Data Archiving and Networked Services

iPRES conference October 8-10 2006, Cornell University

René van Horik
(rene.van.horik@dans.knaw.nl)

DANS is een initiatief van KNAW en NWO
DANS is the Dutch national organisation responsible for storing and providing access to research data from the humanities and social sciences.
DANS is a joint initiative of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO)
Goals DANS

- Take away barriers
  - Permanent access to research data
- Data Archiving
  - Make it easy for depositing new data
  - Fast access to existing data
- Networked Services
  - Networks of cooperation
  - Linking of relevant data
  - TOPical Programs (TOP)
- Development of research infrastructure (also on international level via ESFRI)
Outline of the presentation

• Organization of DANS
• Quality issues concerning digital research data
• Topical Programs (TOP)
• DANS and digital preservation
Organisation

- Staff: ± 15 fte + project resources
- DANS is a network of people and data
  - Cooperation with: researchers, KB, National Archive, faculties, research schools, government bodies, etc.
- Demand-driven
- Trusted digital repository of a number of digital research collections
- Own development team (e.g. development of “EASY”)

Data Archiving and Networked Services
Self archiving is “EASY”

- EASY: *Electronic Archiving SYstem*
- Developed by DANS (Open Source)
- Based on functions described in OAIS model
- Aim: central website for depositing, finding and assessing research data
- In cooperation with
  - Data experts (DANS)
  - Research (e.g. Archeology)
  - Local Data Centers (e.g. VROM)
## Deposit dataset - required elements

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator*</td>
<td>Rutger Kramer</td>
</tr>
<tr>
<td>Title*</td>
<td>Scandinavian Trade, 1980-1990</td>
</tr>
<tr>
<td>Description*</td>
<td>Import and Export figures over a 10 year period for Norway, Sweden and Denmark with countries outside of Scandinavia</td>
</tr>
<tr>
<td>Date created*</td>
<td>1996 (yyyy or yyyy-yyyy)</td>
</tr>
<tr>
<td>Rights</td>
<td></td>
</tr>
<tr>
<td>Access rights</td>
<td>I. everyone, no conditions for use</td>
</tr>
<tr>
<td></td>
<td>II. only scientific research &amp; education, no conditions for use</td>
</tr>
<tr>
<td></td>
<td>III. only scientific research &amp; education, specific group(s), namely:</td>
</tr>
<tr>
<td></td>
<td>IV. only scientific research &amp; education, after permission by depositor</td>
</tr>
<tr>
<td>License</td>
<td>View license agreement (PDF)</td>
</tr>
<tr>
<td>Date available*</td>
<td>2006-07-02 (yyyy-mm-dd)</td>
</tr>
</tbody>
</table>

### Upload dataset

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>File*</td>
<td></td>
</tr>
</tbody>
</table>

---

**http://easy.dans.knaw.nl**
Finding research-data

- Search
  - Easy!
  - Extended (time, location etc.)

- Message to new datasets
- Keep track of usage
# Directly accessible

## Search

Search for [Search]

- Within search results
- In discipline(s):
  - Humanities
  - Social sciences

## Results

Searched for **"amsterdam"**, found 4 results.

<table>
<thead>
<tr>
<th>Title</th>
<th>Creator</th>
<th>Discipline</th>
<th>Created</th>
<th>Submitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open het dorp 1962</td>
<td>Baschwitz instituut, universiteit van Amsterdam</td>
<td>Sociology</td>
<td>1962</td>
<td>2006-05-31</td>
</tr>
<tr>
<td>Straftoemeting van 800 kleine strafzaken 1953-1954</td>
<td>Ch.J. Enschede, Universiteit van Amsterdam</td>
<td>Law</td>
<td>1953</td>
<td>2006-05-31</td>
</tr>
<tr>
<td>The housing market in Amsterdam, 1550-1650</td>
<td>Lesger, C.M.</td>
<td>History</td>
<td>2000</td>
<td>2006-05-31</td>
</tr>
</tbody>
</table>
Data quality issues

Research data must be findable, accessible, usable, reliable and referable

• DANS recommendations concerning:
  • the quality of research data
  • the quality of the usage of the research data
  • the quality of the data storage facility

• (Based on NESTOR memorandum)
Quality of research data

• **Scientific value** → determined by peer-review process
• **Data format**
• **Documentation**
Quality of the usage of the research data

- Open Access (if possible)
- Codes of conduct, concerning
  - privacy
  - scientific practice (original, based on own work)
- Licenses of usage
Quality of the data storage facility

- = “Trusted digital repository”
- Organisational issues
- Technical issues

- Examples:
  - eDepot of National Library (KB)
  - “Big-GRID” of Nikhef (The National Institute for Nuclear Physics and High Energy Physics)
  - DANS data archive
Topical programmes (TOP)

(Implementation of demand-driven approach of DANS)

- Goal: improvement of data infrastructure in a particular discipline or a particular theme
- Initiated by or in cooperation with researchers
- Make databases of national and/or international significance better accessible, and/or enrich them with other databases or by adding data from other databases
- May involve data programs that have been concluded or are still up and running; they may also involve new initiative, or foresight studies
• Raising awareness on importance of digital preservation for researchers and funding bodies (e.g. via quarterly eDATA, see: www.edata.nl)

• Establishment and maintenance of trusted digital repositories

• Development of tools and methods, e.g.
  • Persistent identifiers for research data
  • How to quote research data?
  • Stimulate and facilitate re-use of data
  • Preservation metadata for research data
  • Preferred formats for research data
DANS and digital preservation II

• Support the “emancipation” of databases (recognition of data collector / “citation impact factor”)
• Involvement in projects on digital preservation (e.g. EU Driver project)
• Co-founders of National Coalition on Digital Preservation
Questions?

• More information: www.dans.knaw.nl/en