovenfor
<b>Institutioner, bevillingsgivere og andre</b>	Det Kgl. Bibliotek, LARM.fm, DiGHuMLab, ChaosInsight
<b>1.1: Relaterede politikker</b>	
Hvis der stilles krav til din data management plan, eller krav til data-behandling og -sikkerhed fra bevillingsgivere, institutioner, dataejere, samarbejdspartnere osv. kan du angive dem her:	
<b>Krav til data management planen</b>	DMiP test af langtidsbevaring og data management planlægning. Der er udarbejdet en aftale om brug af brugerskabte data. Se "Vilkår og forudsætninger i forbindelse med LARM.fm".
<b>Krav til data-behandling og -sikkerhed</b>	Nej
<b>1.2: Omkostninger ved håndtering af data</b>	
<b>Omkostninger i forbindelse med håndteringen af dine forskningsdata</b>	"0,00 kr." Udgifter til data i aktiv fase afholdes af DiGHuMLaB. Det Kgl. Bibliotek afholder udgifter til 2027 til bevaring. Derefter er situationen uafklaret.
<b>Opfølgning på din data management plan</b>	Løbende i forbindelse med DMiP projektet.
<b>Dato for første version af projektets datamanagementplan.</b>	1. November 2015. 2. Marts 2016
<b>Dato for seneste reviderede version af projektets datamanagementplan.</b>	Fredag 13/5 2016
<b>Dato for næste planlagte revision af projektets</b>	Løbende

<b>datamanagementplan.</b>	
<b>Hvem har ansvaret for data management?</b>	I denne sammenhæng Jesper B. Thestrup DMiP projektet, da der er tale om et internt Det Kgl. Bibliotek projekt.
<b>1.3: Har du de fornødne ressourcer til at gennemføre din plan?</b>	
Vurder, om du samlet set råder over de fornødne ressourcer som eksempelvis software, hardware og teknisk ekspertise.	
<b>Har du brug for eksterne ressourcer eller bistand, f.eks. hardware, software, særlig ekspertise?</b>	Der er behov for kontakt til ChaosInsight og Det Kgl. Bibliotek IT-udviklingsafdeling. Men de fornødne ressourcer er til rådighed.
<b>1.4: Andres brug af dine data?</b>	
<b>Hvem har ophavsret til data efter projektets afslutning?</b>	Ingen, da der er tildelt en CC0 licens og brugerne har godkendt dette, jævnfør LARM.fm Betingelser: <a href="http://www.larm.fm/Terms">www.larm.fm/Terms</a>
<b>Hvordan skal spørgsmål om ophavsret til nye data, herunder tilladelse til genanvendelse håndteres?</b>	Data er frit tilgængelige for genanvendelse (CC0). Der er dog andre hensyn at tage vedrørende datatilsynet (mulighed for personfølsomme data i de nye data) og ophavsretsloven (data fra Statsbiblioteket, som oprindeligt er fra Ritzau).
<b>Hvordan må tredjepart bruge dine data?</b>	Ikke relevant pga. CC0 licens.
<b>Kan Det Kgl. Bibliotek anvende resultaterne fra forskningsprojektet i andre sammenhænge?</b>	Ja, Det Kgl. Bibliotek kan genanvende de høstede data pga. CC0.
<b>1.5: Ethiske forhold</b>	
Spørgsmål om bevaring, adgang og deling indebærer etiske overvejelser – ud over de juridiske. Du bør vise, at du er opmærksom på mulige etiske problemer og, at du følger din institutions eller relevant myndigheds retningslinjer.	
<b>Har du tilladelse til at gemme og dele data?</b>	Ja, jævnfør "Vilkår og forudsætninger i forbindelse med LARM.fm".

<b>Hvordan vil du sikre anonymitet for f.eks. deltagere i forskningsprojektet, svarpersoner o.l.?</b>	Det Kgl. Bibliotek tester i første omgang, hvad der høstes, for derefter at vurdere om der behov for anonymisering. Det forventes ikke at der er brug for anonymisering.
<b>Hvordan vil du opbevare personfølsomme oplysninger, sikkert?</b>	Det forventes ikke at der vil være personfølsomme data i de høstede data.

## 2: Oplysninger om eksisterende data, der tænkes anvendt.

Ofte baserer et forskningsprojekt sig på allerede eksisterende data (samlinger, databaser eller andet).

<b>Samlingens navn</b>	Det Kgl. Biblioteks radio-og TV-samling + et begrænset udvalg fra DR tilgængeliggjort gennem LARM.fm.
------------------------	---

### 2.1: Juridiske forhold i forhold til indhold og adgang

<b>Indeholder samlingen personfølsomme data?</b>	Radio- og tv-programmer kan indeholde personfølsomme data. Der er givet tilladelse til at disse data kan tilgås af forskere gennem LARM.fm platformen.
--	--

<b>Hvem har ophavsret til de eksisterende data?</b>	Der er tale om en lang række forskellige radio- og tv-programmer med forskellige ophavsret-ejere.
---	---

<b>Hvem kan få adgang til de eksisterende data? Og under hvilke betingelser?</b>	Forskere, der har adgang til LARM.fm, jævnfør "Vilkår og forudsætninger i forbindelse med LARM.fm".
--	---

### 2.2: Oplysninger om eksisterende data

<b>Hvor befinder de eksisterende data sig?</b>	Data findes på Det Kgl. Bibliotek. Metadata findes både hos ChaosInsight og på Det Kgl. Bibliotek.
--	--

<b>Hvilke filformater indgår i de eksisterende data?</b>	audio-filer, video-filer, jpg, pdf og JSON
--	--

<b>Hvorledes vil data blive overført til behandling på en sikker måde?</b>	Der findes en fungerende eksport service for metadata fra Det Kgl. Bibliotek til LARM.fm platformen, hvor forskerne har adgang til data. Data bliver ikke overført.
--	---

<b>Hvor store</b>	Det er kun metadata, der bliver overført. Der findes 2.310.377
-------------------	--

<b>datamængder tænkes overført?</b>	objekter/assets/filer på LARM.fm platformen, men hver af dem meget små. Der er overført ca. 1200 assets/filer indeholdende ca. 12000 annotationer.
-------------------------------------	--

<b>3: Oplysninger om forskningsdata, der skabes og behandles i løbet af forskningsprojektet</b>	
I løbet af et forskningsprojekt skabes en række data. I de følgende punkter bedes du beskrive disse.	
<b>3. 1 Juridiske forhold:</b>	
<b>Hvem har ejerskab/ophavsret til data, der skabes under forskningsprojektet</b>	Brugerskabte data er CC0.
<b>Hvem har adgang til data under forskningsprojektet? Hvorledes vil adgangskontrol blive sikret?</b>	Det er ChaosInsight, der står for LARM.fm platformen og adgangskontrollen. Der er adgang for forskere i projektet og for relevante medarbejdere ved det Det Kgl. Bibliotek.  Til adgangsstyring anvendes <a href="#">WAYF</a> .
<b>Indeholder data personfølsomme oplysninger, mens de behandles?</b>	Ja, det kan de gøre.
<b>Er de nye data personfølsomme?</b>	De forskerskabte data forventes ikke at være personfølsomme, men der er en risiko.
<b>Hvordan vil du sikre anonymitet for f.eks. deltagere i forskningsprojektet, svarpersoner o.l.?</b>	Ikke relevant.
<b>Hvad er forholdet mellem eksisterende data og de nye forskningsdata?</b>	Der er tale om annotationer og nye metadata til eksisterende data.
<b>Indeholder dine data, oplysninger, der skal bruges i forbindelse med</b>	Nej

patentansøgninger?	
<b>3.2: Oplysninger om de data, der skabes i løbet af forskningsprojektet</b>	
Hvor store mængder nye data er der tale om?	Der er tale om manuelt indtastet tekst, forventet maksimalt 10 MB.
Hvilken software tænkes anvendt til at behandle de eksisterende data, som nu er overført til behandling? Hvilken software og filformater tænkes anvendt til at viderebehandle nyskabte data?	Det forventes, at det udelukkende foregår på LARM.fm platformen. Den "nye" version på html5-platformen åbnede 2. december 2015.
Er der specielle krav til software eller hardware?	Ikke relevant
Hvilket operativsystem tænkes anvendt?	Ikke relevant
Overvejes udvikling af ny software?	Der udvikles software til høstning af data fra LARM.fm platformen. En udvidelse af API til høstning fra LARM.fm undersøges. Ambitionen er, at det på sigt skal være muligt at høste nyskabte data, men grundet manglende ressourcer er det sandsynligvis ikke opnåeligt i dette projekt (se endvidere punkt 5.2).
<b>3.3: Organisation af forskningsdata under projektet</b>	
Hvis du forventer at skabe store datamængder eller mange filer er det en fordel at overveje, hvordan du vil organisere dine data i løbet af forskningsprojektet	
Hvilke standarder og metoder vil du anvende?	Ikke relevant
Hvordan vil du strukturere og navngive dine mapper og filer?	Det er muligt at strukturere nye data i navngivne projekter, og det er muligt at tilføje mærkater (tags) til nye data inden for projekterne.
Hvordan vil du håndtere evt. versionering (f.eks.	Det er ikke muligt at versionere de nye data.

<b>benytte versionsstyringsværktøj)?</b>	
<b>3.4: Ikke-digitale data, hvis det er relevant</b>	
Ofte anvendes eller skabes ikke-digitale data. Hvordan forventes disse behandlet?	
<b>Hvor opbevares ikke digitale forskningsdata sikkert?</b>	Ikke relevant
<b>Hvor lang tid vil ikke-digitale forskningsdata blive bevaret sikkert?</b>	Ikke relevant
<b>4.6: Dokumentation og metadata</b>	
Beskriv hvorledes dine data er dokumenteret, således, at udenforstående kan forstå og evt. genanvende dine data. Som minimum skal der være nogle grundlæggende informationer, som kan hjælpe med at finde data, deres ophav og muligheder for at tilgå disse. Dokumentationen skal også gerne inkludere oplysninger om indsamlings- og analysemetode(r), variabeldefinitioner, vokabularer, måleenheder, samt format- og filtype. Anvend gerne eksisterende standarder (DCC vedligeholder en webside med metadata standarder her <a href="http://www.dcc.ac.uk/resources/metadata-standards">http://www.dcc.ac.uk/resources/metadata-standards</a> ).	
<b>Hvilke informationer (dokumentation) er nødvendige for at dine data kan læses og forstås i fremtiden?</b>	Det vigtigste her er persistente ID'er på de programmer, som annotationer og nye metadata tilhører.
<b>Hvilke metadatastandarder vil du bruge og hvorfor?</b>	Her bruges et LARM.fm internt format. Det er vigtigt at dokumentere dette, for at data skal være forståeligt for fremtidens forskere. Der udarbejdes lige nu en vejledning til forskerne, som også vil kun bruges fremadrettet.
<b>4.7: Hvordan vil du opbevare og sikre backup af data medens de behandles?</b>	
<b>Hvordan vil du foretage backup for at sikre mod datatab?</b>	De oprindelige data langtidsbevares på Det Kgl. Bibliotek. Det forventes at LARM.fm platformen sørger for backup af de nye data.
<b>Hvor er data backup placeret?</b>	Hos ChaosInsight, som hoster LARM.fm. ChaosInsight bruger en Amazon storage løsning.
<b>Er lagerplads til backup</b>	Ja



tilstrækkelig?	
Er der sikret mulighed for at genskabe data i tilfælde af et uheld og hvem er ansvarlig for dette?	Ja, på grund af backup og logning.

<b>5: Langtidsbevaring og deling af data</b>	
<i>Både undervejs i og efter et forskningsprojekts afslutning er bevaring af data centralt. Både for at dokumentere og bevare resultater, med henblik på genanvendelse – og til brug for meritering og fagfællebedømmelse. Det handler det følgende om.</i>	
5.1: Jura:	
Hvis dine data er særligt følsomme kan du redegøre uddybende for alle sikkerhedsforanstaltninger og formelle standarder.	Ikke relevant.
<b>5.2: Skal dine data langtidsbevares?</b>	
Ofte vil data skulle opbevares i længere tid. Det kan gælde alle data eller kun udvalgte datasæt. I det følgende har du mulighed for at beskrive dette nærmere.	
Hvilke data skal langtidsbevares og hvilke data skal slettes?	Det Kgl. Bibliotek forpligter sig til at gemme de nye forsknings-data til 2027. Derefter skal der indgås en aftale.
Hvor længe skal data langtidsbevares?	Til 2027 uden beregning, med mulighed for forhandling om forlængelse mod betaling.
Hvor skal data langtidsbevares? Og hvorledes vil indsamlede data blive overført på sikker måde?	På Det Kgl. Bibliotek. De forskerskabte data høstes med en applikation udviklet på Det Kgl. Bibliotek.
Stilles der særlige krav om klargøring af data til langtidsbevaring?	Der skal vælges mellem følgende løsninger. <ol style="list-style-type: none"> <li>1. Opsætning af en løsning, der muliggør en årlig Det Kgl. Bibliotek-høstning af alle brugerskabte annotationer og metadata. Alle brugerskabte data langtidsbevares som et</li> </ol>

	<p>samlet datasæt.</p> <ol style="list-style-type: none"> <li>Opsætning af en løsning, der muliggør en årlig Det Kgl. Bibliotek-høstning af alle projekter og mærkater. Alle projekter bevares som selvstændige datasæt med mærkater og annotationer og metadata på alle de links, der er samlet i projekter.</li> <li>Opsætning af en løsning, der muliggør at en forsker, når hun/han ønsker det, kan høste alle sine egne data (dvs egne projekter, mærkater og annotationer og metadata på alle de links, der er samlet i projekter/mærkater på et givent tidspunkt. Det høstede datasæt kan derefter langtidsbevares.</li> </ol> <p>I alle tilfælde skal beskrivelse af formater også bevares sammen med datasæt.</p> <p>Det kræves desuden, at der er et ophav til data. I de to første løsninger, anbefales det at DigHumLab, eller evt. Iben Have anføres som ophav. I den sidste løsning er forskeren ophav. Her skal sættes henvisning ind til et bilag.</p>
<b>5.3: Skal dine data deles?</b>	
<i>Du bør overveje betingelser for datadeling, både med hvem, internt og eksternt, i hvilket omfang og hvorledes anerkendelsen af dig som ophav til data kan sikres.</i>	
<b>Eksisterer der en formel aftale om deling af data?</b>	Ja, forskerne har underskrevet en CC0 licens.
<b>Hvilke data skal deles?</b>	Det forventes at data kan deles.
<b>Er der restriktioner på deling af data? Har du behov for en periode med ene adgang/embargo?</b>	Vi skal være sikre på, at der ikke er nogen ophavsretlige beskyttede eller personfølsomme data.
<b>Hvordan vil andre kunne finde dine data?</b>	Gennem Det Kgl. Bibliotek's forsknings-data-repository og gennem Den Danske Forskningsdatabase.
<b>Vil du søge at få en persistent identifikation (F.eks. DOI) til dine data?</b>	Ja, gennem Det Kgl. Bibliotek's repository.
<b>Har du brug for at kontrollere adgang til data og hvis ja</b>	Nej.

hvordan?	
----------	--

## Tjekliste for "Probing a Nation's Web Domain"

### Opdateret efter møde fredag 2017-01-27

Bygger datamanagementplanen på eksisterende standarder eller procedurer? Hvis ja, hvilke? Har den institution, som projektet hører under, særlige regler for datasikkerhed? Hvis ja, hvilke (evt. som links)? Stiller samarbejdspartnere, bevillingsgivere eller andre krav til datamanagement? Hvis ja, hvilke (evt. som links)?

#### Stamoplysninger

<b>Projektnavn</b>	Hvis du har søgt en bevilling, brug samme navn som i bevillingsansøgningen	Probing a Nation's Web Domain
<b>ID</b>	Et relevant ID, eventuelt bestemt af bevillingsgiver og/eller institution	
<b>Bevillingsgiver</b>	Navn på bevillingsgiver, hvis det er relevant	Kulturministeriet, DEIC, Netlab
<b>Bevillingsnummer</b>	Reference til en bevillingskrivelse, hvis det er tilgængeligt	FPK.2015-0049
<b>Projektbeskrivelse</b>	Fortæl kort, hvilken type studie, der er tale om, sådan så andre kan forstå formålet med at skabe dine forskningsdata:	<b>Forskningsspørgsmål:</b> Hvordan har .dk domænet udviklet sig over tid fra 2005-2015
	Hvilken type forskningsprojekt er der tale om?	internet-forskning, digital humaniora
	Med hvilket formål bliver data skabt?	For at analysere dem, og besvare forskning-spørgsmål
	Reference til projekt-ansøgning, hvis den er tilgængelig	links Niels Brügger, Professor, Head of the Centre for Internet Studies, and of NetLab/DigHumLab
<b>Primærforsker</b>	Navnet på projektets primærforsker(e)	Janne Nielsen, Assistent Professor, NetLab/DigHumLab  Ditte Laursen, Senior Researcher, SB/Netarkivet
<b>ForskerID</b>	F.eks. ORCID	Niels Brügger <a href="http://orcid.org/0000-0003-1787-1980">http://orcid.org/0000-0003-1787-1980</a>
<b>Kontaktperson</b>	Navn (hvis andet end primærforskere), tlf.nummer, e-mailadresse	Niels Brügger

<b>Organisationer</b>	<i>Involverede organisationer</i>	Netlab, AU, IKK, Center for Internetforskning, SB, Netarkivet
<b>Version</b>	<i>Projektets tidsmæssige længde</i>	open-ended, lige nu med finansiering til sommer 2017.
	<i>Dato for den første version af projektets datamanagementplan</i>	14/9-2015
	<i>Dato for senest ændrede version af projektets datamanagementplanen</i>	tirsdag 7/2-2017
	<i>Dato for næste planlagte version af projektets datamanagementplanen</i>	Der er ikke planlagt flere opdateringer i forbindelse med DmiP-projektet.
<b>Relaterede politikker</b>	<small>Hvis der stilles krav til din datamanagement plan - fra bevillingsgivere, institutioner, dataejere, samarbejdspartnere osv. kan du angive dem her:</small>	
	<i>Bygger datamanagementplanen på eksisterende standarder eller procedurer?</i>	Ja, den bygger på den Tjekliste, som SB opbygger i forbindelse med DmiP-projektet.
	<i>Hvis ja, hvilke?</i>	Der er en række involverede institutioner. Netarkivet stiller særlige krav for datasikkerhed, da der er tale om potentielt personfølsomt og ophavsret-beskyttet data. SB stiller også nogen krav i forbindelse med brug af Kulturarvscluster. Referencer: 1. Retningslinjer adgang/forskere Netarkivet: <a href="http://netarkivet.dk/wp-content/uploads/Retningslinjer-for-adgang-til-Netarkivet.pdf">http://netarkivet.dk/wp-content/uploads/Retningslinjer-for-adgang-til-Netarkivet.pdf</a> 2. Ansøgererklæring til Netarkivet: <a href="http://netarkivet.dk/wp-content/uploads/2016/02/ansoegererklaering-feb2016.pdf">http://netarkivet.dk/wp-content/uploads/2016/02/ansoegererklaering-feb2016.pdf</a> 3. Aftale med Kulturarvscluster.
	<i>Har den institution, som projektet hører under, særlige regler for datasikkerhed?</i>	
	<i>Hvis ja, hvilke (evt.som links)?</i>	
	<i>Stiller samarbejdspartnere, bevillingsgivere eller andre krav til datamanagement?</i>	Nej.
	<i>Hvis ja, hvilke (evt. som links)?</i>	
<b>Data Management</b>	<i>Anslået pris for data management gennem og efter projektet.</i>	0
	<i>Eksempelvis pris for opbevaring</i>	

af data.

### Data

#### Hvilke eksisterende data tænkes anvendt?

Digitale data (samlinger) der eksisterer før forskningsprojektet går i gang.

<i>Samlingens navn</i>	Netarkivet Der skal analyseres på arkiv-data fra Netarkivet  Der skal analyseres på 10 tværsnitshøstninger - 1 for hvert år.
<i>Hvad påtænkes disse data anvendt til ? (uddybende beskrivelse af forskningsspørgsmål og -mål)</i>	Da en tværsnitshøstning i sig selv fylder mere end 30Tbytes er der ikke råd til at lave komplette udtræk af data til langtidsbevaring som forskningsdatasæt og der skal derfor tænkes i baner af eksempelvis at lave et index over en tværsnitshøstning som så kan gemmes som et datasæt en fremtidig forsker kan genbruge for at lave nye analyser på samme datasæt eller for at kontrollere forskningen i dette projekt
<i>Hvem har ansvaret for disse data?</i>	Original-data bor hos Netarkivet på SB og KB, som er data-ansvarlige.
<i>Er der særlige juridiske forhold? (personfølsomme oplysninger, patenter, ophavsret...)</i>	Data i Netarkivet kan være personfølsomme og det er ophavsret-beskyttet, og der er derfor meget begrænset adgang.
<i>Hvem kan få adgang til data? (betingelser)</i>	Man skal have en individuelt udstedt tilladelse for at kunne få adgang.
<i>Hvor befinder disse data sig?</i>	SB og KB
<i>Hvor store datamængder tænkes behandlet (fx antal Gb / Tb samt antal objekter/filer)</i>	Der skal behandles meget store mængder data (10 x 5-30Tbytes) - måske 200 Tbytes
<i>Hvilke filformater indgår i samlingen? (TEI, JP2, MP3, ...)</i>	Principielt kan samlingen indeholde alle kendte fil-formater- pga. Internettets natur er det ikke til at sige.  Original-data fra Netarkivet er lagret i ARC og WARC-filer hvilket betyder at der skal anvendes software der kan forstå disse formater.

Crawllogs, andre metadata-filer, som skal kunne forstås.

*Hvilke værktøjer anvendes (navn, version og systemkrav)? (gerne link til værktøjet)*

Kulturarvscluster v. 1.0  
 primære komponenter:  
 Operativ system: CentOS Linux: CentOS 7.2.1511  
 Cluster software: Apache Spark: Spark 1.6.1  
 Primære værktøjer: RStudio Server 1.0.44

*Hvilken platform og versionsnummer anvendes (windows, linux, mac)?*

Nye data vil bestå af indexer over bestemte corpora udtrukket af Netarkivet.

Den første udfordring vil derfor være at få defineret dette index-format samt at få lave værktøjer der giver en række parametre kan lave et passende udtræk af "en tværsnitshøstning" (som også skal defineres både intellektuelt samt teknisk). Det er valgt at bruge CDXJ-formatet, som ikke er en standard endnu, men en draft, se <https://iipc.github.io/warc-specifications/specifications/cdx-format/openwayback-cdxj/>.

**Hvilke forskningsdata tænkes genereret som en del af projektet?** *Hvad er forholdet mellem eksisterende data og de nye forskningsdata ?*

Andre nye data er

- \* bash-scripts
- \* python-kode
- \* udtrukket statistik i CSV-format
- \* søgninger
- \* søgeresultater fra netarkiv-indexet i JSON-format.

*Skal alle datasæt bevares?*

Indeks skal bevares. Procedurebeskrivelse skal bevares + beskrivelse af anvendte værktøjer. Statistikker og søgeresultater samt beskrivelser af disse formater skal også bevares.

*Hvad påtænkes de nye data anvendt til ? (uddybende beskrivelse af forskningsspørgsmål og tilgangsvinkel)*

De nye data skal bruges til at lave deskriptiv statistik af indholdet i netarkivet. I projektet laver vi 4 probes. Vi henviser til dokumentet "Probes — The Cook Book" for en teknisk forklaring af disse probes.

		Vi forventer også at data skal danne grundlag for yderligere forskning indenfor Digital Humaniora.
	<i>Hvor store mængder nye data er der tale om ? (størrelse og antal objekter/filer)</i>	Under 5 TB
	<i>Har forskningsdata værdi efter projektets afslutning, således at de bør bevares? (i givet fald hvilken værdi og for hvem)</i>	Da projektet bruger mange ressourcer på at udvikle udtræks-algoritmer mm. tænkes de nye datasæt at repræsentere en stor værdi for fremtidige forskere der vil kigge på nogle af de samme afgrænsninger (hele tværsnit af .dk-domænet)
	<i>Hvilke kendte værktøjer tænkes anvendt på de nye data? (gerne link til værktøjet)</i>	R, til deskriptiv statistik ( <a href="https://www.r-project.org">https://www.r-project.org</a> )
	<i>Overvejes udvikling af nye værktøjer?</i>	Ja, det påtænkes at udvikle et ETL værktøj (Extract-Transform-Load), der kan udstille data fra Netarkivet til Kulturarvsclusteret.  Værktøjet/værktøjerne ejes af Det Kgl. Bibliotek med Open Source licens.
	Filformater bør opfylde følgende kriterier:	
<b>Filformat af forskningsdata</b>	<ul style="list-style-type: none"> <li>• Godkendt som en international eller national standard</li> <li>• Almindelig brugt</li> <li>• Accepteret som best practice indenfor fagområdet</li> </ul>	
	<i>Hvilke filformater tænkes de nye data at blive lagret i ?</i>	Et bud på et index-format kunne være det såkaldte CDX-format som i forvejen anvendes som index under WayBack-maskinen. Men der er også andre kandidater (IA har en hel tool-suite der anvender 4 forskellige afledte dataformater: CDX, WAT, WANE, LGA-data)  <a href="https://github.com/vinaygoel/ars-workshop">https://github.com/vinaygoel/ars-workshop</a>  CDXJ er valgt til bevaring: <a href="https://iipc.github.io/warc-specifications/specifications/cdx-format/openwayback-cdxj/">https://iipc.github.io/warc-specifications/specifications/cdx-format/openwayback-cdxj/</a>  Udover index-formatet er der tale om CSV,



		JSON og text.
	<i>Er der specielle software eller hardware krav for at benytte disse data?</i>	Nej, de kan processeres af alm. software/hardware.
	<i>Hvis ja, hvilke?</i>	
<b>Organisation af forskningsddata under projektet</b>	<i>Hvilke standarder og metoder vil du anvende?</i>	Uafklaret
	<i>Hvordan vil du strukturere og navngive dine mapper og filer?</i>	Uafklaret
	<i>Hvordan vil du håndtere versionering (f.eks. benytte versionsstyringsværktøj)?</i>	Uafklaret
	<i>Hvordan vil du kvalitetssikre dine data (f.eks. review hos kollega)?</i>	Uafklaret
<b>Ikke digitale data</b>	<i>Hvor opbevares evt. ikke digitale forskningsdata sikkert?</i>	Der er ingen ikke-digitale forskningsdata.
	<i>Hvor lang tid vil ikke digitale forskningsdata blive bevaret sikkert?</i>	Se ovenstående.
<b>Dokumentation og metadata</b>		
<b>Hvilken form for dokumentation og metadata vil følge dine data?</b>	<i>Hvilke informationer er nødvendige for at dine data kan læses og forstås i fremtiden?</i>	Beskrivelsen af index-filen og beskrivelse af procedure for, hvordan man viser data. Beskrivelser af csv og json data: tekstfiler fx som en README-fil.
	<i>Hvordan vil du skabe den nødvendige dokumentation og metadata?</i>	Dokumentationen vedligeholdes løbende på NetLabs wiki på SBProjects, hvor projektet har sit eget område. De dokumenter, som hører til metadata skal til sidst lægges i et passende format (e.g. en readme-fil) og gemmes sammen med metadata.
	<i>Hvilke metadatastandarder vil du bruge og hvorfor?</i>	Der findes ikke en dedikeret netarkiv-forskning metadata-standard. Der bliver derfor brugt Dublin Core og DataCite.
	<i>Beskriv den type dokumentation der vil følge dine data, sådan at udenforstående kan forstå og genanvende dine data. Som minimum skal der være</i>	Tekstbehandlingsdokumenter, artikler etc. Disse opdateres løbende på projektets wiki-sider.

nogle grundlæggende oplysninger, som vil hjælpe med at finde data, inklusive hvem der har skabt data, en titel, en dato og evt. adgangsbegrænsninger. Dokumentationen skal også gerne inkludere oplysninger om metode, analyse, variabeldefinitioner, vokabularer, målingsenheder, antagelser samt format- og filtype. Overvej hvordan og hvor du vil redegøre for disse oplysninger. Det er en god idé at følge eksisterende standarder.

README-filer følger datasæt.

Metadata tilføjes når data langtids-bevares.

### Etiske og lovgivningsmæssige forhold

#### Hvordan vil du håndtere etiske overvejelser?

Etiske spørgsmål kan have betydning for hvordan du håndterer dine data. Både i forhold til hvordan de opbevares, hvem der har adgang til dem og hvor længe de er bevaret. Du bør vise, at du er opmærksom på evt. etiske problemstillinger, og at du har taget hensyn til retningslinjer udstukket af din institution (også dem der går ud over de rent lovmæssige forhold)

Har du tilladelse til at gemme og dele data?:

Ja, vi må gemme data, men ikke dele data. Se ovenfor. Deling af data og resultater skal overholde persondata- og ophavsret. Statistisk data og "statistiske søgeresultater" kan deles.

Hvordan vil du sikre anonymitet for evt. deltagere i forskningsprojektet?

Hvordan vil du håndtere personfølsomme oplysninger, så du sikrer at de er opbevaret sikkert?

Index ønskes opbevaret i et lukket repository med samme restriktioner som Netarkivet. Alternativt skal de slettes.

#### Hvordan vil du håndtere evt. spørgsmål om copyright og intellektuel ejendomsret?

Hvem skal spørges om lov, hvis data skal genanvendes?

Ikke relevant. Projektet håndhæver ikke copyright over de data, som bliver skabt. Det er Niels Brügger, der er data-ansvarlig.

Er der nogen adgangsbegrænsninger i forhold til genanvendelse af tredjepart?

Vil datadeling blive forsinket på grund af publikationer eller patenter?

Beskriv hvem og hvor mange der skal spørges om lov i tilfælde af, at en

	<i>tredjepart vil genanvende data</i>	
	<i>Er der dansk copyright på data?</i>	
	<i>Er der udenlandsk copyright på data?</i>	
<b>Ejerskab af data</b>	<i>Ejerskab af kilde-data</i>	Ikke relevant.
	<i>Ejerskab af data skabt i projektet (f.eks. et indeks)</i>	
	<i>Ejerskab af metadata skabt i projektet</i>	
<b>Tredje parts data</b>	Hvis andre ejer data, hvordan kom projektet i besiddelse af disse data og hvilke vilkår gælder der for tilgang til disse data.	Ikke relevant.
	<i>Reference nummer til købte eller licenserede data</i>	
	<i>Reference nummer til data, der er opnået gennem en open access licens</i>	
	<i>Beskrivelse af vilkår for data opnået på anden vis</i>	

## 5: Backup, bevaring og deling af data

Både undervejs i og efter et forskningsprojekts afslutning er bevaring af data centralt. Både for at dokumentere og bevare resultater, med henblik på genanvendelse – og til brug for meritering og fagfællebedømmelse. Det handler det følgende om.

### 5.1: Hvordan vil du opbevare og sikre backup af data under dit forskningsprojekt?

**Hvordan vil du foretage backup for at sikre mod datatab?**

Under “Active DM” ingen backup af nye forskerskabte data. I øjeblikket er der ingen backup af clusteret.

**Hvor er data backup placeret?**

**Er lagerplads til backup tilstrækkelig?**

**Er der sikret mulighed for genskabelse af data i tilfælde af et uheld og hvem er ansvarlig for dette?**

Hvis ja, angiv kort hvorledes.

Algoritmer til at hente data fra Netarkivet over på clusteret er under backup og kan køres igen.

Algoritmer lagret på forsker-maskiner kan ligeledes køres igen på Kulturarvsclusteret.

### 5.2: Hvordan vil du håndtere adgang og sikkerhed?

**Er der foretaget risikoplanlægning i forhold til datasikkerhed?** Hvis ja, beskriv kort. Nej

**Hvorledes vil adgangskontrol blive sikret?** Det gælder både uautoriseret adgang og dine samarbejdspartneres adgang. Håndteres af Kulturarvsclusteret (SB).

**Hvorledes vil indsamlede data blive overført på sikker måde?**

**Hvis dine data er særligt følsomme kan du redegøre uddybende for alle sikkerhedsforanstaltninger og formelle standarder.**

### 5.3: Langtidsopbevaring

Ofte vil data skulle opbevares i længere tid. Det kan gælde alle data eller kun udvalgte datasæt. I det følgende har du mulighed for at beskrive dette nærmere.

		Procedurer for at skabe indeks, altså algoritmer skal bevares og deles.
<b>Hvilke data skal bevares og/eller deles?</b>	<i>Vurder hvilke forskningsdata, der har værdi efter projektets afslutning og som derfor bør, bevares. Overvej også gerne hvilke genanvendelsesmuligheder dine data har, både i forhold til at validere dine resultater, til at frembringe nye videnskabelige resultater og til brug i forbindelse med undervisning.</i>	Korpus metadata i CSV-format skal bevares og deles. Beskrivelsen af index-filen og beskrivelse af procedure for, hvordan man viser data skal bevares og deles.
<b>Hvilke data skal bevares eller slettes på grund af lovmæssige, kontraktmæssige eller andre lignende forhold?</b>		Statistiske JSON-søgereultater og beskrivelser skal bevares og deles. Indekser over korpus vil kunne indeholde personhenførbare informationer, og kan derfor ikke deles. Indekser skal i stedet bevares i et lukket repository.
<b>Hvor længe skal data bevares?</b>		10 år

Stilles der særlige krav om klargøring af data til langtidsbevaring?

Angiv kort.

#### 5.4: Datadeling

Du bør overveje betingelser for datadeling, både med hvem, internt og eksternt, i hvilket omfang og hvorledes anerkendelsen af dig som ophav til data kan sikres.

<b>Deling:</b>	<i>Vil du dele dine data?</i>  <i>Ja/nej.</i>	Ja. Nogle data kan deles, også eksternt.
<b>Er der restriktioner på deling af data?</b>	<i>Hvis "ja" til datadeling, hvilke problemer og hvilke løsninger ser du?</i> .	Ja, indekser kan ikke deles. De kan dog bevares, og der kan gives adgang til forskning ligesom for Netarkivet.
<b>Hvis der er et ønske om eneadgang, hvor længe har du behov for eneadgang til data og hvorfor?</b>		
<b>Eksisterer der en formel aftale om deling af data?</b>	<i>Ja/nej.</i>	Hvis nej: er der behov for en sådan aftale?
<b>Hvordan vil mulige brugere af dine data kunne finde og dele dem?</b>	<i>Vil du eksempelvis dele data gennem et repository, eller vil du selv stå for delingen?</i>	Deling af åbne data forventes at foregå gennem SB's Research Data Repository LOAR (Library Open Access Repository).  De data, som ikke er åbne forventes at kunne bevares i SB's repository LCAR (Library Controlled Access Repository), hvor forskere kan søge om adgang til data.
<b>Vil du søge at få en persistent identifikation (F.eks. DOI) til dine data?</b>		Ja, tildeles af repository.
<b>Delingsmetode</b>	<i>Hvordan vil andre modtage/kunne hente og evt. dele dine data?</i>	Data vil kunne downloades fra

**Datakontrol**

*Hvordan vil du kontrollere adgang til data?*

repository.

Adgang til data administreres af repository.

# Probes — The Cook Book

*Probing a Nation's Web Domain — the Historical Development of the Danish Web Domain  
Vers. 01, 160115*

This 'Cook book' describes the dishes that we as researchers would like to make, the ingredients we need, and the ways we must proceed to succeed. The Probes are organised in clusters, each with Probes that are related to each other in one way or another. By and large the columns reflect the phases in the research process. If necessary, each Probe can be described in more detail in a separate document. Legend: T=Technical, L=Legal, D=Data Management.

Probe cluster 1: Counting bytes and files  
Probe cluster 2: Response codes  
Probe cluster 3: Hyperlinks  
Probe cluster 4: Content and metatags

## Probe cluster 1: Counting bytes and files

Research interest	Operationalisation of RQ	Format of corpus	Analytical tool(s)	Visualisation	Dissemination	Storage
<b>Size of all web domains (number of objects, bytes):</b> <ul style="list-style-type: none"> <li>• average</li> <li>• top 100</li> <li>• intervals (1-10, 10-100, 100-1000...)</li> </ul>	T	<p>Preliminary idea: assessment of harvest reports for this purpose. Probably we will be using the crawl-log to establish objects belonging to the site. This is the hard part. The Discovery Path might be useful. The size can be calculated from the bytes field (field 3).</p>	<p>CSV-format. In general a structured format of the metadata information of the crawls.</p>	<p>[beskrivelse af, hvilke visualiseringsværktøjer, der skal bruges]</p>	<p>[beskrivelse af, hvordan resultater skal formidles]</p>	<p>[beskrivelse af bevaring efter projektophør]</p>
	L	Top 100 (domain names have to be screened)	ok			
	D					
<b>The number of specific file types</b> <b>Later: number of references to embedded video/images...</b>	T	<p>Field 7 holds the mime-type of the object. Cf. RQ1 for information on the aggregation function. References to embedded objects might be included in the crawl.log but it depends on the crawl specifics (e.g. the LinkScoper in the Heritrix Frontier) if and how they appear.</p>	<p>CSV-format. In general a structured format of the metadata information of the crawls.</p>	<p>[beskrivelse af, hvilke visualiseringsværktøjer, der skal bruges]</p>	<p>[beskrivelse af, hvordan resultater skal formidles]</p>	<p>[beskrivelse af bevaring efter projektophør]</p>
	L	ok	ok			
	D					



<b>How many bytes is the .dk domain?</b>	T	This follows from RQ1 & RQ2 but needs clarification, i.e. should embedded objects be included in the total?	CSV-format. In general a structured format of the metadata information of the crawls.	[beskrivelse af, hvilke analyseværktøjer, der skal bruges]	[beskrivelse af, hvilke visualiseringsværktøjer, der skal bruges]	[beskrivelse af, hvordan resultater skal formidles]	[beskrivelse af bevaring efter projektophør]
	L	ok	ok				
	D						
<b>What is the average size of a file?</b> <b>What is the average size of specific file types?</b>	T	Field 3, size in bytes, and field 7, mime-type, should allow us to produce statistics like this.	CSV-format. In general a structured format of the metadata information of the crawls.	[beskrivelse af, hvilke analyseværktøjer, der skal bruges]	[beskrivelse af, hvilke visualiseringsværktøjer, der skal bruges]	[beskrivelse af, hvordan resultater skal formidles]	[beskrivelse af bevaring efter projektophør]
	L	ok	ok				
	D						

## Probe cluster 2: Response codes

Research interest	Operationalisation of RQ	Format of corpus	Analytical tool(s)	Visualisation	Dissemination	Storage
<b>Number of in-active/not maintained web domains (404, 500, 503)</b>	T The report seeds-report.txt includes this information per domain. Assessment of the quality of the report is needed. The crawl.log includes the HTTP status code (field 2) and can easily discern in-active domains if the root document / returns 404 but in most cases a landing page will be present and some log	CSV-format. In general a structured format of the metadata information of the crawls.	[beskrivelse af, hvilke analyseværktøjer, der skal bruges]	[beskrivelse af, hvilke visualiseringsværktøjer, der skal bruges]	[beskrivelse af, hvordan resultater skal formidles]	[beskrivelse af bevaring efter projektophør]



### Probe cluster 3: Hyperlinks

Research interest	Operationalisation of RQ	Format of corpus	Analytical tool(s)	Visualisation	Dissemination	Storage
Connectedness to domains, number of links to: <ul style="list-style-type: none"> <li>• .dk</li> <li>• other ccTLDs</li> <li>• other TLDs</li> <li>• social media (FB, YT, Twitter...)</li> <li>• top 100 web domains outside .dk</li> </ul>	T	No idea at the moment! SOLR index to the rescue! :)	Either link extraction of the content in the Netarkiv or utilization of Net Search. The latter would require a definition of how to define such a corpus	[beskrivelse af, hvilke analyseværktøjer, der skal bruges]	[beskrivelse af, hvordan resultater skal formidles]	[beskrivelse af bevaring efter projektophør]
	L	ok (domain names have to be screened)	ok			
	D					

## Probe cluster 4: Content and metatags

Research interest	Operationalisation of RQ	Format of corpus	Analytical tool(s)	Visualisation	Dissemination	Storage
<b>Background colour, font type:</b> <ul style="list-style-type: none"> <li>• top twenty</li> </ul> <b>Page size (length/width in points):</b> <ul style="list-style-type: none"> <li>• average</li> </ul>	T This RQ requires us to analyse the content and even this is difficult since properties like colour can be hard-coded in the HTML or (modern web-era design) defined using CSS either in-line or linked. However, since we also would like the page size we need to render the document using tools like a headless browser.	The content from the Netarkiv. Corpus format/definition TBD	headless browser e.g. PhantomJS	[beskrivelse af, hvilke visualiseringsværktøjer, der skal bruges]	[beskrivelse af, hvordan resultater skal formidles]	[beskrivelse af bevaring efter projektophør]
	L ok	ok				
	D					
<b>Number of servers using specific server software (php, asp...)</b>	T Can be solved in many ways: looking at common file extensions used in the URI (php, asp, aspx, etc.) but this method is not viable since modern web sites do URL-rewriting; HTTP response headers will give a clearer picture but they are stored in the WARC files (per object).	Metadata from the content crawls. Corpus format/definition TBD	[beskrivelse af, hvilke analyseværktøjer, der skal bruges]	[beskrivelse af, hvilke visualiseringsværktøjer, der skal bruges]	[beskrivelse af, hvordan resultater skal formidles]	[beskrivelse af bevaring efter projektophør]
	L ok	ok				
	D					

<p><b>Number of web domains using specific software types (blogger.com, WikiWikiWeb, MediaWiki...)</b></p>	T	<p>Looking at paths in the URL (field 4) specific system paths can be discerned using regular expressions and this should cover default installations of the software, e.g. the wp-content folder signifies WordPress.</p>	<p>The easy: CSV-format. In general a structured format of the metadata information of the crawls. The detailed: Look at HTML metadata in the actual content from the Netarkiv. Format and definition of corpus TBD</p>	<p>[beskrivelse af, hvilke analyseværktøjer, der skal bruges]</p>	<p>[beskrivelse af, hvilke visualiseringsværktøjer, der skal bruges]</p>	<p>[beskrivelse af, hvordan resultater skal formidles]</p>	<p>[beskrivelse af bevaring efter projektophør]</p>
	L	ok	ok				
	D						

# Det Kgl. Biblioteks spørgeramme i forbindelse med at udarbejde datamanagement planer

## 1: BASALE OPLYSNINGER

### ADMIN DETAILS

**Project Name:**

**Principal Investigator / Researcher:**

**Institution:**

### OM PROJEKTET

Projekt navn:

Projekt-id:

Bevillingsgiver:

Bevillingsnummer:

Projektbeskrivelse:

Projektets varighed:

Primærforsker(e):

Primær kontaktperson:

Institutioner, bevillingsgivere mv:

### RELATEREDE POLITIKKER

Bygger datamanagementplanen på eksisterende standarder eller procedurer? Hvis ja, hvilke?

Har den institution, som projektet hører under, særlige regler for datasikkerhed? Hvis ja, hvilke (evt. Som links)?

Stiller samarbejdspartnere, bevillingsgivere eller andre krav til datamanagement? Hvis ja, hvilke (evt. Som links)?

### OMKOSTNINGER VED HÅNDTERING AF DATA

Omkostninger i forbindelse med håndteringen af dine forskningsdata

### OPFØLGNING PÅ DIN DATA MANAGEMENT PLAN

Dato for første version af projektets datamanagementplan.

Dato for seneste reviderede version af projektets datamanagementplan.

Dato for næste planlagte revision af projektets datamanagementplan.

Hvem er ansvarlig for implementering, opfølgning og revision af datamanagement planen?

Hvem er ansvarlig for de enkelte datamanagementaktiviteter?

### HAR DU DE FORNØDNE RESSOURCER TIL AT GENNEMFØRE DIN PLAN?

Har du brug for eksterne ressourcer eller bistand, f.eks. Hardware, software, særlig ekspertise?

Hvor meget koster det at bevare data på bevaringsstedet?

## 2: JURIDISKE OG ETISKE FORHOLD OMKRING DATA

### JURIDISKE OG ETISKE FORHOLD OMKRING DATA

Hvilken ophavsret gælder i forhold til dine data?

Hvem har ophavsret til eventuelle data, der eksisterer inden forskningsprojektet starter?

Hvem har ejerskab/ophavsret til data, der skabes under forskningsprojektet

Hvis andre har ophavsretten, hvordan kom projektet i besiddelse af disse data og hvilke vilkår gælder der for tilgang til disse data?

Hvem har ophavsret til data efter projektets afslutning?

Hvordan skal spørgsmål om ophavsret til nye data, herunder tilladelse til genanvendelse håndteres?

Hvordan må tredjepart – eventuelt - bruge dine data?

Indeholder dine data personfølsomme oplysninger?

Indeholder dine data, oplysninger, der skal bruges i forbindelse med patentansøgninger?

Hvem har adgang til data før, under og efter forskningsprojektet?

### ETISKE FORHOLD

Har du tilladelse til at gemme og dele data?

Hvordan vil du sikre anonymitet for f.eks. Deltagere i forskningsprojektet, svarpersoner o.l.?

Hvordan vil du opbevare personfølsomme oplysninger, sikkert?

## 3: OPLYSNINGER OM EKSISTERENDE DATA

### OPLYSNINGER OM EKSISTERENDE DATA

Samlingens navn

Indeholder samlingen personfølsomme data?

Indeholder samlingen oplysninger, der skal anvendes i forbindelse med en påtænkt patentansøgning?

Hvem har ophavsret til de eksisterende data?

Hvem kan få adgang til disse data? Og under hvilke betingelser?

Hvor befinder disse data sig?

Hvor store datamængder tænkes behandlet?

Hvilke filformater indgår i samlingen?

Hvilket operativsystem anvendes?

Hvilken software anvendes til at behandle data?

## 4: OPLYSNINGER OM FORSKNINGSDATA, DER SKABES OG BEHANDLES I LØBET AF PROJEKTET

### FORHOLDET MELLEML EKSISTERENDE DATA OG DE NYE FORSKNINGSDATA

Hvad er forholdet mellem eksisterende data og de nye forskningsdata?

Skal alle datasæt bevares?

### OPLYSNINGER OM DE DATA, DER SKABES I LØBET AF FORSKNINGSPROJEKTET

Hvor store mængder nye data er der tale om?

Hvilken software anvendes til at behandle data?

Overvejes udvikling af ny software?

### HVILKE FILFORMATER ANVENDES DER OG SKABES

Hvilke filformater tænkes de nye data at blive lagret i?

Hvilken software anvendes til at behandle data?

Er der specielle krav til software eller hardware

### ORGANISATION AF FORSKNINGSDATA UNDER PROJEKTET

Hvilke standarder og metoder vil du anvende?

Hvordan vil du strukturere og navngive dine mapper og filer?

Hvordan vil du kvalitetssikre dine data?

Hvordan vil du håndtere evt. Versionering (f.eks. Benytte versionsstyringsværktøj)?

**IKKE-DIGITALE DATA**

Hvor opbevares ikke digitale forskningsdata?  
Hvor lang tid vil ikke-digitale forskningsdata blive bevaret sikkert?

**DOKUMENTATION OG METADATA**

Hvilke informationer er nødvendige for at dine data kan læses og forstås i fremtiden?  
Hvordan vil du skabe den nødvendige dokumentation af metadata?  
Hvilke metadatastandarder vil du bruge og hvorfor?

**5: BACKUP, BEVARING OG DELING AF DATA****HVORDAN VIL DU OPBEVARE OG SIKRE BACKUP AF DATA UNDER DIT FORSKNINGSPROJEKT?**

Hvordan vil du foretage backup for at sikre mod datatab?  
Hvor er data backup placeret?  
Er lagerplads til backup tilstrækkelig?  
Er der sikret mulighed for genskabelse af data? Hvem er ansvarlig for dette?

**HVORDAN VIL DU HÅNDTERE ADGANG OG SIKKERHED?**

Er der foretaget risikoplanlægning i forhold til datasikkerhed? Hvis ja, beskriv kort.  
Hvorledes vil adgangskontrol blive sikret?  
Hvorledes vil indsamlede data blive overført på sikker måde?  
Hvis dine data er særligt følsomme kan du redegøre uddybende for alle sikkerhedsforanstaltninger og formelle standarder.

**LANGTIDSOPBEVARING**

Hvilke data skal bevares og/eller deles?  
Hvilke data skal bevares og/eller deles? Evt. Slettes hvilke data skal bevares eller slettes på grund af lovmæssige, kontraktmæssige eller andre lignende forhold?  
Hvor længe skal data bevares?  
Hvor skal data langtidsbevares?  
Stilles der særlige krav om klargøring af data til langtidsbevaring?

**DATADELING**

Vil du dele dine data?  
Er der restriktioner på deling af data?  
Hvor længe har du behov for ene adgang til data og hvorfor?  
Eksisterer der en formel aftale om deling af data?  
Kan det kgl. Bibliotek anvende resultaterne fra forskningsprojektet i andre sammenhænge?  
Hvornår vil du gøre data tilgængelige?  
Hvordan vil mulige brugere af dine data kunne finde og dele dem?  
Vil du søge at få en persistent identifikation (f.eks. Doi) til dine data?  
Hvordan vil du kontrollere adgang til data?



[Denne modelaftale kan danne udgangspunkt for forskningssamarbejder involverende datasamlinger og brug af IT-infrastruktur. Der er indsat kommentarer i aftalen markeret med gult. Aftalen skal altid gennemgås og evt. justeres i henhold til det konkrete projekt og gældende bevillingsvilkår og /eller særregler for pågældende forskning/data. Aftalens anvendelse sker på parternes eget ansvar.]

### Modelaftale om samarbejde om forskningsdata mellem

[Forskningsinstitution]

[CVR nr.]

[Adresse]

[postnummer by]

(herefter benævnt som "ZZ")

Og

[Forskningsinstitution]

[CVR nr.]

[Adresse]

[postnummer by]

(herefter benævnt som "XX")

Og

[Forskningsinstitution]

[CVR nr.]

[Adresse]

[postnummer by]

(herefter benævnt som "ZZ")

[I det omfang, der indgår virksomheder/kommercielle parter i samarbejdet skal aftalen eventuelt tilpasses, herunder eksempelvis i kommercielle adgangsrettigheder til resultater]

Ovennævnte er i denne aftale i fællesskab benævnt "Parterne" og hver for sig benævnt "Part".

## Indhold

1.	Definitioner .....	3
2.	Aftalens baggrund og formål .....	4
3.	Økonomi.....	4
4.	Organisering.....	4
5.	Ledelsesgruppen.....	5
6.	Parternes bidrag og ansvarsområder .....	6
7.	Forpligtelser vedrørende data i Projektet .....	7
8.	Ejendoms- og/eller ophavsrettigheder .....	7
9.	Adgangsrettigheder til Projektets gennemførelse .....	7
10.	Adgangsrettigheder efter Projektets gennemførelse.....	8
11.	Hemmeligholdelse.....	8
12.	Publikation.....	9
13.	Brug af ekstern it-infrastruktur .....	10
14.	Ikrafttræden og ophør.....	10
15.	Misligholdelse.....	10
16.	Erstatningsansvar.....	11
17.	Indtræden og udtræden .....	12
18.	Tvister.....	13
19.	Bilag.....	13

## 1. Definitioner

- 1.1. "Aftalen": Denne aftale om samfinansieret forskning med tilhørende bilag.
- 1.2. "Baggrundsviden": Viden i form af Data, knowhow eller anden særlig viden, uanset om denne er beskyttet eller kan beskyttes og alle immaterielle rettigheder knyttet hertil, som den pågældende Part ønsker at selv at anvende eller som stilles til rådighed for de øvrige Parter med henblik på udførelse af Projektet.

Baggrundsviden omfatter blandt andet følgende:

### [Oplisting af specifik Baggrundsviden]

- 1.3. "Bevillingstilsagn": Skrivelsen fra Bevillingsgiver af [dato] jf. bilag 3. Enhver reference til Bevillingstilsagnet inkluderer en reference til dens bilag og de dertilhørende bevillingsvilkår.
- 1.4. "Bevillingsgiver": [Navn på bevillingsgiver]
- 1.5. "Code of Conduct": Danish Code of Conduct<sup>1</sup> udgivet af Uddannelses- og forskningsministeriet, November 2014.
- 1.6. "Data": Enhver form for data herunder også fysisk og/eller digital data.
- 1.7. "FAIR-principperne": som vedtaget anvendt for EU's Horizon 2020 projekter<sup>2</sup> (Findable, Accessible, Interoperable and Reusable).
- 1.8. "Forgrundsviden": Al den viden, inklusiv enhver form for Data og resultater uanset form og uanset om det er beskyttet eller kan beskyttes og alle immaterielle rettigheder knyttet hertil, som hidrører fra og er skabt under samarbejdet om Projektet.

Forgrundsviden omfatter blandt andet følgende:

### [Oplisting af specifik Forgrundsviden]

- 1.9. "Fortrolig Information": Baggrundsviden og Forgrundsviden som er tydeligt markeret som værende fortroligt eller som i øvrigt er af åbenbar fortrolig karakter og ikke omfattet af punkt 11.3
- 1.10. "Infrastrukturudbyder": [Navn på infrastrukturudbyder]
- 1.11. "Ledelsesgruppe": Den til enhver tid udpegede gruppe af personer med overordnet ansvar for Projektet som anført i punkt 5.4.
- 1.12. "Projektet": Det mellem Parterne under denne Aftale aftalte Projekt beskrevet i bilag 1.
- 1.13. "Projektleder": Den til enhver tid udpegede person udpeget til at varetage den daglige ledelse af Projektet som anført i punkt 4.3.

---

<sup>1</sup> Jf. publikationer på UFM.dk. Se link: <http://ufm.dk/publikationer/2014/filer-2014/the-danish-code-of-conduct-for-research-integrity.pdf>

<sup>2</sup> Jf. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

## 2. Aftalens baggrund og formål

Denne Aftale har til formål at regulere Parternes indbyrdes rettigheder og forpligtelser i forbindelse med samarbejde om indsamling, bearbejdning, publicering, anvendelse af data og IT-infrastruktur i forbindelse med Projektet. Parterne har interesse i gennem forskning at udvide deres viden inden for [indsæt fagområder]. Parterne har på denne baggrund en fælles interesse i at samarbejde om Projektet.

## 3. Økonomi

- 3.1. Projektet er samfinansieret mellem Parterne [idet Parterne ligeledes har opnået finansielt tilskud fra Bevillingsgiver]. Parterne har sammen udarbejdet et budget for Projektet, jf. bilag 2 til Aftalen. Budgettet er udarbejdet i henhold til Finansministeriets budgetvejledning. Parternes kontante eller in-kind bidrag til Projektet fremgår af bilag 3. [Bevillingsgiver har givet tilsagn om medfinansiering, som angivet i Bevillingstilsagnet, jf. bilag 3 Alle Parterne anerkender og accepterer at være bundet af bestemmelserne i Bevillingstilsagnet.]
- 3.2. Såfremt det mellem Parterne aftalte budget ikke kan overholdes, og det ikke skyldes fejl eller forsømmelser fra en af Parternes side, foretager Parterne i fællesskab en fornyet vurdering af Projektet og omkostningerne til Projektets færdiggørelse. Parterne træffer i Ledelsesgruppen herefter beslutning om, hvorvidt Projektet skal føres til ende med en yderligere arbejdsindsats, et højere kontantbidrag m.v. eller indstilles på det foreliggende grundlag.
- 3.3. [[Part YY] er bemyndiget til at modtage bevillingsmidler fra Bevillingsgiver, og ansvarlig for at distribuere bevillingsmidler, der er udbetalt i overensstemmelse med Bevillingstilsagnet, til de respektive Parter i overensstemmelse med denne Aftale. YY er i intet tilfælde forpligtet til at betale beløb, der ikke er modtaget fra Bevillingsgiver. Parterne skal fremsende faktura indeholdende beskrivelse af og dokumentation for afholdte udgifter samt indbetalingsoplysninger til YY. Betaling af afholdte udgifter sker bagudrettet og efter, at Bevillingsgiver har godkendt og betalt de pågældende udgifter.]

## 4. Organisering

- 4.1. Projektet er et forpligtende samarbejde mellem uafhængige Parter. Projektet er organiseret med en forskningsgruppe, en Projektleder og en Ledelsesgruppe. Parterne har udpeget NN som Projektleder, som er ansvarlig for den daglige ledelse af Projektet. I relation til Projektet vil Projektlederen fungere som kontaktperson for Infrastrukturudbyderen [og Bevillingsgiver].
- 4.2. Bortset fra Projektledere og Ledelsesgruppen skal Parterne være berettiget til at indsætte andre medarbejdere end nævnt i bilag 1, såfremt dette kan ske uden ulempe for Projektets gennemførelse og uden overskridelse af fastsatte tidsfrister. Parterne skal tidligst muligt informere hinanden via Ledelsesgruppen om de påtænkte udskiftninger.

- 4.3. Projektlederen varetager den daglige ledelse af Projektet. Projektlederen er herunder ansvarlig for
- [Kontakten til Bevillingsgiver, herunder indsendelse af faglige/evalueringer rapporter samt regnskaber].
  - Kontakten til Infrastrukturudbyderen, herunder ansvarlig for aftaleindgåelse
  - [Fordeling af modtagne betalinger fra Bevillingsgiver]
  - [Opkrævning af Parternes kontante medfinansiering jfr. budgettet,]
  - Indkaldelse til og forberedelse af møder i Ledelsesgruppen
  - [Anmeldelser til Datatilsynet, samtykkeerklæringer [godkendelser fra Videnskabsetisk Komité efter delegation fra Ledelsesgruppen]
  - Udarbejdelse og løbende ajourføring af data management plan for Projektet
  - Planlægning og budgettering for evt. brug af repository.
  - [Andet]
- 4.4. Projektlederen holder Ledelsesgruppen orienteret om Projektets fremgang og status. Orienteringen foregår på møderne i Ledelsesgruppen.
- 4.5. Projektlederen skal sørge for at sikre sig et mandat fra Parterne til indgåelse af aftaler med tredjemand, der også forpligter eller i øvrigt virker bebyrdende for de øvrige Parter. Projektlederen skal derfor forelægge endeligt udkast til sådanne aftaler for Ledelsesgruppen til godkendelse. [Parterne er ved Aftalens indgåelse enige om, at YY indgår aftale med Infrastrukturudbyderen og de øvrige Parter bekræfter herved at være ansvarlig for denne Infrastrukturaftale og de til enhver tid gældende generelle vilkår med senere ændringer.]

## 5. Ledelsesgruppen

- 5.1. Ledelsesgruppen har det overordnede ansvar for projektledelse, økonomistyring, monitorering og opfølgning på projektplanen for at sikre Projektets fremdrift.
- 5.2. Ledelsesgruppen er herudover ansvarlig for
- Forudgående godkendelse af anmeldelser til Datatilsynet og andre pligtige anmeldelser, tilladelser, godkendelse, samtykkeerklæringer m.v.
  - Optagelse af nye parter i Projektet og betingelserne herfor
  - Udarbejdelse af indstillinger til Parterne om en Parts udtræden af Projektet
  - Udarbejdelse af retningslinjer og tidsfrister for Parternes bidrag til rapporter i henhold til Bevillingsgivers krav
  - Godkendelse af faglige rapportertil indsendelse af Projektlederen til Bevillingsgiver
  - Godkendelse af data management plan og senere opdateringer.
- 5.3. Ledelsesgruppen træffer alle væsentlige beslutninger, som ikke er henlagt til Parterne eller Projektlederen.
- 5.4. Ledelsesgruppen har følgende medlemmer:
- NN, repræsentant fra YY
  - NN, repræsentant fra XX
  - NN, repræsentant fra ZZ
- 5.5. NN er formand for Ledelsesgruppen og leder møderne. Formanden kan beslutte at invitere andre personer til at deltage i møderne uden stemmeret. Projektlederen

deltager i møderne uden stemmeret. Infrastrukturudbyderen kan deltage i møderne uden stemmeret.

- 5.6. Medlemmerne forventes at deltage i møderne. Ledelsesgruppen er beslutningsdygtig, når alle medlemmer er til stede.
- 5.7. Parterne skal af hensyn til kontinuiteten og kvaliteten i Projektet i videst muligt omfang undgå at udskifte medlemmerne. Er en Part undtagelsesvis nødsaget til at udskifte et medlem, skal Parten sikre at det nye medlem mindst har samme kvalifikationer som det udtrådte medlem. Udskiftning af medlemmer skal uden ugrundet ophold meddeles formanden.
- 5.8. Ledelsesgruppen afholder møde efter behov, dog mindst én gang årligt. Formanden indkalder til møderne med mindst 15 kalenderdages varsel og påser, at dagsorden og alt nødvendigt materiale fremsendes inden samme frist. Ekstraordinære møder afholdes, når formanden finder det fornødent, eller mindst en tredjedel af medlemmerne forlanger det. Formanden fastsætter tid og sted for de ekstraordinære møders afholdelse.
- 5.9. Formanden har ansvaret for at udarbejde et beslutningsreferat fra møderne. Referatet sendes til medlemmerne. Hvis der ikke er modtaget indsigelser eller ændringsforslag til referatet inden for 15 kalenderdage, anses referatet for at være en korrekt gengivelse af de beslutninger, der blev truffet på mødet. I tilfælde af indsigelser eller ændringsforslag skal en ny version af referatet så hurtigt som muligt sendes til Ledelsesgruppens medlemmer. Herefter gælder den samme 15 kalenderdages ret til indsigelser og ændringsforslag. Beslutningsreferatet skal dog godkendes senest på det efterfølgende møde. Når beslutningsreferatet er godkendt, sendes det til medlemmerne.
- 5.10. Hvert medlem af Ledelsesgruppen har én stemme. Ledelsesgruppen træffer sine beslutninger ved afstemning med kvalificeret stemmeflertal på 2/3 flertal, idet beslutninger i enighed tilstræbes. Alle parter har veto angående beslutninger, som påvirker partens økonomi eller forpligtelser.
- 5.11. Punkter til beslutning skal fremgå af den udsendte dagsorden før mødet - medmindre der er enstemmighed om beslutningen på det pågældende møde, og forudsat at alle medlemmerne er til stede eller repræsenteret af stedfortræder med skriftlig fuldmagt.
- 5.12. Formanden kan beslutte, at møder afholdes som video- eller telekonference, hvis der foreligger særlige omstændigheder.

## **6. Parternes bidrag og ansvarsområder**

- 6.1. Uagtet at Projektlederen har det daglige ansvar for Projektet, jf. punkt 4.3, er Parterne enige om følgende arbejdsfordeling af opgaver udover forskningsaktiviteter [/ALTERNATIVT arbejdsfordeling som beskrevet i Projektbeskrivelsen]:
  - NN fra XX's bidrag er følgende: [Er dataansvarlige for Projektet, og dermed ansvarlig for anmeldelser til Datatilsynet og for indgåelse af nødvendige databehandleraftaler]
  - NN fra YY's bidrag er følgende: []

## 7. Forpligtelser vedrørende data i Projektet

- 7.1. Alle Parter forpligter sig til i fornødent omfang at instruere egne og andre parter ansatte involveret i Projektet i gældende ret for datahåndtering – herunder ophavsretsloven og Code of Conduct.
- 7.2. Hver Part forpligter sig til at sikre, at partens databehandling sker i overensstemmelse med de til enhver tid gældende relevante lovregler, forordninger og principper, herunder vejledninger vedrørende behandling af persondata. Såfremt en Part ønsker at overlade persondata til behandling hos en eller flere Parter i Projektet, er vedkommende Dataansvarlig for disse data og den/de øvrige part(er) er databehandlere. Overladelse af persondata kan ikke ske, før indgåelse af separat databehandleraftale parterne mellem dataansvarlig og databehandler(e).
- 7.3. Alle Parter forpligter sig til at behandle Baggrundsviden og Forgrundsviden som er fysiske, biologiske og/eller humant materiale i overensstemmelse med gældende regler og retningslinjer, herunder [eksempelvis regler om nano-matriale, nagoya-protokollen, gmo-materiale, persondataloven (biobankregler)].

## 8. Ejendoms- og/eller ophavsrettigheder

- 8.1. Aftalen medfører ikke, at der sker en overgang af rettigheder til Baggrundsviden, medmindre dette konkret er aftalt og fremgår af bilag 4. Der henvises i øvrigt til en uddybning i Bilag 4.
- 8.2. Fysiske materialer genereret eller indsamlet under Projektet tilhører den/de Part(er), som genererede og/eller indsamlede den i forbindelse med Projektet
- 8.3. Øvrig Forgrundsviden tilhører den/de Part(er), som genererede den i forbindelse med Projektet. Forgrundsviden, som skabes af Parterne i fællesskab, ejes af Parterne i fællesskab med forholdsmæssige andele (pro rata) baseret på deres intellektuelle bidrag til den udviklede Forgrundsviden. Hvis flere Parter har genereret Forgrundsviden, og det ikke er muligt med rimelighed at fastslå deres respektive andele deraf, fordeles ejerskabet af denne Forgrundsviden ligeligt mellem Parterne med lige andele. Parterne skal indgå en aftale, som fastlægger de nærmere bestemmelser vedrørende håndtering af Forgrundsviden, som ejes af Parterne i fællesskab, herunder i forbindelse med direkte udnyttelse, patentering og anden beskyttelse deraf i overensstemmelse med ophavs- og ansættelsesretlig samt videnskabelig sædvane.
- 8.4. Forfatterskab reguleres af principperne fastlagt i "the Danish Code of Conduct for Research Integrity"<sup>3</sup> og hver Parts ansatte opnår ophavsretten til vedkommendes værker jfr. punkt 12.

## 9. Adgangsrettigheder til Projektets gennemførelse

- 9.1. Parterne giver i Projektets løbetid hinanden vederlagsfri adgang i videst muligt omfang til at benytte deres respektive Baggrundsviden, herunder datasamlinger, og Forgrundsviden med henblik på udførelse af Projektet. Såfremt en Part giver lovlig adgang til sin Baggrundsviden til en eller flere af de øvrige Parter, skal den

---

<sup>3</sup>Jf. publikationer på UFM.dk. Se link: <http://ufm.dk/publikationer/2014/filer-2014/the-danish-code-of-conduct-for-research-integrity.pdf>

afgivende Part loyalt informere om alle juridiske begrænsninger, som Parten er eller bør være bekendt med, for de øvrige Parter adgang til og brug af Baggrundsviden under Projektets gennemførelse og medtage disse oplysninger i Bilag 4 til denne Aftale. Adgangsretten er udelukkende knyttet til arbejdet i forbindelse med Projektet og må ikke udnyttes kommercielt eller overdrages til tredjemand.

- 9.2. Uanset punkt 9,1, kan der såfremt en Part inddrager Baggrundsviden belagt med tredjeparts rettigheder eller forpligtelser og/eller myndighedskrav, kun gives adgang til de øvrige Parter i det opfang tredjeparts rettigheder og/eller myndighedskrav giver mulighed herfor.
- 9.3. Hver Part er selv eneansvarlig for at sikre, at den har de licenser i forhold til immaterielle rettigheder tilhørende tredjemænd og/eller nødvendige myndighedsgodkendelser, som måtte være nødvendige for at Parten kan få adgang til eller bruge pågældende Forgrundsviden og/eller Baggrundsviden i eller efter Projektet, uanset om sådan Forgrundsviden ejes af Parten selv, eller den har fået adgangsrettigheder dertil i henhold til denne Aftale.

## 10. Adgangsrettigheder efter Projektets gennemførelse

- 10.1. Parterne stiller deres respektive Forgrundsviden, der ikke er software, patenterbare opfindelser eller fysiske materialer, åbent til rådighed for de øvrige projektparter efter Projektets afslutning medmindre konkrete lovgivningsmæssige eller kontraktmæssige forpligtelser eller offentligretlige forskrifter forhindrer det. Dog er hver Part altid ansvarlig for egen brug af Forgrundsviden, jf. punkt 16.3.
- 10.2. Forgrundsviden stilles til rådighed for offentligheden efter FAIR principper om Open Access i H2020<sup>4</sup>, dog først efter udløbet af en periode på 8 måneder fra Projektets afslutning hvorunder Parterne har mulighed for at publicere egen Forgrundsviden. I løbet af de 8 måneder har hver Part adgang til intern brug af de øvrige Parter ikke-publicerede Forgrundsviden.
- 10.3. [NB: Hvis der indgår opfindelser og/eller EDB-programmer, skal aftalen indeholde en særskilt regulering af eksklusiv kontra ikke-eksklusiv kommerciel brug i henhold til lov om opfindelser ved offentlige forskningsinstitutioner og ophavsretslovens § 59]

## 11. Hemmeligholdelse

- 11.1. Fortrolig Information, som en Part modtager fra en anden Part i forbindelse med Projektet, må alene udnyttes af førstnævnte Part inden for rammerne af Projektets formål og kun i overensstemmelse med denne Aftale.
- 11.2. Parterne er forpligtede til at hemmeligholde al Fortrolig Information modtaget i forbindelse med Projektet, således at ingen Fortrolig Information hverken helt eller delvist, direkte eller indirekte på noget som helst tidspunkt kommer eller er i fare for at komme til tredjemands kundskab. Parterne skal sikre, at personer, som involveres i Projektet, pålægges samme hemmeligholdelsesforpligtelse som Parterne selv.

---

<sup>4</sup> Jf. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)



- 11.3. Hemmeligholdelsespligten omfatter ikke Fortrolig Informationer, som
- på tidspunktet for Partens modtagelse var lovligt offentliggjort eller på anden måde var lovligt tilgængelige for offentligheden;
  - efter Partens modtagelse er blevet lovligt offentliggjort eller er blevet tilgængelige for offentligheden på anden lovlig måde, uden at dette skyldes den modtagende Parts tilsidesættelse af hemmeligholdelsesforpligtelser under denne Aftale;
  - allerede på modtagelsestidspunktet var retmæssigt i den modtagende Parts besiddelse uden hemmeligholdelsesrestriktioner af nogen art;
  - er modtaget fra en tredjemand, som bevisligt fremstod som berettiget til at videregive informationen;
  - er udviklet uafhængigt af Projektet af den modtagne Part;
  - er eller bliver omfattet af love eller bekendtgørelser, offentligretlige afgørelser, domme, kendelser m.v., som pålægger den modtagende Part at videregive informationen helt eller delvist.
- 11.4. I tilfælde af uenighed mellem Parterne påhviler det den modtagende Part at bevise, at modtaget materiale fuldt ud er omfattet af punkt 11.3 i denne Aftale.
- 11.5. Hemmeligholdelsespligten ophører tre (3) år efter, at Projektet er ophørt.
- 11.6. Uanset (i) at information omfattes af undtagelsesbestemmelserne i punkt 11.3 i denne Aftale, eller (ii) at hemmeligholdelsesperioden er udløbet, jf. punkt 5 i denne Aftale, er den modtagende Part ikke berettiget til at gøre brug af det modtagne materiale i strid med andre (herunder immaterielle) regler og/eller aftaler mellem Parterne.

## 12. Publikation

- 12.1. Hver Part er berettiget til at offentliggøre pågældende Parts Forgrundsviden, såfremt en af Partens ansatte vil kunne anses som forfatter jfr. principper i the Danish Code of Conduct for Research Integrity<sup>5</sup>. Forgrundsviden, som Parterne ejer i fællesskab, kan offentliggøres af den ene Part, såfremt den anden Part ikke ønsker at deltage i offentliggørelsen.
- 12.2. I forbindelse med publikation skal hver Part gøre sin publikation tilgængelig i overensstemmelse med pågældende Parts retningslinjer, eksempelvis på Partens egne portaler og/eller Open Access [samt Bevillingsgivers eventuelle krav om Open Access og/eller Open Science].
- 12.3. Den Part, der ønsker offentliggørelse af Forgrundsviden genereret gennem projektarbejdet som nævnt i punkt 12.1, skal mindst 30 dage før det påtænkte tidspunkt for offentliggørelsen orientere de andre Parter herom med fremsendelse af den tekst og eventuelt yderligere materiale, der ønskes offentliggjort. Indtil 30 dage efter modtagelsen af orienteringen kan hver af de modtagende Parter kræve, at offentliggørelsen udskydes i op til 3 måneder fra modtagelsen, såfremt den pågældende Part godtgør, at udskydelsen har betydning for Partens mulighed for at opnå immaterialretlig beskyttelse af den viden, der ønskes offentliggjort.

---

<sup>5</sup> Jf. publikationer på UFM.dk. Se link: <http://ufm.dk/publikationer/2014/filer-2014/the-danish-code-of-conduct-for-research-integrity.pdf>

- 12.4. Offentliggørelse af viden skal altid ske med respekt af pligten til hemmeligholdelse i punkt 11.

### **13. Brug af ekstern it-infrastruktur**

- 13.1. I forbindelse med projektet er der Data, som skal behandles via en ekstern it-infrastruktur. Det drejer sig om de Data, som er beskrevet i bilag 4. Ansvar for databehandlingen via ekstern it-infrastruktur påhviler den Part, der er ansvarlig for brug af de pågældende Data jf. bilag 4.
- 13.2. Databehandlingen sker i henhold til kontrakt, herunder de til enhver tid gældende vilkår, som er vedlagt i bilag 5.

### **14. Ikrafttræden og ophør**

- 14.1. Nærværende aftale træder i kraft ved Parternes underskrift heraf og vedvarer indtil Projektet afsluttes i overensstemmelse med det i tidsplanen fastlagte.
- 14.2. Uanset aftalens ophør gælder de punkter fortsat, som udtrykkeligt efter deres indhold forudsættes at gælde efter Aftalens ophør.

### **15. Misligholdelse**

- 15.1. I tilfælde af at en Part væsentligt eller gentagne gange misligholder sine forpligtelser efter Aftalen og det forhold, der konstituerer misligholdelsen, ikke er blevet bragt til ophør inden for [x] dage efter opfordring hertil fra Ledelsesgruppen, kan Ledelsesgruppen indstille til Parterne at hæve Aftalen over for den misligholdende Part. Beslutning om at hæve Aftalen over for en misligholdende Part kræver enstemmighed mellem de andre Parter.
- 15.2. Såfremt en Part hindres i at opfylde sine forpligtelser efter Aftalen som følge af udefrakommende, ekstraordinære forhold, som Parten ved Aftalens indgåelse ikke burde have forudset (force majeure), anses dette ikke for misligholdelse, men de andre Parter skal i sådanne tilfælde være berettiget til at opsige Aftalen over for den Part, der hindres i at opfylde sine forpligtelser, såfremt hindringen medfører væsentlig forsinkelse af Projektets gennemførelse. En forsinkelse på mere end [x] måneder i forhold til den af Parterne aftalte tidsplan, jf. bilag 1 anses altid for væsentlig.
- 15.3. Såfremt Aftalen hæves over for en misligholdende Part, kan de andre Parter kræve det af misligholdelsen følgende tab erstattet efter bestemmelserne i punkt 16.
- 15.4. Hæves Aftalen over for en misligholdende Part, skal hver enkelt af de øvrige Parter for hvem den misligholdendes deltagelse i Projektet er af væsentlig betydning være berettiget til at udtræde af Aftalen. Samme ret tilkommer en Part, der har stemt for at hæve Aftalen over for en misligholdende Part, såfremt den i 15.1 krævede enstemmighed bag en beslutning om at hæve Aftalen ikke har kunnet opnås.

## 16. Erstatningsansvar

- 16.1. Parterne giver ingen indeståelse og kan ikke gøres ansvarlige for, at deres ydelser til gennemførelse af Projektet fører til et bestemt resultat.
- 16.2. Parterne skal udføre deres arbejde til Projektets gennemførelse efter bedste evne og under overholdelse af god skik for videnskabeligt arbejde. En Part er erstatningspligtig for sin groft uagtsomme eller forsætlige tilsidesættelse af pligter efter Aftalen.
- 16.3. En Part har i enhver henseende det fulde ansvar for sin egen brug af en anden Parts Forgrundsviden, Baggrundsviden og/eller Fortrolig Information modtaget af førstnævnte Part i forbindelse med Projektet. Den modtagende Part kan ikke på nogen måde gøre krav gældende mod den anden Part baseret på den modtagende Parts egen brug heraf, dog med forbehold for den afgivende Parts iagttagelse af sin loyale oplysningspligt i bilag 4 vedrørende restriktioner på Baggrundsviden, jf. også punkterne 9.1 og 9.2. Denne ansvarsbegrænsning omfatter blandt andet, men ikke udelukkende, manglende funktionalitet, ansvar for personskade eller skade på ejendom eller krænkelse af tredjemands immaterielle rettigheder.
- 16.4. Parterne er i øvrigt erstatningsansvarlige for egne medarbejderes skadegørende handlinger og undladelser og for farlige egenskaber ved Parternes ydelse (produktansvar) efter dansk rets almindelige regler.
- 16.5. Ingen af Parterne er erstatningsansvarlig for manglende opfyldelse af sine forpligtelser i henhold til Aftalen, såfremt den manglende opfyldelse skyldes force majeure som angivet i punkt 15.2.
- 16.6. Parternes erstatningspligt i henhold til punkt 16.2 er undergivet de begrænsninger, der fremgår af punkter 16.7 [og 16.8].
- 16.7. Bortset fra tilsidesættelse af pligten til hemmeligholdelse, jf. punkt 11, omfatter en Parts erstatningspligt over for den anden Part ikke tab ved følgeskader så som produktionsforstyrrelser og andet driftstab, mistet omsætning / avance eller anden indirekte skade.
- 16.8. [Parternes erstatningspligt er maksimeret til DKK [beløb] [En Parts erstatningspligt overfor de øvrige Parter er samlet begrænset til det mindste beløb af én gang den pågældende Parts andel af de samlede omkostninger til Projektet eller DKK [beløb].]
- 16.9. [I tilfælde af at Bevillingsgiver har rejst et krav, inklusiv et krav om tilbagebetaling af udbetalte bevillingstilskud, mod YY som bevillingsansvarlig Part, der skyldes enten misligholdelse eller et andet forhold, jf. Bevillingstilsagnet i bilag 2, som YY ikke er ansvarlig for, skal YY være berettiget til at kræve kompensation (regres) for ethvert tab fra den/de misligholdende/ansvarlige Part(er). Hvis ingen af Parterne er ansvarlige for Bevillingsgivers krav, skal YY være berettiget til at kræve forholdsmæssig kompensation for hver af de øvrige Parter. Ansvarsbegrænsningerne i punkt 16. 7 og 16.8 finder ikke anvendelse på Parternes ansvar som følge af Bevillingsgivers krav i henhold til dette punkt.]
- 16.10. [I tilfælde af at Bevillingsgiver reducerer bevillingstilskuddet som følge af misligholdelse eller øvrige forhold, jf. Bevillingstilsagnet i bilag 3., som YY ikke er ansvarlig for, skal YY være berettiget til at regulere udbetalingen til den/de misligholdende/ansvarlige Parter tilsvarende eller kræve forholdsmæssig betaling fra de pågældende Parter i overensstemmelse med reduktionen. Hvis ingen af

Parterne er ansvarlige for reduktionen, vil YY være berettiget til at regulere udbetalingen forholdsmæssigt eller kræve forholdsmæssig tilbagebetaling for hver af de øvrige Parter. Ansvarsbegrænsningerne i punkt 16. 7 og 16.8 finder ikke anvendelse på Parternes ansvar som følge af Bevillingsgivers krav i henhold til dette punkt.]

- 16.11. [YY skal, i tillæg til kompensation efter punkter 16.9 og 16.10, være berettiget til at kræve erstatning fra den/de misligholdende Part(er) for yderligere tab i forbindelse med misligholdelsen, jf. punkterne 16.1-16.7 [og 16.8].]

## 17. Indtræden og udtræden

- 17.1. I tilfælde af at en Part udtræder af projektsamarbejdet ved misligholdelse, eller hvor der i øvrigt måtte opstå behov herfor, kan de andre Parter gennem enstemmig beslutning på baggrund af Ledelsesgruppens indstilling jfr. punkt 5.2 lade en ny Part indtræde i Projektet. En ny Part skal underskrive denne Aftale, og alle Parter skal godkende de ændringer af bilagene der må foretages som følge af den ny Parts indtræden.
- 17.2. Viden, der af de oprindelige Parter i Aftalen er genereret efter Projektets start, men før en ny Parts indtræden, er i forhold til den nye Part omfattet af denne Aftales regler om Baggrundsviden.
- 17.3. Hvis særlige forhold bevirker, at en Part ikke kan fortsætte som deltager i Projektet, skal de andre Parter via Ledelsesgruppen [samt Bevillingsgiver] straks informeres herom. Informationen skal indeholde en nærmere redegørelse for årsagerne til, at Parten må udtræde af Projektet før dets afslutning, jf. punkt 14, og varsle tidspunktet for Partens udtræden, som i videst mulig udstrækning bør fastsættes under hensyntagen til de andre Parter's interesser og Projektets gennemførelse.
- 17.4. De andre Parter forventes at ville give samtykke til en Parts udtræden før Projektets afslutning, såfremt de af Parten angivne grunde findes at dokumentere det urealistiske i Partens fortsættelse som deltager i Projektet. Enhver af de andre Parter kan dog gøre sit samtykke betinget af, at Partens udtræden ikke fører til en til Projektet meddelt Bevillingstilsagns bortfald eller nedsættelse, ligesom samtykket kan gøres betinget af den udtrædende Parts opfyldelse af bestemte vilkår, f.eks. en udskydelse af tidspunktet for Partens udtræden, fortsat brug i Projektets løbetid af den udtrædende Parts Baggrundsviden og Forgrundsviden eller den udtrædende Parts betaling af en kompensation for de ekstra omkostninger, som påføres Projektet ved Partens udtræden før dets afslutning.
- 17.5. Såfremt en Part udtræder af Projektet, skal Ledelsesgruppen afgive en indstilling til de tilbageværende Parter om grundlaget for en aftale, hvorved enten den udtrædende Parts skyldige bidrag til Projektet overtages af en eller flere af Parterne eller af en ny Part, eller projektbeskrivelsen ændres, alt med henblik på i videst mulige udstrækning at fastholde de opstillede mål for Projektet.
- 17.6. [Såfremt en Parts ansatte ophører med sin deltagelse af Projektet, uden at Parten udtræder af Projektet, er den ansatte berettiget til at modtage en kopi af de data vedkommende har genereret under Projektet. Den ansatte kan under iagttagelse af nærværende Aftales adgangsrettigheder jfr. afsnit (adgangsret til forgrundsviden) anvende data til forskningsformål efter Parternes publikation af Projektets resultater baseret på pågældende data, dog tidligst 8 måneder efter Projektet gennemførelse.]

## 18. Tvister

- 18.1. Tvister mellem Parterne i forbindelse med samarbejdet skal søges afgjort i mindelighed. Såfremt dette ikke er muligt, skal tvisten afgøres ved de almindelige danske domstole.
- 18.2. Nærværende Aftale skal fortolkes og anvendes i henhold til dansk ret.

## 19. Bilag

Denne Aftale omfatter dette dokument og vedhæftede bilag.

Bilag 1: Projektbeskrivelse og tidsplan

Bilag 2: Budget

[Bilag 3: Bevillingstilsagn]

Bilag 4: Baggrundsviden med oplysning om vilkår for brug

Bilag 5: Aftale med udbyder af it-infrastruktur, herunder databehandleraftale

## 20. Underskrifter

**For ZZ,  
Forskningsinstitution:**

Dato:

---

Navn:

Titel:

**For ZZ,  
Forskningsinstitution:**

Dato:

---

Navn:

Titel:

**Bilag 4**

Parterne har bragt nedenstående baggrundsviden ind i projektet. I skemaet er oplysninger om rettigheder, betingelser, adgangs- og brugsmuligheder. Parterne indestår for nedenstående oplysninger.

Vedrørende part xx baggrundsviden

Hvilken baggrundsviden drejer det sig om	Hvem har ophavs- og ejendomsrettigheder til baggrundsviden.	Har øvrige parter adgang til baggrundsviden i projektet og i givet fald på hvilke vilkår?	Har øvrige parter adgang til baggrundsviden efter projektet og i givet fald på hvilke vilkår ?
[fx Data fra tidligere projekt]	[Ophavsrettighederne tilhører ZZ]	[Ja]	
[fx Kopi af data fra Statsbiblioteket]	[Ophavsrettigheder tilhører tredjemand]	[Nej – det er en personlig adgang]	

Vedrørende part yy baggrundsviden

Hvilken baggrundsviden drejer det sig om	Hvem har ophavs- og ejendomsrettigheder til baggrundsviden.	Har øvrige parter adgang til baggrundsviden i projektet og i givet fald på hvilke vilkår?	Har øvrige parter adgang til baggrundsviden efter projektet og i givet fald på hvilke vilkår ?
[fx Data fra tidligere projekt]	[Ophavsrettighederne tilhører ZZ]	[Ja]	
[fx Kopi af data fra Statsbiblioteket]	[Ophavsrettigheder tilhører tredjemand]	[Nej – det er en personlig adgang]	

# DM plan: Kommunalvalgsundersøgelsen 2013

---

DM plan er fundet på: <https://sbprojects.statsbiblioteket.dk/pages/viewpage.action?pageId=30638350>

## Admin Details

**Project Name:** Kommunalvalgsundersøgelsen 2013

**Principal Investigator / Researcher:** Christian Elmelund-Præstekær

**Description:** This data management plan will cover the research project on the Danish 2013 Municipality Elections.

**Institution:** Rigsarkivet / The Danish National Archives

## Data Collection

### What data will you collect or create?

In this research project data will be collected through a combination of telephone and webbased interviews. The population is Danish residents aged 18 or older. The interviews will be conducted by TNS Gallup on behalf of the Institute of Political Science, The Southern University of Denmark.

The collected data will be entered into a statistical software package for further analysis. The chosen software package will be able to output data in formats that can be read and entered into other software packages. This will make data shareable with others who use different software packages as well as guaranteeing long-term access to data.

### How will the data be collected or created?

Data will be collected by TNS Gallup through a combination of telephone and webbased interviews on TNS Gallups webpanel, "GallupForum."

## Documentation and Metadata

### What documentation and metadata will accompany the data?

A methodology rapport, a documentation rapport, the questionnaires used for the interviews as well a list of variables/codebook will accompany the data.

Data will be documented in the DDI-L 3.1 (Data Documentation Initiative, Lifecycle) metadata standard by The Danish National Archive.

## Ethics and Legal Compliance

### How will you manage any ethical issues?

TNS Gallup is responsible for the data collection. Anyone who participates in an interview with TNS Gallup do so anonymously and the answer they give are therefore not available for themselves or others to see, including the researchers.

The answers are anonymized after the interview where they are combined with the other respondents answers for analysis. According to the law TNS Gallup may not disclose the respondents answer alongside their personal information and therefore the identities of the respondents are always anonymized before the results are passed on to third parties.

### How will you manage copyright and Intellectual Property Rights (IPR) issues?

Data is owned by the primary researchers/The Faculty of Political Science, The University of Southern Denmark.

## Storage and Backup

### How will the data be stored and backed up during the research?

Data will be used on the researchers own personal computers. Data is anonymized so there are no inherent risks that any personal information will be disclosed.

Data will be regularly backed up on several external storage devices to ensure that data can be recovered in the event of an incident.

### How will you manage access and security?

Data will only be accessible by people involved in the research project.

## Selection and Preservation

### Which data are of long-term value and should be retained, shared, and/or preserved?

The primary dataset as well as any documentation needed in order to understand and interpret data will be retained.

### What is the long-term preservation plan for the dataset?

A copy of data and documentation will be handed over to The Danish National Archive for long-term preservation. The Danish National Archive commits to retain the data indefinitely.

## Data Sharing

### How will you share the data?

Data will be preserved and disseminated at The Danish National Archive where data from earlier projects also is available. The Danish National Archive will make data available through their search engine where they will provide a detailed description of the research project and data. A DOI will also be provided.

### Are any restrictions on data sharing required?

Data will be made freely available.

## Responsibilities and Resources

### Who will be responsible for data management?

The primary researcher will be responsible for data management.

### What resources will you require to deliver your plan?

Given the scope of the project no additional specialist expertise or training will be required. The data will be preserved at The Danish National Archive free of charge.



## DMP title

**Project Name:** CALPIU

**Project Identifier:** CALPIU

**Grant Title:** 09-065696

**Principal Investigator / Researcher:** Hartmut Haberland

**Project Data Contact:** Anne Fabricius (fabri@ruc.dk)

**Description:** See project website: <http://calpiu.dk/ruc/Home.html//> Funded by: The Danish Council for Independent Research | Humanities (FKK)

**Institution:** Roskilde Universitet / Roskilde University

## Data Collection

### What data will you collect or create?

- Audio and video recordings of naturally occurring interaction (classroom interaction, student group project work).
- Video files are .mov; audio files are .wav
- Audio and video recordings of research interviews (individual as well as group interviews).
- Total 360 hours of recording. Of these approximately 28 hours have been transcribed using linking software (CLAN).
- The CLAN transcription files are part of the data set.
- No existing data available.

### How will the data be collected or created?

- Data has been collected by researchers associated with the CALPIU research centre, and student assistants hired by the research centre.
- Files are named according to the CALPIU file naming grid (available here: #insert link).
- Transcripts have been checked by at least two student transcribers.

## Documentation and Metadata

### What documentation and metadata will accompany the data?

- Log files for each recording session (with notes on who created or contributed to the data, its title, date of creation and under what conditions it can be accessed).
- Transcripts (partial) with information about participants and transcribers.
- No particular metadata standards have been adopted, since this was considered too costly to implement.

## Ethics and Legal Compliance

### How will you manage any ethical issues?

- Informed consent from all participants, also for sharing anonymised audio data.
- Names and any other identifying information anonymised in transcripts, and taken out in audio files used for wider sharing.
- Screenshots used for publication will be blurred.
- Data are stored securely at Roskilde University servers and transported on encrypted HDDs only.

### How will you manage copyright and Intellectual Property Rights (IPR) issues?

- Data owned by individual researchers.
- For restrictions on the re-use of data, see Data Sharing below.
- Data sharing will be gradually implemented.
- No third-party data in the storehouse.

## Storage and Backup

### How will the data be stored and backed up during the research?

- Data will be stored on a server (NAS), bought as part of the project.
- Daily, automatised back-ups to separate NAS.
- Data set-up located at Roskilde U and administered by Campus-IT.
- Campus-IT at RU is responsible for backup and recovery.
- Recovery possible from complete backup of main server.

### How will you manage access and security?

- Access to NAS administered by PI team (with technical assistance from Campus-IT)
- The server can only be accessed on campus or via VPN.

- In the field, data are transported on encrypted discs.
- With these procedures in place, the risks to data security are considered to be minimal.

## **Selection and Preservation**

### **Which data are of long-term value and should be retained, shared, and/or preserved?**

- All recordings will be valuable to retain
- Retaining data is important for the purpose of validating research findings and facilitating further research in the area.
- The data will enable longitudinal studies and comparative studies across different time periods.
- As far as possible, data will be retained indefinitely.

### **What is the long-term preservation plan for the dataset?**

- For now, the data set is stored at Roskilde University
- No immediate solutions for long term archiving exist.
- We have not costed in time and effort to prepare the data for sharing / preservation, but we hope to be able to find additional funding for this, when appropriate solutions for long term preservation exist.

## **Data Sharing**

### **How will you share the data?**

- Potential users will be able to read about the storehouse on the centre's website.
- The data will be shared with other bona fide researchers, subject to specific agreements.
- Ideally, data sharing will be through a repository with full control on our part of granting rights.
- The data will be made available, under the conditions specified above, as soon as possible.
- We will pursue getting persistent identifiers for our data!

### **Are any restrictions on data sharing required?**

- We are past the point where the research centre has a need for exclusive use of the data, but sharing is made difficult by the fact that the data are, perhaps, sensitive.

## Responsibilities and Resources

### Who will be responsible for data management?

- Hartmut Haberland, Anne Fabricius and Janus Mortensen are responsible for implementing the DMP, and ensuring it is reviewed and revised.
- RU was responsible for data storage in the centre's phase.
- Data ownership and responsibilities for RDM rest with the group of researchers mentioned above.
- Plan for long term responsibility for the data is under development.

### What resources will you require to deliver your plan?

- Legal advice needed to move on from here.
- Student assistants needed to anonymise data. Training for this purpose required.
- Renewed software licenses might be necessary for anonymization purposes.
- Will charges be applied by data repositories?



# PROJECT MANAGEMENT PLAN GALAXY

GUT-AND-LIVER AXIS IN ALCOHOLIC LIVER FIBROSIS  
GRANT NUMBER 668031

***DELIVERABLE NUMBER: D8.1***

***DELIVERABLE DUE DATE: APRIL 1<sup>ST</sup> 2016***

***COMPLETION DATE OF DELIVERABLE: MARCH 18 2016***

***DISSEMINATION LEVEL: PUBLIC***

DOCUMENT MAIN AUTHOR: Project manager Linda Sevelsted Møller

DOCUMENT SIGNET OFF BY: Project coordinator Aleksander Krag

## **TABLE OF CONTENTS**

<b>1. INTRODUCTION</b>	<b>4</b>
<b>2. LEGAL BASIS</b>	<b>4</b>
<b>3. PROJECT OBJECTIVES</b>	<b>4</b>
<b>4. PROJECT STRUCTURE</b>	<b>5</b>
4.1 CONSORTIUM PARTNERS	5
4.2 ORGANIZATIONAL STRUCTURE OF GALAXY	5
4.2.1 EUROPEAN COMMISSION	5
4.2.2 THE STEERING COMMITTEE	6
4.2.3 THE COORDINATOR	6
4.2.4 WORK PACKAGE LEADERS	7
4.2.5 ADVISORY BOARD	7
4.2.6 SPECIAL MANAGERS	8
4.3 MANAGEMENT STRUCTURE IN GALAXY	8
4.3.1 PROJECT COORDINATOR	9
4.3.2 PROJECT MANAGER	9
4.4 CONFLICTS	10
4.5 PROJECT CHANGES	10
<b>5. WORK BREAKDOWN STRUCTURE</b>	<b>10</b>
<b>6. PROJECT DELIVERABLES &amp; MILESTONES</b>	<b>11</b>
<b>7. COMMUNICATION AND DISSIMINATION PLAN</b>	<b>12</b>
7.1 INTERNAL COMMUNICATION STRATEGY	12
7.1.1 EMAIL CORRESPONDENCES	12
7.1.2 VIDEO CONFERENCES	13
7.1.3 MEETINGS	13
7.1.4 MINUTES OF MEETINGS	14
7.1.5 WEBSITE FORUM & PRIVATE PROJECT FORUM	15
7.1.6 TEMPLATES	15
7.1.7 FILE SHARING	15
7.2 EXTERNAL COMMUNICATION AND DISSIMINATION STRATEGY	15
7.2.1 PUBLIC WEBSITE	15
7.2.2 PRESS STRATEGY	16
7.2.3 GALAXY FLYER	16
7.2.4 PUBLIC SYMPOSIA & CONFERENCES	16
7.2.5 EDUCATIONAL WORKSHOPS AND COURSES	16
7.2.6 PUBLICATIONS	16
7.2.7 VIDEO AND INFOGRAPHICS	17
7.2.8 SOCIAL MEDIA	17
7.2.9 DISSIMINATION OF RESEARCH RESULTS TO SCIENTIFIC COMMUNITY AND PUBLIC HEALTH ORGANISATIONS	17



<b>8. PROJECT REPORTING</b>	<b>18</b>
8.1 INTERNAL PROGRESS REPORTING	18
8.2 REPORTING TO THE EUROPEAN COMMISSION	18
8.2.1 REPORTING PERIODS	18
<b>9. PROJECT FINANCING</b>	<b>19</b>
<b>10. AUTHORIZATION</b>	<b>19</b>

## 1. INTRODUCTION

The EU Framework Programme for Research and Innovation Horizon 2020 has granted the GALAXY project 6.3 million € to investigate the gut-and-liver axis in alcoholic liver fibrosis under the grant number 668031.

GALAXY was initiated on January 1<sup>st</sup> 2016, and will continue until January 1<sup>st</sup> 2022.

The purpose of this management plan is to outline the overall infrastructure of the management approach for the project. It serves as a central guide in regards to daily managing of the overall project progression, reporting, meeting schedules, templates and communication between GALAXY members.

## 2. LEGAL BASIS

The GALAXY project operates within the Horizon 2020 Framework Programme.

Responsibilities towards the funding party, the European Commission, are stated in the Grant Agreement Nr 668031.

The roles and responsibilities of the parties towards each other are stated in the Consortium Agreement that has been signed by all partners.

The GALAXY Consortium Agreement (CA) and Grant Agreement (GA) are at all times superior to this management plan.

## 3. PROJECT OBJECTIVES

The objective of the GALAXY project is to effectively employ integrated system modeling of multi-omics data together with clinical data, to develop prevention and treatment programs for alcoholic liver fibrosis (ALF). By use of systems biology approach the project investigates the role of the gut microbiome in alcoholic liver fibrosis.

The specific aims are:

1. To define specific gut microbiome and associated host genome, transcriptome and metabolome key networks and risk signatures that predict the development and progression of ALF.
2. From these signatures to develop novel 'omics-based biomarkers tailored for early diagnosis, prognosis and treatment monitoring. The goal is to enable timely interventions and prevent advanced disease.
3. Verify discovered key networks in animal models and validate biomarkers in an interventional clinical study.
4. Evaluate interventions based on microbiome modulation to prevent ALF progression



## 4. PROJECT STRUCTURE

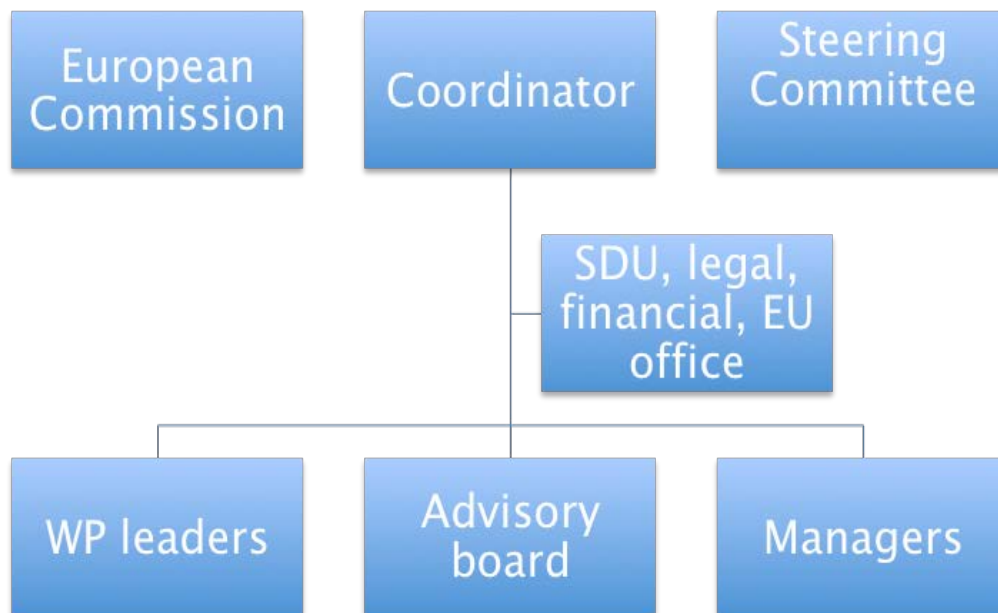
### 4.1 CONSORTIUM PARTNERS

The GALAXY consortium consists of 10 partners in 4 countries.

#	PARTNER NAME	COUNTRY
1	University of Southern Denmark (coordinator)	Denmark
2	Odense University Hospital	Denmark
3	European Molecular Biology Laboratory	Germany
4	University of Copenhagen	Denmark
5	Biomedical Research Foundation, Academy of Athens	Greece
6	Universitätsklinikum Bonn	Germany
7	Steno Diabetes Center A/S	Denmark
8	Nordic Bioscience A/S	Denmark
9	University of Oslo	Norway
10	Nordic Rebalance A/S	Denmark

### 4.2 ORGANIZATIONAL STRUCTURE OF GALAXY

The organizational structure of GALAXY is shown in the diagram below.



#### 4.2.1 EUROPEAN COMMISSION

The European Commission is the grant giver. All responsibilities towards the grant giver are stated in the Grant Agreement.

#### **4.2.2 THE STEERING COMMITTEE**

The Steering Committee is GALAXY’s central managerial body and responsible for decision-making within the GALAXY consortium. Project Coordinator Aleksander Krag chairs the Steering Committee. The committee is responsible for any decisions regarding alterations in the forwarded strategy of the project, budget alterations or changes in allocations of resources available.

Voting rules are described in detail in the Consortium Agreement. Each member has one vote.

During meetings any partner can participate with additional representative from their institutions, as long as voting rules are respected.

The composition of the Steering Committee is designed to cover all the essential activities in GALAXY, consisting of all work package leaders and special managers. All partners are represented in the Steering Committee and the partner representative in the Steering Committee is authorised, on behalf on their institution, to make decisions on all matters of discussion in the Steering Committee. All partners can submit resolutions for consideration by the Steering Committee.

Following representatives constitutes the Steering Committee:

<b>NAME</b>	<b>INSTITUTION</b>
Prof Krag	University of Southern Denmark
Dr Madsen	Odense University Hospital
Dr Arumugam	University of Copenhagen
Prof Hansen	University of Copenhagen
Prof Oresic	Steno Diabetes Center A/S
Prof Trebicka	Universitetsklinikum Bonn
Prof Melberg	University of Oslo
Prof Bork	European Molecular Biology Laboratory
Prof Anastasiadou	Biomedical Research Foundation, Academy of Athens
Dr Thiele	Odense University Hospital
Prof Karsdal	Nordic Bioscience A/S
Dr Israelsen	Nordic Rebalance A/S

The Steering Committee is directly advised from the Advisory Board. The dissemination manager (prof Trebicka) together with the project coordinator (prof Krag) will be responsible for the overall contact with the advisory board. Plans for meeting activities of the Steering Committee is outlines in section 7.1.3.

#### **4.2.3 THE COORDINATOR**

The Coordinator is the link between the grant giver – the Commission – and the grant takers – the GALAXY consortium. The Coordinator oversees project progression, and the roles are clarified in section 4.3.1.

#### 4.2.4 WORKPACKAGE LEADERS

GALAXY is divided into 8 specific work packages (WP). A person in charge has been appointed for each work package.

WP #	WP TITLE	LEAD BENEFICIARY	WP LEADER	START MONTH	END MONTH
WP1	Clinical studies	OUH	M. Thiele	1	72
WP2	Multi-omics studies of host and gut microbiome	EMBL	P. Bork	1	57
WP3	Metabolomics and biochemical assays	Steno A/S	M. Oresic	1	57
WP4	Bioinformatics and systems modeling	EMBL	P. Bork	6	60
WP5	Animal studies and hepatic fibrogenesis	UKB	J. Trebicka	1	48
WP6	Biomarker development: evaluation and validation	NB	M. Karsdal	28	66
WP7	Socio-economic evaluations	UO	H. Melberg	36	72
WP8	Project management, dissemination, communication	USD	A. Krag	1	72

The WP leader is responsible for overseeing the progress of their respective work package and ensures the scientific quality of the work performed. They furthermore hold responsibility for the preparation of both internal progress reports within the scope of the project, and provide reports for the project management office in order to carry out the financial and periodic reports to the European Commission. The WP leader is responsible for notifying both the project management office and the Steering Committee in case of delays in fulfillment of deliverables and milestones. The WP leader holds responsibility for identifying and managing the risks within their work packages.

#### 4.2.5 ADVISORY BOARD

GALAXY has installed an external advisory board with the main purpose to provide GALAXY with a continuous ‘fresh’ and independent input on its activities and results. This external advisory board does not have decision authority in the project, but will provide the GALAXY Steering Committee with ex-ante advice and ex-post feedback on the activities and results of GALAXY, from their respective backgrounds, interests and expertise. They will have full confidential access to all the GALAXY documentation they need in order to form their expert opinions and will be invited to selected project meetings. The Advisory Board will consist of key opinion leaders in the field of gut-liver axis and research in alcoholic liver fibrosis. Additional

members may come during the projects life span. The current Advisory board consists of:

NAME	INSTITUTION
Prof Szabo	University of Massachusetts
Prof Schnabl	University of California
Prof Shah	Mayo Clinic
Prof Lotersztjan	University of Paris
Prof Wiest	Bern University Hospital

#### **4.2.6 SPECIAL MANAGERS**

Special managers have been appointed to set attention to areas of special interest in order to maximize impact of work. The managers are overall responsible for progression of their special area (e.g. dissemination). Compilation of management plans describing the action for the particular area (e.g. exploitation plan) are supported by the project management office, where the project manager will assist in writing up the plan, and the appointed manager takes active part in the strategy of this plan. All partners are expected to participate in maximizing input of the work, and both information on e.g. risks and dissemination activities will be a standard topic in the internal progress reporting.

The appointed managers and their primary tasks are:

MANAGEMENT AREA	NAME	PRIMARY TASK
<b>Dissemination</b>	J. Trebicka (UKB)	Results broadcasted to stakeholders identified as health care providers, industry, policy makers, patients, scientific community
<b>Risk</b>	T. Hansen (UCPH)	Assisting WP leaders in identification and resolution of critical risks, conflict resolution.
<b>Innovation</b>	P. Bork (EMBL)	Supporting WP leaders in innovative tasks
<b>Exploitation</b>	M. Karsdal (NB)	Development of exploitation strategy
<b>Ethics</b>	B. Madsen (OUH)	Overseeing ethical standards

### **4.3 MANAGEMENT STRUCTURE IN GALAXY**

The Project Coordinator, Aleksander Krag, and Project Manager, Linda Sevelsted Møller, has the overall authority and responsibility for managing and executing all management related issues of the project according to this Management Plan. They constitute the Project Management Office, as part of work package 8. The Project Management Office has assistance from legal and financial personal at the University of Southern Denmark, and is further supported from appointed special task managers, as elaborated in section 4.2.6.

The main task of the Project Management Office includes development of internal technical and financial progress monitoring, reporting, planning of meetings and delivery of financial-administrative and managerial support in order to disseminate.

The Project Management Office will be a matrix in that the *appointed leaders from each work package must continue to report to their project progression throughout the duration of the project*. The project manager is responsible for communicating with work package leaders on the progress and performance of each project resource.

#### **4.3.1 PROJECT COORDINATOR**

Professor Aleksander Krag  
Department of Gastroenterology and Hepatology  
University of Southern Denmark  
Sdr. Boulevard 29  
5000 Odense C  
Denmark  
E-mail: [coordinator@livergalaxy.eu](mailto:coordinator@livergalaxy.eu)

The coordinator will be exclusively in charge of the following:

1. Chairs all meeting of the Steering Committee (can be delegated to another member of the Steering Committee).
2. Prepares the agenda for meetings.
3. Monitor that the action is implemented properly.
4. Is primary contact person between the GALAXY consortium and the European Commission.
5. Request and review any documents or information required by the European Commission and verify their completeness and correctness before passing them on to the European Commission (e.g. certificates of Financial Statements)
6. Submit deliverables and reports to the commission.
7. Administers the financial contribution from the European Commission to partners.

#### **4.3.2 PROJECT MANAGER**

Dr. Linda Sevelsted Møller  
Department of Gastroenterology and Hepatology  
University of Southern Denmark  
Sdr. Boulevard 29  
5000 Odense C  
Denmark  
E-mail: [projectmanager@livergalaxy.eu](mailto:projectmanager@livergalaxy.eu)

The project manager works with all resources to perform project planning. All project and subsidiary management plans will be reviewed and approved by Coordinator.

The project manager is responsible for:

1. Supporting the Coordinator to ensure implantation of action, by tracking the progress of deliverables at all times.
2. Keeps updated list of all key contact personal.

3. Receives reports from WP leaders and writes up quarterly progress reports for the overall project as well as reports to the European Commission to be approved and submitted by the Coordinator.
4. Arrange meetings for the Steering Committee, and track minutes of meetings.
5. Administers the website.
6. Responsible for internal communication forum in GALAXY by providing templates, uniform logo, progress report templates etc.
7. Drafts plans for project infrastructure, communication, dissemination, exploitation and risk supported by appointed special managers.

#### **4.4 CONFLICTS**

Disputes within the project should be settled as “local” as possible. For practical purposes this will be in the hands of the respective work package leader. Assistance towards conflict resolution can be obtained from the Project Management Office with assistance from the risk manager. Conflict that cannot be solved in this manner should be raised before the Steering Committee. The Steering Committee has the authority to pass a solution by the voting rules outlined in the Consortium Agreement.

#### **4.5 PROJECT CHANGES**

Changes to the project must be carefully considered and the impact of the change must be clear in order to make any type of approval decisions. Changes to the project plan must ultimately be approved by the funding authority. Proposal for changes must be submitted and approved by the Steering Committee.

Any Party may submit a change request for the project to the Steering Committee. All change requests will be logged by the Project Manager and tracked through to completion whether approved or not.

### **5. WORK BREAKDOWN STRUCTURE**

GALAXY is subcontracted into 8 work packages. The milestone plan includes the schedule baseline for monitoring progress of the project. The Management Office will use various management tools for schedule baselines (Grant charts ect.) to oversee progression of the project as well as deadlines for reporting.

The project manager will also determine the impact of changes or delays to the project progression.

The work packages are divided by the following overall tasks:

1. Clinical studies: all necessary samples and clinical data are collected from a running cross-sectional study of patients with alcoholic liver disease together with healthy controls. In addition, two interventional studies are included to validate the therapeutic effect of an antibiotic, Rifaximin, and a synbiotic, Profermin.
2. Multi-omics; metatranscriptomics, metagenomics, genomics, transcriptomics and data mining are applied to a selection of samples to gather an exhaustive, comprehensive host/microbiome data warehouse.

3. Metabolomics; metabolites play a crucial role during the pathogenesis, therefore metabolites, including bile salts, are analysed and added to the GALAXY data warehouse.
4. Systems modelling; systems medicine tools will be developed to analyse the data warehouse and extract key networks during alcoholic liver disease.
5. Animal models; key networks will be validated and further analysed in ALF animal models to elucidate causal mechanisms.
6. Biomarker development; building upon identified key networks and causal mechanisms, we will identify biomarkers based on host and microbiome -omics characteristics. Biomarkers will be validated in the Profermin study.
7. Socio-economic analysis; cost-benefit analysis will be performed to analyse opportunities and identify high potential biomarkers and treatments to enter the clinic.
8. Project management; providing all work packages with support in all management related issues.

## **6. PROJECT DELIVERABLES & MILESTONES**

Deliverables for the individual work packages are specified in annex 1 of the grant agreement. The work package leader has the overall responsibility of assuring that deliverables are completed on time. Any delays must be reported to the Project Management Office in the quarterly progress reports (see section 9.1), but preferably as soon as possible. It is also required that a clarification as to why delays have occurred is attached, as well as a proposed solution to the delay.

The chart below lists the major milestones for GALAXY. This chart is comprised only of major project milestones such as completion of a project phase. Smaller milestones are not included on this chart. In case of any scheduling delays that may impact a milestone or delivery date, the Project Management Office must be notified immediately so proactive measures may be taken to mitigate slips in dates.





23. Gantt chart of GALAXY. Stars represent milestones as indicated in Table 3.2a.

## 7. COMMUNICATION AND DISSIMINATION PLAN

The purpose of this section is to set the framework for communication for the project by defining and elaborating the communicative requirements both in the daily communication between partners and to external bodies. The goal is to ensure that partners and interest groups receive the information they need to know at the appropriate time, as well as to identify roles and expectations for communication.

The project manager serves as communication manager, with the lead role in overseeing communication in the projects, e.g. by planning meetings, conferences and templates as communicative tools in the project.

### 7.1 INTERNAL COMMUNICATION STRATEGY

#### 7.1.1 EMAIL CORRESPONDENCES

Correspondences should be kept professional and as brief as possible. Please initiate emails with GALAXY in the topic field, followed by a short description of the subject of the correspondence. To ease email correspondences regarding management issues, email accounts for the coordinator and project manager has been set up:



Project Coordinator Aleksander Krag: [coordinator@livergalaxy.eu](mailto:coordinator@livergalaxy.eu)

Project Manager Linda Sevelsted Møller: [projectmanager@livergalaxy.eu](mailto:projectmanager@livergalaxy.eu)

Upon request, similar email addresses can be setup to respective work packages.

An updated contact list for partners, members of the Steering Committee and key associates can be found at the private project forum at the website [www.livergalaxy.eu](http://www.livergalaxy.eu).

It is imperative that changes in contact information are notified to the project manager immediately.

### **7.1.2 VIDEO CONFERENCES**

Several meetings will be planned as videoconferences to avoid unnecessary time and expenses for travel activity. Hence, all partners, and most importantly all WP leaders and members of the Steering Committee should setup an account for the videoconferences and notify the project manager of username. The project manager, will prior to the first videoconference, give notification to attendees on the platform used and instructions to setup.

### **7.1.3 MEETINGS**

This section describes the meeting cycle, responsibilities and persons involved. For public meetings including workshops, symposia and conferences please go to 7.2 on external communication and dissemination.

#### **Project team meetings:**

As GALAXY involves a number of research groups, the partner and/or WP leader plan meetings within the individual project teams. The frequency of the meetings will vary among partners. Track of meeting minutes is recommended, but is not required.

#### **Internal WP meetings:**

Collaborations between the individual work packages in GALAXY are a key aspect. The work package leaders are responsible for the arrangement of these meeting when needed. Minutes of meetings should be tracked and send to the project manager as part of the quarterly progress reporting (section 7.1.4).

#### **Steering Committee (SC) meetings:**

Meetings within the Steering Committee will take place quarterly, alternating between video conferences (section 7.1.2) and physical meetings. The overall purpose of the meetings is to review project status for the previous 3 months and plans for the following year.

The Management Office will outline meeting dates 3 months in advance; the agenda for the SC meeting will be distributed 4 weeks in advance, for comments and suggestions. The finalized agenda will be available 1 week before the meeting.

When needed, additional video – or phone conferences can be scheduled and requested by any

partner. Notification of additional meetings should be given at least 14 days prior to the meeting.

The project manager tracks meeting minutes, and will distribute these to the partners no later than 14 days after the meeting.

Advisory board meetings:

Meetings with the advisory board will take place once or twice yearly. The dissemination manager and the project coordinator are responsible for planning meetings with the advisory board, and also to invite the members to selected GALAXY meetings, e.g. workshops and symposia. The project manager tracks meeting minutes, and distributes to partners no later than 14 days following the meeting.

Meeting Type	Description	Reporting	Frequency	Participants/ Distribution	Responsible
<b>Project Team Meeting</b>	Meeting to review action register and status	Not required	Ad hoc	Project Team	Partner
<b>Internal WP meetings</b>	Collaborations between work packages	Minutes of meeting	Ad Hoc	WP leaders, project team members	WP leader
<b>Steering Committee Meeting</b>	Review project status	Minutes of meetings	3 months	SC members	Coordinator/ Project manager
<b>Advisory Board Meeting</b>	Review of designs or work associated with the project	Minutes of Meetings	6 months	Project Team	Dissemination Manager, Coordinator

#### 7.1.4 MINUTES OF MEETINGS

Meeting minutes record decisions taken and the actions required by partners. Ultimately the chair of the meeting is responsible to pointing out the referent for taking of the minutes.

A template for minutes will be available at the GALAXY website. If not used, the minutes of meeting should include: date, location, attendees, agenda and action.

The meeting minutes must be sent to the project manager, who will create a private project filestore as a permanent record of the meetings.

### **7.1.5 WEBSITE FORUM & PRIVATE PROJECT FORUM**

The GALAXY website [www.livergalaxy.eu](http://www.livergalaxy.eu) has been created and will be fully operational by July 1<sup>st</sup> 2016. The website consist of both a public platform as well as a private site for partners. To access the private site, partners and project team members must register on the log in site at the first visit, and will hereafter, upon approval from the project manager, be granted access. The private forum will give project members access to all necessary templates, contact information etc. Project members will here also have opportunities to post messages of interest to all partners. The login site serves the purpose of sharing private, but not strictly confidential information.

### **7.1.6 TEMPLATES**

We encourage all partners to use the GALAXY templates for posters and PowerPoint presentations upon presentations related to the project. This is not a requirement, whereas the use of EU and GALAXY logos are.

Templates for meeting minutes and quarterly progress reports will also available at the project forum at the website before July 1<sup>st</sup> 2016.

### **7.1.7 FILE SHARING**

Operational tools and rules for sharing data among partners will be elaborated in a separate data management plan. Due to the importance of this plan, it serves as a separate deliverable, which will be distributed to all partners and disseminated public on the website within the deliverable due date on July 1<sup>st</sup> 2016.

## **7.2 EXTERNAL COMMUNICATION AND DISSIMINATION STRATEGY**

Upon external commucation from GALAXY all partners are obligated to use the EU emblem and the GALAXY logo. The logos are available for download from the log in site at the GALAXY website.

It must furthermore be stated that the project is supported from the EU by use of the phrase “*This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 668031*”.

### **7.2.1 PUBLIC WEBSITE**

The GALAXY website [www.livergalaxy.eu](http://www.livergalaxy.eu) serves as an important tool for all external communication from the project. It introduces the project to the public and provides information on the individual work packages and links to partners. Dissemination activities will also be announced here, as well as publications and deliverables that must be disseminated on a public level. Relevant links to other GALAXY profiles, e.g. on social medias, can be found at the website.

The website is administered by the Project Manager with technical support from an appointed Webmaster. All partners are encouraged to give input to website, in the form on news, activities and pictures.

### **7.2.2 PRESS STRATEGY**

For the announcement of the Horizon 2020 grant each partner must plan how they wish to do the announcement, e.g. by sending out a press release from their respective institutions.

Press releases regarding the project should be notified to and coordinated with the Project Management Office. The coordinator will also evaluate the press releases, and results of major public or scientific interest will be approved by the Steering Committee. The Consortium Agreement states the access right of the results.

The Project Management Office will also take initiative to press releases.

### **7.2.3 GALAXY FLYER**

The GALAXY flyer will be available for download for all interested (public and partners) by July 1<sup>st</sup> 2016. We encourage all partners to use the flyer in dissemination activities to promote the knowledge to the project.

### **7.2.4 PUBLIC SYMPOSIA & CONFERENCES**

GALAXY will host a number of public symposia during the project's lifespan.

The Project Management Office will take initiative to plan and execute the symposia in collaboration with all partners.

GALAXY will host at least 2 symposia and 1-2 conferences. To maximize impact, GALAXY will collaborate with other H2020 projects to host the conferences.

### **7.2.5 EDUCATIONAL WORKSHOPS AND COURSES**

GALAXY will host 2-3 workshops throughout the project period. The purpose of the workshop will be to promote further collaborations among partners, especially at the level of young investigators. The workshops will be hosted by different partners in the project, but planned in collaboration with the Project Management Office. The overall responsibility for initiation and planning of these workshops lies with the dissemination manager and the coordinator. The workshops will be target-oriented and serve as an important educational platform within the project.

Furthermore, GALAXY will host 1-2 courses in article writing, with young investigators as the target audience.

### **7.2.6 PUBLICATIONS**

GALAXY is obligated to disseminate results according to the Grant Agreement. Hence all publications from the project must be publicly available. To ensure open access, publications will be uploaded to the webpage.

The GALAXY consortium has agreed on the following principles regarding authorship:

- 1) Duty of work package (WP) leader to look for opportunities for dissemination
- 2) Main papers (analyses, drafting, etc.) under directorship of WP leader
- 3) All authors must meet authorship criteria
- 4) WP leader takes ultimate responsibility for deciding authorship within these principles
- 5) WP leader takes ultimate responsibility for deciding authorship order

- 6) Coordinating center recognized in all papers
- 7) Generally up to one author per country/network on main papers
- 8) Recognition of rest of platform, where appropriate

Decisions made in consultation with coordinator in event of controversy/conflict

### **7.2.7 VIDEO AND INFOGRAPHICS**

GALAXY will set up a YouTube channel in order to upload education videos and tutorials. All partners are strongly encouraged to send in material. This may include tutorials in ultrasonography and elastometric examinations, laboratory procedures and techniques. Selected presentations from key note speakers at the public GALXAXY symposia will also be uploaded, given approval from the presenter.

### **7.2.8 SOCIAL MEDIA**

All partners should contribute to keeping social media profiles updated. News and updates can be notified to the project manager, who will update the profiles.

GALAXY is on Facebook:

<https://www.facebook.com/Galaxy-514907385336140/>

The Facebook site will be used for public announcement regarding symposia and news from the project.

A LinkedIn profile will be set up within the project's first 6 month as a collaboration platform between both GALAXY partners and external collaborators.

### **7.2.9 DISSIMINATION OF RESEARCH RESULTS TO SCIENTIFIC COMMUNITY AND PUBLIC HEALTH ORGANISATIONS**

This will include:

- Attendance at international conferences.
- Participation at Alcohol awareness week.
- Close communication with patient associations for idea sharing. Collaboration with the European Liver Patient Association (ELPA) has already been established.
- Collaborations with scientific communities for dissemination of the project. GALAXY will in particular be collaborating with the European Association for the Study of the Liver (EASL)
- As part of work package 1 with clinical studies, a patient advisory board will be established. This will consist of 5-6 patients with alcoholic liver fibrosis and 1-2 nurses specialized in alcoholic liver disease, who will advise GALAXY on quality-of-life matters and patient reported outcomes.

## 8. PROJECT REPORTING

### 8.1 INTERNAL PROGRESS REPORTING

All partners must submit quarterly internal progress reports. Templates for this has been distributed to all partners, and is available at the website forum.

The reports must be sent to the project manager no later than 14 days after the end of each quarter. Partners, whose work packages are initiated later in the project, do not need to submit progress reports until their work begins. However they should provide a brief reporting in case of attendances in meetings and dissemination activities.

The project manager will, based on the individual reports submitted by the partners, draft up a quarterly project report for the project as a whole.

The quarterly progress reports will include sections on project status, actions and objective for the quarter to come, deviations from the plan, dissemination activities, meeting attended and resources employed.

### 8.2 REPORTING TO THE EUROPEAN COMMISSION

Reporting periods for the European Commission are scheduled every 18 months. The Project Management Office will submit the report, as required, within 60 days following the end of these reporting periods. The reports are based on the reports received from the partners.

Reports to the European Commission consist of periodic technical reports and periodic financial reports.

Technical reports covers the work carried out, overview of the progress, a summary for publication by the Commission and answers to the questionnaire, covering issues related to implementation and the economic and social impact.

The Financial reports covers individual financial statements from each partners, and explanation to use of resources and a periodic statement. To track personal resources used, time registrations must be completed. To ease the process, templates for time registration will be available at the private forum at the website.

Templates and instructions for these reports can be found at participant portal:

[http://ec.europa.eu/research/participants/portal/desktop/en/funding/reference\\_docs.html](http://ec.europa.eu/research/participants/portal/desktop/en/funding/reference_docs.html)

Partners must also fill out Certificates on the Financial statements (CFS) and Certificate on the methodology (CoMUC). Templates can be found at:

[http://ec.europa.eu/research/participants/portal/desktop/en/funding/reference\\_docs.html#h2020-call\\_ptef-ef](http://ec.europa.eu/research/participants/portal/desktop/en/funding/reference_docs.html#h2020-call_ptef-ef)

#### 8.2.1 *REPORTING PERIODS*

GALAXY reporting periods are scheduled at:

1. Period month 1-18 (Jan 1<sup>st</sup> 2016 – July 1<sup>st</sup> 2017)
2. Period month 19-36 (July 1<sup>st</sup> – Jan 1<sup>st</sup> 2019)
3. Period month 37 - 54 (Jan 1<sup>st</sup> – July 1<sup>st</sup> 2021)

4. Period month 55 – 72 (July 1<sup>st</sup> 2021- Jan 1<sup>st</sup> 2022)

## 9. PROJECT FINANCING

The budget for the each partner can be found in the Grant Agreement. A more detailed table for the distribution of payments from the European Commissions to partners will shortly be available at the private project forum at the GALAXY website.

All partners are responsible for notifying the project manager immediately in case of changes in bank account information.

## 10. AUTHORIZATION

Approved by the Project Coordinator:



March 18 2016

---

Aleksander Krag



# INITIAL DATA MANAGEMENT PLAN

## GALAXY

GUT-AND-LIVER AXIS IN ALCOHOLIC LIVER FIBROSIS  
GRANT NUMBER 668031

**DELIVERABLE NUMBER: D8.3**

**DELIVERABLE DUE DATE: JUNE 30<sup>TH</sup> 2016**

**COMPLETION DATE OF DELIVERABLE: JUNE 29<sup>TH</sup> 2016**

**DISSEMINATION LEVEL: PUBLIC**

### DOCUMENT HISTORY

Issue date	Version	Reasons for this issue
12 June 2016	v0.1	First draft
	v1.0	Final version

DOCUMENT MAIN AUTHOR: Dr Maja Thiele (OUH), PhD

DOCUMENT SIGNET OFF BY: Project Manager Linda Sevelsted Møller, MD, PhD

© COPYRIGHT 2016 UNIVERSITY OF SOUTHERN DENMARK.

This document has been produced within the scope of the GALAXY project. All rights reserved.



## TABLE OF CONTENTS

<b>1. INTRODUCTION</b>	<b>3</b>
<b>2. DATA SET REFERENCE AND NAME</b>	<b>3</b>
<b>3. DATA SET DESCRIPTION</b>	<b>3</b>
3.1 DATA SET DESCRIPTION – GENERAL PRINCIPLES	4
3.1.1 ORIGIN	4
3.1.2 NATURE	4
3.1.3 SCALE	4
3.1.4 USEFULLNESS	5
<b>4. STANDARDS AND METADATA</b>	<b>5</b>
4.1 STANDARDS AND METADATA – GENERAL PRINCIPLES	5
4.1.1 DATA CAPTURE	5
4.1.2 DATA QUALITY	5
4.1.3 DATA FORMAT	5
4.1.4 METADATA	5
4.2 STANDARDS AND METADATA – SPECIFIC PRINCIPLES FOR CLINICAL STUDY DATA	5
4.2.1 DATA CAPTURE AND STORAGE OF RAW, ORIGINAL DATA	5
4.2.2 DATA QUALITY	6
4.2.3 META DATA	6
<b>5. DATA SHARING</b>	<b>6</b>
5.1 DATA SHARING – GENERAL PRINCIPLES	6
5.1.1 DATA WITH UNIQUE PATIENT IDENTIFIERS	6
5.1.2 PSEUDO-ANONYMISED DATA	7
5.1.3 ANONYMISED DATA AND ANIMAL DATA	7
5.1.4 STUDY PROTOCOLS, RESULT DATA AND PUBLICATIONS	8
<b>6. ARCHIVING AND PRESERVATION</b>	<b>8</b>
6.1 ARCHIVING AND PRESERVATION – GENERAL PRINCIPLES	8
6.1.1 DATA VOLUMNES, PRESERVATION AND DISCOVERY	8
6.1.2 DATA SECURITY	9
6.1.3 DOCUMENTATION AND PRESERVATION PLAN	9

## 1. INTRODUCTION

The GALAXY project (Gut-and-liver axis in alcoholic fibrosis) is a part of the EU Framework Programme for Research and Innovation Horizon 2020.

The purpose of this data management plan (DMP) is to outline how the GALAXY consortium will manage data generated by the project.

The document reports the initial DMP for GALAXY, which can also be assessed online at the [Danish Digital Curation Centre](#).

The DMP is not a fixed document; it evolves during the lifespan of the project to fine-tune it to the data generated and the uses identified by the consortium. The initial DMP will therefore be supplemented with a Mid-term Review DMP and a Final Review DMP.

A DMP is a key aspect of any research project where large scale, high throughput data are generated. GALAXY is such a project.

The GALAXY project will generate research data from work packages 1-7. In total, we estimate that the GALAXY project will generate around 1 terabyte of data.

The Initial DMP for GALAXY has five sections: (1) Data Set Reference and Name, (2) Data Set Description, (3) Standards and Metadata, (4) Data Sharing, (5) Archiving and Preservation.

The sections are subdivided in general principles regarding the overall GALAXY project that apply to data management across the consortium partners and to the specific principles for clinical study data from Work Package 1. Additionally, we describe data management for raw data, aggregated data and publications.

The DMP may be updated outside the mid-term and final review to fine-tune it to the data generated and the uses identified by the consortium since not all data or potential uses are clear from the start. Therefore, the latest version of the DMP are available online from <https://dmponline.dtu.dk>.

## 2. DATA SET REFERENCE AND NAME

GALAXY-IDMP.

## 3. DATA SET DESCRIPTION

In this DMP, the term "data" cover raw data generated from investigations in humans and animals, human and animal tissue, aggregated data, result data, data documentation and publications related to the GALAXY project.

The GALAXY project consists of 10 partners (University of Southern Denmark, European Molecular Biology Laboratory, University of Copenhagen, Biomedical Research Foundation Academy of Athens, University of Bonn, Steno Diabetes Center, Nordic Bioscience, Nordic Rebalance, University of Oslo and Odense University Hospital). All partners will generate data to the project.

### **3.1 DATA SET DESCRIPTION – GENERAL PRINCIPLES**

#### **3.1.1 ORIGIN**

The data origins from four prospectively conducted human studies at Odense University Hospital (OUH) and animal studies at University of Bonn (UKB).

The data underpins planned publications within the GALAXY consortium.

#### **3.1.2 NATURE**

The data consists of the following quantitative, raw data: (1) clinical study logs, (2) patient history, (3) clinical investigations, (4) questionnaires, (5) outcome event data, (6) existing data from the Danish unique personal id registries, (7) biobank logs, (8) data generated from quantitative analyses of human liver tissue, blood, feces, urine, saliva, hair and sigmoid tissue, (9) rodent logs, (10) data generated from quantitative analyses of rodent liver and colonic tissue, blood and feces.

Generated data - except (6) - will be original.

Quantitative data formats are continuous, ordinal, nominal and binary.

Few qualitative and quantitative data, such as signed patient consent forms and questionnaires, exists in paper format. Some result data may also exist in sound, image and video format.

#### **3.1.3 SCALE**

The four human studies consists of: (A) 400 patients in the GALA-ALD study, a prospective cohort of alcohol overusing patients; (B) 100 participants in the GALA-HP study, a cross-sectional study of healthy controls, matched for age and gender with GALA-ALD; (C) 136 patients in the GAB-ALD study, recruited from GALA-ALD. GAB-ALD is a randomised controlled trial of the gut-selective antibiotic Rifaximin vs. placebo for alcoholic liver fibrosis; (D) 80 patients in the SYN-ALD study, a randomised controlled trial of the synbiotic Profermin vs. placebo for patients with a dysbiotic microbiome and alcoholic liver fibrosis.

The rodent studies explore two strategies: prevention of fibrosis by intervening concomitantly with gut-selective antibiotic Rifaximine with ethanol diet, and treatment of fibrosis with Rifaximine by shorter use under ethanol diet. Different groups of mice (15 animals per group), animal models and genotypes are used: (A) Lieber-DeCarli diet (wild type germ-free, specific-pathogen-free and Ago2-Flag-HA-k.i. mice); (B) High percentage ethanol diet (cannabinoid receptor-1 (CB1) and receptor-2 (CB2) knockout and wild type mice). (C) Controls (wild type, Ago2-Flag-HA-k.i. and germ-free mice, bile duct ligation, carbon tetrachloride and thioacetamide models as controls for cirrhosis.

We estimate the full dataset to be 1 TB large for which participating centres have adequate storage capacity.

### **3.1.4 USEFULLNESS**

The dataset may be useful for comparable projects in other liver disease aetiologies (e.g. non-alcoholic fatty liver disease), other fields of medicine (e.g. endocrinology) or for other microbiome and multi-omics studies.

## **4. STANDARDS AND METADATA**

### **4.1 STANDARDS AND METADATA – GENERAL PRINCIPLES**

#### **4.1.1 DATA CAPTURE**

Participating partners will capture quantitative data electronically. Files and folders will be named according to local custom in clearly marked GALAXY folders. Datasets will be named with date and version.

#### **4.1.2 DATA QUALITY**

Quality and consistency will be controlled through standardised data capture, data entry validation, representation with controlled vocabularies and peer review of data.

#### **4.1.3 DATA FORMAT**

Data will be stored according to local preference in accordance with the UK Data Service guidelines <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats>. Tabular data are stored in .csv, .tab, .sav, .dta, .mdb, .acddb, .por, command ('setup') files (SPSS, Stata, SAS, etc.), or structured text or mark-up file of metadata information (DDI XML file). Textual data are stored as .rtf, .txt or .doc/.docx files. Image, audio and video data are stored as .tif, .flac and .mp4. Documentation and scripts are stored as .rtf, .pdf, .xhtml, .htm or .odt files.

#### **4.1.4 METADATA**

We will follow principles described by CERIF (Common European Research Information Format; developed and maintained by EuroCRIS, <http://www.eurocris.org/>) for recording metadata. See more here: <http://www.dcc.ac.uk/resources/metadata-standards/cerif-common-european-research-information-format>.

Converis, Pure or Symplectic Elements may be used as research information system tools for implementing the CERIF metadata standards.

### **4.2 STANDARDS AND METADATA – SPECIFIC PRINCIPLES FOR CLINICAL STUDY DATA**

#### **4.2.1 DATA CAPTURE AND STORAGE OF RAW, ORIGINAL DATA**

Four systems will be used for data capture of raw, original data: (1) The REDCap database system hosted by Odense Patient data Explorative Network (OPEN), (2) the

SurveyXact Database hosted by Rambøll and licenced to Region of Southern Denmark, (3) Secure SharePoint drives hosted by the Region of Southern Denmark and (4) the COSMIC electronic patient file system hosted by the Region of Southern Denmark. One database will be generated per clinical study.

Clinical study logs, patient history data, clinical investigations and electronic questionnaires are entered directly into the databases by study investigators, project nurses and study participants. Except two questionnaires (48 hour recall and Food Frequency Questionnaire) that are stored in paper form to be machine-read at the end of the study. Biochemical data, histological data and outcome event data are stored in the electronic patient files, to be electronically copied to the study database after study completion.

Data files will be named according to GALAXY study ID, type of data and date of creation/revision; e.g. "galaald\_complete\_yymmdd.dta" or "gabald\_biochemistry\_yymmdd.dta". Data variables in individual files will be given an abbreviated variable name, together with a longer, explanatory variable label; e.g. "S-bilirubin (umol/L)"

All project data will be stored in accordance with the approval given to the individual projects by the Danish Data Protection Agency under the collective permission given to the Region of Southern Denmark.

Odense Patient data Explorative Network ([OPEN](#)) is responsible for secure storage and back up on the databases in REDCap and OPEN Projects (registry of biological material). The Region of Southern Denmark is responsible for the storage and back up on secure SharePoint drives. Rambøll is responsible for the storage and back up of data in SurveyXact.

#### **4.2.2 DATA QUALITY**

Database audits by study investigators and OPEN data managers prior to data collection will ensure that the database is complete and consistent.

#### **4.2.3 META DATA**

Metadata are stored in the REDCap system as the "Data Dictionary" and in SurveyXact under analyses/dataset. Metadata can be exported from both programmes as a .csv file.

## **5. DATA SHARING**

### **5.1 DATA SHARING – GENERAL PRINCIPLES**

Data will be shared through the GALAXY consortium, with participating partners accept and in compliance with applicable legislation.

#### **5.1.1 DATA WITH UNIQUE PATIENT IDENTIFIERS**

Patient identifiable data connected to the participants in the four human studies will be stored at Odense University Hospital in accordance with the approval given to the

individual project by the Danish Data Protection Agency under the collective permission given to the Region of Southern Denmark.

We will not share unique identifiers with project partners outside the Odense University Hospital, unless specific approval are granted from Danish Data Protection Agency under the collective permission given to the Region of Southern Denmark.

### **5.1.2 PSEUDO-ANONYMISED DATA**

Pseudo-anonymised data can be shared between project partners if a Data Processing Agreement is signed by both partners and if sharing is in accordance with the approval given to the individual project by the Danish Data Protection Agency under the collective permission given to the Region of Southern Denmark.

Transfer of pseudo-anonymised data between GALAXY partner institutions will follow local protocols in accordance with the Data Processing Agreement. The giving and receiving partners guarantee that adequate security, hardware, software and bandwidth are available for any large volume data movement.

### **5.1.3 ANONYMISED DATA AND ANIMAL DATA**

We consider patient data fully anonymised when – by the data controller or by any other person – it is not possible to single out an individual from the data, it is not possible to link records relating to an individual, and information concerning an individual cannot be inferred from the data.

In accordance with the Opinion 05/2014 on “Anonymisation Techniques” ([European Union article 29 on the General Data Protection Regulation](#)), we will use a combination of anonymisation techniques (e.g. noise addition, permutation, differential privacy, aggregation, k-anonymity, l-diversity and t-closeness) to ensure that data are anonymised.

Human anonymised data and data from the animal studies are shared within the GALAXY consortium in agreement with the Consortium Agreement.

Raw anonymised data and related metadata, provided they are not subject to terms of confidentiality and/or IP protection, will be made available to the public by three methods:

- Registration of study protocols with OPEN. By request, OPEN can make study data available to outside researchers, if agreed upon by the GALAXY Steering Committee.
- Sharing of datasets and metadata to the Zenodo website (<http://zenodo.org/>) or similar data repositories.
- Long-term storage of datasets, related metadata and study protocols by the Danish National Archived after completion of the project period (see section 6)

Anonymised raw datasets will only be shared with the public when GALAXY has published results from the data (or pending an accepted publication, in accordance with regulations by some scientific journals).

#### **5.1.4 STUDY PROTOCOLS, RESULT DATA AND PUBLICATIONS**

Data documentation, data descriptions and protocols for clinical studies, which are not subject to terms of confidentiality and/or IP protection, will be published in three places to enhance further research collaboration:

- The OPEN website will publish project summaries for each individual clinical study, with the possibility of sharing detailed protocols, data documentation and data sets by request, after acceptance from the GALAXY Steering Committee.
- The GALAXY website will publish a general project description of the GALAXY project and published results with the possibility of sharing detailed protocols, data documentation and data sets by request, after acceptance from the GALAXY Steering Committee.
- Institutional repositories such as the University of Southern Denmark's Publication & REsearch platform ([PURE](#)) will be used to publish project descriptions, links to shared datasets and published results.

Study protocols for the randomised studies (GAB-ALD and SYN-ALD) will be registered with EUDRA-CT, clinicaltrials.gov or a similar internationally recognised study registration site.

Publications (peer reviewed journal articles, non-peer reviewed articles, books and book chapters, conference papers etc.) will be shared with the public in accordance with copyright agreements with the publisher.

We will make published results available for download from institutional repositories (e.g. [PURE](#)) and the GALAXY website using Green Open Access.

## **6. ARCHIVING AND PRESERVATION**

### **6.1 ARCHIVING AND PRESERVATION – GENERAL PRINCIPLES**

Data from the GALAXY project will be stored for long-term preservation at the earliest five year after project completion, in accordance with the Grant Agreement. Data will only be entered into long-term storage if it is in agreement with the GALAXY Steering Committee, or when the relevant Data Protection Agency approval expires.

#### **6.1.1 DATA VOLUMNES, PRESERVATION AND DISCOVERY**

The Danish Data Repository hosted by The Danish National Archives ([Rigsarkivet](#)) will be used for long-term archiving of human study data. Archiving is without an end-date and is free of charge to the GALAXY consortium. The National Archives performs data curation, does regular backup of data and guarantees adequate storage facilities.

Data can be searched and discovered by other researchers using metadata <http://dda.dk/simple-search>. Any publication using data from the Danish Data Repository require permission from the donor.



### **6.1.2 DATA SECURITY**

Data security will conform to national Data Protection guidelines and the signed Data Processing Agreements.

The security of electronic data described is guaranteed by the host institution and will follow national and European Data Protection Acts. All data in paper format will be kept in accordance with Data Protection guidelines (e.g.

<https://www.ukdataservice.ac.uk/manage-data/store/security>).

### **6.1.3 DOCUMENTATION AND PRESERVATION PLAN**

Data and metadata will be prepared for preservation in accordance with standards described in 4.1 and 4.2. Data and metadata files will be stored together with study protocols, publications relating to the project and standard operating procedures.



## DMP

**Title Project Name:** Teknisk støtte til at operationalisere dele af CAFF's Terrestrial Circumpolar Biodiversity Monitoring Programme (CBMP)

**Principal Investigator / Researcher:** Knud Falk

**Project Data Contact:** Søren Møller

**Description:** More than 35 years of monitoring data of Peregrine Falcons in South Greenland. Part of CAFF's Terrestrial Circumpolar Biodiversity Monitoring Programme. Reproduction, Territory occupancy, Banding, Eggshell thickness

**Institution:** Roskilde Universitet / Roskilde University

## Data Collection

### What data will you collect or create?

Existing data from more than 30 years of research:

- Reproduction: MySQL database on vandrefalk.dk (100 kB), OA
- Phenology: MySQL database on vandrefalk.dk (100 kB), OA.
- Territory occupancy: MySQL database on vandrefalk.dk (100 kB), OA
- Banding: MySQL database on vandrefalk.dk (100 kB), OA
- Locations: MySQL database on vandrefalk.dk (100 kB), Access only upon request.
- Eggshell thickness: Microsoft Excel (.xlsx, 100kB). Currently only on request. Aiming at OA.
- Environmental chemistry: Chemical analysis of POP's in addled eggs (100kB). Currently only on request. Aiming at OA.
- Geo locator data: (100 kB), Access only upon request.
- Prey density: (10kB), OA.
- Video on YouTube (2 GB). OA
- Raw Video material (20GB, mp4). Only on request to protect the peregrine population.
- Photos: © Knud Falk og Søren Møller, 2009 a few OA, thousands in private collection.

The data will be updated through current research.

### How will the data be collected or created?

Old and new data are or will be created according to this method:

The project is designed as a "lean" field programme to be conducted annually by two persons in 21-30 days. Small dinghies or Zodiacs are used to navigate the fjords between camp sites, from where the field team hikes to the selected standard monitoring Peregrine sites spanning the coastal and inland areas (see map, right). All

field work is based on basic monitoring parameters sampled at selected sites every year in the core survey area and includes:

- Nest success and productivity - 3 parameters: proportion of occupied sites producing young, number of young reaching banding age per occupied and successful site. Data are compared to “critical thresholds” (USFWS 2003).
- Breeding phenology: Date of first hatching in each nest – measured by standard aging catalogue and wing length or egg weight/measurements.
- Samples: Addled eggs collected for contaminant analyses. Eggshell fragments from hatched eggs – for monitoring change in eggshell thickness as a proxy for DDT/DDE contamination. Moulted feathers for mercury and other heavy metals.

The new (2012) migration study applies miniature (1.9 g) archival light level data loggers (“geolocators” – GLs) providing daily locations almost year round. Adult females are (re)captured at the breeding site by standard methods we have applied for many years when studying adult turnover. In addition, from 2013 we also collect data on prey density by recording passerines on line transects along the hikes to/from Peregrine nesting sites (and other trips). We identify all species and age (adult or fledgling) and count all birds within 50 m horizontal distance from the observer path. This is a rough method providing an index for comparing changes over the coming years.

## Documentation and Metadata

### What documentation and metadata will accompany the data?

#### Reproduction and phenology

When data are made available we will use:

- Metadata: Dublin core (title, creator, date, subject)
- File format: Plain CSV containing headings. One line per nest visit (date, location, no. of eggs, no. of young, age, comments)

#### Territory occupancy

When data are made available we will use:

- Metadata: Dublin core (title, creator, date, subject)
- File format: Plain CSV containing headings. One line per site (year, location, no. adults, comments)

More descriptions will follow.

## Ethics and Legal Compliance

### How will you manage any ethical issues?

Location data will only be shared to acknowledged researchers and government authorities.

Sharing method will be decided later and will depend on repository options.

### **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

Data will be owned by Knud Falk and Søren Møller.

All data except location data will be publicly accessible. Publications based on these data will require owners consent and proper credit. There will be no embargo period.

## **Storage and Backup**

### **How will the data be stored and backed up during the research?**

Sufficient storage available.

Currently several copies. Database at vandrefalk.dk hosted at One.com (professional backup and data recovery).

### **How will you manage access and security?**

Location data currently protected by username/passwords issued by owners.

In the future location data will also be considered to be made available to relevant departments of Naalakkersuisut and Danish organizations.

## **Selection and Preservation**

### **Which data are of long-term value and should be retained, shared, and/or preserved?**

All data collected during this project will be retained and made OA as far as possible and for as long as possible.

The data will be used in the circumpolar terrestrial biomonitoring program (Dancea/CAFF).

### **What is the long-term preservation plan for the dataset?**

Repository/archive to be decided.

No costs expected.

Time and effort included in grant.

## **Data Sharing**

### **How will you share the data?**

Data advertised on:

vandrefalk.dk

CAFF website

All data except location data will be publicly accessible. Publications based on these data will require owners consent and proper credit.

Data will be shared in a repository.

Data will be available as soon as possible.

We will pursue a persistent identifier.

**Are any restrictions on data sharing required?**

No actions will be taken to overcome restrictions to location data in order to protect the peregrine falcons.

No exclusive time needed.

Publications based on these data will require owners consent and proper credit.

**Responsibilities and Resources**

**Who will be responsible for data management?**

Søren Møller and Knud Falk will be responsible for all aspects.

**What resources will you require to deliver your plan?**

No extra resources needed.

## TYPE 1: DATA SET PART OF A PUBLICATION– *EUROPEAN WIND ATLAS*

### DMP TITLE

#### ADMIN DETAILS

**Project Name:** Type 1: Data set part of a publication

**Principal Investigator / Researcher:** Niels Gylling Mortensen

**Description:** Sample DMP for data set published as part of a publication; here the European Wind Atlas (DTU Orbit ID: 5916560)

**Institution:** Danmarks Tekniske Universitet / Technical University of Denmark

### DATA COLLECTION

*DESCRIBE THE DATA THAT WILL BE COLLECTED.*

- Observed and generalised wind climates, see Appendix D in the European Wind Atlas.
- Observed data (OWC) collected by project participants from EU12 countries.
- Modelled data (GWC) analysed by Risø National Laboratory, now DTU Wind Energy.
- Files are ASCII text and formats are described in Appendix D of the European Wind Atlas. OWC data are in TAB file format, GWC files are in LIB file format.
- Size of data set on disk is 1.60 MB (402 files in 13 folders)
- Files are organised in folders according to country. Files have names according to MS-DOS/MS Windows LOCATION.XXX; there is a ToC in the root folder.
- No version control; data are as published in European Wind Atlas (1989).

*DESCRIBE ANY RESTRICTIONS TO THE DATA.*

No specific restrictions on data usage

- Data are in public domain since 2015 and can be downloaded from dtu.dk.
- No specific ethical or legal issues to be considered.

General requirements for use of data

- Data should be referenced according to current best practice.
- Data should be copied and distributed as a complete data set only.

### DATA STORAGE

*DESCRIBE THE IT INFRASTRUCTURE TO BE USED.*

- Data set is stored in Media Library on DTU Sitecore server.
- Data are backed up by AIT with Sitecore Media Library.
- No access control with download. Access to original data set by DTU staff only.
- Data was distributed within project upon request to Risø.
- Data are not sensitive and no specific security is applied.
- Data set is distributed as ZIP-file to protect data integrity.

## DOCUMENTATION

*DESCRIBE THE METADATA TO BE ASSOCIATED WITH THE DATA.*

- Metadata are given in European Wind Atlas (DTU Orbit ID: 5916560).
- Publication defines metadata requirements and propose standard metadata.
- Metadata are generated by project participants, authors and editor.

*DESCRIBE THE TYPES OF DOCUMENTATION THAT WILL ACCOMPANY THE DATA.*

- Data set is documented in [European Wind Atlas](#) (DTU Orbit ID: 5916560).
- Complete methodology in publication, including definitions and assumptions.
- Software tools used to generate modelled data set (WASP) published at the same time. WASP software is proprietary; all versions still available at DTU Wind Energy.

## DATA SHARING

*DESCRIBE WHICH DATA WILL BE SHARED.*

- Data set is one reference data set for European wind climatology.
- Data set can be visualised and explored using WASP software as reader ([www.wasp.dk](http://www.wasp.dk)). Data can be read by any software reading ASCII text files.
- Complete data set can be shared and is available for download.
- Data set is now in public domain on Zenodo.org.

*DESCRIBE HOW THE DATA WILL BE SHARED FOR POSSIBLE REUSE.*

- Data set has been commercially available since publication of European Wind Atlas in 1989.
- Data set is presently freely available from Zenodo.org.
- Data set discoverable through web search of publication or data set.
- No data set identifier available; only publication ISBN 87-550-1482-8.

## LONG-TERM PRESERVATION

*DESCRIBE HOW DATA WILL BE ARCHIVED BEYOND THE SCOPE OF THE RESEARCH PROJECT.*

DTU Wind Energy has no detailed strategy for long-term archiving of data or code.

- Criteria: published data set should be archived as long as publication exists, i.e. permanently.
- Archive: no specific IT-infrastructure exist for this type of data set at DTU. Archived at Zenodo.org.
- Readability: ASCII text data, but in the hands of the current data set custodian.
- Destruction: no part of the data set needs to be destroyed.
- Preservation: presently no one; preferably DTU Library (or, as a last resort, IT service).
- Archiving period: permanently, as long as publication exists.
- Funding: publisher (DTU) and custodian (DTU Library) should finance preservation.

## **TYPE 2: EXPERIMENTAL DATA NOT PART OF A PUBLICATION - ØRESUND EXPERIMENT DATA BANK**

### **DMP TITLE**

#### **ADMIN DETAILS**

**Project Name:** Type 2: Experimental data not part of a publication

**Principal Investigator / Researcher:** Niels Gylling Mortensen

**Description:** Sample DMP for data set published separately and not part of specific publication; here the Øresund Experiment Data Bank (experimental data)

**Institution:** Danmarks Tekniske Universitet / Technical University of Denmark

### **DATA COLLECTION**

*DESCRIBE THE DATA THAT WILL BE COLLECTED.*

1. Meteorological measurements from the Øresund Experiment, see DTU Orbit ID: 6283243
2. Data collected during the Øresund Experiment, see Bull. Amer. Meteor. Soc., 66(1985), 1403-1407
3. Files are ASCII text and format is IOC general magnetic tape format (GF-3)
4. Size of data set on disk 208 MB (456 files in 11 folders)
5. File names according to MS-DOS/MS Windows XXXXXXXX.GF3; ToC in file 1515\_002.GF3
6. Version control by compiler of the GF3 data bank; current and final version is 2.0

*DESCRIBE ANY RESTRICTIONS TO THE DATA.*

1. No restrictions apply. Data have been available from Risø National Laboratory / DTU Wind Energy / Zenodo.org since 1987
2. No specific ethical or legal issues to be considered
3. Data should be referenced according to current best practice and should be distributed as one complete data set only

### **DATA STORAGE**

*DESCRIBE THE IT INFRASTRUCTURE TO BE USED.*

1. Data set is stored in personal PC of Niels G. Mortensen, DTU Wind Energy
2. Data are backed up by Niels G. Mortensen, DTU Wind Energy
3. Access to master data set by Niels G. Mortensen, DTU Wind Energy only
4. Data was distributed within project upon request to Risø; data set is distributed as ZIP-files to protect data integrity
5. Data are not sensitive and no specific security is applied

Data bank has now been uploaded to Zenodo.org.

**DOCUMENTATION**

*DESCRIBE THE METADATA TO BE ASSOCIATED WITH THE DATA.*

1. Metadata according to GF3 standard are part of the data bank. Additional metadata in The Øresund Experiment. Data bank report (DTU Orbit ID: 6283243)
2. GF3 standard defines metadata requirements and propose standard metadata
3. Metadata are generated by project participants, authors and data bank compiler

*DESCRIBE THE TYPES OF DOCUMENTATION THAT WILL ACCOMPANY THE DATA.*

1. Data set is documented in data bank itself. Additional documentation in The Øresund experiment. Data bank report (DTU Orbit ID: 6283243)
2. Complete documentation is part of data set according to GF3 standard
3. Software tools to read and extract data set published in ZIP archive with data bank

**DATA SHARING**

*DESCRIBE WHICH DATA WILL BE SHARED.*

1. Master data set for the Øresund Experiment is unique and has been shared many times with researchers and modellers since the publication
2. Data set can be extracted and explored using simple software tools that read ASCII text files. Sample Fortran procedures are included in data set
3. Complete data set can be shared and is available for distribution
4. Data set is shared upon request to DTU Wind Energy; no restrictions apply

*DESCRIBE HOW THE DATA WILL BE SHARED FOR POSSIBLE REUSE.*

1. Data set has been available since publication of the Øresund Experiment Data Bank in 1987
2. Data set is presently available from DTU Wind Energy upon request (to Niels G. Mortensen)
3. Data set discoverable through web search of publications describing data set, The data bank report can be found in DTU Orbit (ID: 6283243) and has ISBN 87-550-1592-1

Data set has now been uploaded to Zenodo.org.

**LONG-TERM PRESERVATION**

*DESCRIBE HOW DATA WILL BE ARCHIVED BEYOND THE SCOPE OF THE RESEARCH PROJECT.*

DTU Wind Energy has no detailed strategy for long-term archiving of data or code.

1. Criteria: published data set should be archived as long as requested for study, i.e. permanently
2. Archive: no specific IT-infrastructure exist for this type of data set, stored in personal PC
3. Readability: in the hands of the current data set custodian (Niels G. Mortensen)
4. Destruction: no part of the data set needs to be destroyed
5. Preservation: presently no one; preferably Library (or, as a last resort, IT service)
6. Archiving period: permanently, as long as publications referring to data set exist
7. Funding: publisher (DTU) and custodian (DTU Library)



## **TYPE 3: MODELLED DATA NOT PART OF A PUBLICATION - WIND RESOURCE MAP FOR THE WASA DOMAIN**

### **DMP TITLE**

#### **ADMIN DETAILS**

**Project Name:** Type 3: Modelled data not part of a publication

**Principal Investigator / Researcher:** Niels Gylling Mortensen

**Description:** Sample DMP for data set published separately and not part of specific publication; here the Wind Resource Map for the WASA Domain (modelled data)

**Institution:** Danmarks Tekniske Universitet / Technical University of Denmark

### **DATA COLLECTION**

*DESCRIBE THE DATA THAT WILL BE COLLECTED.*

1. Modelled wind speed, wind power density, terrain elevation and ruggedness index
2. Data modelled using WASP 11 (Frogfoot) at DTU Wind Energy
3. Files are ASCII text and format is ESRI ArcGIS ASC grids
4. Size of data set on disk 92 MB (8 files) in eight ZIP files
5. Files generated in MS Windows with accompanying metadata document in PDF
6. Version control by date of generation; three versions have been published, last version valid only

*DESCRIBE ANY RESTRICTIONS TO THE DATA.*

1. No restrictions apply. Data are in public domain and can be downloaded from project download site <http://wasadata.csir.co.za/wasa1/WASADData>. Last published data set is the one in effect
2. No specific ethical or legal issues to be considered. Metadata document contains disclaimer
3. Data should be referenced according to current best practice

### **DATA STORAGE**

*DESCRIBE THE IT INFRASTRUCTURE TO BE USED.*

1. Data set is stored in project download site, <http://wasadata.csir.co.za/wasa1/WASADData>, in project SharePoint site at DTU and in pc's at DTU Wind Energy (Niels G. Mortensen has master data set)
2. Data are backed up by AIT (SharePoint site) and by Niels G. Mortensen in DTU Wind Energy
3. Access to **master data set** via Niels G. Mortensen only
4. Data shared within project via SharePoint site at DTU
5. Data are not sensitive and no specific security is applied

**DOCUMENTATION**

*DESCRIBE THE METADATA TO BE ASSOCIATED WITH THE DATA.*

1. Metadata are project-specific and given in separate, non-published document (four-page PDF)
2. Document defines metadata requirements and propose standard metadata
3. Metadata are generated by project participants, authors and data set compiler

*DESCRIBE THE TYPES OF DOCUMENTATION THAT WILL ACCOMPANY THE DATA.*

1. Data set is documented in metadata document (unpublished) and several project reports and conference papers (all available in DTU Orbit, e.g. ID: 89873100 and ID: 89873158)
2. Complete methodology in publications, including definitions and assumptions + metadata PDF
3. Software tools used to generate modelled data set are available; methodology described in the literature and in project reports

**DATA SHARING**

*DESCRIBE WHICH DATA WILL BE SHARED.*

1. Data set is one reference data set for South African wind climatology
2. Data set can be visualised and explored using common GIS software, including open source
3. Complete modelled data set can be shared and is available for download
4. Data set is now in public domain on project download site **upon registration**

*DESCRIBE HOW THE DATA WILL BE SHARED FOR POSSIBLE REUSE.*

1. Data set has been available since end of Wind Atlas for South Africa project, Phase 1
2. Data set is presently available from project download site in South Africa  
<http://wasadata.csir.co.za/wasa1/WASAData>
3. Data set discoverable through web search of publications or project web sites. No data set identifier is currently available

**LONG-TERM PRESERVATION**

*DESCRIBE HOW DATA WILL BE ARCHIVED BEYOND THE SCOPE OF THE RESEARCH PROJECT.*

DTU Wind Energy has no detailed strategy for long-term archiving of data or code

1. Criteria: published data set should be archived as long as publications exist, i.e. permanently
2. Archive: no specific IT-infrastructure has been identified for this type of data set; project download site is temporary, DTU Sharepoint site is also temporary
3. Readability: data in industry-standard formats. Long-term: in the hands of the data set custodian
4. Destruction: no part of the data set needs to be destroyed
5. Preservation: presently no one; for the future, preferably Library (or, as a last resort, IT service)
6. Archiving period: permanently, or as long as publications exist
7. Funding: publisher (DTU) and custodian (DTU Library)

## Tek\_DTUB\_DTU Space

**Project Name** DTU Space - Geomagnetic Ground Stations

**Project Identifier** Tek\_DTUB\_DTU Space

**Principal Investigator / Researcher** Anna Willer

**Project Data Contact** anna@space.dtu.dk

**Description** Pilot project for Data Management i Praksis: Geomagnetic data from observatories in Denmark and Greenland. DTU Space is monitoring the Earth's magnetic field with a network of ground stations in Greenland, Denmark (Brorfelde) and the Atlantic Ocean (Tristan da Cunha). There are 17 in total, of which 3 are observatories (with building / housing installations): Qaanaaq (1956), Qeqertarsuaq (1926) and Narsaruaq (1983). Measurements are a continuous record of variations in the Earth's magnetic field. A detailed description can be found on the group's homepage:

[http://www.space.dtu.dk/english/Research/Scientific\\_data\\_and\\_models/Magnetic\\_Ground\\_Stations](http://www.space.dtu.dk/english/Research/Scientific_data_and_models/Magnetic_Ground_Stations)

**Institution** Danmarks Tekniske Universitet / Technical University of Denmark

### Data Collection

#### Describe the data that will be collected.

- The data that will be collected contain measurements of the variations in the geomagnetic field. Each file contains data for one station over one day with a resolution of 1 second.
- Each file is approximately 1 MB. The estimated total amounts of data are several GB.
- For most of the stations, digital data is available dating back to around 1980.
- The data are automatically sent in ascii format, then transformed to cdf format and stored on a local server at DTU Space.
- The data are manually quality controlled and calibrated one day and one station at a time. This is very time consuming (and at the moment it is very delayed), but generally rather common in geomagnetic ground data quality control to guarantee that nothing naturally is removed. However, we would like to make an intelligent automatic quality control but we do not have the manpower to develop that at the moment.
- Formats and structure for the quality-controlled data are the same as for the raw data.

The files are structured by:

1. IAGA-code (which is a three letter code, unique for each station) + one digit.
2. date: year,month,day

For example: THL6\_20160318.cdf

The data are arranged in two arrays, "time" and "HEZ" (H = local magnetic north, E = local magnetic east, Z = vertical down), where the latter contains data the three component vector (HEZ(\*,1) gives H-data, HEZ(\*,2) gives E data and HEZ(\*,3) gives Z-data). The data format is cdf, where information about the data is embedded in the file, such as for example units. In the near future these files will also include the so-called quiet-levels, which will mean that each file will include one more \*x3 matrix of data.

The quality-controlled and calibrated data will finally be uploaded to a server, where user can have access to them via a new webpage.

In addition, we would like to have a near-realtime-plot of the raw data on the webpage, preferably plotted with D3.json.

#### Describe any restrictions to the data.

The data are the intellectual property of DTU Space.

Data policy:

1. The geomagnetic data from the ground magnetic stations, in the following referred to as 'the data', are acquired with the magnetic stations operated by the Technical University of Denmark, National Space Institute (DTU Space). The data are the intellectual property of DTU Space.
2. The data may be freely used for the purpose of illustration, for teaching, and for scientific research. DTU Space encourages users of the data to collaborate with the Project Scientist at DTU Space\*. Users are also encouraged to consider a joint authorship with the Project Scientist in those cases where the data contributes significantly to a publication. DTU Space must be properly acknowledged when the data is used in publications\*\* and presentations\*\*\*. Kindly inform the Project Scientist at DTU Space when using the data in

- presentations, publications or reports and send a copy of all published reports and articles to the Project Scientist at DTU Space to facilitate the documentation of the data's usefulness. When data of the stations MCN, MCG, MCE, MCW, or MCR are used, the National Science Foundation (NSF) and the University of Michigan must also be acknowledged.
3. The data must not be passed on to others or must not be included in a distributed data base without explicit agreement from DTU Space.
  4. The data or products derived from the data must not be used for commercial purposes or in the development of commercial services without the explicit agreement from DTU Space\*.
  5. The data and products derived from the data are supplied without warranties, expressed or implicit. In no event shall the DTU Space be held liable for any damage or loss of any kind resulting from the use of the data or derived products.

*	Rico Behlke, ( <a href="mailto:rico.behlke@space.dtu.dk">rico.behlke@space.dtu.dk</a> )
**	E.g.: 'Magnetometer data from the Greenland Magnetometer Array were provided by the National Space Institute at the Technical University of Denmark (DTU Space)'.
***	Refer to DTU Space as the source of the data.

## Data Storage

### Describe the IT infrastructure to be used.

Raw data and results are stored on the department's O-drives that are shared within specific user-groups. Quality-controlled data should additionally be stored on a ftp-server or similar, where users can get access to them via our webpage. The server should preferably be located at DTU Space and have automatic backups.

Backup: the department provides backup of the O-drives.

## Documentation

### Describe the metadata to be associated with the data.

The data are stored in cdf-files, which are self-describing (each file contains data and description of data, for example name of array and units). Each filename is unique and contains IAGA-code to identify which station it is and year, month and day to identify the date. This metadata are generated automatically.

### Describe the types of documentation that will accompany the data.

Data will be described on the webpage where the users can see the near-realtime plots of the raw data and download quality-controlled data in cdf-files. The description will include the definition of the variables, the instrumentation and the location of the measurements. There will also be quality-control reports available for download.

## Data Sharing

### Describe which data will be shared.

We want to share the best possible data, which is the quality-controlled data including quiet-levels. Near-realtime plots of the raw data are visible on the webpage for the public - but the numerical data will not be accessible from our webpage until it is quality-controlled. The raw data will not be shared.

For example C, FORTRAN, Java, Perl, C#/Visual Basic, IDL, MATLAB and Python can read cdf files.

The data may be freely used for the purpose of illustration, for teaching, and for scientific research. We would like to have a system where each user types their name, email-adress, institute (<--optional) and the datarequest (e.g. start/end date and name of station), whereafter they get access to the data and that the email-adresses will be stored (together with the download-date) in order to get an overview of the number of users.

The data or products derived from the data must not be used for commercial purposes or in the development of commercial services without the explicit agreement from DTU Space.

We would like to have a near-realtime data illustration from all our stations. It would therefore be ideal to have an interactive map that covers Greenland, Faroe Islands and Denmark with a small black dot at every station. The main page represents the near-realtime data with a transparent circle for every station with a radius that represents the amplitude of the data. The circles will thereby "pulsate" with time.

5 buttons are always present in the bottom of the page: information, instrumentation, contact, data request and disclaimer.

if you hover a station the data will be shown in a marked area over the buttons. Three components are shown (or maybe 4...) around their meanvalues.

if you click on a station you will be able to choose if you want to see the data, information about the station, access to the data archive, or long term statistics.

We would prefer that it is made by D3js.

**Describe how the data will be shared for possible reuse.**

The near-realtime plot of the raw data will be visible in near-realtime. The quality-controlled data will typically be available after a couple of months. We would like to share the data via our webpage.

**Long-term Preservation**

**Describe how data will be archived beyond the scope of the research project.**

Geomagnetic data from ground-based stations in Greenland and Denmark are of great value for the scientific community and indirectly to the public, as it helps us understand and model the Earth's magnetic field.

We want two sets of data to be archived:

1. Raw data including flagged quality-controlled data - in the long-lived file format cdf.
2. Quality-controlled data including quiet levels - in cdf format.

No data has to be destroyed.

The data should be preserved permanently.

The data can possibly be hosted permanently at DTU Space.

## KASOC Data Management Plan

**Project Name** KASOC

**Project Identifier** 1

**Principal Investigator / Researcher** Jørgen Christensen-Dalsgaard, Hans Kjeldsen, Aarhus University

**Project Data Contact** Rasmus Handberg, rasmush@phys.au.dk, +4587155851

**Description** This data management plan covers data from the Kepler Asteroseismic Investigation part of the NASA Kepler project. The Kepler project launched a satellite in 2009 with the primary goal to search for extrasolar planets, meaning planets outside our solar system. The current research project uses data from the Kepler telescope to characterize planet-hosting stars. Satellite data from NASA are imported and processed for asteroseismic purposes. Apart from regular publications, the outcome of the project is a large amount of KASOC data products that will be of value for researchers many years ahead. The data management plan is focused towards facilitating that long-term preservation.

**Institution** Det Kongelige Bibliotek / The Royal Library

### Data Collection

**What data will you collect or create?**

The KASOC database ([Kepler Asteroseismic Science Operations Center](#)) consists of research data related to the [Kepler mission](#) and the associated research organization. The data is being made accessible through the [KASOC website](#) for the members and working groups of KASC (Kepler Asteroseismic Science Consortium). The KASOC database and website is being operated by the Stellar Astrophysics Center at Aarhus University. The operation of KASOC is covered by regular university funding and is not dependent on project funding.

The datasets are based on time-series of observations of light emissions from stars, performed by the Kepler satellite, launched on March 6, 2009 by NASA. There are two kinds of observations represented in the datasets: 4000 short-cadence target stars, having been observed once per minute in a period up to 4.5 years. And 160.000 long-cadence target stars, having been observed once every 30 minutes for a period of up to 4.5 years.

The *Kepler* mission was extended in 2012 and currently (Jan. 2016) continued as the K2 mission, still collecting data from the Kepler satellite and building on these data. This data management plan comprises data from both the *Kepler* and K2 missions. The extended K2 mission is expected to last as long as the Kepler satellite remains operational, probably several years. KASC is expected to exist at least for that period of time, KASOC (the database and website) possibly longer.

The data is consisting of

- **Raw data**, as supplied by NASA, copied from the [Mikulski Archive for Space Telescopes \(MAST\)](#). The raw data consists of both pixel data and lightcurves (extracted brightness as a function of time). The KASOC database is not a primary archive site for this data.
- **Processed data**, consisting of corrected data products, produced by researchers associated with KASC and its working groups. The processed data contain extracted time series of stellar brightness, frequency power spectra, as well as extracted parameters. The KASOC database is the primary archive for all the processed data from KASC.
- **Documentation**, describing the target star selection as well as the datasets available at the KASOC site.
- **Publications**, by KASC members and relating to the Kepler datasets.
- **Groundbased data**, records of star data observed by telescopes from earth and

supplied by third parties.

- **Release notes**, describing a given release of a data product and its dependencies.
- **Scripts** that have been used for data processing.

All data are stored in a database, and datasets and documents are under version control.

The main storage format for the data is [FITS \(Flexible Image Transport System\)](#) which is a standard exchange and archival format used for images and complex data in astronomy. FITS is an old and well established format and is being extensively used by astronomers for image data and related information. The format is [standardized](#), independent of suppliers, well supported by software and well documented. For a detailed description and evaluation of FITS for archival purposes, see the Library of Congress [format description page](#). FITS is extensively used in the astronomer and astrophysicist community and therefore excellently suited for data sharing.

The FITS files constitute the primary body of files. Apart from the FITS files, the database holds an ASCII file for each FITS data file - with the exception of pixel files - with header and time series information extracted in a tabular format. Some of the earliest files, however, are only produced as ASCII files.

Documentation and publications are in PDF format.

The total volume of data associated with the KASOC database and website is currently around 100 TB.

### **How will the data be collected or created?**

Raw pixel and lightcurve data are downloaded in a FITS file format from the MAST archive at regular basis and saved as part of KASOC. Single files are downloaded via FTP, whereas complete data packages (5-6 TB) are delivered by NASA on physical media. Each FITS file has a checksum written into the file header. The checksum is not being validated upon reception of data, but in practice no subsequent data processing will be able to work with corrupted data. The metadata is kept in a MySQL database, whereas data files are saved onto the large storage disks at Aarhus University IT Centre.

The MAST archive only delivers the latest version of any file or document, even though NASA manages the files under versioning control. In case KASOC needs to fetch an older version of a file, this can be done by personal inquiry to MAST. Different versions of NASA files will not be discernable through file names, but only through header data referring to the release notes covering the version in question. The KASOC database holds a record for each version of the file with an added KASOC version number and a reference to the proper file in question. File names as such are unchanged. On the KASOC website, data files are sorted according to the different data releases.

The KASC working groups prepare new derived data, as well as corrected data, based on the imported NASA data. The deposition workflow prescribes that each new data file uploaded to KASOC must be accompanied by a single ASCII file, listing which data files were used to generate the new data. The upload to KASOC must also contain a release note describing the correction procedures that have been performed on the data.

Data files from researchers are being retrieved by KASOC that in turn process the data into data products in a uniform way. Files are being accepted by KASOC in either FITS or ASCII format. There are no specific requirements for the exact format of the files being delivered to KASOC or for the transmission protocol (could be FTP, HTTP or similar).

All files and documents produced by KASC and its working groups, are under version control. Version numbers are positive integers incremented by one for each new version. The version number is indicated in file headers for FITS files (header field VERSION) and is added as the last part of the file name.

## **Documentation and Metadata**

### **What documentation and metadata will accompany the data?**

In order to interpret and reuse KASOC data, the following documentation and metadata will be in place:

1. Metadata fields are present in the FITS/ASCII file header, describing details of the observation as related to the *Kepler* mission. The header metadata in KASC produced files are documented in *KASC User Requirements Specifications: Archiving of corrected data in KASOC. March 2015*. This metadata is supplied either by NASA (for original *Kepler* data) or by the KASC researcher uploading the data file to KASOC.

2. Data release notes supplied by either NASA (for files from the NASA MAST archive) or KASC. Data release notes from NASA are organised according to Quarters (*Kepler*) or Campaign (K2). The release notes document what kind of processing has been done on the files covered by the release.
3. Dependency information: Each datafile in KASOC is accompanied by information on which earlier data releases from *Kepler* or KASC it builds upon. The dependency information is supplied by KASC members as part of the upload workflow.

Further documentation on the algorithms behind the processing is mostly only available in published articles. Related publications are indicated through their Bibcode (reference to the [SIMBAD](#) and [ADS](#) astronomical databases).

The general NASA documentation for the *Kepler* mission, such as the Data Processing Handbook and Data Characteristics Handbook can be accessed from the [KASOC Documentation page](#).

All metadata about the files, including versioning information, is being stored in the KASOC relational database. The data model for metadata storage is self-developed by the research team. Apart from the KASOC web interface, metadata can be searched as a [Virtual Observatory resource](#) through a [Virtual Observatory TAP API](#).

## Ethics and Legal Compliance

### How will you manage any ethical issues?

All data in KASOC are open data based on observations of stars by the Kepler spacecraft. There are no issues of person-related data, anonymisation, consent or the like.

To the best of knowledge, there is no data which is either confidential or controversial to publish, with respect to ethical issues of any kind. All the original *Kepler* data are treated by NASA as completely open data, and in line with this, all data in KASOC are open data.

### How will you manage copyright and Intellectual Property Rights (IPR) issues?

Original *Kepler* data (raw data) are retrieved from the NASA MAST archive under the [MAST Data Use Policy](#). The policy states that:

*Data retrieved from the Mikulski Archive for Space Telescopes (MAST) are part of the public domain and need no explicit permission for use... Please acknowledge NASA and MAST when using data retrieved from MAST.*

In addition to original *Kepler* data, the KASOC database also contains KASC corrected/modified data, publications and other related data. Everything produced by KASC members and published through KASOC is considered to be owned by the author(s). Only KASC members can submit data or publications to KASOC, and by doing that and by signing up for KASC membership, researchers accept the KASC Data Policy. Although ownership remains with the individual author(s), the latest version of the KASC data policy, found in [KASC Strategies and Policies in the Extended Kepler Mission](#), states that *All data are public-* being in line with the NASA open access policy. KASOC helps authors with distribution as content providers, but claims no ownership.

The data on KASOC is being published under the KASC data policy, without any license associated with it. In addition to this, individual authors may impose policies for their particular contribution, with specific requirements for attribution and the like.

## Storage and Backup

### How will the data be stored and backed up during the research?

Metadata are being stored in a MySQL relational database, located at the Stellar Astrophysics Centre, Aarhus University. The database is accessible to researchers through the KASOC web site. The datafiles themselves are being stored in a file system and referred from the database.

All codes for the KASOC website and related systems are stored in an external SVN repository under full version control. The MySQL database is being backed up to disks every few hours. Data files are being backed up routinely by the AU IT department.

There is currently no mirror site or external backup beyond Aarhus University of the MySQL database and KASOC website. There used to be a mirror site/file system copy, and it is planned to recreate a mirror site at a different physical site than Aarhus. There are currently



talks with the [ERDA datacenter at Copenhagen University](#) about them hosting a new mirror site.

### **How will you manage access and security?**

Access to data are granted through the KASOC website only. Access to the website is regulated by login, requiring each researcher to register for one of the KASC work groups. Upon login, members of KASC are granted full access to all data of the KASOC website and database.

### **Selection and Preservation**

#### **Which data are of long-term value and should be retained, shared, and/or preserved?**

The data produced and corrected by the KASC members and workgroups is expected to remain relevant for research until at least 2065. To ensure that data remain accesible for reuse for that entire period, independantly of the continued availability of the KASOC website and database, data sets of suitable structure, content and documentation will be extracted from KASOC and deposited in a long-term preservation repository.

There is a commitment from NASA to store original Kepler data at the MAST archive for 10 years after the end of the mission. This is only a commitment to store the final processing, not intermediate versions that may have been referred to by articles. There is a possibility that data at MAST will be retained after the 10 years period is ended.

However the requirement for a 50 years preservation of KASOC data implies that the original NASA data that is currently part of KASOC, and that is the foundation for KASC research and data products, should be included in the KASOC preservation plan. Also for the future usability of the KASOC data, this is strongly recommended.

The guiding principle for structuring KASOC data into preservation packages, is to enable reuse of the data for research. The most likely unit for future research on this material, is the study of one star. It can be expected that the data can be used for research that builds on the ongoing KASC research, but also for research questions or even disciplines which are quite different. In many cases future researchers can be expected to trust the KASC supplied data products and build directly on those sources. In other cases it may be necessary for researchers to re-process original data for studying of different properties or signals.

For this reason the preservation data packages will contain all levels of data products present in the KASOC database, from NASA-supplied pixel and lightcurve data to the most elaborate physical models.

There will be one preservation package for each star, containing all the data files from KASOC for that particular star. This includes up to five levels of data or data products, covering different levels of processing and researcher involvement:

1. *Kepler* original pixel data (as retrieved directly from the MAST archive)
2. Light curves (time series) originating from NASA as well as from KASC
3. Power spectres
4. Analyses, based on light curves or power spectra (e.g. stellar oscillation frequencies and stellar properties). These analyses are typically published in an article.
5. Physical models of the stars, concluding on star mass, size and age, based on observed oscillation frequencies. Inferences drawn from these models are typically published in an article, but the models themselves are typically not.

The preservation packages will serve as self-contained data packages, and should present themselves in a way that allows future researchers to understand and assess the contents of the package without relying on external contextual knowledge. This calls for a reasonable level of documentation and metadata for the datafiles to help overview the observed target and the observation circumstances, the provenance and processing of the data, the dependencies on other data products, and references to possible articles through the NASA abstract index. For this purpose, the package will contain an annotated list of the package content, as well as links to relevant documents. The generic documentation will be collected into a special documentation package and preserved alongside with the data packages

Files can exist in several versions, each identified by version number. The preservation package for a given star will contain all files in all versions, as published research can refer to a specific version of the data. As the production of new data files and new versions of

existing files is ongoing, the generation and deposition of preservation files need to be repeated at regular interval for the active lifetime of the KASOC website. In each case of depositing, the full package with all versions of all files for the star in question, will be contained in the deposition package.

### **What is the long-term preservation plan for the dataset?**

The goal is to secure preservation for minimum 50 years, which is a lot longer than sustainability models for almost any data repository. For this reason negotiations are currently made with select data archives that are thought to provide the best possible sustainability for storage and accessibility of the data.

The data will be collected into packages, each containing all the data files relating to observations of one star, with metadata and documentation references. All the generic documentation will be collected into one specific documentation package.

Preservation packages will be stored in a [BagIt format](#) and each package will be provided with at metadata package in a standardised format. The metadata should facilitate the anticipated discovery functionality, such as searching on celestial coordinates for a point or an area on the sky. This points towards the Virtual Observatory definitions, particularly [Resource Metadata for the Virtual Observatory](#). There will need to be one metadata record per data bag, as that will be the unit for discovery and access to the data repository.

The bagit packages will be generated by the KASOC database, using an automatic procedure. Packages will need to be deposited to the data repository on regular basis as long as data production to KASOC is ongoing.

## **Data Sharing**

### **How will you share the data?**

Data are currently shared through the KASOC website. Here, members of KASC have full and unrestricted accessibility to all data for all stars, that are prepared and made available through the KASOC website. Data as such are open and not accompanied by any restrictions of distribution or use. Everybody can sign up as a member of KASC, but will by signing up agree to follow the [KASC Strategies and Policies in the Extended Kepler Mission](#).

Data downloads through KASOC is password controlled and data logged.

Data in KASOC are persistently identified by a generated unique ID [FILEID] assigned to each file as an integer number value, that uniquely identifies the data file in KASOC and makes the origin of each file traceable. A version control number is also generated, indicating the current version of the particular data product for the particular star. Generated version numbers start from 1 and will be incremented by 1 each time a new version of the data is uploaded.

For the long-term preservation repository copy of files, it is anticipated that they will be freely accessible, and that each data package (containing all KASOC files relating to one star) will be identified by a DOI. Details of the discovery, access interface and identification will depend on the functionality of the chosen data repository.

### **Are any restrictions on data sharing required?**

All data shared through KASOC are open, by the KASC and NASA data policies.

## **Responsibilities and Resources**

### **Who will be responsible for data management?**

Although KASC is an international collaboration and consortium of researchers, its management and IT operations are based at Stellar Astrophysics Centre, Aarhus University, Denmark. The responsible party for implementing and maintaining all phases of the DMP, is Aarhus University. The Royal Library will assist the creation and review of the initial plan.

For the long-term preservation activity, the chosen data repository will be responsibility for continued storage, access and possibly additional preservation actions.

As all data in KASOC are open, there are no ownership or IPR contracts or agreements anticipated. There may however be some requirements by a preservation repository for some sort of legal agreement or licensing. The responsibility for taking care of this will be with Aarhus University.

### **What resources will you require to deliver your plan?**

**what resources will you require to deliver your plan:**

The KASC has assigned a postdoc as technical responsible person for running and developing the KASOC website and database. The data centre facilities are already in place and running, including the necessary hardware and software. The necessary competences are also available, including how to process and ingest new data files and documentation. This is so closely related to the scientific methods and procedures being implied, that it needs to be taken care of by a scientist with very good IT and specifically database competences, as the responsible postdoc also serves as Database Administrator. In order to ensure continued operations, if that postdoc quit his position, it could be worth considering training one or two more people at the KASOC site.

If a new mirror site for KASOC will be established, there will be work involved in the creation, both at KASOC and the host site, to install and configure the necessary infrastructure (servers, network, storage, duplication mechanism). There will likely be an additional running fee to pay to the external host for operations.

With respect to deposition at a long-term repository, there will be some work involved, as well as some additional skills requirements. The repository may need to develop new functionality to be able to better support astronomical data and to be able to accept, handle and expose the data packages. That development may be project funded by the *DEFF Data management i praksis* project, if one of the project partners will serve as data repository.

Some technical development will be needed at the KASOC site to automatically generate the preservation data packages. On top of that an ingest workflow will need to be developed, in order to transfer the data to the repository at a regular basis. This will involve KASOC, but may also involve the data repository.

Particularly with respect to organising and designing the long-term preservation procedures, some additional library and archiving skills are needed. This includes structuring the data packages and metadata according to best practice and to best support researchers needs, persistent identification, and finding the best possible data repository.

## Technical setup for DataFlowToolkit.dk

DataFlowToolkit is divided into two systems:

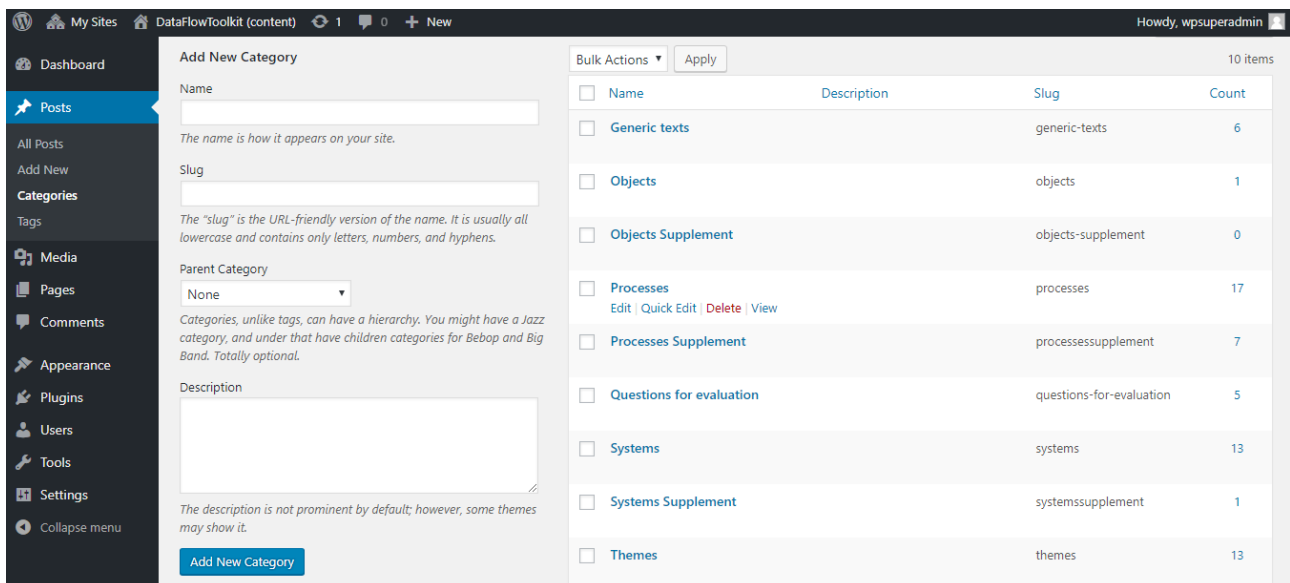
1. Backend for managing the content
2. Frontend for presentation for content

### Backend for managing the content

The backend for the system is based on a standard WordPress installation, currently running version 4.8.2. This way the toolkit can utilize user and content handling with standard WordPress functionality without writing any code, and use other plug-ins for adding additional functionality from the broad user community. Besides the WordPress standard installation, there is added a single plug-in named JSON API (<https://wordpress.org/plugins/json-api/>) for easy access to the content in JSON format. This plug-in exposes all public content as JSON by adding `json=1` as a parameter in the url when requesting single posts, posts in category, posts with tags or posts in a category with specific tags. This follows standard WordPress url structure. E.g. <https://dataflowtoolkit.dk/wordpress/?p=261&json=1> for a single post with id 261.

### Backend structure for content

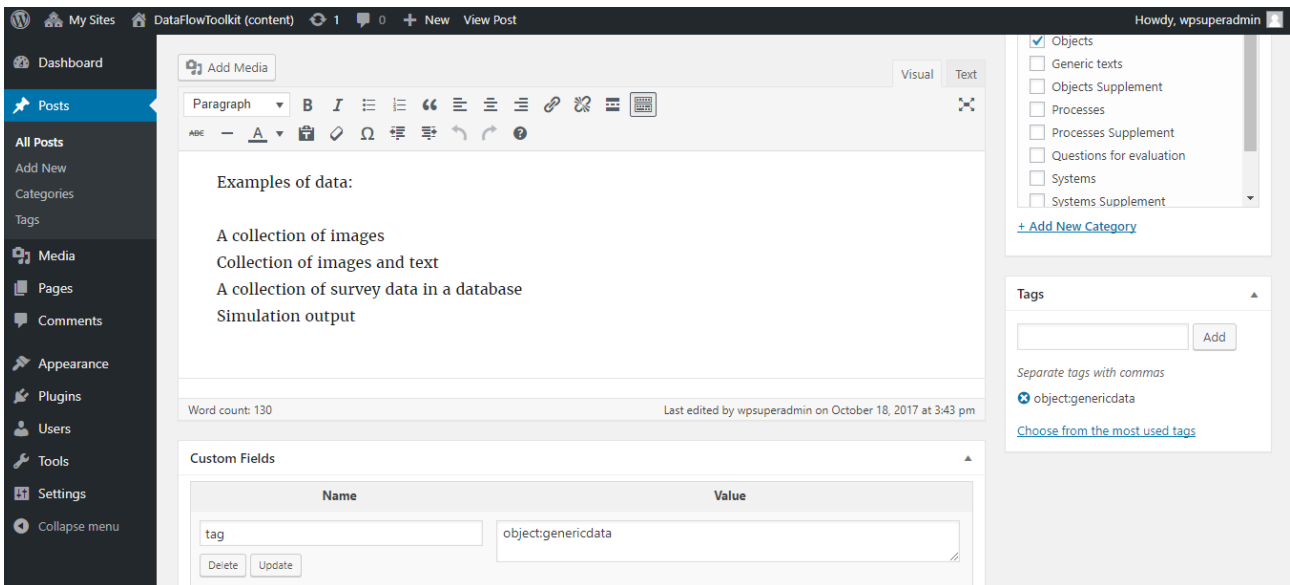
The content is structured using categories and tags. The main parts of the toolkit (data objects, processes, systems and themes) have individual categories, with categories for supplement information. This is for adding information about the a specific theme for a specific process, like security for the process create.



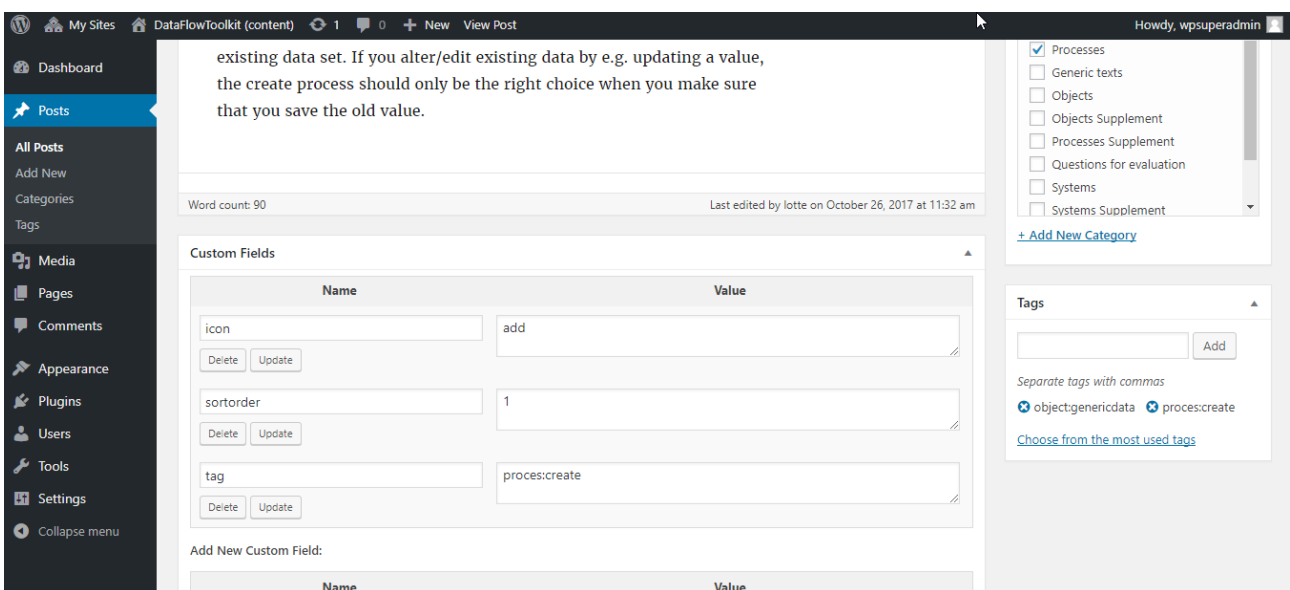
The screenshot shows the WordPress admin dashboard for 'DataFlowToolkit (content)'. The left sidebar contains navigation menus for Dashboard, Posts, Categories, Tags, Media, Pages, Comments, Appearance, Plugins, Users, Tools, Settings, and Collapse menu. The main content area is split into two panels. The left panel is the 'Add New Category' form, which includes fields for Name, Slug, Parent Category (set to None), and Description. The right panel displays a table of existing categories with columns for Name, Description, Slug, and Count. The table lists categories such as 'Generic texts' (6 items), 'Objects' (1 item), 'Processes' (17 items), 'Systems' (13 items), and 'Themes' (13 items).

Name	Description	Slug	Count
<input type="checkbox"/>	Generic texts	generic-texts	6
<input type="checkbox"/>	Objects	objects	1
<input type="checkbox"/>	Objects Supplement	objects-supplement	0
<input type="checkbox"/>	Processes	processes	17
<input type="checkbox"/>	Processes Supplement	processessupplement	7
<input type="checkbox"/>	Questions for evaluation	questions-for-evaluation	5
<input type="checkbox"/>	Systems	systems	13
<input type="checkbox"/>	Systems Supplement	systemssupplement	1
<input type="checkbox"/>	Themes	themes	13

The system can provide several guides for multiple data object types. In order to bind all this together, we utilize tags in WordPress to tie things together. Object has a custom field and tag prefixed “object:” that defines text of the tag, to use for processes, systems and themes for the specific kind of object. In this screenshot the object is named “object:generic”.



Each theme, object, process and system have a single post. The properties (like icon, sorting order, tag for relating) is defined in custom fields, and tags. In the example below, we see (1) the post for the process create for generic data and (2) the post for describing the theme metadata related to the process create for generic data using tags.

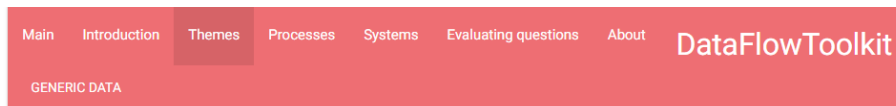


The screenshot displays the WordPress admin interface for editing a post. The left-hand navigation menu includes options like Dashboard, Posts, All Posts, Media, Pages, Comments, Appearance, Plugins, Users, Tools, and Settings. The main editing area shows the post title 'When you create ...', a text area containing the letter 'p', a word count of 3, and a 'Custom Fields' section with a table for adding new fields. The right-hand sidebar contains 'Categories' and 'Tags' sections. Under 'Categories', 'Processes Supplement' is checked. Under 'Tags', 'object:genericdata', 'proces:create', and 'them:metadata' are listed.

## Frontend for presentation

The frontend is responsible for presenting the content for the end user. This is done through custom PHP based scripts running on a standard webserver that is configured like most webhotels without any specific addons. Only requirement is to allow fetching of file contents from urls.

The screenshot shows the frontend of the DataFlowToolkit website. At the top, a navigation menu includes 'Main', 'Introduction', 'Themes' (which is highlighted), 'Processes', 'Systems', 'Evaluating questions', and 'About'. The 'DataFlowToolkit' logo is on the right. Below the navigation is a red banner with the text 'GENERIC DATA'. The main content area has a large heading 'Themes' and a paragraph: 'Both data, processes and systems are described on the basis of a number of common themes.' Below this are two dark grey boxes: 'Backup and restore' with the text 'It is important to have backup of your data in case of events t...' and 'Ethics and privacy' with the text 'There is no simple definition of research ethics and privacy, a...'. Each box has a 'MORE ABOUT' link at the bottom.



## Theme » Ethics and privacy

There is no simple definition of research ethics and privacy, and this guide is not a complete overview of ethics and privacy issues. Ethics and privacy go hand-in-hand and are interconnected. Ethics relate to your ethical conduct when working within a specific research area, and privacy refers to the need and obligation to keep research data protected, so that the subjects are protected.

Working with personal data is guarded with much control from authorities, ethics boards etc. You should always consult those funding your research project, your university and legislation on the matter in order to clarify what rules that apply to you.

The webpage utilizes the Materialize Framework by Google for presentation. The content is fetched from the WordPress site as JSON, which can be easily parsed, embedded and presented.

The source code is available upon request.