The International Journal of Digital Curation
Volume 7, Issue 2 | 2012

Peer-Reviewed Open Research Data: Results of a Pilot

Marjan Grootveld and Jeff van Egmond,

Data Archiving and Networked Services (DANS)

Abstract

Peer review of publications is at the core of science and primarily seen as instrument for ensuring research quality. However, it is less common to independently value the quality of the underlying data as well. In the light of the "data deluge" it makes sense to extend peer review to the data itself and this way evaluate the degree to which the data are fit for re-use. This paper describes a pilot study at EASY - the electronic archive for (open) research data at our institution. In EASY, researchers can archive their data and add metadata themselves. Devoted to open access and data sharing, at the archive we are interested in further enriching these metadata with peer reviews.

As a pilot, we established a workflow where researchers who have downloaded data sets from the archive were asked to review the downloaded data set. This paper describes the details of the pilot including the findings, both quantitative and qualitative. Finally, we discuss issues that need to be solved when such a pilot is turned into a structural peer review functionality for the archiving system.

Introduction

Increasingly, research data are considered as important scientific results in their own right; no less so, or even more so, than publications. The amount of data becoming available also increases at an impressive speed, referred to with notions like "data deluge" (Hey & Trefethen, 2003; The Economist, 2010). As the report by the High Level Expert Group on Scientific Data (2010) puts it, this calls for action to "develop and use new ways to measure data value, and reward those who contribute it."

Peer review is the standard way to assess the quality of research publications. In a similar vein, data can be peer reviewed. To some extent this already happens as part of peer review of publications. The current paper reports on a pilot study carried out at our data archive EASY in a Web 2.0 style. Characteristic of Web 2.0 is the fact that content consumers – in our case, research data consumers – also *produce* content, which is available to other internet users – in our case, this means feedback on the data. Typically, websites that support online transactions, such as booking a hotel room or buying books, feature review mechanisms.

It is to be expected that peer reviews of a data set provide useful feedback to the depositor, possibly even a kind of reward (Pfeiffenberger & Carlson, 2011). Moreover, metadata enriched by some kind of 'rating' or review will help others to better assess the relevance of a data set and stimulate discourse within the community. Furthermore, peer review of data is one of the three quality improvement methods recommended by Waaijers and van der Graaf (2011). Waaijers and van der Graaf investigated the operational aspects of the concept of quality for the various phases in the life cycle of research data: production, management, and use/re-use. They tested a list of nine potential methods with nearly four hundred representatives of three disciplinary domains: Physical Sciences and Engineering, Social Sciences and Humanities, and Life Sciences. Despite differences between the disciplines, the authors recommend three encompassing quality improvement measures, one of which is "Promote the provision of quality-related user comments on datasets" (ibid.).

In the remainder of this paper we start with the mission of DANS as research archive, and the role of data peer reviews therein. Second, the pilot setup is described. Next, we analyse the outcome of the pilot, draw lessons from it and illustrate how the ratings are made visible in EASY. The final section highlights some issues that need to be solved when DANS moves from a pilot phase to regular functionality.

DANS and the EASY Archive

DANS is the largest national data archive in the Netherlands in the social sciences and humanities, although its archiving task is not limited to these fields. Funded as a public institution by the Netherlands Organisation for Scientific Research (NWO) and the Royal Netherlands Academy of Arts and Sciences (KNAW), its mission is to promote sustained access to digital research data. "Digital research data" is meant in a wide sense. The data come in forms such as specific databases, spreadsheets, text, images, audio, video, and other multimedia formats. "Data" also extends to (digital) publications, including preprints and reports, as far as those are coupled to data.

Moreover, research information as provided by NARCIS[1] is part of the digital research data container. In order to carry out its mission, DANS encourages scientific researchers to archive and reuse data in a sustained manner, for instance by means of EASY, the Electronic Archiving System.

EASY[2] is open for self-archiving. After registration as user, a set of interfaces leads the researcher along a pathway of data archiving while at the same time collecting the metadata and provenance data information. The typical unit in EASY is the so-called data set. Central to sustained data storage is that the data should be traceable, accessible and usable at all times. Upon submission, archivists therefore check whether the files are in a sustainable, so-called "preferred", format and whether the necessary Dublin Core[3] metadata elements have been provided and fit the data. Furthermore, the Data Seal of Approval[4], which was originally developed by DANS, is used as a criterion for traceability, accessibility and usability. EASY has acquired this Data Seal. As part of sustained accessibility, data sets are provided with a persistent identifier. At DANS, different forms of peer reviews of data have been discussed in order to increase the visibility of data sets and thereby their accessibility. Furthermore, peer review is regarded as additional quality control for data sets archived in EASY.

As a consequence of DANS's origins, the focus in EASY is on data sets from the social sciences and humanities. EASY stores many archaeological data sets as well. The primary target group for DANS consists of scientific researchers. Students who are being prepared for doing research and researchers in training are also part of this group. In addition, the interested general public is also welcome to use the services.

Pilot Setup

The primary approach of this pilot was to rely on the "active" community of EASY users. In particular, we focused on re-users of data, not on self-archivers or domain experts. In other words, we contacted data set consumers. For this pilot, people who have downloaded a data set from the archive between October 2009 and February 2012 were asked via email to fill in an online questionnaire to review this particular data set. Frequent downloaders received up to three email requests. In total 3631 emails have been sent out in three rounds: in November 2010 (Dutch only), in April 2011, and in March 2012 (email and questionnaire in both Dutch and English).

As a questionnaire tool SurveyMonkey[5] was used. The respondents were asked to answer 11 questions and, depending on their answers, up to five follow-up questions such as "*Why not*?"[6]. The primary aim was to get a rating and comments for a specific data set and not a qualitative, in-depth reviewing report. Secondary goals were to achieve information about why people download data sets and how well the EASY website supports findability of data sets. Accordingly, questions were arranged in

---

[1] NARCIS: http://www.narcis.nl/?Language=en

[2] EASY: https://easy.dans.knaw.nl/ui/home

[3] Dublin Core metadata elements: http://dublincore.org/specifications/

[4] The Data Seal of Approval: http://datasealofapproval.org/

[5] SurveyMonkey: http://nl.surveymonkey.com/

[6] The questionnaire is available from the authors on request.

three sections: data set aspects, research aspects, and the EASY website. In the next sections, we detail the questions and responses.

Some of the questions allowed to enter free text, but most of the questions were five point scales ranging from "bad" (1) to "very good" (5). An example for such a rating is the question to evaluate the quality of the downloaded data. Such questions also offered the option "not applicable". Open questions were used to ask for comments and for keywords or tags; these questions were optional. The survey therefore yields both quantitative and qualitative information.

Intentionally, no definitions or examples of "quality" or other concepts were provided in the questionnaire; interpretation was left to the respondents, and so to the norms and standards of their corresponding communities. This runs counter to the study by Waaijers and van der Graaf (2011), who distinguish quality during the data production phase, quality of data management, and scientific or scholarly quality. However, it is in line with the archiving process at DANS, which only checks relatively external quality aspects (such as preferred file formats and metadata) but refrains from evaluating the scientific quality of the submitted data. This policy implies that DANS is relying on quality standards in the communities from which users submit. To ensure the formal quality of data submissions, DANS also offers training and manuals for various research disciplines, but the pilot did not relate to this.

Pilot Findings

A total of 573 persons have responded, which is a response of 15.8%. However, nearly 30% of the respondents did not finish the survey. Furthermore, many questions were optional and follow-up questions of the "*Why not*?" type were only asked if relevant. For these reasons the number of responses per question varies.

The respondents were asked to select a job title from a short list provided in the questionnaire. 49% of the respondents are researchers, 19% archaeologists, 11% students, 5% policy makers, and 17% hold other positions, such as teacher.

Currently in EASY, the reviews of a particular data set are made public in an anonymous way. In that presentation, partly a selection and partly an aggregation of survey-based answers is displayed. We discuss this in more detail in the section "Responses Presented in EASY", which also contains a screenshot (Figure 2). However, in the following subsections the responses to the survey questions are aggregated over the data sets. The three subsections present the findings for groups of questions addressing features of the data sets, features of research aspects around the data sets, and features of the EASY website, respectively.

Data Sets

The first question (further referred to as Q1) was:

> "'Data set' refers to the data files that together constitute one study, even if you may have downloaded just a single file. How would you judge the downloaded data on the following aspects?"

The aggregated scores can be seen in Table 1.

Aspect	Very Good (5)	Good (4)	Neither good nor bad (3)	Insufficient (2)	Bad (1)	N/A (0)	Average rating	N=
Data quality	110	283	36	3	0	45	4.16	432
Quality of the documentation	113	253	61	16	2	32	4.03	445
Completeness of the data	105	247	61	9	1	54	4.05	423
Consistency of the data set (if applicable)	67	196	48	4	0	162	4.03	315
Structure of the data set (if applicable)	59	216	50	12	1	139	3.95	338
Usefulness of the file formats	126	257	39	13	8	35	4.08	442

Table 1. Aggregated scores for data set aspects (Question 1).

For this question the respondents were offered five scores (from *Very good* to *Bad*) and the option "not applicable" (N/A). The table contains an aggregation of all respondents selecting one of the options. Overall, this question was answered by 477 respondents, but for each aspect a different number of respondents chose the N/A option and so N (which can be obtained by 477-[N/A]) varies for the different aspects. The average rating per aspect is based on its particular N.

Among the respondents, the group of researchers (N=201) is somewhat more positive than the total group of respondents (this is not visible in Table 1).

From the answers to the open question (Q2) "*What do you like about the data set*?" (answered by 55% of the respondents) two topics emerge after manual inspection. First, that the data set is complete[7], large, extensive, covers a long time span, or forms a large sample, in short, that it is comprehensive. Second, many responses refer to the (online) availability and accessibility of the data. The respondents were also asked what they were "*not satisfied with*" (Q3). Here 19% of the respondents gave answers like "no documentation in English" and "it would be easier if the file had not been split into two periods", to give two examples.

Common on the Web are user-assigned tags. Although there is no sharp distinction between tags and keywords, a difference in practice is that tags are often thought up by users, whereas keywords are often selected from a vocabulary. Tags can serve different purposes, such as organising one's personal information or advising others. Therefore, we specified a purpose in the open question (Q4), but used the term "keyword" because it is more usual in research: "*Which keywords would you assign to the downloaded data set such that it is found more easily by other researchers?*" 54% of the respondents provided on average 2.9 tags.

---

[7] The majority of the open text responses and tags cited in this paper were originally given in Dutch.

Figure 1. Tag cloud of the top 35 tags assigned to data sets in the pilot (Q4).

Almost all tags refer to the content ("excavation", "energy consumption"), whereas opinion tags (e.g. "easy to use") are very few. As an impression, Figure 1 is a cloud of the most frequently assigned tags, after manual editing. A larger font reflects higher frequency.

The final question (Q5) in this section pointed to quality: "*Would you recommend this data set to other users?*" This question was answered affirmatively by 92% of the respondents. Figure 2 later illustrates how EASY, for a single data set, presents the response to questions Q1 and Q5.

Research

The question: "*What was the most important reason for downloading this data set?*" (Q6) was presented as an open question. The 352 responses were manually classified as research (62%), "out of interest" (9%), "for study or educational purposes" (9%), and miscellaneous (20%). Examples are "GIS survey into election results of populist parties" and "my library does not have the relevant papers".

70% of the respondents found the data set "*helpful in answering [their] research questions*" (Q7). 12% answered in the negative and 18% opted for "not applicable". 50 respondents elaborated on their negative answers by selecting, for instance, the options: "*the contents of the data set are not what I expected*" (34% of them) or: "*not relevant enough*" (38%).

One way to operationalise the value of a data set is when it is used for a publication. 15.3% of the respondents have "*used the data set for a publication*" (Q8). 59% indicated to "*intend to use the data set for (another) publication*" (Q9). Figure 2 shows how EASY presents the information from Q8 and Q9 for a single data set.

EASY Website

In a similar vein as the data sets, the EASY website was rated. Table 2 aggregates the responses to Q10, which are somewhat lower than the ratings for the data sets in Table 1, but once again close to four on a five-point scale. Finding the data is the least satisfactory aspect. Consequently, a follow-up question soliciting comments and suggestions for improvement yielded mainly answers with regard to search functionality and metadata.[8]

---

[8] There has been a new release of EASY between survey rounds two and three, based among other things on various kinds of user feedback. The responses show no effect of this change.

Aspect	Very Good (5)	Good (4)	Neither good nor bad (3)	Insufficient (2)	Bad (1)	N/A (0)	Average rating	N=
Clarity	70	234	99	21	3	10	3.81	427
Information about the data	65	278	70	14	0	10	3.92	427
Finding the data	57	204	108	50	8	10	3.59	427
Consistency of the data set (if applicable)	82	247	68	24	7	10	3.88	427

Table 2. Aggregated scores for website aspects (Question 10).

Almost by definition, "The User" is essential in a Web 2.0 environment. This could be interpreted as an argument for presenting information about a reviewer along with her review. However, on the other hand users should feel free to review data sets anonymously; after all, EASY also allows anonymous downloading. The final question in the pilot questionnaire (Q11) concerned this issue. It turned out that 54% of the respondents preferred "*remaining anonymous*" to the alternative of having EASY show their "*name and organisation as used when registering with EASY*".

Responses Presented in EASY

From the outset of the pilot it was clear that not all responses would be published in EASY immediately. The website ratings do not relate to specific data sets and will be used only within DANS internally. Furthermore, we have no convincing design yet for presenting the open text responses, and therefore present only quantitative responses, as can be seen in Figure 2.
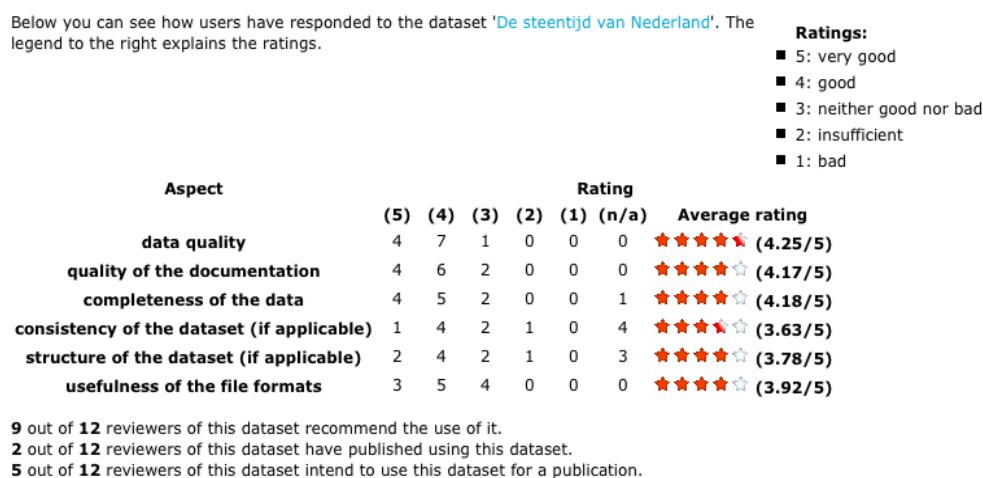
Figure 2. Screenshot with ratings, based on 12 reviews of the same data set.

An EASY user sees the following information:

- Top left, they are informed that this page concerns user responses about the data set "The stone age of The Netherlands".

- In the middle, the six aspects of data sets from question Q1 are listed, providing the ratings and their frequencies.

- At the right-hand side the average ratings are visualised in a star notation as well as a fraction, both on a five point scale.

- The ratings are explained at the top right.

- Finally, the three text lines at the bottom state how many reviewers recommend the use of the data set (Q5) and how many have used it for publication (Q8) or intend to do so (Q9), respectively.

Analysis

How appropriate is this peer review process? Given that this is a first explorative experiment, we are quite satisfied with the outcome. First of all, it was a great learning experience and an interesting implementation experiment of data set peer review in practice. The response rate of 15.8% is acceptable. Furthermore, the high ratings, the relevance of tags and comments given by the respondents, and the huge willingness to recommend the data sets are convincing signals. This is probably not surprising given that the respondents voluntarily contributed to the questionnaire, so we can assume some satisfaction with the data sets. The format of filling in an online questionnaire also seems to work well and may improve when the reviewing functionality will be more closely integrated within the EASY environment. Still, one should keep in mind that, first, not all available review data is yet shown in EASY (see Figure 2) and second, that the part that is visible has not been evaluated by the community. Furthermore, no attention has been paid to the non-response and its potential explanations.

With these provisos, what do we learn from the pilot?

- Reviewers rate data sets positively (around four on a five-point scale) and over 90% of them would recommend the data set to others. Nevertheless, the qualitative responses show that they are critical.

- 70% of the respondents found the data set helpful in answering their research questions. Especially given that for nearly 20% of the respondents this question was not applicable, this is quite satisfying.

- From the many answers to "*What do you like about the data set*?" an interpretation of "quality" of data sets from the user perspective can be derived, as we saw that the respondents set great store by comprehensiveness and online availability.

- More than half of the respondents intend to publish (again) using the downloaded data set. This seems to be an interesting indicator for the data's value, but probably Nielsen's first rule of usability applies: "Definitely don't believe what people predict they *may* do in the future"

(Nielsen, 2001). Although requested in the user's licence, DANS seldom receives references to publications that make use of downloaded data sets.

- Presenting quantitative feedback in EASY is much more straightforward than presenting qualitative feedback or tags. For this reason only the former is currently visible in EASY, on condition that minimally two reviews are available (see Figure 2).

- To achieve more quantitatively interpretable answers some open questions should be rephrased as closed (multiple choice) questions, for instance the question about the main reason for downloading a data set.

- Nearly 50% of the respondents selected the job title of researcher, which seems surprisingly low in light of the fact that DANS is a service provider for research and science. However, a large share of EASY data sets belong to the (non-academic) archaeological domain and the job title of archaeologist was only added in the second round of the survey (April 2011; selected by 19% of the total respondent population). When the reviews continue, the numerical balance between (self-indicated) archaeologists and (self-indicated) researchers is therefore expected to change.

- So far roughly 70 people preferred the English questionnaire to the Dutch one. Consequently, their remarks and tags are in English.

Considerations for Future Development

The amount and quality of the reviews are such that we have begun to develop a structural peer review process for data sets in EASY. In this section we discuss several issues that need to be taken into account, clustered in three subsections. The content of the review as well as the review process will be redesigned in the near future. We end with some suggestions for a more distant future.

Content and Presentation of the Reviews

No information about someone's research domain is required for registration in EASY. It is plausible that researchers mainly re-use data sets from their "own" domain, but we do not know exactly to what extent downloaders stick to "their" domain or cross borders. To find out more about this we will add a question to the survey about the respondent's research domain, although one should keep in mind that so far half of the respondents do not qualify themselves as researchers. We see no need for other new questions.

Currently, EASY shows quantitative scores when at least two reviews of a data set are available. Two is admittedly an arbitrary number, but we see no better or less arbitrary one and will therefore continue this way. All numerical information is simply added and where relevant averaged, i.e. reviews from key researchers, students, or archaeologists working in the private sector have the same weight.

First reactions from depositors confirm our conviction that the archive should also present qualitative feedback and tags, but critical mass is needed before this is reliable

(in a non-statistical, Web 2.0 sense). Furthermore, we must decide on the language(s) for presentation. The feedback is predominantly in Dutch; presenting comments in two languages is unproblematic, but how informative would a bilingual tag cloud be? Some moderation or harmonisation seems inevitable for effective communication. Another tag issue to be solved is where tags are most useful: at the level of individual data sets, aggregated to tag clouds at the level of research domains, or even at the level of the archive. The cloud in Figure 1 contains tags for data sets from various disciplines and would therefore fit the third option.

Review Process

In principle, frequent downloaders might receive a request for reviewing each time. It is crucial to minimise the effort of reviewing and not "bother" our customers. While the data set questions and the research questions are relevant every time, questions about the website should only be asked once. Also, an easy opt out procedure is needed. In the long run it might be worthwhile to store some information, such as the domain of research, and next time support the reviewer with pre-filled answers.

Exactly when to send a request for reviewing is a matter of further experimentation. Basically, the archival system will have features that resemble a hotel booking site: Shortly after an online transaction – downloading a data set as if booking a hotel room – customers are requested to express their satisfaction. Obviously, for data this time span must be longer than for a hotel room. In the pilot this time span was at least a month. Perhaps the scores for "Not applicable" in Table 1 might be interpreted as: "it is too early for me to evaluate the data set", but we have received no comments about this. On the other hand, in the third survey round a few people have indicated that they found the time passed since downloading (up to ten months) too long to evaluate the data set.

Distant Future

A reviewer of an earlier version of this paper mentioned the situation in which a data set that initially looks promising, turns out to be disappointing some months later and vice versa. Although this is an interesting use case, for the time being we will not design functionality that enables one to revise one's earlier review. Nor will we shortly enable spontaneous reviews, i.e., reviews not triggered by DANS. Peer reviewing of research data is still in its infancy and DANS opts for a cyclic approach of gaining experience and enhancing the review system, rather than trying to support all kinds of use cases from the beginning. Another feature we foresee for a later cycle is the possibility that reviewers contact depositors or the other way round. Online contact within a community is common on the Web, but it clearly has ramifications for the anonymity aspect, and we find the idea of, for instance, "hiding behind" nicknames in a research environment unattractive.

To conclude: the peer-reviewing process at DANS and elsewhere certainly has some way to go, but already the pilot results have convinced us that peer review of open data in an archival context is feasible and yields valuable information for a large audience.

References

Hey, A.J.G. & Trefethen, A.E. (2003). The data deluge: An e-Science perspective. In F. Berman, G.C. Fox & A.J.G. Hey (Eds.) *Grid Computing - Making the Global Infrastructure a Reality*. Chichester: Wiley and Sons.

High Level Expert Group on Scientific Data. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data – Final report.* Retrieved from http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707

Nielsen, J. (2001). *First rule of usability? Don't listen to users.* Retrieved from http://www.useit.com/alertbox/20010805.html

Pfeiffenberger, H. & Carlson, D. (2011). "Earth System Science Data" (ESSD) — A peer reviewed journal for publication of data. *D-Lib magazine 17* (1/2). doi:10.1045/january2011-pfeiffenberger

The Economist. (2010). The data deluge: Businesses, governments and society are only starting to tap its vast potential. *The Economist*. Retrieved from http://www.economist.com/node/15579717

Waaijers, L. & van der Graaf, M. (2011). Quality of research data, an operational approach. *D-Lib magazine 17*(1/2). doi:10.1045/january2011-waaijers