

Library Lab Report: Who's Using Data and Why? (Dataverse Network Project)

Project Summary:

The Dataverse Network Project is an open-source application for publishing, citing, and discovering research data. For this project, we added Guestbook functionality to the Dataverse Project to allow data owners to gather information on the people who use their data (via download, analysis, visualization, etc).

Accomplishments:

We added the ability for a data owner to create a "Guestbook" form that would be filled out by end users when they went to download or otherwise use the data. This form, by default, asks for basic identity information: first name, last name, email, institution, and position. The data owner can also add custom questions, either as free form text (e.g. "What will you use the data for?") or with answers as radio button options (e.g. "Gender?"). All of these fields, including the defaults, can be set to be required or optional.

We also added a summary page where a user can review the gathered info. In addition to the questions asked above, we store the associated study, file, and task (whether they were downloading a file, running a analysis, viewing a visualization, etc.). The data owner also has the ability to download this data as a CSV (comma separated values) file, which they could then open in Excel, in order to run further metrics.

Challenges:

During development, we had the usual challenges. While a simple concept at first, identifying and coding for all the edge cases caused the development of this feature to span two releases.

Next Steps:

There are many next steps for the Dataverse Network Project, of course. Related to Guestbook, we would like expand the reporting page so that in can include many of the basic metrics that the data owners will want (right now the data owner can download the csv file and gather these metrics, but it would make sense to allow them to do this on the Dataverse site directly. We also plan on getting further feedback from our library partners (e.g. Diane Sredl) to determine other next steps.

Code:

The code is open source and available as part of the Dataverse Network Project code on SourceForge: <https://sourceforge.net/projects/dvn/>

Budget Spent:

This project was granted \$30,000, and we spent nearly the full amount, \$29,922.

Publicity:

We participated in the video interviews for the Library Lab Projects. Also, we continue to do our standard publicity for Dataverse and, through our support email, had received several requests (including from libraries at other universities) for this type of functionality. We responded to those emails, reporting that it was now available and that they should try it. Lastly, we have a twitter feed, where we announced the release of the new version of DVN and linked to a page with new features, including the Guestbook functionality: <http://thedata.org/pages/latest-release>

Presentations:

While we have not given many presentations specifically for the Guestbook functionality, it is usually included in the presentations we give for the Dataverse Network Project as a whole. We have met with several groups within Harvard to promote DVN (e.g. the libraries) as well as external groups that are interested in running their own installations of Dataverse (e.g. representatives of the University of Connecticut, also representatives from Fudan University in China). We also conduct regular Dataverse training in collaboration with the libraries.