Data storage in the Netherlands, a joint effort

Datamanagement Workshop
Swedish National Data Services

Henk van den Hoogen
May 13th 2016
h.vandenhoogen@maastrichtuniversity.nl
Overview

Collaboration supporting datamanagement

1. DataverseNL
   • Why
   • What
   • How

2. RDNL: Dutch data infrastructure

3. SURF: National Research Data Management Platform

4. Positioning RDM at Maastricht University
Driver for collaboration in NL

• Discovery of fraud / sloppy science in Social Psychology

The Utrecht train station, site of one of Stapel’s most famous fake experiments.

The University of Groningen, where Stapel conducted many of his fraudulent experiments.
and again......
First steps of collaboration

University libraries of Utrecht, Rotterdam, Tilburg and Maastricht: worked with Dataverse

NEREUS- (EU-)project: economists’ data storage and sharing;

Experiences:
- Proven technology:
  developed at Harvard University;
  open source
- User friendly
  researcher is ‘the driver’
  good disclosure options (metadata fields)
- Multi purpose for end users:
  during research process: storage and sharing
  after research: archiving, sharing and linking

But:..............................
Doubts with Dataverse

Reliability and secureness of data storage in the cloud at Harvard
Solution: Dutch Dataverse

Dataverse hosted in the Netherlands
University of Utrecht

Cloud storage of datasets in the Netherlands
via Vancis/Surfsara (redundant storage)

Connection via federated authentication and authorisation
Surfconext, an infrastructure for online collaboration giving users secure access through their institutional account

Organisation: collaboration based on equality
request for joining on assent base of existing members
no strict, predefined requirements from the members: we were all looking for options
What is Dataverse?

‘Private’ virtual archive with possibilities to:

- **Store** research data (during and after research);
- **Share** research data;
- **Publish** research data;
- **Link** research data to publications.

Open-source Web application for hosting several Dataverses
Infrastructure Dutch Dataverse
Functional issues

• Upload capacity < 8 GB
  limitations of the web

• Webbased,
  access anywhere, anytime

• Researcher in control
  rights management during research

• Persistent URL
  linking to publications

• Positioning: storage and sharing
  during research (dynamical data)
  up until 5-10 years after ending research (static data)
Position Dataverse for Dutch University Libraries

DataVerse Network focussing on long tail target group, with services:
- **During** research process: dynamic data storage and/or sharing (remote access)
- **After** research: static data storage (5-10 years)
- **Visibility/findability**: metadata>Description
- **Availability** (linking data - full text)

=> Cooperation on national level/ sharing expertise
Embedding Dutch Dataverse in national infrastructure

DANS becomes the hosting partner

*Data Archiving in the Humanities and Social Sciences*

DANS

- Institute of Dutch Academy and Research Funding Organisation (KNAW & NWO) since 2005
- Mission: promote and provide permanent access to digital research information
- First predecessor dates back to 1964 (Steinmetz Foundation), Historical Data Archive 1989

Maastricht University
DANS - services

**EASY**: dataset self-deposit service for long time archiving

**NARCIS**: national portal for scientific info (publications, datasets, research projects)

**DataverseNL**: dataset deposit service during and after (max 10 years) research
Positioning DataverseNL

- Dutch RDM landscape
- Organisation
## Position of the members of DataverseNL

<table>
<thead>
<tr>
<th>Location</th>
<th>During research</th>
<th>After research for first 10 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groningen</td>
<td>Dedicated services based on DVNL (among others)</td>
<td>Own repository and DVNL</td>
</tr>
<tr>
<td>Maastricht</td>
<td>DVNL + UM server capacity for larger data sets</td>
<td>DVNL (registration and storage datasets) + local storage capacity</td>
</tr>
<tr>
<td>Rotterdam</td>
<td>DVNL + SURFsara/Lisa (very large data sets)</td>
<td>DVNL + SURFsara/Lisa (large data sets)</td>
</tr>
<tr>
<td>Tilburg</td>
<td>DVNL</td>
<td>DVNL</td>
</tr>
<tr>
<td>3TU</td>
<td>DVNL as datalabs</td>
<td>3TU.Datacentrum</td>
</tr>
<tr>
<td>Utrecht</td>
<td>Own service on the base of iRods; DVNL</td>
<td>DVNL</td>
</tr>
<tr>
<td>VU</td>
<td>Own service</td>
<td>DVNL for compliance purposes (open data); Own service for larger data sets and back up Offline data vault (DANS)</td>
</tr>
</tbody>
</table>

**DataverseNL used for:**
- Registration
- Dynamic storage and sharing
- Static storage

*But not all members use DataverseNL in the same way*
DataverseNL: challenges 2016

- **Policy**
  - jointly, supported (by the members and hosting organisation)

- **Operational**
  - SLA
  - Membership fee

**Roadmap 2016**
- DataverseNL functional requirements make the choice for upgrade to version 4
- SLA
- Linking DataverseNL to long term preservation archives (DANS/3TU to start with)

Repercussions for membership fee!
Research Data Netherlands: RDNL

RDNL: joined forces (2013) in the area of long-term data archiving via an alliance between 3TU.Datacentrum, Data Archiving and Networked Services (DANS) and SURFsara.

**Mission:** promotion of sustained access and Responsible re-use of research data

RDNL uses the Front Office-Back Office model
RDNL offers several Services
RDNL: front office – back office

A unified back-office: The partners in RDNL fulfil a back-office function in the process of data curation.

The front offices: university libraries, for example deal directly with researchers and can advise them in ensuring that research data is accommodated in one of the archives.
Services RDNL

• **Long-term access to research data**
  Researchers can find data from previous research in the RDNL partners’ long-term archives and can deposit data there themselves.

• **Data management knowledge and advice**
  RDNL is a knowledge centre and advisory body to support the design and implementation of policies for good research data management and data stewardship before, during and after research.

• **Data Management course**
  RDNL organises a course for front office data supporters, called ‘Essentials 4 Data Support’

• **The Dutch Data Prize**
  RDNL awards the Dutch Data Prize to researchers who bring together data, document data and make data accessible to others.
National Research Data Coördination Platform

VSNU asked SURF to facilitate (coordinate?) a national strategy for research data management

The platform will
- support the development and implementation of research data management policy for scientific research in the Netherlands and
- stimulate the cooperation among the various stakeholders.

Via:
- collecting and disseminating knowledge and information,
- bringing together the right people,
- facilitating existing initiatives,
- providing access to expertise,
- monitoring developments
Themes coordinated by the National platform

- Facilities and data infrastructure
- Legal aspects and ownership
- Financial issues
- Research(er) support
- RDM awareness

⇒ Working groups per theme, covering players from data management field

⇒ Time frame: October 2015 – December 2017
Maastricht University and RDM

• General info about UM

• How the Library engage researchers
  - in general
  - with respect to RDM
Very young university: officially opened her doors in 1976

6 faculties:
FHML / Hospital, FPN,
FASoS, SBE, FL, FHS (humanities and sciences)

2015:
• students: 16.350, 47% foreign
• phd degrees: 250
• academic staff: 1.830 fte (3.500 persons), 39% international
THE Rankings

*World university ranking*
101th position (Edinburgh 36, Manchester 52, Leicester 199)

*Social Sciences: 62th position*

*Clinical, pre-clinical and health universities: 45th position*

*Most international universities: 17th*

*100 under 50 (years old): 4th position*
University Library

Central library with two main locations; no real deposit library, study landscape, emphasis on study environments

http://tools.maastrichtuniversity.nl/campustour/
Library and researchers

“Library is a **good content provider**, with additional services in finding and storing literature.”

=> Researchers *do not have to come to library*

=> Other services are not –automatically– requested
Opportunities

Funding
Collecting
Analysing
Writing
Publishing

The Research Cycle

Identify research area
Design research study
Publish research results
Carry out research
Analyse research results
Challenge

“Finding ways to meet researchers’ needs for support”

- at the right moment
- with tailormade support
- in their preferred working places
A new service concept is needed

• Characteristics:
  - personalized
  - faculty—overarching
  - using existing resources

• (permanent) Investigation
  - preferences / needs
  - possibilities / chances
UML + You
the plus side of your library

Starting even before the first day of work at the university; the researcher can expect the support of the library at the various stages of his/her process.

https://vimeo.com/51138293
Steps to fulfill this ambition

Programme manager research support
- responsible for a coherent plan of service activities

Meeting researchers:
- strategic / tactical
- operational

Research Support team
Stakeholders

- Executive Board
- Research Committee (faculty board members responsible for research/valorisation, rector, library)
- Faculties / ResearchGraduate Schools, UMC
- Institutions (e.g. Valorisation Center, Contract Research Center en ICTS)
- Policy advisors
- PhD-coördinators
- Researchers
Consequences for the Library organisation

- Policy widely supported by Management

- Dedicated –extra- workforce

- Change in organisation culture:
  - flexible attitude
  - dealing with uncertainties and risks
  - success is not guaranteed
2015/2016
Datamanagement support

- Library initiated *RDM-policy discussions* (after bad practice);

- *UM code of conduct for RDM* (2014)
  => verification of research
  Storage of data linked to publications for 10 years
  Central registration of data sets

- *RDM infrastructure*
  - Dutch Dataverse member
  => road shows (policy makers per faculty/school);
  => pilots for registration tooling,
  resulting in the use of *Dataverse as central registration tool*
  - additional project for ICT-infrastructural service
RDM support (2)

- RDM-supporting service:
  - Data specialist & IT-tooling expert
  - Helpdesk (UM-wide specialists)
  - Meet and Greet sessions
  - Training/awareness/workshops
  - RDM-support Website
  - DataverseNL for storage during and after research and registration

⇒ Several Schools requested for support!

⇒ Personal assistance (grant application, DMP, publishing data,...);
Datamanagement: challenges

General:
  implementation per discipline / faculty
  tuning with faculties

Library:
  advocacy role ?
  developing services: initiative?
  =>defining position / role