Data Archiving and Networked Services

Peer-reviewed open research data: results of a pilot

Marjan Grootveld & Jeff van Egmond

DANS is an institute of KNAW and NWO

Driven by data
Selling Products? Use Product Reviews to Increase Conversions!

User-generated content increases conversion, doubles page views for Halfords.

Product Reviews Increase Conversion Rates

Consumer reviews are significantly more trusted -- nearly 12 times more -- than descriptions that come from manufacturers, according to a survey of US mom Internet users by online video review site EXPO. (eMarketer, February 2010)

LEVERAGING USER GENERATED CONTENT
Agenda

• Who is DANS?
• Peer review pilot study: What we did
• What we got
• What it looks like
• What now?
About DANS

- Promote sustained access to digital research data
- Data archiving in Social Sciences and Humanities
- On-line repository “EASY” for self-archiving
About DANS

• Explore and research new approaches to providing access to data

• Enhanced Publications

• International research infrastructures
Data quality at DANS EASY

• Check at deposit
  – Data expert examines data as it comes in
  – Metadata, data set and sources are checked
  – Correctness is up to depositor

• Extend peer reviewing to data in their own right

• Engage our users
  – There are 8000 of them
  – What do they think about our collections?
Selected a target group
  - People who have downloaded data from October 2009 onwards

Set up a simple questionnaire using SurveyMonkey

Sent out 1937 e-mail invitations

Asked for a rating 1-5 and comments on data set and website
What we got

- Response: 279 out of 1937 (14.4%)
What we got: Data set

Researchers
All respondents

Data Quality
Quality of the documentation
Completeness of data
Consistency of dataset
Structure of dataset
Usefulness of file formats
What we got: Data set

• “Quality = Comprehensiveness + Accessibility”
• 91% would recommend the data set to others
• 51% assigned tags to the data set
What we got: Research

• Goal:
  - 69% find the data set helpful for research question
  - 16% have used the data set for a publication
  - 58% intend to use it for (another) publication
What we got: Website

Clarity Information about the data Finding the data Availability of the data

Researchers
All respondents
Showing the results

Dutch
The C
do of the
major
gover
the N
tion
from
Kingd
Sport
Meth
dicomp
Numt
A ven

strongly restrictive rules at Statistics Netherlands. The questionnaires (both Dutch and English) used in this survey are available on www.dans.nl. A full codebook will become available in November 2007.

NKO website

Voor deze dataset zijn enkele reviews beschikbaar.
Detailed reactions for 'Nationaal Kiezersonderzoek, NKO 2006'

Reviews

Below you can see how users have responded to the dataset 'Nationaal Kiezersonderzoek, NKO 2006'. The legend to the right explains the ratings.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Rating</th>
<th>Average rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>data quality</td>
<td>4</td>
<td>★★★★★ (4.67/5)</td>
</tr>
<tr>
<td>quality of the documentation</td>
<td>4</td>
<td>★★★★☆ (4.17/5)</td>
</tr>
<tr>
<td>completeness of the data</td>
<td>4</td>
<td>★★★★☆ (4.67/5)</td>
</tr>
<tr>
<td>consistency of the dataset (if applicable)</td>
<td>4</td>
<td>★★★★★ (4.8/5)</td>
</tr>
<tr>
<td>structure of the dataset (if applicable)</td>
<td>4</td>
<td>★★★★★ (4.67/5)</td>
</tr>
<tr>
<td>usefulness of the file formats</td>
<td>3</td>
<td>★★★★☆ (4.17/5)</td>
</tr>
</tbody>
</table>

5 out of 6 reviewers of this dataset recommend the use of it.
1 out of 6 reviewers of this dataset has published using this dataset.
2 out of 6 reviewers of this dataset intend to use this dataset for a publication.
Below you'll find all datasets that have been reviewed by two or more users who have downloaded that set. Click on the name of a dataset to see more detailed ratings. As more EASY-users share their opinion on the dataset(s) they have downloaded this list will grow. Please note that these ratings are not by DANS.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of reactions</th>
<th>Average rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>De steentijd van Nederland</td>
<td>12</td>
<td>(4.02/5)</td>
</tr>
<tr>
<td>WoON2009: release 1.2 - Woononderzoek Nederland (voor overheid en universiteiten)</td>
<td>8</td>
<td>(4.19/5)</td>
</tr>
<tr>
<td>Relief in Tijd en Ruimte. Interdisciplinair onderzoek naar bewoning en landschap van Oost-Nederland tussen vroege prehistorie en middeleeuwen.</td>
<td>7</td>
<td>(3.77/5)</td>
</tr>
<tr>
<td>Nationaal Kiezersonderzoek, NKO 2006</td>
<td>6</td>
<td>(4.51/5)</td>
</tr>
<tr>
<td>NLGis shapefiles</td>
<td>5</td>
<td>(4.3/5)</td>
</tr>
<tr>
<td>Culturele Veranderingen in Nederland 2006 - CV’06</td>
<td>4</td>
<td>(3.71/5)</td>
</tr>
<tr>
<td>Bestand bodemgebruik 2006 - BBG’06</td>
<td>4</td>
<td>(3.7/5)</td>
</tr>
<tr>
<td>Harinxveld-Giessendam - De Bruin</td>
<td>4</td>
<td>(4.1/5)</td>
</tr>
</tbody>
</table>
What did depositors say?

• “I like those scores!”

• “Any tips on how to improve my rating?”
Results

• What we sought:
  – Basic ratings for various data sets
  – Estimate of feasibility of ratings in repository

• What we got:
  – Rather positive feedback
  – Diminishing “deposit-and-forget” effect
    • What you did matters to others
    • People are willing to share their experiences
  – Many new questions...
So...

- Turn pilot into a structural data review process
- How to present qualitative feedback and tags?
- Critical mass versus harassment
- How to grow it into a community platform?
Questions and suggestions

Marjan Grootveld – marjan.grootveld@dans.knaw.nl

http://easy.dans.knaw.nl
http://datareviews.dans.knaw.nl/?l=en&/