



Werken met onderzoeksdata
Services en repositories voor de onderzoeker in Nederland

Ana van Meegen, Vrije Universiteit

Leon Osinski, TU Eindhoven

Henk van den Hoogen, Universiteit Maastricht

Maarten van Bentum, Universiteit Twente





Algemeen

- Doel: overzicht van research data repositories en services beschikbaar voor onderzoekers in Nederland
- Doelgroep: ondersteunende medewerkers van de instellingen
- Resultaat: Nog beperkt overzicht, dynamisch document



Definities

- Research data repository: een gestructureerd opslagsysteem voor (onderzoeks)data waarbij onder meer door middel van metadatering de vindbaarheid en preservering gefaciliteerd wordt.
- Data services: tools en vormen van ondersteuning die voor onderzoekers beschikbaar zijn ten behoeve van het werken met onderzoeksdata.



Scope

- Repositories die onderzoekers kunnen gebruiken om hun data te archiveren, beschikbaar te stellen en eventueel te publiceren.
- Services die onderzoekers kunnen gebruiken voor analyse van (big) data. Software applicaties worden niet meegenomen.
- (Op maat) ondersteuning voor het analyseren van onderzoeksdata waar specifieke infrastructuur voor nodig is.





Afbakening

- niet-commerciële research data repositories en services
- Nederland, Europese Unie
- beschikbaar voor onderzoekers in Nederland en waar zij zelf hun data kunnen deponeren
- geen data repositories en diensten van onderzoeksinstituten (alleen toegankelijk voor onderzoekers van de betreffende instelling)



Resultaat

Het overzicht bestaat uit twee lijsten:

- Data repositories, ('niet-afgebakende' overzichten in
 - Registry of research data repositories: [re3data](#)
 - open access directory: <http://oad.simmons.edu>
 - [Databib](#) of [DataCite](#))
- Data services, (zoals overzicht:
 - [Research Data Services - EUI Library](#))





Resultaat: data repositories (12)

Kenmerken

- Naam repository
- Organiserende instelling
- Website
- Termijn data archivering (korte, middellange en lange termijn)
- Gewenste dataformats
- Gebruikte metadataformat
- Zoekmogelijkheden (discoverability)
- Toegankelijkheid (open, semi-gesloten, gesloten, combinatie)
- Mogelijkheid om gevoelige data op te slaan
- Beschikbaarheid voorwaarden voor het bewaren van data
- Back-up beschikbaar
- Kosten voor gebruik
- Beschikbaarheid ondersteuning
- Discipline(s)
- Extra belangrijke informatie





Een aantal beschreven data repositories

- DANS
- 3TU.Datacentrum
- SurfSara
- Pangaea
- Figshare
- B2Share
- Zenodo



Disciplinespecifieke repositories (zelf data deponeren)

- Voordeel: zoekmogelijkheden en bewaarfunctionaliteiten toegespitst op doelgroep.
- Nadeel in geval van wetenschappelijke communities: mogelijk niet toekomstvast omdat ontstaan vanuit kleine projecten met beperkt en tijdelijk budget
- Voorbeelden:
 - Pangaea voor aard- en levenswetenschappen,
 - ProteomeXchange voor proteomica (onderdeel van systeembio) en
 - The Language Archive (TLA) en Laudatio voor linguïstiek.



Vertrouwelijkheid

Vertrouwelijke data: DANS al mogelijk, 3TU.Datacentrum mee bezig

Niet-vertrouwelijke data: DDN

Hoe is voorziening bij instellingen?



Infrastructuur en tools

- Reken capaciteit voor grote hoeveelheid data (SurfSara)
- Grid infrastructuur (SurfSara, eScienceCenter)
- High Performance Computing (SurfSara)
- Data Visualisatie (eScienceCenter i.s.m. SurfSara)
- Discovery services (EUDat, DANS, 3TU.Datacentrum, DDN)
- Data replicatie (EUDat)
- Discipline-specifieke applicaties voor het verzamelen, delen en/of verwerken van data (TLA – The Language Archive, Life Sciences Grid van SurfSara).



Resultaat: data services (20)

Kenmerken

- Naam
- Organisatie
- Website
- Soort dienst
- Korte beschrijving (overgenomen van de website van de organisatie)
- Hoe wordt de dienst aangeboden (cloudoplossing,...)
- Hoe wordt toegang tot de dienst aangeboden (abonnement, offerte, subsidieaanvraag)
- Discipline(s)
- Kosten





Een aantal beschreven data services

- DANS data portal
- DANS data management plan
- BiGGrid SurfSara
- Datalab 3TU.Datacentrum
- B2Find EUDAT
- B2Share EUDAT
- DDN
- MyExperiment
- NLeSc



Vragen van onderzoekers

Binnen eigen instelling: eenvoudige vragen over beheer en opslag van onderzoeksdata (ook via hun website en op afspraak).

Buiten eigen instelling: website van de dienstverlenende organisatie (e-mail: vragen stellen en advies op afspraak).



Ondersteuning en advies

Soort ondersteuning:

- vinden van datasets in verschillende repositories (DANS),
- schrijven datamanagementplannen (DANS, 3TU.Datacentrum)
- bewaren en toegankelijk maken onderzoeksdata (DDN, DANS, 3TU.Datacentrum, TLA)
- transporteren, analyseren en archiveren van data (SurfSara, eScienceCenter)
- opstellen grid voor samenwerking tussen onderzoeksinstellingen en systemen (SurfSara)
- high performance computing (SurfSara).