New Developments in Data Management Support
for Universities, Research Institutes and Higher Education in The Netherlands

Peter Doorn, Vyacheslav Tykhonov & Marion Wittenberg (DANS)

IASSIST 2017, Lawrence, Kansas, May 23-26, 2017
Session B2: Repository Strategies across Communities
Wednesday, 24 May 2017, 10:30 – 12:00, Room Big 12
Data Archiving & Networked Services

Mission: promote and provide permanent access to digital research resources

Institute of Dutch Academy and Research Funding Organisation (KNAW & NWO) since 2005

First predecessor dates back to 1964 (Steinmetz Foundation), Historical Data Archive 1989
DANS core data services

EASY: long-term Electronic Archiving System for self-deposit

DataverseNL for short/intermediate term storage (including RDM during research)

NARCIS: Gateway to scholarly information in the Netherlands

500,000 Open Access (40% → 25%)
Three additional services

http://www.datasealofapproval.org/

http://www.persid.org/

Persistent Identifier
URN:NBN resolver

Cradle of the
Data Seal of Approval

Training & Consultancy

http://datasupport.researchdata.nl/
Three recent collaborative services

**Mendeley Data**
https://data.mendeley.com/

**Background Archive**

**ELSEVIER**

**Software Heritage Archive**
https://www.softwareheritage.org/

**Research Data Journal**
http://www.brill.com/rdj
Why DataverseNL?

• Growing requirements for data managements by Research Funders & Institutions
• Institutions want to provide a system to support RDM and data sharing for their staff
• Changing role of Research/University libraries: finding new roles in information/data services
• Alternatives for institutions:
  1. Develop local data services including data repository → Leiden University
  2. Use a commercial supplier of repository services/systems → University of Amsterdam
  3. Use existing national/international data repository services → Continuation of existing practices
  4. “Shared Services” in a public cooperation → Most Dutch universities → DataverseNL
DataverseNL partners (15)

(most) Universities
(several) Research Institutes
Higher Education Institutions

Several more on the way…
Cooperative Model

- Collaboration agreement + S.L.A.
- Shared costs (service membership + storage)
- Advisory Board consisting of partner representatives decides on the general policy and direction
- Administrators committee to discuss technical and functional issues
- DANS provides and manages the system & storage, organizes the meetings (back-office)
- Research institutions run their own RDM-support for their staff (front-office)
- Every partner is responsible for their own content
- Work in progress to transfer data from Dataverse to certified long-term storage (DANS EASY or 4TU.Data Centre)
Advantages for partners

- Efficient: low costs for partners
- Build common RDM support experience
- Advisory Board and Administrators Committee provide platform for exchange and learning from each other
- One instance: search across partner holdings
- Dataverse (developed at Harvard IQSS) is one of the most used data repository systems
  - Re3Data: 43 occurrences, see: http://www.re3data.org/
  - Dataverse.org: 23 installations
- Dataverse is Open Source
World-Wide Dataverse Installations

Source: http://dataverse.org/
Some peculiarities of DataverseNL

• Dataverse is essentially a rigorous repository/archiving system, it is not as free & flexible as a “Dropbox” or VRE
  • but this exactly helps to guarantee the integrity of data during research; in a VRE changes to the data can go undocumented and unnoticed
  • Track data provenance
• Why is DataverseNL a solution for RDM during research and for short-term archiving?
  • Partners are responsible for the content and RDM policies → DANS cannot guarantee the long-term preservation
  • Data during research is often changed and does not always comply with long-term preservation criteria (e.g. privacy)
Dataverse advantages and added functions

• Excellent integration in institutional websites
• Federated Login for Dutch universities
• Reporting tool on content: dashboard for administrators
• Interactive data exploration and analysis tools, visualisation and geographic mapping of data
• Metadata formats for various disciplines (social sciences, humanities, astronomy, bio-sciences) compliant with DDI
• Compliant with data citation and data publishing guidelines
• Integration with e-journals (OJS)
• Privacy tools for sharing sensitive data (DataTags)
• Dataverses can reflect the structure of the organisation (University, Faculty, Department, Research Group, Project)
• Work on implementation of FAIR principles in progress
• “Big data” solutions
DataverseNL integrated in University websites

VU, Amsterdam

Tilburg University
Real-time metrics: Dashboard

- **Dataverses Added Over Time**
  - March 2016: 6 New Dataverses
  - Number of Dataverses: 148

- **Dataverses by Category**
  - Research Projects (29.6%)
  - Organizations Institutions (28.4%)
  - Researchers (27.2%)
  - Journals (6.0%)
  - Other (3.6%)

- **Datasets Added Over Time**

- **Files Added Over Time**

Data Archiving and Networked Services

DataVerseNL
Data explorations and visualisations

TwoRavens, which integrates with Zelig statistical framework)
Work in progress on FAIR Data Assessment and GDPR compliant Data Tags

Harvard DataTags supporting HIPAA and FERPA (and DUAs)
http://datatags.org/

<table>
<thead>
<tr>
<th>Level</th>
<th>DUA Agreement Method</th>
<th>Authentication</th>
<th>Transit</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Risk</td>
<td>None</td>
<td>None</td>
<td>Clear</td>
<td>Clear</td>
</tr>
<tr>
<td>Minimal</td>
<td>None</td>
<td>Email or OAuth</td>
<td>Clear</td>
<td>Clear</td>
</tr>
<tr>
<td>Shame</td>
<td>Click Through</td>
<td>Password</td>
<td>Encrypted</td>
<td>Clear</td>
</tr>
<tr>
<td>Civil Penalties</td>
<td>Sign</td>
<td>Password</td>
<td>Encrypted</td>
<td>Encrypted</td>
</tr>
<tr>
<td>Criminal Penalties</td>
<td>Sign</td>
<td>Two Factor</td>
<td>Encrypted</td>
<td>Encrypted</td>
</tr>
<tr>
<td>Max Control</td>
<td>Sign</td>
<td>Two Factor</td>
<td>Double Encryption</td>
<td>Double Encryption</td>
</tr>
</tbody>
</table>

Data Seal of Approval

1. Archivist Assessment
3. User Reviews
16 Downloads
Disadvantages/weaknesses/challenges?

- Collaborative model is less business-like and straightforward:
  - Some institutions may prefer other business models:
    - outsourcing to a commercial provider
    - developing and managing in-house solutions
- How to serve best universities with different priorities and RDM policies with one system and under one agreement?
  - It is possible, but at a cost: it takes time to agree on common wishes (e.g. future functional requirements)
  - But: does it make sense that RDM policies and practices vary among institutions?
  - Some partners have stricter requirements for Service Level Agreements than others
    - Solution: incorporate in the business model \( \rightarrow \) the higher the demands, the higher the membership fee
DataverseEU for new CESSDA partners?
CESSDA SAW “Strengthening and Widening” EU Project

- Hosting DataverseEU for starting data archives
- Support in setting up national Dataverse instances for starting data archives
Questions?

Watch our videos and webinars on YouTube
https://www.youtube.com/user/DANSDataArchiving