

Pacific University

CommonKnowledge

Volume 10 (2010)

Interface: The Journal of Education, Community
and Values

5-1-2010

Initial Development of an Online Repository to Facilitate Data Sharing and Collaborative Research: A Berglund Fellowship Report

Isaac Gilman
Pacific University

Daniel Munoz
Pacific University

Recommended Citation

Gilman, I., & Munoz, D. (2010). Initial Development of an Online Repository to Facilitate Data Sharing and Collaborative Research: A Berglund Fellowship Report. *Interface: The Journal of Education, Community and Values* 10(4). Available <http://bcis.pacificu.edu/journal/2010/04/article.php?id=669>

This Article is brought to you for free and open access by the Interface: The Journal of Education, Community and Values at CommonKnowledge. It has been accepted for inclusion in Volume 10 (2010) by an authorized administrator of CommonKnowledge. For more information, please contact CommonKnowledge@pacificu.edu.

Initial Development of an Online Repository to Facilitate Data Sharing and Collaborative Research: A Berglund Fellowship Report

Rights

Terms of use for work posted in CommonKnowledge.

Initial Development of an Online Repository to Facilitate Data Sharing and Collaborative Research: A Berglund Fellowship Report

Posted on **May 1, 2010** by **Editor**



By Isaac Gilman, MLIS

Pacific University Libraries and

Daniel Munoz, Ph.D.

Pacific University School of *Professional Psychology*

Summary: *At Pacific University (Oregon), a project funded by the Berglund Center for Internet Studies was developed to introduce the idea (and benefits) of data sharing to university faculty engaged in research. Pacific is a small, private doctoral/research university (Carnegie Basic Classification) with four campuses in western Oregon and four colleges: College of Arts & Sciences, College of Education, College of Health Professions and College of Optometry. This project systematically evaluated several different approaches to encourage use of data sharing.*

Introduction

Research, as defined by the United States Code of Federal Regulations, is “a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.”[1] This contribution to knowledge has always been a primary purpose of scientific research and implies the need to communicate the results of research so that others may learn from, and build upon, that work. The traditional model for communicating research has been to analyze collected data and then summarize key points of that analysis for publication in scholarly journals. [2][3] In an era of paper journals, the necessity of distilling results and only sharing data that was necessary to illustrate or support conclusions was clear – there was a finite amount of space to share each study. However, in today’s digital

world, where online journals have rendered page fees meaningless and there is seemingly no limit to the available space for sharing information, the possibilities for sharing the full extent of research results (and data) are virtually limitless. [4]

Data sharing is not a new practice. For example, professional associations, like the American Psychological Association, encourage their members to make their data sets available to other professionals for re-analysis and in fact codify that aspiration in their code of ethics. [5] Some journals, such as the *American Economic Review*, have explicit policies about requiring published authors to make their data available to other researchers. [6] In the last decade, however, there has been an increased call for research data sets to be made openly available online without the mediating step of requesting access from the data owner(s). Researchers from both the biomedical and social sciences have communicated the necessity of open access to data sets for reasons including “accelerated scientific progress” [7] and the need to ensure the “transparency of analytic work” [8] by providing an additional level of peer review. [9] Recent professional meetings in Toronto and Rome addressed the importance of both prepublication and post-publication data sharing and the need for standards that would make it easy to identify and access available data sets. [10][11] The benefits to researchers from sharing their data, including potential increased citation rates for associated articles, have also been examined. [12]

Many disciplines have made efforts to provide the necessary infrastructure to encourage data sharing. This infrastructure usually takes the form of a data repository where researchers can deposit their data sets, code books and other materials necessary for others to replicate their original analyses. Examples of disciplinary repositories include arXiv.org, the ICPSR data archive, PsychData, Protein Data Bank, GenBank, PANGAEA and the World Data System. [13] The availability of data sets from researchers around the world in these repositories encourages collaborative work across institutional and national boundaries and is fundamental to researchers’ abilities to solve data-intensive problems that exceed the capabilities of individual scientists, labs or research groups. [14]

For some researchers, contributing data to large disciplinary repositories with global users may initially seem daunting. The benefits of possible collaborations or initiating new projects from existing data can be overshadowed by fears of others’ capitalizing on their results for publication glory or of re-analysis revealing mistakes in their methodology or initial analysis. Additionally the need to completely de-identify data that may contain sensitive human subjects or proprietary information (and possibly receive permission from institutional review boards or other oversight bodies to release the data) can present an ethical dilemma for researchers. [15][16] Of these potential impediments to sharing, the last is certainly valid—but even that may be ameliorated (in most cases) through careful coding and removal of unique identifiers. Education and support is vital to help researchers navigate these concerns about data sharing and encourage them to take the first steps to contributing their information to an openly available collection of research data.

Data Sharing in a Local Context: Motivating Factors

At Pacific University (Oregon), a project funded by the Berglund Center for Internet Studies was developed to introduce the idea (and benefits) of data sharing to university faculty engaged in research. Pacific is a small, private doctoral/research university (Carnegie Basic Classification) with four campuses in western Oregon and four colleges: College of Arts & Sciences, College of Education, College of Health Professions and College of Optometry.

As an institution, Pacific is committed to interdisciplinary and collaborative teaching, service and research. Though the 380 faculty across the four colleges represent a diversity of research interests, experiences and sources of data, there are faculty in different schools and departments who share research interests and who represent potential collaborations. However, there is no central location within the University that faculty (or students seeking advisors/mentors) can access to identify potential collaborators. To meet that need, this project focused on the creation and initial promotion of a searchable database that would house metadata and data sets resulting from faculty research. Such a database will allow faculty and students to identify not only colleagues on campus with similar research interests, but even more micro-level items such as specific protocols or survey questions that may be of interest. By sharing descriptive information about faculty data sets, the database goes beyond simply identifying research interests to identifying aspects of faculty research that would not be captured in a simple “research interest(s)” label on a departmental website. For example, if faculty members identify themselves as having a research interest in “eating disorders”, it is likely that others outside of psychology might not envision potential collaborative opportunities with that faculty member. However, if the faculty member has datasets that, in addition to eating disorder data, includes measures of depression, health behaviors (e.g. smoking, drinking, exercise) and medical history that are searchable through the database’s metadata, then the relevance of that faculty member’s work expands to a greater proportion of faculty and students. That faculty member and their available research data would be of interest not only to colleagues interested in eating disorders, but also to those with interests in depression, health and medical behaviors—which could attract interdisciplinary collaboration (e.g. with pharmacy, physical therapy, physician assistants, public health, etc.).

A secondary motivator in creating a centralized database for faculty research data is the cost of initiating research. Research at the university level, particularly in the social and health sciences, is dependent on a myriad of funding sources such as institutional, corporate or government grants. In institutional terms, Pacific currently only provides a competitive \$3,500 faculty development grant where basic research must compete with applications for travel or materials unrelated to basic research. Corporate foundation and governmental grants (e.g. National Institutes of Health grants) often require a level of infrastructure (i.e. course release time, start-up funding) that is not easily available at Pacific University. Furthermore, to be competitive for foundation or government grants, researchers typically need to demonstrate a consistent and productive output of published research, which, again, may be hampered by the aforementioned limitations. Finally, grant providers are increasingly requiring a multidisciplinary component as part of a fundable proposal. At Pacific University, the time and work demands of each individual discipline often make it difficult for faculty to have the opportunity to initiate conversations that

may lead to collaborative research ideas. The database initiated through this project will facilitate access to existing datasets for faculty and students from different disciplines. Analyzing and re-purposing existing data both saves on the costs of initiating new research and also increases the publication productivity that makes researchers more viable prospects for external funding.

It is anticipated that faculty at Pacific will, on the whole, share similar concerns to other researchers related to providing widespread access to their research data. By focusing primarily on these local motivating factors and not pushing prematurely for faculty to deposit full data sets in disciplinary repositories and other widely used venues, the initial objectives of this project have been to introduce faculty to the benefits of openly sharing this type of information and to create a venue that makes it easy to share that information.

Materials & Methods

As technology options, knowledge and opportunities have shifted since the project's conception, the data sharing venue has evolved accordingly and will continue to evolve as participation—and the nature of participation—increases/changes in coming years. This report focuses on the two initial phases of the project—exploration and development—and will provide guidance for further growth/use of the database at Pacific moving forward.

Initial Database Platform

As noted in the introduction, online repositories for sharing datasets and related information are fairly common across multiple disciplines and it would make little sense to reinvent the repository when there are existing platforms available for use. One of the best examples of an existing platform that was created for this purpose is the Dataverse Network software developed at Harvard University (<http://thedata.org>). Dataverse Network software is open source, and available for local installation. [17] However, the free, hosted version of the software was utilized for this project. The hosted service is currently used by researchers at large universities, as well as by government agencies, and provides excellent support and services without the commitment of time and resources required by a local installation. Hosted “Dataverses” provide nearly all of the benefits of a local installation, and hosted data is backed up on a daily basis, with the last 6 months of data available on Harvard servers.

The Dataverse Network software is best explained by its creators from the Institute for Quantitative Social Science at Harvard University:

Individual Dataverses are self-contained virtual data archives [...] The extensive digital library services of each Dataverse include data archiving, preservation formatting, cataloging, data citation, searching, conversion, subsetting, online statistical analysis, and dissemination. Each Dataverse presents a hierarchical organization of data sets, which might include only studies produced by the Dataverse creator (such as for an author or research project), those associated with published work (such as replication data sets for journal articles), or data sets collected for a particular community (such as for a journal's replication

archive, or a college class or subfield). [18]

The first step in this project was configuring a hosted Dataverse for Pacific University (<http://dvn.iq.harvard.edu/dvn/dv/pacificu/>). Configuration of the Dataverse included branding with Pacific University colors and logo, the creation of hierarchical collections for all colleges and schools within the University, and the creation of indexing templates to guide the input of data set metadata. Currently, the Dataverse is not openly accessible and requires an account/password. [19] Once a Dataverse is “released” (made openly accessible), all Dataverse Network users are able to see search results from that Dataverse (though restrictions may be placed on the ability to download materials).

Exploration: School of Professional Psychology Pilot

After configuring the Dataverse, a pilot study to gauge faculty/student interest and response was conducted within the School of Professional Psychology (SPP). SPP was considered an ideal site for the pilot due to the relatively high output of data by its students and faculty. Students in SPP are expected, as part of their thesis and dissertation requirements, to seek out and initiate research projects, at least one of which typically generates quantitative data. Faculty within SPP represents a wide variety of research interests and many diverse research programs are available for student participation. However, there are barriers to connecting interested students with appropriate faculty researchers. There is no centralized list of specific faculty projects, only broad topic areas, and students typically need to initiate contact with each faculty to discover a) whether the faculty member has an active research program and b) if, so, what projects are currently ongoing. In addition, students within SPP may have research interests that do not coincide with active faculty, and thus need to look outside the department; unfortunately, there is no easily accessible database of University-wide research interests and projects. For SPP faculty who may wish to broaden their own research programs (and create opportunities for students) by seeking collaborators from across the College of Health Professions or the University at large, there is no centralized repository of cross-discipline research information.

The first step in the SPP pilot was collection of available data set information from faculty volunteers within SPP. Several emails were sent to the SPP faculty describing the project and interested respondents were directed to a secure online spreadsheet where they entered descriptive information about their respective datasets (i.e. subject population, assessment protocols, summary of findings, etc.). To mitigate possible concerns about sharing actual data, only descriptive metadata about existing data sets was collected for ingest into the Dataverse. The Dataverse records about the data sets provided descriptions of sample size, population and assessment instruments used in the studies. All fields in the database records were searchable, allowing users to look for information about data sets relevant to their interests (and search terms). In total, information about 6 studies from 7 participants was initially available in the Dataverse (as the project has progressed, information about additional data sets has been added).

Concurrent with the collection of data set information, an anonymous survey was sent via SurveyMonkey to all SPP faculty and doctoral students. The purpose of the survey was to determine a) interest in intra- and interdisciplinary collaboration within SPP and b) the current level of research activity within SPP. Two weeks following the initial survey, a second survey was sent to request feedback on the usability of the interface and to determine whether or not any collaborative ideas had been generated through use of the Dataverse. Prior to sending the second survey, Dataverse accounts were created for all SPP faculty and doctoral students to enable them to access the database and potentially upload information. Survey responses are discussed in the Results section of this report.

Development: Usability Testing

Following the SPP pilot, the investigators originally intended to implement a similar study across the College of Health Professions to see if the results from the SPP pilot were confirmed or challenged by faculty and students in different disciplines. However, developments in the available technological resources at the University necessitated a change in the planned methodology.

In June 2009, shortly after this project began, the Pacific University Library purchased a digital repository platform. This platform, Digital Commons®, from Berkeley Electronic Press (bepress), provides a means of preserving and disseminating online Pacific University faculty and student work. *CommonKnowledge* (the chosen name for Pacific's Digital Commons® site) is envisioned as the central portal for users both within and outside the University who would like to access scholarly and creative works created faculty, staff and students. *CommonKnowledge* collections (<http://commons.pacificu.edu>) are primarily open access, which means anyone with an Internet connection can discover and download materials that have been uploaded to the site. Bepress hosts *CommonKnowledge* (though it is configured to meet local needs), ensures that materials are indexed in online search engines (e.g. Google™), and provides comprehensive usage statistics for items posted on the site.

With the availability of *CommonKnowledge* and its shared objectives with the project (open sharing of information), it became logical to consider *CommonKnowledge* as a possible alternative venue for hosting data set information. Several immediate benefits to using *CommonKnowledge* were readily apparent: as a hosted product for which the University Library pays (as opposed to a free hosted Dataverse site), there is greater flexibility in configuration and available dedicated technical support; faculty across the University were already being exposed to *CommonKnowledge* and receiving education about how to contribute their work to the site; *CommonKnowledge* would not require users to create accounts/log in to view descriptive information about others' data sets (as with the Dataverse Network, which does not allow sites to be public unless they host actual data files); and, finally, *CommonKnowledge* provided a simpler user interface for both searching and uploading information about data sets. Though the first three benefits are fairly clear, the latter required testing of the hypothesized improvement in user experience.

Website usability testing can vary in complexity but, at its core, is usually concerned with the user's ability to comprehend, navigate and utilize functionality on a given site. The usability testing conducted for this project was comparative in nature, asking participants to complete the same tasks on both the *Dataverse* and *CommonKnowledge* sites, provide feedback as to the usability of each site through an assessment measure and interviews, and finally decide on a preferred site.

A convenience sample of five College of Health Professions faculty members and five College of Health Professions students was selected to participate in the usability testing. SPP students or faculty who had previously participated in the evaluation of the *Dataverse* site were excluded from participation. The order of use of both sites within each testing session was alternated (first participant used the *Dataverse* first; second participant used *CommonKnowledge* first, etc.) in an attempt to control for any possible carryover effect/bias that could occur by examining one site first (or last).

After providing basic demographic information to potentially aid in analysis, each participant was taken through a series of four tasks on each website. During the tasks, a think-aloud protocol was used (participants were asked to verbalize their thought processes and reasons for their actions/movements on the site). Immediately after completing the tasks on each site, participants were asked to complete an adapted version of the 10-item System Usability Scale© (SUS), an instrument developed for the purposes of evaluation industrial systems. [20] Though not originally designed for use with websites, it provides a standardized assessment of usability factors that are relevant to website use and design. At the end of the session, each participant was asked to comment on likes/dislikes of each site before providing a final conclusion as to the preferred site (*Dataverse* or *CommonKnowledge*).

Results

SPP Survey Results

Seventy-one faculty members and students (13 & 58 respectively) responded to the initial survey. Table 1 summarizes findings of interest. In summary, while many respondents report interest in conducting research and engaging in interdisciplinary collaboration, actual engagement was less.

Table 1.

Survey Item	% Respondents
Average number of papers published/presentations made (2008-2009 academic year)	1.36 (responses from 0-7)
% "somewhat" or "very" interested in conducting research	70.4%
% "somewhat" or "very" interested in interdisciplinary collaboration	56.3%

% engaged in research collaboration with colleagues in other disciplines (2008-2009 academic year)	23.9%
--	-------

% discusses research with colleagues in discipline (“sometimes” or more)	80.3%
--	-------

% discusses research with colleagues in other disciplines (“sometimes” or more)	35.2%
---	-------

For both faculty and students, time was a significant barrier in conducting research. Students also reported difficulties in finding faculty with similar research interests; difficulties in securing populations of research interest and problems with developing a topic.

Four faculty members and 19 students responded to the follow-up survey assessing use of the Dataverse site. Table 2 summarizes their responses.

Table 2.

Survey Item	% Respondents
Mean rating (out of 10) on ease of use of DVN	8(43.5%)
% that found data of interest on DVN	56.5%
% (of the 4 who contacted a researcher found on DVN) engaged in research as a result of that contact	75%
% (of the 4 who contacted a researcher found on DVN) engaged in research related to DVN data	66.7%
% willing to add their data to DVN	87%
% faculty that would use DVN to identify future research collaborations (n=3)	100%
% students that would use DVN to develop thesis/dissertation ideas (n=18)	100%

Usability Testing Results

Participants in the usability testing were predominantly female (the College of Health Professions student body is more than 60% female; faculty gender ratios within the University are closer to even), and ages ranges were between the twenties to fifties (ranges, not exact ages, were reported). All participants indicated using computers on a regular basis for both personal and professional purposes.

Participants showed slightly more difficulty completing tasks using the Dataverse site, but there were specific points of confusion indicated with using both sites, so the error rate is not a significant comparative factor when looking at the two sites. Participants *did*, however, express greater frustration with difficulties experienced on the Dataverse site, which included not being able to determine how to submit information to the database, and difficulty using the search/advanced search functions to locate information about data sets in the database.

The most significant results were found in the SUS scores and the post-test interviews. SUS scores are based on 10 Likert items, with a total (best) possible score of 100. The mean SUS score for the Dataverse site was 51.75, while the mean SUS score for *CommonKnowledge* was 86 (a statistically significant difference: $t(9) = 6.13$, $p < .001$). In addition to this discrepancy in scores, 100% of participants indicated that they preferred the *CommonKnowledge* interface, finding it more “intuitive”, navigable and generally easier to use when attempting to search for information or add information to the database. Importantly, most participants also encountered a similar point of confusion on the *CommonKnowledge* site, which is easily rectified and which will make the site even easier to use going forward.

Discussion

The results of the pilot study indicate that for students and faculty within SPP (and, potentially within the College of Health Professions, though a more representative sample would be needed to extrapolate to that population), there is a disparity between their interest in research and interdisciplinary collaboration and their actual engagement in these activities. Availability of resources (especially time) was a barrier to engagement, but difficulty in finding colleagues with similar interests was also noted, which indicates a possible use for a centralized database of faculty research data and information. This was confirmed by the results of the follow-up survey assessing the Dataverse site, in which 100% of respondents indicated interest in utilizing the Dataverse to identify potential ideas or collaborators, and nearly 90% of respondents were willing to contribute information about their own research to the database. These results suggest that a centralized database of research data sets/information may help attenuate barriers (such as time, cost, and knowledge of existing research) and assist in the development of new research collaborations.

While the results of the SPP surveys indicate the conceptual use/viability of a local repository for data sets/research information, the usability testing results suggest that *CommonKnowledge* may be the most appropriate platform for this repository. Though the *CommonKnowledge* platform is not as robust as the Dataverse site in terms of data-related functionality (Dataverse includes built-in analytical functions, more advanced search options and significantly more detailed/complex metadata options), selecting a platform that faculty and students find simple to use is an incredibly important factor in soliciting/sustaining participation in data sharing activities across the University. Using *CommonKnowledge* as the portal for research data information also means that such information will be collocated with faculty publications, presentations and other work – creating a more complete picture of faculty interests and activities that can be searched/accessed by students and colleagues through a single interface.

It is important to note, however, that *CommonKnowledge* does not immediately address some of the key issues identified by data sharing advocates. Altman & King [21] identified six elements that should comprise a basic citation to a data set available online:

1. Author(s)
2. Date of publication/date data was publicly released
3. Title
4. Unique global identifier
5. Universal Numeric Fingerprint
6. Bridge service

The first three elements are easily included in a *CommonKnowledge* metadata record; the last three are not (at present). A unique global identifier (e.g. a Digital Object Identifier (DOI®)) provides a persistent identifier/access point for a digital object, irrespective of whether the URL of the object's hosting website changes or the location of the object changes. Altman & King [22] recommend the Handle (HDL) System®, upon which DOI® is based, for assigning unique identifiers. HDL identifiers must be assigned by an organization that has registered as a Resolution Services Provider (RSP) [23]; currently, neither Berkeley Electronic Press nor Pacific University is a RSP.

While a unique global identifier provides a persistent means of accessing a digital object, a Universal Numeric Fingerprint (UNF) addresses the issue of changes in/multiple versions of a data set. A UNF is "is an algorithmic tool used to verify that a data matrix or related digital object that is produced in one software environment and/or format has been correctly interpreted when moved to a different environment and/or format." [24] Put simply, a UNF is automatically generated by software that examines the contents of a data set and calculates a "string of numbers and characters that summarizes all the content." [25] The UNF for a given data set will never change; if any data elements within the set are changed, and a UNF for that set is generated, it will be different from the original UNF. [26] For researchers who may be concerned about whether the data set that is cited somewhere is exactly the same as the one they are examining, a comparison of the UNFs can provide confirmation that they have accessed the correct version.

Finally, Altman & King [27] recommend a bridge service as part of the citation to online data sets. A bridge service provides a means for web browsers to recognize (and thus, access) unique global identifiers by adding a URL to the front of the global identifier. For example, the DOI® for Altman [28] is 10.1007/978-1-4020-8741-7_57. Without a bridge service to translate the DOI®, a web browser would not be able to access this digital object. However, with the addition of a bridge service (in bold), the object can be browsed to: http://dx.doi.org/10.1007/978-1-4020-8741-7_57.

CommonKnowledge is not able to provide unique global identifiers (with the accompanying bridge service) or UNFs for data sets it may hold. Dataverse sites, however, will automatically generate

both unique global identifiers and UNFs for data sets that are uploaded. This is a clear benefit to the Dataverse platform. Altman & King [29] are not alone in raising issues related to clear and consistent data citation; recent discussions of data sharing have discussed similar issues. [30][31][32] In particular, the ability to provide a persistent link to available data sets, and the ability to confirm that the data accessed is identical to the data used/cited in publications is of great importance for researchers who may wish to attempt to replicate or validate others' results. A combination of the usability of *CommonKnowledge* and the robust functionality of the Dataverse Network would be ideal for ensuring that efforts within Pacific University to share data continue to grow.

Conclusions/Future Directions

This present project represents a starting point for data sharing initiatives at Pacific University. The primary and secondary research involved in this study provide the information necessary to guide an informed effort to encourage faculty and student participation in sharing research data and in utilizing data shared by others. The identification of potential barriers and benefits to faculty involvement will inform the development of educational materials and outreach for faculty, while the identification of a preferred platform for accessing information about research data provides a clear technological direction for this initiative.

Several directions for future work across the University are evident from this project. First, given the barriers that exist for researchers who are considering sharing their data sets online, it may be unreasonable to promote this within Pacific University as the only viable option for data sharing. The creation of a searchable database of metadata/descriptive information about existent data sets within the University would be a significant achievement in and of itself. Successfully encouraging faculty to take the time to contribute this information should be the initial priority (though the possibility of uploading/sharing actual data sets should always be an option as well).

Second, *CommonKnowledge* should be promoted as the repository for information about research data from across the University, and faculty should be provided assistance in learning to add their information to the repository. This assistance will be made available through online tutorials (linked from the *CommonKnowledge* site, <http://commons.pacificu.edu/pudata/>), information sessions and individual meetings, as appropriate.

Third, in anticipation of potential faculty willingness to share actual data sets, the Dataverse site developed for this project should be maintained and should be used as the site for data storage. In this way, unique identifiers and UNFs may be generated for uploaded data and users will have the option of utilizing the full functionality of the Dataverse. However, *CommonKnowledge* will still be promoted as the portal/search interface for all research data information; links in metadata records will allow users to easily access actual data sets (if available) in the Dataverse.

Fourth, dialogue with the Pacific University Institutional Review Board should be started to discuss data sharing in general as well as potential ways for investigators submitting proposals to

the IRB to proactively identify a plan to share resulting research data (including any plans to de-identify the data before sharing, if necessary). Finally, it would be beneficial to explore the possibility with promotion and tenure committees of formally recognizing the merits of data sharing. With proper citation, use of data sets can be tracked and recognized. However, even the act of contributing data (in a form that is usable by other researchers) should be recognized as a significant contribution to a faculty member's profession. [33][34][35]

Sharing research data (or, at the very least, information that the data exists) is a way not only for Pacific faculty and students to engage in intra- or interdisciplinary collaborations, but to also share their activities and expertise with the broader research community. Pacific faculty author dozens of publications every year as contributions to their respective disciplines; sharing the data associated with those publications presents both another contribution to disciplinary knowledge and another opportunity for faculty to be recognized for the work that they do. By providing resources like *CommonKnowledge* and *Dataverse*, and promoting their use, the University will enable its faculty to take advantage of these opportunities.

Acknowledgements

The authors wish to gratefully acknowledge the enthusiastic (and financial) support of the Berglund Center for Internet Studies at Pacific University, which made this project possible. The authors were co-recipients of a Berglund Fellowship for 2009-2010.

Endnotes

[1] United States Department of Health and Human Services, *Code of Federal Regulations, Title 45: Public Welfare, Part 46: Protection of Human Subjects*, Revised January 15, 2009: <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm> (accessed April 13, 2010).

[2] Jim Gray, "Jim Gray on eScience: A Transformed Scientific Method," edited by Tony Hey, Stewart Tansley and Kristin Tolle. In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley and Kristin Tolle (Redmond, Washington: Microsoft Research, 2009), xix-xxxiii.

[3] Toronto International Data Release Workshop Authors, "Prepublication Data Sharing," *Nature* 461, no. 10 (2009): 168-170.

[4] Jeremy Freese, "Replication Standards for Quantitative Social Science: Why Not Sociology?" *Sociological Methods & Research* 36, no. 2 (2007): 153-172.

[5] American Psychological Association, "Ethical principles of psychologists and code of conduct," *American Psychologist*, 57, no. 12 (2002): 1060-1073.

[6] Freese, 2007.

- [7] Heather A. Piwowar, Michael J. Becich, Howard Bilofsky and Rebecca S. Crowley, "Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers," *PLOS Medicine* 5, no. 9 (2008): <http://dx.doi.org/10.1371/journal.pmed.0050183> (accessed September 28, 2009), 1315.
- [8] Freese, 2007, 154.
- [9] Glenn Firebaugh, "Replication Data Sets and Favored-Hypothesis Bias," *Sociological Methods & Research* 36, no. 2 (2007): 200-209.
- [10] Toronto, 2009.
- [11] Paul N. Schofield, Tania Bubela, Thomas Weaver, Lili Portilla, Stephen D. Brown, John M. Hancock, David Einhorn, Glauco Tocchini-Valentini, Martin Hrabe de Angelis and Nadia Rosenthal, "Post-Publication Sharing of Data and Tools," *Nature* 461, no. 7261 (2009): 171-173.
- [12] Heather A. Piwowar, Roger S. Day and Douglas B. Fridsma, "Sharing Detailed Research Data is Associated with Increased Citation Rate," *PLOS ONE* 2, no.3 (2007): <http://dx.doi.org/10.1371/journal.pone.0000308> (accessed September 28, 2009).
- [13] Bryn Nelson, "Empty Archives," *Nature* 461, no. 7261 (2009): 160-163.
- [14] Tony Hey, Stewart Tansley and Kristin Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Redmond, Washington: Microsoft Research, 2009).
- [15] Piwowar et al., 2007.
- [16] Freese, 2007.
- [17] Gary King, "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing," *Sociological Methods & Research* 36, no. 2 (2007): 173-199.
- [18] *The Dataverse Network Project*, <http://thedata.org/#section-3> (accessed April 15, 2010).
- [19] The Dataverse Network will not allow a Dataverse to be "released" (made public) if there are no data files attached to metadata records in the account. This ensures that Dataverse Network users are only searching through records that contain openly available data sets. There are currently no data sets associated with records in the Pacific University Dataverse.
- [20] John Brooke, "SUS – A Quick and Dirty Usability Scale," *UsabilityNet*, www.usabilitynet.org/trump/documents/Suschapt.doc.
- [21] Micah Altman and Gary King, "A Proposed Standard for the Scholarly Citation of Quantitative

Data,” *D-Lib Magazine* 13, no. 3/4 (2007):

<http://www.dlib.org/dlib/march07/altman/04altman.html> (accessed March 29, 2010).

[22] Ibid.

[23] Ibid.

[24] Micah Altman, “A Fingerprint Method for Scientific Data Verification,” in *Advances in Computer and Information Sciences and Engineering*, edited by Tarek M. Sobh (Springer, 2008): 311-316.

[25] Altman and King, 2007.

[26] Ibid.

[27] Ibid.

[28] Altman, 2008.

[29] Altman and King, 2007.

[30] “Credit where Credit is Overdue,” *Nature Biotechnology* 27, no. 7 (2009): 579.

[31] Nelson, 2009.

[32] Schofield et al., 2009.

[33] Piwowar et al., 2008.

[34] Toronto, 2009.

[35] Credit, 2009.

This entry was posted in Uncategorized by **Editor**. Bookmark the **permalink** [<http://bcis.pacificu.edu/interface/?p=3789>] .

15 THOUGHTS ON “INITIAL DEVELOPMENT OF AN ONLINE REPOSITORY TO FACILITATE DATA SHARING AND COLLABORATIVE RESEARCH: A BERGLUND FELLOWSHIP REPORT”

best tankless water heater

on **February 2, 2014 at 8:32 AM** said:

Wow, this paragraph is nice, my younger sister is analyzing these kinds of things, so I am going to inform her.

best garbage disposal

on **February 3, 2014 at 8:18 PM** said:

I am in fact happy to glance at this weblog posts which includes lots of helpful information,
thanks for providing such information.

vacuum sealer reviews

on **February 3, 2014 at 9:50 PM** said:

Good post. I learn something new and challenging on sites I stumbleupon everyday. It will always be interesting to read through content from other writers and use a little something from their web sites.

best humidifier

on **February 3, 2014 at 10:35 PM** said:

We are a gaggle of volunteers and opening a new scheme in our community. Your web site provided us with helpful information to work on.

You have done a formidable activity and our entire group might be grateful to you.

robotic pool cleaner reviews

on **February 4, 2014 at 1:18 AM** said:

Does your website have a contact page? I'm having trouble locating it but, I'd like to send you an email. I've got some creative ideas for your blog you might be interested in hearing.

Either way, great blog and I look forward to seeing it improve over time.

best jig saw

on **February 4, 2014 at 2:43 AM** said:

hello!,I really like your writing so much! percentage we be in contact extra about your post on AOL? I need an expert on this space to solve my problem. Maybe that is you! Looking ahead to see you.

rifle scope reviews

on **February 4, 2014 at 4:03 AM** said:

Hey! Would you mind if I share your blog with my twitter group?

There's a lot of folks that I think would really enjoy your content. Please let me know. Many thanks

best espresso machine

on **February 4, 2014 at 4:07 AM** said:

Hello would you mind letting me know which webhost you're using? I've loaded your blog in 3 completely different internet browsers and I must say this blog loads a lot faster then most. Can you suggest a good internet hosting provider at a fair price?

Thank you, I appreciate it!

best crossfit shoes

on **February 4, 2014 at 4:39 AM** said:

I tend not to drop a great deal of responses, but after browsing some of the remarks on Initial Development of an Online Repository to Facilitate Data Sharing and Collaborative Research:

A Berglund Fellowship Report | Interface. I do have a couple of questions for you if it's alright. Could it be only me or do a few of these remarks appear as if they are written by brain dead visitors?

😊 And, if you are posting on other places, I would like to follow everything new you have to post. Would you list of the complete urls of your social networking sites like your linkedin profile, Facebook page or twitter feed?

table saw reviews

on **February 4, 2014 at 5:06 AM** said:

I appreciate, cause I found exactly what I used to be taking a look for. You've ended my 4 day long hunt! God Bless you man. Have a nice day. Bye

best folding bike

on **February 4, 2014 at 5:42 PM** said:

I'll immediately snatch your rss feed as I can't in finding your e-mail subscription hyperlink or e-newsletter service. Do you've any? Kindly let me recognise in order that I could subscribe.

Thanks.

best clothes steamer

on **February 5, 2014 at 1:17 AM** said:

Hi, I do believe this is an excellent web site. I stumbled upon it 😊 I will revisit once again since
i have bookmarked it. Money and freedom is the greatest way to change, may you be rich and continue to help other people.

best electric tea kettle

on **February 5, 2014 at 2:39 AM** said:

It's really very complex in this full of activity life to listen news on TV, thus I just use the web for that purpose, and get the latest news.

best induction cooktop

on **February 5, 2014 at 3:35 AM** said:

Hey I know this is off topic but I was wondering if you knew of any widgets I could add to my blog that automatically tweet my newest twitter updates. I've been looking for a plug-in like this for quite some time and was hoping maybe you would have some experience with something like this. Please let me know if you run into anything. I truly enjoy reading your blog and I look forward to your new updates.

Laurice Melchiorre

on **February 6, 2014 at 8:49 AM** said:

Hey! Ich bin Angela. Registriert offene Diskussion zu haben und erfahren Sie mehr über eine Karriere in tech. Ich werde Entsendung und Fragen ziemlich viel, wie ich für das, was die Welt für mich auf Lager vorzubereiten.