



Fase 1 rapport for projekt

Udvikling af DM Guide

DM Forum 17. maj 2017

Christina Myrup

Indhold

AFSNIT 1 FORMÅL, LEVERANCER OG DELTAGERKREDS	4
1.1 Baggrund og formålet med projektet.....	4
1.2 Projektets hovedleverancer.....	4
1.3 Status for udviklingsprojektet	5
1.4 Projektets deltagere	5
AFSNIT 2 FORELØBIGE LEVERANCER I PROJEKTET	7
2.1 Kortlægning af lignende / relevante løsninger i forhold til dm guide	7
2.1.1 Sammenhæng til øvrige projekter.....	7
2.3 Løsningsbeskrivelse for DM guiden	9
2.4 Overblik over relevante gældende regler og lovgivning.....	11
2.5 Etik og forskningsintegritet	11
2.6 Mock-up	11
2.7 Specifikation af kravene til det interaktive værktøj.....	14
2.8 Forslag til organisering af DM guiden	14
3. OVERSIGT OVER RESSOURCEBEHOV TIL FÆRDIGGØRELSE	15
BILAG	17
Bilag a: Note on RDM guides best practices	17
Bilag b: Overblik over RDM tiltag på danske universiteter	23

BILAG C: DM FORUM ACTIVITY: DEVELOPMENT OF A DIGITAL DATA MANAGEMENT GUIDE	27
Is your data quantitative or qualitative?	27
Can your data be reproduced?	27
What is the origin of the data?.....	28
How are the data collected?	28
What is the intended use of the data?	29
Which medium are the data in?	29
Do the data contain sensitive information?	29

Afsnit 1 Formål, leverancer og deltagerkreds

1.1 Baggrund og formålet med projektet

Datamanagement får i de kommende år en endnu større betydning for forskere, blandt andet fordi finansierende fonde i stigende grad vil kræve, at der foreligger datamanagement planer (DMP) som en del af ansøgningen til fondene. I DMP skal forskerne redegøre for, hvilken type data der tilvejebringes, hvor det opbevares, og hvem der har ansvaret for data. En anden grund til at datamanagement fremover får større betydning er, at flere videnskabelige tidsskrifter nu forlanger åben adgang til de data, der ligger til grund for publikationer. Endeligt ændres EU og national lovgivning omkring persondata, således at kravene til behandling af personfølsomme data bliver skærpet.

Formålet med datamanagement-guiden er at vejlede forskeren i processen med at vurdere, hvordan behovet for datamanagement løses.

Den digitale datamanagementguide (DM guide) skal understøtte forskere inden for alle discipliner i at håndtere forskningsdata korrekt og effektivt. Guiden skal støtte forskerne i håndtering af forskningsdata med det formål, at:

- Anvise lovgivning og politikker der skal overholdes
- Anvise aftaler, der skal indgås
- Anvise dækkende tekniske løsninger, dybdegående guides og mulighed for support
- Identificere eventuelle etiske problemstillinger til overvejelse.

1.2 Projektets hovedleverancer

Projektet skal levere:

- En datamanagement-guide. Guiden skal være anvendelig for forskere inden for alle discipliner og skal derfor forefindes i det nødvendige antal varianter. Guiden udvikles, så den let kan opdateres og tilpasses lokale behov. Guiden udvikles kun i en engelsksproget version.
- En overordnet specifikation af *kravene* til et digitalt interaktivt værktøj til understøttelse af guiden.
- En vurdering af om guiden skal suppleres med mulighed for at få rådgivning og support.

- En overordnet plan for udbredelse af guiden til de danske universiteter og andre forskningsinstitutioner, inklusive en redegørelse for, hvordan guiden samvirker med andre nationale eller lokale initiativer vedr. datamanagement.
- En skitse til, hvordan tilpasning og videreudvikling af guiden skal forankres organisatorisk.

1.3 Status for udviklingsprojektet

Projektet har været aktivt i perioden maj 2016 til december 2016. Fra januar 2017 har arbejdet været præget af mangel på ressourcer/deltagere til projektarbejdet. Styregruppen for projektet besluttede derfor på mødet 16. marts 2017, at projektet sættes på pause indtil 1. september, hvor styregruppen forventer, at det igen er muligt at tilvejebringe de nødvendige ressourcer (fase 2).

1.4 Projektets deltagere

Projektets styregruppe består af

- Klaus Kvorning Hansen (KU) (formand)
- Anne Marie Pahuus (AU)
- Claus Vesterager Pedersen (RUC)
- Mette Hall-Andersen (SA)
- Flemming Koch (AAU)
- Kurt Gammelgaard Nielsen (SDU)
- Mogens Sandfær (DTU)
- Lene Offersgaard (KU)
- Lars Nondal (CBS)

Projektgruppen består af:

- Christina Myrup (KU) (projektleder)
- Jens Worm (SA)
- Lærke Friis Neergaard (KU) (på barsel)
- Anne-Marie Bach (AU)
- Karsten Kryger Hansen (AAU)

- Marieke Buus (CBS) (på barsel)
- Falco Jonas Hüser (DTU)
- Mikael Karstensen Elbæk (DTU) (Udgået af gruppen)
- Anja Holm Lundeman (RUC) (udgået af gruppen)
- Ene Rammer (RUC)
- Asger Væring (SDU)
- Arne Mark Jørgensen (KU)
- Jonas Ladegaard Pedersen (ITU)

Afsnit 2 foreløbige leverancer i projektet

2.1 Kortlægning af lignende / relevante løsninger i forhold til dm guide

En tidlig opgave i projektet var at kortlægge, hvorvidt der nationalt eller internationalt findes løsninger, tilsvarende DM guiden. Konklusionen er, at der ikke er udviklingsarbejde i gang, der svarer til det, projektgruppen arbejder med i DM guide (se bilag a: Note on RDM guides best practices). Projektgruppen har endvidere kortlagt danske institutioners arbejde generelt (bilag b).

2.1.1 Sammenhæng til øvrige projekter

DataFlowToolkit

I regi af DEFF-projektet "Data Management i Praksis" er der udviklet et koncept for undervisningsmateriale - og potentiel udvikling til et interaktivt værktøj. Projektet søger som sådan ikke at skabe en guide til data management, men fungerer ud fra konceptet om at brugeren laver et flow-diagram for data, og derigennem opnår indsigt i specifikke problemstillinger relateret til vedkommendes håndtering af data. Der er således ikke tale om en guide, men et værktøj med hjælp-til-selv-hjælp, og der indgår eksempelvis ikke beskrivelse af processer som eksempelvis at skrive en data management plan. Dele af DataFlowToolkit vil kunne indbygges i en guide eller vice-versa, eksempelvis dele af indholdet til forklaring af de forskellige termer mv.

Sammenhæng Dmponline og DM guiden

Til hjælp i forbindelse med forskernes udarbejdelse af DMP findes der skabeloner og værktøjer, enten af generisk art eller tilpasset forskellige fondes og institutioners politikker og formelle krav til DMP. DMPonline er et værktøj, der er blevet udviklet af Digital Curation Centre i Edinburgh baseret på økonomisk støtte fra JISC siden 2009. Værktøjet, som nu er i version 4, har de seneste år været frit tilgængeligt som open source og benyttes ud over i UK af en række institutioner globalt som hjælp til forskerne. DMPonline er således velunderstøttet af et aktivt community, mailinglister osv. Der er blevet etableret en dansk installation, som p.t. tilpasses de danske behov (DEFF projekt). Hvor DM guiden giver hjælp og vejledning i forbindelse med forskerens generelle håndtering af forskningsdata, kan forskeren ved hjælp af DMPonline få genereret en datamanagementplan. Det vil derfor være relevant, at DM guiden linker til DMPonline.

Der er hjælpetekster i DMPonline, men til trods for dette kan det være vanskeligt for forskere, der ikke er vant til at navigere i datamanagement verdenen, at udfylde skabelonerne. Det vil derfor ligeledes være relevant med links til DM guiden, der kan uddybe begreberne og henvise til øvrigt relevant indhold.

2.2 Oversigt over typer af data

For at kunne strukturere data således, at det kan indgå i en guide, er det relevant at inddele i datatyper. Dette arbejde har krævet en del tid og ressourcer i projektgruppen. Der blev udviklet en oversigt (se bilag c: DM Forum Activity: Development of a Digital Data Management Guide), som siden blev testet på en workshop med deltagelse af forskere. På baggrund af denne workshop blev udviklet nedenstående kategorier, som udgør fundamentet for løsningsbeskrivelsen for DM guiden.

Reviderede datatyper efter workshop (input til løsningsbeskrivelsen)

For a given dataset:

Overall type of data

Is your data quantitative or qualitative?

Reproducible

Can your data be reproduced?

- Yes / No / Maybe, under certain circumstances

Origin

What is the origin of the data?

- I collect it myself.
- I receive it from a third party.
 - o Research collaborator
 - o External company
 - o Public authority
 - o Subcontractor

Collection method

How are the data collected?

- Measured with special equipment
- Obtained in interviews or surveys and observation of facts
- Aggregated from published sources, e.g. literature or databases
- Simulated from models
- Processed from other data
- Combined from different sources

Intended use

What is the intended use of the data?

- Description of other data or primary materials
- Input for further processing, e.g. analysis
- Publication without further processing

Medium

Which medium are the data in?

- Physical assets
 - o Books and Notes
 - o Photographs, Videos, Audio recordings
 - o Physical samples
- In digital form on own storage devices
- In digital form on external or not directly accessible storage devices

Sensitivity

Do the data contain sensitive information?

- Public data
- Internal data
- Person-related data
- Person-sensitive data
- Confidential data (e.g. data for patent applications, trade secrets, contractual obligations)

2.3 Løsningsbeskrivelse for DM guiden

Siden projektopstart har en del af arbejdsgruppens fokus været på strukturen for guidens indhold. Idéen om et spørgetræ har været udgangspunktet, men har vist sig ikke at rumme kompleksiteten i forskning. Strukturen i spørgetræer vil kunne bruges til meget begrænsede emner, eksempelvis i

afklaring af klassifikationen af et givet datasæt på et givet tidspunkt, men ikke som sådan som indgang til en overordnet guide for data management.

Tanken er nu i stedet en løsning der baseres på opbygning af en vidensbank om data management, som tilgås af forskeren baseret på vedkommendes forskningsmæssige profil. Denne er opbygget ved identifikation af en række karakteristika ved forskeren og dennes forskning og data. Efter at forskeren har udfyldt en profil (eller indlæst fra tidligere), får vedkommende målrettet, specifik hjælp hentet fra vidensbanken. Men kun den information, der er identificeret til at stemme overens med forskerens profil.

Dataprofil:

- Kvalitativ vs kvantitativ
- Reproducerbarhed af data
- Oprindelse af data
- Indsamlings-/genereringsmetode
- Formål
- Form; fysisk vs digitale data
- Klassifikation

Forsker- og projektprofil:

- Institution
- Ansættelse
- Bevillingsgiver
- Samarbejdsrelation

Efter at forskeren har udfyldt sin profil, vil guiden være struktureret omkring en forsimplet livscyklusmodel for forskningsdata bestående af (1) planlægning, (2) gennemførelse og (3) deling.



Det er også tanken, at man skal kunne tilgå informationen uden at skulle lave en profil, men det er ikke afklaret, hvorledes dette skal laves.

Guidens indhold er fortsat med fokus på juridiske og etiske problemstillinger og praktisk understøttelse af de problematikker forskerne oplever i forhold til håndtering af deres data.

Guiden skal opbygges særdeles modulært, og vil kunne implementeres såvel centralt i en national udgave, men vil også let kunne implementeres på de enkelte institutioner, dersom der er behov for dette.

2.4 Overblik over relevante gældende regler og lovgivning

Projektet har fokus på at danne overblik over relevant lovgivning på området. Der er udarbejdet et arbejdspapir, som er kommenteret af juridisk netværk under DeIC samt af enkeltpersoner med ekspertise på området (jf. notat *Gældende forskningslovgivning fordelt på områder, maj 2017*). Guiden skal så vidt muligt vejlede forskerne, også på det juridiske område. Det er dog meget begrænset hvad man håndfast kan skrive, idet svar på konkrete juridiske spørgsmål som oftest er kontekstafhængige. Guiden kan først og fremmest pege på hvilke problemstillinger, forskerne skal være opmærksomme på, og hvor de kan hente yderligere support.

2.5 Etik og forskningsintegritet

Guiden skal også indeholde vejledning vedrørende centrale problemstillinger på området for forskningsetik og forskningsintegritet. Projektgruppen er gået i gang med at se på hvilke emner / spørgsmål, det kan være relevant at give svar på i guiden i forhold til disse emner (jf. Bilag d: Research and Ethics). Det kan være vanskeligt at afgrænse i forhold til det juridiske på området.

2.6 Mock-up

Arbejde med udkast til guiden og specifikation af kravene til et interaktivt værktøj er foregået løbende og i sær i forbindelse med arbejdet med løsningsbeskrivelsen. I figuren nedenfor vises eksempel på, hvordan forsiden på guiden kan komme til at se ud. Forskeren kan vælge at logge ind på en eksisterende profil eller anvende guiden uden at logge ind. Forskeren kan også vælge at ændre i en eksisterende profil.

Velkommen

Data Management Guide DK er en guide til håndtering af forskningsdata.

Indstil forskerprofil

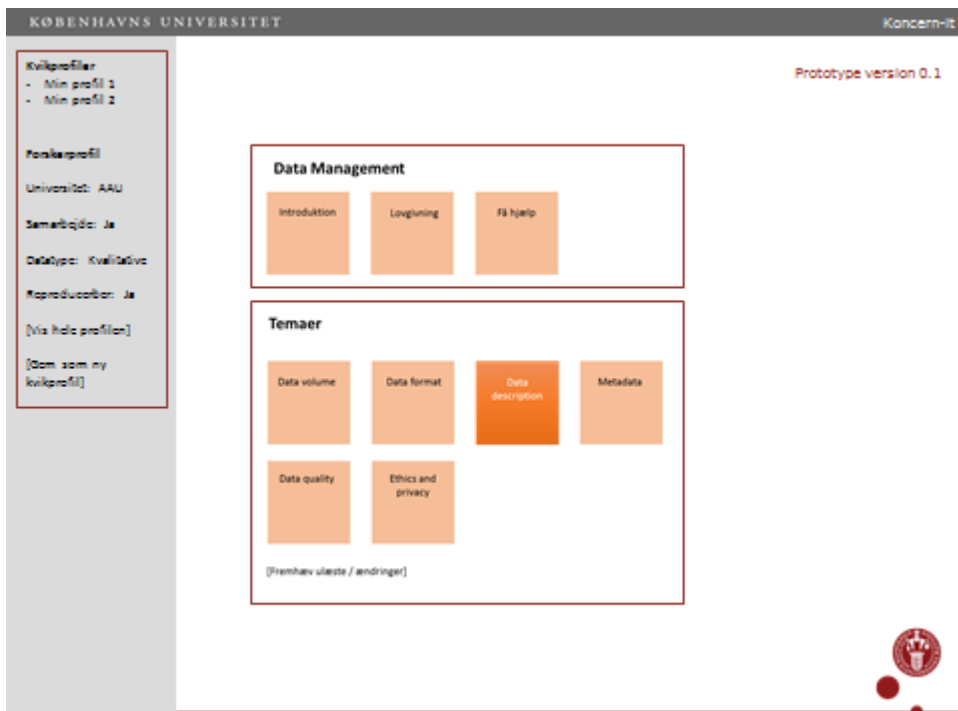
Login til eksisterende forskerprofil

Brug uden forskerprofil



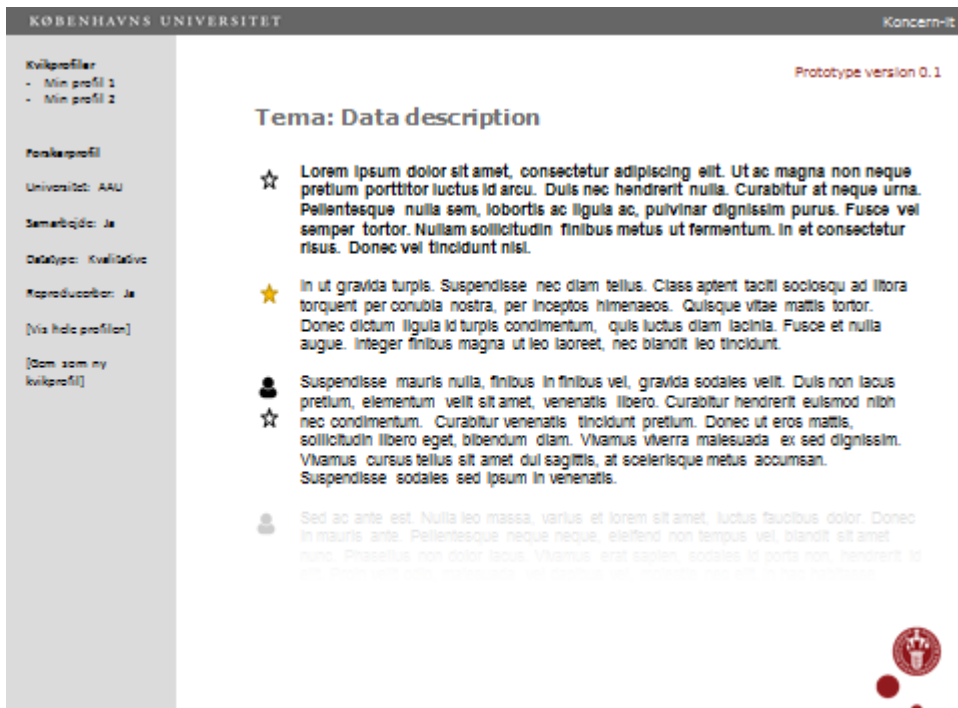
Figur 1: Side 1 i DM guiden

Efter at forskeren har udfyldt en kvikprofil vil guiden foreslå hvilke temaer og emner, det er relevant for forskeren at orientere sig i. Forskeren vil også selv kunne søge på temaer og emner for at finde svar på spørgsmål.



Figur2: Temaer i DM guiden

Såfremt forskeren søger svar på et specifikt emne, vil guiden udover at tilbyde hjælp om dette emne også foreslå beslægtede emner, som kunne have forskerens interesse.



Figur3: Eksempler på tekst og supplerende tekst

2.7 Specifikation af kravene til det interaktive værktøj

DM Guiden skal have en lagdelt og fleksibel arkitektur, når det kommer til den tekniske løsning, og i videst muligt omfang være baseret på åbne standarder og på en teknologi, der er åben for adgang til data og logik fra andre systemer via API'er.

Fleksibiliteten i løsningen skal også tilgodese, at de deltagende institutioner skal kunne præsentere egne tekniske løsninger, kontaktmuligheder mv i guiden



baseret på forskerprofilen. Alternativt skal institutionerne kunne benytte dele af guiden som teknisk platform til at lave deres egen udgave af guiden. Der stilles derfor krav om gode muligheder for versionering af guidens indhold.

- **Vidensbasen** oprettes i en løs struktur, hvor indholdselementer ligger i et fleksibelt markupsprog, eksempelvis reStructuredText. Alle elementer i vidensbasen er hostet i en GitHub-lignende struktur, der muliggør eksempelvis let versionering, eller at en institution kan lave sin egen udgave af vidensbasen, der er delvist synkroniseret til den fælles vidensbase.
- **Logik/knowledge broker** er ansvarlig for at sammensætte guidens indhold på baggrund af input til data-, forsker og projektprofil. Der opbygges en standard til repræsentation af en data-, forsker- og projektprofil, der kan returnere henvisning til hvilke elementer fra vidensbasen dette skal inkludere.
- **Præsentationen** kan foregå i et fælles system, eller hver enkelt institution kan bygge sit eget præsentationslag, men samtidig udnytte logik/knowledge brokeren samt egen og fælles vidensbase.

2.8 Forslag til organisering af DM guiden

Arbejdet med at beskrive forslag til organisering af DM guiden udestår. Der skal tages stilling til en samlet driftsorganisation og/eller lokale ansvarlige enheder.

2.9 Forslag til implementeringsplan

Arbejdet med at beskrive forslag til en implementeringsplan udestår.

3. Oversigt over ressourcebehov til færdiggørelse

Projektet har været aktivt i perioden maj 2016 til december 2016. Fra januar 2017 har arbejdet været præget af mangel på ressourcer/deltagere til projektarbejdet. Styregruppen for projektet besluttede derfor på mødet 16. marts 2017, at projektet sættes på pause indtil 1. september, hvor styregruppen forventer, at det igen er muligt at tilvejebringe de nødvendige ressourcer (fase 2). I tabellen nedenfor beskrives hvilke ressourcer der skal til for at færdiggøre udviklingsprojektet.

Tabel 1. Resterende indsatser, kompetencer og ressourcer

Indsats	Kompetencer (timer)	Ressourcer (timer i alt)
Videreudvikling af 1. udkast til prototype	It medarbejder (37), it arkitekt (2), 3 medarbejdere med kendskab til data management og guiden (37)	150
Validering af prototype (workshops og interviews)	8 forskere fra forskellige områder (3), 3 medarbejdere med kendskab til dm og arbejdet med guiden (5)	39
Tilretning af prototype	IT medarbejder (5) og 1 medarbejder med kendskab til dm og arbejdet med guiden (5)	10
Kravspecifikation til system	IT medarbejder (20) og 1 medarbejder med kendskab til dm og arbejdet med guiden (20)	40
Overslag udviklingsomkostninger til system og driftsomkostninger	IT medarbejder (4)	4
Videreudvikling af indhold (dataopbevaring, jura, forskningsetik)	Juridisk (100), dm kendskab (100), it-medarbejder (50)	250

Tilpasning af indhold og indtastning i system	Juridisk (100), kommunikationsmedarbejder (100), 2 medarbejdere med kendskab til dm og arbejdet med guiden (a 100)	400
Endelig færdiggørelse af prototype	It-medarbejder (10), generalist (10)	20
Forslag til af organisering /drift af guiden	5 Projektgruppemedlemmer (a 10)	50
Forslag til Implementeringsplan	5 Projektgruppemedlemmer (a 10)	50
Projektledelse		100
SUM timer		1.113

Estimaterne i skemaet er vurderet ud fra at deltagerne arbejder koncentreret om de enkelte indsatser i afgrænsede tidsrum, og ikke spreder ressourcerne over lange perioder.

Der efterspørges i alt 1.113 timer til at gøre arbejdet færdigt. Dette er mindre end i det oprindelige budget, hvor der blev afsat i alt 3.486 timer, hvoraf der kun er brugt 1900 timer i alt d.d.

Bilag

Bilag a: Note on RDM guides best practices

SAGSNOTAT

21. NOVEMBER 2016

Vedr. Best practices RDM guides

Sagsbehandler Lærke Friis Neergaard

Method

The method of finding best practices for RDM guides has been to google and search university pages online for RDM guides and sending request to multiple institutions worldwide on the JISC mailinglist for international RDM community and subjects.

The feedback was then critically and thoroughly reviewed. The primary criteria for usefulness to this project has been 1) if the guides have some interactive modules, 2) if it is aimed directly at the researchers in various fields, and 3) if a single guide covers more than one aspect of RDM.

Results

Just three course stands out and meets all three main criteria as mentioned above.

The three top guides/courses are:

- 1) **MANTRA**
- 2) **MOOC on Research Data Management and Sharing**
- 3) **RDM boot camp**

Following is a short description of the courses.

MANTRA

The first is the MANTRA course. It is an online course, well known and recommended within the RDM field. It is widely used in Great Britain, where it was developed by University of Edinburgh, but have also spread to an international audience.

MANTRA is aimed at researchers in various fields and even research students and RDM-advisers. It is covering all aspects of the data lifecycle.

It has 7 units on different aspects of RDM, a definition of research data and a list of tutorials: 1) Data management plans, 2) Organising data, 3) File formats & transformation, 4) Documentation, metadata, citation, 5) Storage & security, 6) Data protection, rights & access, and 7) Sharing, preservation & licensing.

It has interactive modules and video, but more in the forms of validating quizzes or exercises to confirm that the reader has understand the main points of a given module. Example of interactive modules below:

Storage and security Summary

You have now been introduced to the storage, backup and security issues in data management.

As a summary task please drag and drop the words at the bottom of the page into the gaps left in the text.

Hint: If the word you select is not the correct word you won't be able to drop it into the gap.

Through the course of your research you must ensure that you store your research data in a [] way and have [] copies in at least [] locations that are [] regularly.

You can store your research data, for example on:

- the central or College or School [] drives (highly recommended)
- [] computers and laptops (these should not be used for storing [] copies of your data)
- and [] storage devices (not recommended for the [] term storage of your data, particularly your master copies).

[backup] [master] [networked] [secure] [external]

[personal] [three] [maintained] [long]

The MANTRA course is free and online available at all time: <http://datalib.edina.ac.uk/mantra/>

MOOC on Research Data Management and Sharing

MOOC is just a general term for Massive Open Online Course. This RDM MOOC was created by the University of North Carolina at Chapel Hill in collaboration with the University of Edinburgh. The course is aimed at researchers in various fields. It covers all aspects of the data lifecycle and

has 5 modules: 1) Understanding Research Data, 2) Data Management Planning, 3) Working with Data, 4) Sharing Data, and 5) Archiving Data.

The material for the course is interactive in that it contains quizzes and video and more importantly is an actual class with teachers and other participants.

The course requires enrollment, but is free of charge. Thus, it is only available on specific dates and the modules are scheduled for 5 weeks, so it is not possible to just jump around in the material.

For an additional fee, the researcher can get a diploma for completing the course.

The course is highly recommended by the researchers who have attended the course and they give excellent comments and high ratings on the course web page.

More information and enrollment is here: <https://www.coursera.org/learn/data-management>

RDM bootcamp

The University of Bristol have collected RDM course material and made it available as an online tutorial.

The course is very broad and basic and will thus be useful for both researchers and RDM support staff, but is primarily aimed at researchers in all fields. It covers all aspects of the data lifecycle and lists 4 areas: 1) Planning and creating data, 2) Organising, structuring and documenting data, 3) Backing up and storing data, and 4) Preparing data for analysis, to share with others or to preserve for the long term.

It also contains an introduction and a section on the definition of research data and a section that motivates the researchers on why they should manage their research data.

The interactive modules are a few optional quizzes that test the reader's insight into aspects and facts of RDM at Bristol and in general. In this respect, the interactive part is inspired a lot by MANTRA.

Here is an example:

1. Which of the following would not generally be considered to be research data?

- A. Files generated by an electronic laboratory notebook (ELN)
- B. A research progress report, submitted to the funder of the research project
- C. A digitised audio recording of an interview
- D. An electronic database of results

When answered, the correct answer will be revealed:



1. Which of the following would not generally be considered to be research data?

- A. Files generated by an electronic laboratory notebook (ELN)
- B. A research progress report, submitted to the funder of the research project
- C. A digitised audio recording of an interview
- D. An electronic database of results

Answer: B is an example of administrative data

The answers are not recorded.

The course is available at: <https://data.bris.ac.uk/bootcamp/>

UC3: Ongoing project

Feedback from the RDM community through the JISC mailinglist revealed, that UC Curation Center (UC3) is also working on a RDM guide for researchers.

They have a lot of interesting point of views and have gotten feedback from the community already, that might be helpful.

The project has so far been described in two blogpost (the latest from November 2016):

Part one: <https://datapub.cdlib.org/2016/09/12/building-a-user-friendly-rdm-maturity-model/>

Part two: <https://datapub.cdlib.org/2016/11/08/building-an-rdm-model-for-researchers-what-weve-learned-so-far/>

The person in contact was John Borghi, PhD, UC Curation Center at California Digital Library

Conclusion

To make an interactive RDM guide for researchers is to break new ground internationally.

There is no current example of a fully interactive RDM guide for researchers, but there is a few excellent guides/courses with partly interactive modules, that should be examined before deciding a frame for the RDM guide.

Additional, it would be highly useful to examine the blogposts and follow the progress of UC3's RDM guide.

There is a vast variety of RDM guides available online, but they are not interactive and is often times very specifically aimed at researchers in relation to the institution who has made the guide, and/or limited to specific national conditions affecting RDM.

These guides will not be of any significant relevance for this project, however, they can be an inspiration on how to structure the guides on themes and data lifecycles.

Below is a list of examples of just a few of these guides:

- 4TU, DANS and SURF SARA has made an online course Essential 4 Data Support. It is aimed at RDM support staff, but is a great inspiration on how to structure a guide. It has limited interactive elements in that, it is possible to leave comments, has videos and links to further reading: <http://datasupport.researchdata.nl/en/>
- The UK Data Archive's guide on best practices on RDM for researchers: <http://www.google.dk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiT34X33LnQAhUDBywKHVc6CaoQFggZMAA&url=http%3A%2F%2Fwww.data-archive.ac.uk%2Fmedia%2F2894%2Fmanagingsharing.pdf&usg=AFQjCNFSrcgQKbVetekSP7CoUzSZTcgyPA&bvm=bv.139250283,d.bGg>
- European University Institute Library's guide on RDM for researchers: <http://www.eui.eu/Documents/Research/Library/ResearchGuides/Economics/PDFs/EUI-Research-Data-Services-Guide-4.0-2016.pdf>
- Harvard University's guide on RDM for researchers working with sensitive data: <https://catalyst.harvard.edu/pdf/regulatory/Investigators%20Guide%20to%20RDM%20practice.pdf>
- Berkeley University's guide on RDM best practices: <http://researchdata.berkeley.edu/data-best-practices>
- Cambridge University's guide on RDM for researchers: <http://www.data.cam.ac.uk/data-management-guide>
- Washington University's guide on RDM for researchers: <http://guides.lib.uw.edu/research/dmg>
- 10 simple rules for RDM for researchers: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003542>
- Although interactive in some way, this USGS' guide on RDM is more a online lecture, but is inspirational in that it is very focused on research practice and provides a lot of examples and

exercises. It is aimed at geological research mainly and is a research department under the U.S. government:

<https://www2.usgs.gov/datamanagement/training/modules.php>

Bilag b: Overblik over RDM tiltag på danske universiteter

SAGSNOTAT

6. DECEMBER 2016

Vedr. Forskningsdatamanagement på danske universiteter

Baggrund: Formålet er, at skitsere status for forskningsdatamanagement (RDM) på danske universiteter.

Følgende universiteter har pt konkrete tiltag indenfor RDM-området og vil blive præsenteret i det følgende: Aalborg Universitet, Dansk Teknisk Universitet, Copenhagen Business School, Syddansk Universitet og Århus Universitet og Københavns Universitet.

ITU og RUC har hverken strategi, organisering eller konkrete tiltag indenfor RDM-området og er ikke medtaget. RUCs bibliotek er dog involveret i udarbejdelse af en dansk template til DMPOnline¹ og overvejer at implementere DataCite.

Alle universiteter i DK er medlem af de nationale fora for [RDM i DeIC](#). Se f.eks. [medlemsliste](#) for DM Forum.

Metode: Al information er hentet via institutionernes hjemmesider og oplæg i DM guide P projektgruppen. Der kan være aktiviteter, organisation og værktøjer, som ikke er medtaget pga. sagsbehandlers manglende kendskab.

- **Danmarks Tekniske Universitet (DTU):**

Enheder og organisation: DTU har en enhed på 3 ansatte, der beskæftiger sig med bredt med RDM og yder support til forskerne. Enheden er centralt placeret som en del af Afdeling for Innovation og Sektorudvikling.

¹ Et frit tilgængeligt online værktøj til at udforme templates og udfylde Data Management Planer (DMP'er) for forskningsdata. DMP'er er et krav i Horizon2020 og hos andre, især engelske funders. Udviklet af Digital Curation Center (DCC) i Storbritannien. Læs mere: <http://www.dcc.ac.uk/dmponline>

I forbindelse med pilotimplementeringen af DTUs nye RDM politik på 3 institutter, er det hensigten at etablere et forum for samarbejde mellem bibliotek, it og forskersupport.

Formidling og undervisning: DTU har en hjemmeside specifikt for RDM og RDM-enheden:

<http://www.dtu.dk/forskning/om-dtus-forskning/forskningsdata-management>

Politikker og retningslinjer: DTUs RDM politik er blevet endeligt godkendt og vil snarest blive offentliggjort på deres hjemmeside (se bilag). Det har taget et år at udarbejde og få godkendt politikken og det er gjort som led i implementeringen af Danish Code of Conduct.

Ud over politikken er der formuleret en række procedure, der er mere fleksible for tilpasning på det enkelte institut.

DTU har været inspireret af Royal Melbourne Institute of Technology i Australien, hvor der både er policy og procedures.

Værktøjer: DTU er dansk distributør af DataCite til DOI'er, et værktøj der sikrer at forskningsdata overholder FAIR-principperne.²

Derudover har DTU udviklet en DTU-template til DMPOnline: <https://dmponline.dtu.dk/>

Et igangværende projekt er, at finde en arkivløsning til forskningsdata, der kan overholde alle krav. Senest har de undersøgt det tyske Radar.

Et andet igangværende projekt undersøger hvilke Electronic Lab Notebook (ELN) løsning der skal implementeres på DTU. Det er noget som deres egen undersøgelser peger på, at forskerne er interesserede i og efterspørger.

Andet: DTU har generelt lavet mange undersøgelser om RDM-behov og ønsker blandt deres forskere og på deres institutter. De har derfor også indirekte skabt interesse og ambassadører for RDM på DTU.

- **Aalborg Universitet (AAU):**

² FAIR-principperne for forskningsdata er bredt anerkendte, f.eks. i Horizon2020. FAIR står for Findable, Accessible, Interoperable og Re-usable: <http://www.datafairport.org/fair-principles-living-document-menu>

Enheder og organisation: [Karsten Kryger Hansen](#) er netop blevet ansat i en nyoprettet stilling som Datamanagementkoordinator på AAUs bibliotek. Der er nyligt oprettet et strategisk råd for forskning og innovation, der bl.a. bliver tovholder på RDM aktiviteter.

Der er planer om at oprette et RDM storcenter. Der er ligeledes planer om at etablere en RDM support, som et samarbejde mellem funding (forskingsupport), it og biblioteket.

Formidling og undervisning: RDM fylder ca. 1½ time af et obligatorisk 1-dags kursus for alle ph.d.-studerende.

Politikker og retningslinjer: Det forventes, at der startes tiltag til at udarbejde en RDM politik for DTU ultimo februar 2017. Der findes allerede politikker for Open Access og OpenData og politikken bliver formentlig en viderebygning af disse, samt inspireret af DTU's RDM politik og forlæg med at supplere politikken med processer.

Værktøjer: Biblioteket arbejder på at kunne tilbyde DataCite.

Andet: RDM er blevet et fokusområde på AAU og der er opmærksomhed på området. Flere dekaner har udtrykt ønske om at blive opkvalificeret og opdateret indenfor RDM. Der er både interesse og engagement i området, vurderer Karsten Kryger.

- **Syddansk Universitet (SDU):**

Enheder og organisation: Der findes 2 følgegrupper for RDM, én for administrative enheder og én med deltagelse af forskere fra alle fakulteter. Der afholdes koordinerende møder mellem de to følgegrupper.

Formidling og undervisning: Der afholdes obligatoriske 2 dages RDM kurser for samtlige ph.d.-studerende. SUND har opdelt deres i et kvalitativt og et kvantitativt:

http://www.sdu.dk/en/Forskning/PhD/Phd_skoler/PhdSkolenSundhedsvidenskab/PhD_Students/PhD_Courses/Courses/Folk_data_documentation og

http://www.sdu.dk/en/Forskning/PhD/Phd_skoler/PhdSkolenSundhedsvidenskab/PhD_Students/PhD_Courses/Courses/Folk_responsible_management

Politikker og retningslinjer: SDU er påbegyndt processen med at formulere en RDM politik. Der hentes inspiration fra DTU, både i indhold og i adskillelsen mellem politik og processer.

- **Copenhagen Business School (CBS):**

Enheder og organisering: CBS har pt en fuldtidsstilling som forskningsdatamanager, besat af [Mareike Buss](#). Der er planer om at nedsatte et fast organ for RDM på CBS, dvs. en følgegruppe, som der er på andre universiteter, så snart en RDM politik er vedtaget.

Politikker og retningslinjer: En RDM politik er snart helt færdig bearbejdet.

- **Århus Universitet (AU):**

Enheder og organisation: Der er nedsat et e-Science udvalg med 4 medlemmer. Udvalget skal i første omgang drøfte og formulere udkast til RDM politik for AU. Der er nedsat en Følgegruppe for datamanagement på AU, med medlemmer fra både AU og Statsbiblioteket.

Politikker og retningslinjer: Arbejdet med RDM er planlagt, men endnu ikke påbegyndt.

Politikken vil formentlig blive inspireret af DTU's. Politikken forventes at være færdig i juni 2017.

- **Københavns Universitet (KU)**

Enheder og organisering: SUND har på det organisatoriske plan nedsat en DM Task Force bestående af både VIP'er og TAP'er.. Fakultetet har en hjemmeside for RDM og obligatorisk undervisning i RDM for ph.d.'er.

HUM har nedsat en gruppe, der er i gang med at analysere RDM på fakultetet.

Politikker og retningslinjer: SUND har formuleret en RDM politik og fakultetet er i gang med at anskaffe og implementere ELN's.

Bilag c: DM Forum Activity: Development of a Digital Data Management Guide

Questions and issues regarding research data management depend strongly on the nature of the data that has to be managed. This overview lists different possibilities of how to classify research data in all relevant categories. Due to the complex nature of research data, some categories are not completely independent from each other and might overlap. A full classification, that means assigning one attribute in each category, should give a complete description of a given data set and trigger all relevant guidance.

Is your data quantitative or qualitative?

Quantitative data:

Numerical data or any data that can be converted into numbers. Any data that can be measured.

Qualitative data:

Descriptive data and plain observations that cannot be converted into numbers.

The same object usually contains both quantitative and qualitative data. The contents of a text file might be of descriptive nature, whereas the number of words and the file size are represented by numbers (and units), for example. It depends on the aims of the research project, which of these data are used.

Can your data be reproduced?

Yes.

Some data can be reproduced exactly under the same conditions. This is often the case for output from computer simulations when using the same input and code (and possibly the same hardware) and for experiments in the laboratory, when identical experimental setups and methods give the same results. Also data that is derived from other data is reproducible as long as the original source still exists. Statistical uncertainties may occur.

No.

Data obtained from observations at a given place and time are usually not reproducible, e.g. behavioral studies or collections of data in real time. The same holds when the original conditions cannot be restored or when the original source is no longer available, e.g. when an experiment destroys or alters the sample or when the study itself might influence the participant.

Maybe, under certain circumstances.

Data can be reproduced in principal, but it would take an unreasonably large effort, since the collection is extremely costly and time-consuming. Or there are other external factors that can have an influence on whether or not results are reproducible, e.g. when it cannot be guaranteed that the original source is accessible.

What is the origin of the data?

I collect it myself.

This means that you are responsible for all steps of the data collection and documentation. Possibly with the help of colleagues and partners but you have the main overview. This includes gathering of data that has previously been published.

I receive it from a third party.

You work with data that someone else has collected and prepared. You have an informal agreement or a contract regulating the use of these data. A third party could be:

- Research collaborator
- External company
- Public authority
- Subcontractor

How are the data collected?

Measured with special equipment.

The data are generated by an instrument in a specific file format.

Obtained in interviews or surveys and observation of facts.

The data are collected “by hand”, e.g. written down in a notebook or a template.

Simulated from models.

The data are outcome of a computer simulation with a given input and software.

Aggregated from published sources, e.g. literature or databases.

The data are a compilation of previous data that are (openly) available.

Processed from other data.

The data are the results of an analysis or editing of other data. The difference to “aggregated data” is that the act of processing implies a significant creative contribution.

Combined from different sources.

The data are collected with the main purpose of studying connections and correlations between different data sets that each fall into one of the other categories.

What is the intended use of the data?

Description of other data or primary materials.

This includes any kind of documentation, digitalization of physical objects and copies of existing data sets.

Input for further processing, e.g. analysis.

The data are actively used for research. Any kind of creative input (beyond bare description) will be performed on the data set.

Publication without further processing.

No further scientific work needs to be done on the data set. Additional metadata and documentation might still be required.

The same data set might be used for more than one purpose.

Which medium are the data in?

Physical assets.

- Analogue books and notes.
- Analogue photographs, videos, audio recordings.
- Physical samples.

In digital form on own storage devices.

The data are stored in files on a medium that you have full access to, e.g. own computers, institutional storage systems, portable devices, etc.

In digital form on remote or not directly accessible storage devices.

The data can only be stored at another institution or are locked inside a certain environment, e.g. in a closed software.

Do the data contain sensitive information?

Public data

Data that do not contain any sensitive information can be made openly available with unrestricted access for everyone. These data can be published on homepages, in public databases and journals, for example. The reuse of these data might be limited, however, through copyright licenses.

Internal data

Internal data may only be read and used by a well-defined group of people, e.g. a particular research group and collaborators. Security measures and managed access control procedures have to be in place to prevent unauthorized access. This is typically research data in a pre-publication state.

Person-related data

Data that does not contain personal information explicitly but allows for conclusions about the original source, e.g. through correlation with other (publicly available) data. This is particularly the case for interviews, observations and surveys. Which data may be published and how is often settled in personal agreements between the researcher and the participant.

Person-sensitive data

Data that contain personal information, such as names of participants, addresses, CPR numbers, etc., and clinical data. These data are protected by the Danish law on “The Act on Processing of Personal Data”. Person-sensitive data may not be published unless they are completely anonymized.

Confidential data

Other data that may not be made openly available or only after an embargo period, e.g. data for patent applications and data from industrial partners that contain trade secrets. Details are usually regulated in contracts with the respective companies or authorities.