Users and uses of a digital data archive: A case study of DANS

Christine L. Borgman, *University of California, Los Angeles*
Ashley E. Sands, *University of California, Los Angeles*
Milena S. Golshan, *University of California, Los Angeles*

Available at: https://works.bepress.com/borgman/402/
Users and uses of a digital data archive: A case study of DANS

Christine L. Borgman, Ashley E. Sands, Milena S. Golshan

Project collaborators:
DANS: Andrea Scharnhorst, Henk van den Berg
DANS visiting scholars
ANDS: Andrew Treloar
LANL: Herbert van de Sompel

KNOWeSCAPE 2017, Sofia, Bulgaria
February 22, 2017

https://knowledgeinfrastructures.gseis.ucla.edu

Data Archiving and Networked Services

https://dans.knaw.nl/en
Overview of talk

- Motivation for study of DANS users and uses
- Research questions and methods
- Findings from interviews
- Discussion of results
- Implications for knowledge infrastructures
Knowledge Landscapes and Knowledge Infrastructures

- **Knowledge landscapes**: “large knowledge spaces and systems that organize and order them”
  
  KNOWeSCAPE. http://knowescape.org/about/

- **Knowledge infrastructures**: “robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds”

Sustaining Access to Scholarship depends upon building better Knowledge Infrastructures
Open access policies

- Research Councils of the UK
- European Union
- Australian Research Council
- U.S. Federal research policy
- Taiwan, China, India...
- Individual countries, funding agencies

Policy RECommendations for Open Access to Research Data in Europe
When to invest in data?

http://www.lib.uci.edu/dss/images/lifecycle.jpg
Research Questions

1. What are the roles of data archives in knowledge infrastructures?

2. How do stakeholder roles in digital archives vary?
   a. Contributors to a digital data archive
   b. Consumers of data from a digital archive
   c. Archivists who manage the digital data archive
DANS as a Digital Data Archive

- Netherlands government funding; about 50 staff
- Data Seal of Approval
- Participates in international research infrastructure projects
- 50+ years of social science and humanities data
- Published datasets in EASY: 32,000+
- About 3.5 million files; total 5TB
Research Methods

• Document analysis
• Ethnography
• Mining transaction logs
• Interviews
  – DANS data contributors
  – DANS data consumers
  – DANS archivists and staff

A. Scharnhorst, C. Borgman, DANS 2014
## DANS Interviews 2015-2016

<table>
<thead>
<tr>
<th>Stakeholders/Participants</th>
<th>Number of Interviews</th>
<th>Domain Expertise</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data contributors</td>
<td>9</td>
<td>Archaeology, History, Paleogeography (3)</td>
<td>Academic staff (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Labor Economics (1)</td>
<td>Cultural institution staff (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linguistics (1)</td>
<td>Private company staff (1 interview with 2 staff)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral Histories (1)</td>
<td>Unaffiliated (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Information Science (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theology (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biology (1)</td>
<td></td>
</tr>
<tr>
<td>Data consumers</td>
<td>8</td>
<td>Archaeology, History (6)</td>
<td>Academic staff (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Political Science, Sociology, Public Administration (3)</td>
<td>Cultural institution staff (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Citizen scientists (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Students (2)</td>
</tr>
<tr>
<td>DANS staff</td>
<td>10</td>
<td>Archaeology and humanities (6)</td>
<td>Archivists</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IT Development (4)</td>
<td>Project Managers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IT developers</td>
</tr>
</tbody>
</table>
Findings

• Data types
• Data contributors
• Data consumers
• DANS archivists and staff
• Infrastructure roles of data archives
DANS data

Wilgen, L. R. van (SOB Research) (2104). DANS. https://doi.org/10.17026/dans-z5y-tdb6

Bosch, J. E. van den (SOB Research) (2102). DANS. https://doi.org/10.17026/dans-2cg-tg88

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:63812/tab/2

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:62234/tab/2

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:65225/tab/2
Who contributes data to DANS?

- Archaeologists
  - Companies that conduct site surveys
  - University researchers
- Researchers in other fields
  - Public policy, economics, geography
  - Sociology, linguistics, oral history...
- Staff on behalf of data collectors
  - Institutional librarian
  - Head of the department

Bosch, J. E. van den (SOB Research) (2102). DANS. https://doi.org/10.17026/dans-2cg-tg88
Why do they contribute data to DANS?

• Meet legal requirements
• Get credit for data
• Share data with others
• Preserve data for long term
• Control access to their data

Wilgen, L. R. van (SOB Research) (2104). DANS. https://doi.org/10.17026/dans-z5y-tdb6
Controlling access to deposited data

• Contributor goals
  – To identify consumers of their data
  – To control how their data are used

• Consumer requirements
  – Register with DANS
  – Request permission from data contributor

Wilgen, L. R. van (SOB Research) (2104).DANS. https://doi.org/10.17026/dans-z5y-tdb6
## Types of Access

<table>
<thead>
<tr>
<th>Access</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open (everyone)</td>
<td>825</td>
</tr>
<tr>
<td>Open (registered users)</td>
<td>22522</td>
</tr>
<tr>
<td>Restricted ('archaeology' group)</td>
<td>7613</td>
</tr>
<tr>
<td>Restricted (request permission)</td>
<td>2625</td>
</tr>
<tr>
<td>Other</td>
<td>468</td>
</tr>
</tbody>
</table>

https://easy.dans.knaw.nl/ui/browse
Granularity of access

You need to log in to be able to view/access (some of) the files. **Log In**

You need to have special permission to be able to access (some of) the files. You can request permission after logging in.

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Accessible</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEM growth and yield data Monocultures - Silver Birch - plot data book.pdf</td>
<td>353253</td>
<td>Yes</td>
</tr>
<tr>
<td>FEM growth and yield data Monocultures - Silver Birch - plot information.csv</td>
<td>6282</td>
<td>Requires granted permission request</td>
</tr>
<tr>
<td>FEM growth and yield data Monocultures - Silver Birch - plot level.csv</td>
<td>21235</td>
<td>Requires granted permission request</td>
</tr>
<tr>
<td>FEM growth and yield data Monocultures - Silver Birch - tree level.csv</td>
<td>190875</td>
<td>Requires granted permission request</td>
</tr>
<tr>
<td>FEM growth and yield data Monocultures - Silver Birch - tree maps atlas.pdf</td>
<td>1017162</td>
<td>Yes</td>
</tr>
<tr>
<td>Read me - FEM growth and yield data Monocultures - Silver Birch.pdf</td>
<td>205772</td>
<td>Yes</td>
</tr>
</tbody>
</table>

https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:63812/tab/2
Who consumes data from DANS?

- Archaeologists
  - Companies that conduct site surveys
  - University researchers
- Researchers in other fields
  - Public policy, economics, geography
  - Sociology, linguistics, oral history...
- Students, teachers, visiting scholars
- Local history guides, genealogists
- Museum curators, amateur scientists...

http://www.volkstellingen.nl/nl/vt_bundel/index.jpg
How do users search in DANS?

- Browse topics
-Browse spatial regions
-Search for place names
  - Archaeology sites
  - Building sites
- Few known item searches
- Open access datasets are most used

https://easy.dans.knaw.nl/ui/?wicket:interface=:1:1::
How are DANS data used?

• To assess geographic regions
• To use map data in other research
• To compare to other data
• To create new products, e.g., local history guidebooks

http://www.intlconnect.illinois.edu/jobsearch
DANS as Invisible Infrastructure

- DANS data delivered via external websites
- DANS as “back office-front office” services
- DANS as metadata harvesting source

http://adst.org/oral-history/oral-history-interviews/
Visible roles of archivists

• Acquire data
  – Work with contributors
  – Seek useful data
• Disseminate data
  – Assist users in searching
  – Outreach to communities
• Staff help desk
Invisible roles of archivists

• Curate data
  – Ingest, clean, verify anonymity
  – Migrate data to other formats
  – Describe, document, add metadata
  – Preserve the data in long term

• Bring expertise
  – Subject domain experts
  – Metadata and cataloging experts
  – Design and software engineering
  – Statistics and data analysis

#archivist
Mediation and Stakeholders

- Archivists
- Contributors
- DANS
- Consumers

Science research → Data contribution → Archive/Repository → Data use → Science education
Mediation and Stakeholders

Department manager/Project manager → Institutional librarian

Department manager/Institutional librarian → DANS data Manager (Data and metadata check)

Researchers/Data producers → DANS: Special projects - EASY

Institute/university → (Back office-front office)
Discussion of findings

• Data archive content is community driven
• Archives are a source of trust in data
• Archive uses and users are diverse
• Access to data is mediated by
  – Archivists
  – Contributors who retain control over data
  – Institutional policies for contribution
  – Services that deliver data
Implications for Infrastructures

• Invisible work shapes knowledge spaces

• Interdisciplinary data archives
  – Acquire a disparate array of datasets
  – Attract a diverse array of contributors and consumers
  – Are difficult to search due to their diversity

• Incentives to contribute data vary by stakeholder
  – Legal responsibility for archaeological reports
  – Assure preservation and access
  – Transfer responsibility while maintaining control

• Interoperability of data archives with publishers, libraries, project sites, could increase access
Acknowledgements

Christine Borgman

Peter Darch

Ashley Sands

Irene Pasquetto

Bernie Randles

Milena Golshan

Data Archiving and Networked Services

UCLA Center for Knowledge Infrastructures