

DM Forum møde 2.10.2018

Mødelokalitet: Rigsarkivet, Tøjhusgade 1, 1218 København, mødelokale "Harsdorf Salen"

Deltagere:

Anne Sofie Fink Kjeldgaard
Anne-Marie Bach
Asger Væring Larsen
Bo Bai
Christian Storm Pedersen

Falco Hüser
Karsten Kryger Hansen
Katrine Düring Davidsen
Katrine Flindt Holmstrand
Lene Offersgaard

Mareike Buss
Mette Hall-Andersen
Michael Svendsen
Paula Martinez Lavancy
Susanne den Boer

Sekretariat
Anders Sparre Conrad
Rene Belsø
Diba Markus (referent)

Gæster
Birgitte Vedel Thage

Dagsordenpunkt

01 - Godkendelse af dagsorden

Dagsorden blev godkendt.

02 - Godkendelse af referat

Referatet blev godkendt.

03 - Præsentation af infrastrukturpilotprojektet 'ScienceCloud4Cities' (SC4C) v/Per Sieverts

Infrastrukturen består overordnet af to dele:

- **SciCloud** - en privat cloud infrastruktur bestående af fysisk infrastruktur, netværk, sikkerheds- og identitetsstyring (rollebaseret) og en open source cloud computing management platform (OpenNebula) og med Windows og Linux baserede VM images inkl. præinstallerede, skræddersyede softwarepakker med analyseværktøjer, fx R, Python, Pandas, Scikit-learn, og data management relaterede værktøjer, fx PostgreSQL, MySQL, MongoDB, Influxdb.
- **Data infrastrukturen** der er deployeret i SciCloud, og som understøtter Smart City forskningsprojekterne. Infrastrukturen består af en open source software stack til data management, processing og analyse i skyen, og er delt i to lag:
 - Datalaget - kun tilgængelig for en administrator, som håndterer alle processer omkring dataindsamling (energidata fra indtil videre 5 danske byer) og transformationer (oprensning, anonymisering mv.) af rådata og efterfølgende lagring af de præprocesserede data. Her er open source systemer og værktøjer PostgreSQL, Apache Zeppelin, ownCloud og Zenodo deployeret - sidstnævnte for de åbne, anonymiserede data; øvrige data kræver autentifikation og autorisation.
 - Analyselaget - hvor forskere via web kan lave analyser på og samarbejde omkring de præprocesserede data vha. Jupyterhub og RStudio.

Kildekoden til SciCloud er frit tilgængelig på <https://github.com/xiufengliu/scicloud> og et antal datasæt er publiceret på Zenodo. Dvs. projektet ikke blot opfylder kravet om brug af open source værktøjer, men bidrager også selv med open source og åbne data.

SciCloud bruges både i undervisning og forskningen på partnerinstitutionerne og i store forskningsprojekter CITIES og Smart Cities Accelerator (EU projekt). Infrastrukturen forventes at danne basis i flere kommende offentlig-privat samarbejder, og er budt ind som infrastruktur for adskillige kommende, store forskningsprojekter. Papers er publiceret om infrastrukturen, og adskillige publikationer er baseret dataanalyse foretaget i SciCloud.

Slides [SciCloud_Presentation_Oct2 v3.pptx] @ sciencedata.dk/shared/DM_Forum (mappen 'Slides mm fra møde 2018 10 02')

04 - Q&A/diskussion: SC4C's værdiskabelse

Per Sieverts præsentation blev meget vel modtaget og gav anledning til stor spørgelyst. Fx spurgte Bo Bai om hvor de største udfordringer lå - i teknik eller data - og svaret var her dataaftalerne. Tillid fra dataleverandørerne er lav pga. myter om data sikkerhed og konkurrencefokus. Grundig formidling af data- og sikkerhedsmodel er således et must. Udfordringerne til trods, udtrykte Per Sieverts, at infrastrukturen er godt positioneret i en europæisk kontekst, hvor det er "rimelig i forefront" (citater).

05 - 4 x Show and tell: DMF aktiviteterne leverancer - siden sidst og til næste gang v/aktivitetslederne

Alle slidesæt @ sciencedata.dk/shared/DM_Forum (mappen 'Slides mm fra møde 2018 10 02') med navn 2018-10-02 Show and Tell.

Data Management by Design:

Siden sidst: Workflows for arkivering på <https://github.com/Data-Management-by-Design/DMbD>

Til december: Andre workflows på GitHub

FAIR på tværs:

Siden sidst bl.a.:

- 8 Data Pilots, herunder interviewguide og "talk kits"
- oversigt over disciplinspecifikke dataservices
- En TtT: Exploring tools for FAIR data (16.-17. august)
- Første udkast til FAIR toolbox

Disse ting offentliggøres på <https://vidensportal.deic.dk/da/FAIR>

Til december:

- Konferencen 20.11.2018
- Den endelige FAIR Toolbox
- Kommunikation og vidensdeling af projektets resultater

Processer og Organisation:

KU har lige overtaget ledelsen af aktiviteten, og en restrukturering er i gang. Nogle leverancer udgår, og AU er ikke længere deltager, men SDU og ITU forsøges inddraget mere.

Fokus på supportfunktion og kommunikation

Til december:

- Canvas modellen og andre modeller testet for nogle institutioner
- Intern interessentanalyse der deles i aktivitetsgruppen
- Organisationsdiagram for mulig /eksisterende supportorganisation på medvirkende institutioner
- Liste af the "main issues" fra de enkelte institutioner
- Liste af processer der mangler (fokus)

Derudover foreslås to tematiserede netværk etableret i 2019:

- Data management planning - faciliteret med et nyhedsbrev og mailingliste
- Data management competenceopbygning

Danish eLearning module for Research Data Management:

Siden sidst:

- Final script of manuscript
- Visual design
- End-User review
- Promotion strategy

Til december:

- Promotion strategy
- LMS mailing list
- Individual university promotion plan
- Emails
- Trailers
- List of proposed new modules
- Other material
- Maintenance quote from COBL
- Masterclass as webinar late Nov. – Open Science MOOC + promotion
- Hopefully a near final course to preview

06 - Strategi for national forskningsinfrastruktur V/Lars Nondal

Punktet udgik på anmodning fra Danske Universiteter, da DeiCs bestyrelse officielt endnu ikke har fået udmelding fra ministeriet.

07 - FAIR data action plan v/Anne Sofie Fink

I forventning om at Rigsarkivet skal lede arbejdet med etablering af en national RDA node, hvor FAIR data naturligt er i absolut fokus, havde DM sekretariatet forinden mødet bedt Anne Sofie sætte sig ind i den publicerede, på det tidspunkt foreløbige 'FAIR data action plan' og i det hele taget være DM Forums "informationsofficer" på dette område.

Der to væsentlige papirer var FAIR Data Action Plan <https://doi.org/10.5281/zenodo.1285290> og den fulde baggrundsrapport <https://doi.org/10.5281/zenodo.1285272> [NB: 23.11.2018 udkom den endelige version <https://doi.org/10.2777/1524>]

Anne Sofie introducerede de 34 anbefalinger, der fordeler sig på en række stakeholders, og der falder indenfor

- Primary Recommendations and Actions
- FAIR Data Policy
- FAIR data culture
- Technology for FAIR
- FAIR metrics
- Cost and investments in FAIR

Slides [DMF 20181002 FAIR Data Action Plan.pptx] @ sciencedata.dk/shared/DM_Forum (mappen 'Slides mm fra møde 2018 10 02')

08 - Indledning til planlægning af 2019 DM Forum aktiviteter v/formanden

Anders fremlagde sit og sekretariatets syn på DM Forum 2019 aktiviteter, herunder præmisserne for dem:

- Vi skal kun bruge tid på noget som gør en forskel
- Når der mangler ledelsesfokus (ressourcer) er vi nødt til at arbejde kreativt med hvad vi har
- Vi skal gøre de ting sammen, der bringer os længere end hvis vi gør dem hver for sig

"Rules of engagement" er, at aktiviteterne skal være:

- Inden for en FAIR eller EOSC ramme, dvs. understøtter FAIR data action plan
- 100% produktorienteret
- Forankrede! Dette gælder også produkterne!
- Inden for eksisterende aktivitetsmodel
- Inddrage forskere!

Anders kom med sit og sekretariatets bud på overskrifter på 3 aktiviteter:

1. FAIRify eksisterende data samlinger (der er tale om "legacy data" og dermed en "Federate the gems" indsats, hvor hver deltagende institution i aktiviteten minimum skal have "FAIRified" mindst én samling inden 2020)
2. Findable - PIDer på fremtidige/nye data
3. National koordinering og akkreditering af en data steward uddannelse

Tanken med dette oplæg var at sætte rammerne for den efterfølgende diskussion (pkt. 09 Fra Open Floor til Common Ground)

Slides [DM aktiviteter 2019.pptx] @ sciencedata.dk/shared/DM_Forum (mappen 'Slides mm fra møde 2018 10 02')

09 - Planlægning af DMF aktiviteter 2019 #1 - Fra Open Floor til Common Ground

Én DM Forum deltager havde forberedt et forslag hjemmefra. Ift. de tre fremlagte forslag fra formand og sekretariat, vil det være fair at opsummere det sådan, at der var flere der stillede spørgsmål og så problemer, end der meldte sig under fanerne.

Den første runde diskussion førte til en bruttoliste af overskrifter:

1. FAIRify - GEMS & PUBLIC DATA
2. FINDABLE - PIDs on future data
3. National Koordinering af Data Stewards
4. FAIR datatype / From DMP to data protocols
5. Lær forskere at bruge licenser
6. Data Science
7. Follow up på e-learning modulerne

10 - Planlægning af DMF aktiviteter 2019 #2 - Gruppearbejde

Bruttolisten (pkt. 09) dannede basis for gruppearbejde mhp. efterfølgende fremlæggelse af ideer.

11 - Planlægning af DMF aktiviteter 2019 #3 - Fremlæggelse og diskussion i plenum

3 af ideerne fra bruttolisten (pkt. 09) blev lagt frem i plenum:

FAIRify - Gems & Public Data

- Lead: Bo and Mette
- Aktivitetens leverancer kunne være:
 - Guidelines omkring hvor meget data skal aggregeres for ikke at være følsomme
 - FAIRified datasæt
- Handlinger/næste skridt: Møde før næste DM Forum møde i december

Licenses in practise

- Lead: KU (?)
- Aktivitetens leverancer kunne være:
 - Afdækning af nuværende praksis?
 - Best practise guide omkring valg af licens (flowchart med spørgsmål)
 - for hver institution en aftaleskabelon til brug for at lave aftaler mellem PI og institution (ekspliciterer hvem der har rettighederne og ansvaret - ikke ejerskabet)
 - FAQ / DMP online udvidelse
- Handlinger/næste skridt: Zoom møde før næste DM Forum møde i december

National koordinering og akkreditering af Data Steward uddannelse

- Lead: Diba (foreløbig)
- Aktivitetens leverancer kunne være:
 - Definerer af DS rollen og behovet for dem
 - Facilitere (koordinere) DS training initiativer
 - Forslag til et curriculum
 - Plan for forankring
- Handlinger/næste skridt: Foreløbig aktivitetsbeskrivelse før næste DM Forum møde i december

De øvrige forslag er ikke glemt eller udelukket. Det kan bare konstateres, at der ikke rigtigt er ressourcer til at udforme aktivitetsforslag før de nuværende aktiviteter er afsluttede.

12 - Hvordan CESSDA ERIC arbejder med FAIR v/direktør Ron Dekker og senior adviser Ivana Versic

Ron Dekker og Ivana Versic gav to meget veloplagte oplæg, der gav anledning til ivrig diskussion. Indlæggene drejede sig om CESSDA ERIC og samfundsvidenskabelige behov for forskningsinfrastruktur, om de EU-projekter CESSDA indgår i, og om hvordan man fra CESSDA's side tænker at indgå i EOSC.

Der er to slidesæts [CESSDA at National Forum Denmark.pdf] og [CESSDA Projects Copenhagen.pptx] @ sciencedata.dk/shared/DM_Forum (mappen 'Slides mm fra møde 2018 10 02')

13 - Eventuelt

N/A