Purdue University

Purdue e-Pubs

Charleston Library Conference

Opportunities and Challenges of Data Publication: A Case from Purdue

David Scherer
Purdue University, daschere@andrew.cmu.edu

Lisa Zilinski
Purdue University, ldz@andrew.cmu.edu

Courtney Matthews
Purdue University, courtneyearlmatthews@gmail.com

Follow this and additional works at: https://docs.lib.purdue.edu/charleston

Part of the Library and Information Science Commons, and the Transportation Commons

An indexed, print copy of the Proceedings is also available for purchase at:

http://www.thepress.purdue.edu/series/charleston.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information

Sciences. Find out more at: http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences.

David Scherer, Lisa Zilinski, and Courtney Matthews, "Opportunities and Challenges of Data Publication: A Case from Purdue" (2013). *Proceedings of the Charleston Library Conference.*
http://dx.doi.org/10.5703/1288284315319

Opportunities and Challenges of Data Publication: A Case from Purdue

David Scherer, Scholarly Repository Specialist, Purdue University
Lisa Zilinski, Data Specialist and Assistant Professor of Library Science, Purdue University
Courtney Matthews, Digital Data Repository Specialist, Purdue University

Abstract

Beginning in 2011, there have been several policy changes directly affecting the management, preservation, and accessibility of publically funded research and resulting research data in the United States. On January 18, 2011 the National Science Foundation (NSF) required data management plans to be submitted with all grant proposals (National Science Foundation, 2013). On February 22, 2013, the Office of Science and Technology Policy of the President of the United States extended a similar requirement to all federal agencies that have a research and development budget of more than $100 million (Holdren, 2013). These requirements illustrate the need for further coordination and management of data as scholarship and traditional scholarship in integrated publishing solutions.

The case study presented in this paper will illustrate an early initiative at Purdue University to integrate discrete data publications with traditional scholarly publications through leveraging new and existing repository platforms and services (Zilinski, Scherer, Bullock, Horton, & Matthews). Examination of the case study will involve a discussion of workflow integration between Purdue's data repository, the Purdue University Research Repository (PURR); its traditional press, the Purdue University Press (PUP); and the institutional repository, Purdue e-Pubs.

Publications and Data: Changes in Public Policy

Changes in Public Policy

There have been several recent policy changes directly affecting the management, preservation, and accessibility of publically funded research and resulting research data in the United States. In March 2002, the National Institutes of Health (NIH) shared a draft statement on sharing research data and invited comments on the statement. In February 2003, the NIH finalized their statement on research data, stating all investigator initiated applications with direct costs greater than $500,000 in any single year will be expected to address data sharing in their application (National Institutes of Health, 2003). Furthermore, in 2008, the NIH released its Public Access Policy which requires funded research papers to be publically accessible on PubMed Central (National Institutes of Health, 2008). In January 2011, the National Science Foundation (NSF) required data management plans to be submitted with all grant proposals (National Science Foundation, 2013). In February 2013, the

OSTP of the President of the United States extended a similar requirement to all federal agencies that have a research and development budget of more than $100 million (Holdren, 2013). Currently, agencies are crafting their responses to this memo and are asking for feedback from their communities. These requirements illustrate the need for further coordination and management of data as scholarship and traditional scholarship in integrated publishing solutions.

Case Study: Purdue Libraries and the Joint Transportation Research Program (JTRP)

A Service Model of Collaboration

Purdue recognized the early need to develop distinct repositories that were focused on the specific and unique needs of content ranging from archival materials, traditional institutional repository materials, and data. With this in mind, Purdue created three distinct institutional repositories: e-Archives, the traditional archival repository; Purdue e-Pubs, the traditional institutional repository; and the Purdue University

Research Repository (PURR), the research data repository and collaboration platform.

Formed in 2005, the Purdue e-Pubs repository is the traditional institutional repository of Purdue University, as well as an online publishing platform for the Libraries Publishing division. Built upon the Digital Commons platform from bepress, the Purdue e-Pubs repository provides free global online open access to scholarship and research authored by members of the Purdue faculty, staff, and students.

Made operational in fall 2011 and available to Purdue users in January 2013, PURR is a research collaboration and data management solution for Purdue researchers and their collaborators. PURR allows researchers and graduate students to collaborate on research and publish datasets online. Examples of research data may include spreadsheets, instrument or sensor readings, software source code, surveys, interview transcripts, or images and audiovisual files.

Introducing JTRP

Formed in 1937, the Joint Transportation Research Program (formerly called the Joint Highway Research Project) is a collaboration between Purdue University and the Indiana Department of Transportation. JTRP's main aim is "developing the best methods of improving and maintaining the highways of the state and the respective counties thereof" (Joint Transportation Research Project, n.d.).

Today, JTRP is supported through $5.3 million worth of annual investments into around 270 students, faculty, and staff. Among other things, JTRP has produced over 1,500 technical reports since the 1950s that cover a wide range of topics related to Indiana's highways.

Since the program is a partnership with the state, one of its major goals is to provide increased access to state-funded research, with an aim to decrease investment in duplicative research both nationally and globally.

JTRP was one of the earliest adopters of e-Pubs. The relationship began in 2006 with the initial batch transfer of their born digital and digitized

technical reports. This relationship then evolved to the digitization of their backlog of technical reports going back to the 1950s.

Unique Publication Workflows

Traditional Scholarly Publications: JTRP Technical Reports

In 2012, a Newton et al. report stated that by creating a workflow model that leveraged the publishing and repository services, resources, and expertise of the libraries and press, JTRP researchers were able to increase discoverability and add value to the reports through the press' manageable review and production process and through the use of the Digital Commons repository software's publishing and manuscript management capabilities (Newton et al., 2012). This workflow provided researchers an opportunity to modernize the report production process. The press and library partners increased the value of the workflow by assigning CrossRef Digital Object Identifiers (DOIs), which increased the discoverability and visibility of the technical reports. The increased discoverability has allowed for broader dissemination of the JTRP scholarly research and has provided an early example of the benefits of integrating library and publishing colleagues into research center activities. The Purdue Libraries continues to increase its publishing capabilities by providing a much more collaborative service model to meet the needs of campus partners.

In March 2012, the Purdue University Press became more strategically aligned with the Purdue Libraries by forming a new divisional unit, the Libraries Publishing Division. This new alignment was created to form a stronger partnership for campus publishing activities. The Purdue Libraries Publishing Division is comprised of two distinct, yet parallel, units: the Purdue University Press and Scholarly Publishing Services, which is where Purdue e-Pubs is organizationally contained. The mission of the division is to maximize access to scholarly publications in financially sustainable open models for the benefit of users across the state, nation, and world. Through this partnership, the Libraries now

offers a range of publishing services that are beneficial to the university and stakeholders.

With the evolution of the publication workflow and libraries-based service models, the Libraries has been able to provide JTRP a means to evolve their dissemination activities from a model with limited visibility and presence to a model with increased global reach and dissemination capability through the implementation of production and collection processes and activities for JTRP content, improved management and structure for dissemination purposes (JTRP Collection and Series), and increased visibility for access and discoverability through the employment of Open Archival Initiative, or OAI, Google Scholar indexing, and RSS feeds interlinking their personal web presence to their collection in Purdue e-Pubs.

All of these improvements have assisted in driving the impact of the JTRP and of the partnership with the libraries and the services offered by the Libraries. JTRP now has around 1,760 items available in Purdue e-Pubs (out of ca. 36,000). In the last year, JTRP materials have been downloaded over 285,000 times. Since the inception of the collection, JTRP materials have been downloaded over 694,000 times. The average number of downloads per month the collection sees continues to increase by an average of 8.5% per month. JTRP materials have also been accessed from many countries worldwide, most notably India, China, South Korea, Italy, and Belgium.

Evolving Collaborations Toward Data

In February 2013, an "all-hands" meeting was held within the Libraries to discuss the existing Libraries relationship with JTRP and to explore how this relationship could be extended. During this meeting it was concluded that the JTRP group should be made aware that (1) a new repository for data (PURR) at Purdue existed and (2) that the Libraries would work with JTRP to leverage PURR for the publication of their technical report research data.

During the early part of 2013, a series of meetings were held between the Libraries and JTRP, and

they were tasked with identifying mutual goals and to extend the Libraries' relationship with JTRP to include dataset publishing and the development of an integrated data and technical report publishing workflow. The goals outlined were:

1. Identification of a use case;

2. Prototype a workflow where discrete but related JTRP data and JTRP technical reports are linked, enabling greater discovery and reuse;

3. Expose data to demonstrate compliance with good research data management practice and meet funding agency requirements and federal government directives; and

4. Provide an example of an integrated scholarly and data publication workflow.

Data Publication Workflows: JTRP Data

PURR provides the user with a data management workflow that is accessible at multiple points. Purdue principle investigators, or PIs, can add PURR boilerplate text to their grant application to meet the DMP requirements of grant funding agencies, for example, NSF. Purdue PIs can create a project space to upload, store, share, and manage their data. PIs can then package, describe, and submit files for publication with a DOI. The publication package is then reviewed by the DDRS and the disciplinary librarian prior to final publication.

Once reviewed, the dataset is published and presented publicly at purr.purdue.edu and in the library's catalog for at least 10 years. During this period, PIs receive a monthly report detailing monthly and total downloads and monthly and total page views. Datasets are preserved for at least 10 years. At the end of the 10-year period, the datasets are again reviewed by the disciplinary librarian, the DDRS, and a Digital Archivist to determine whether they will be maintained in the Library's permanent collection. By providing these services, PURR meets the previously unmet need for the publication of datasets as unique scholarship.
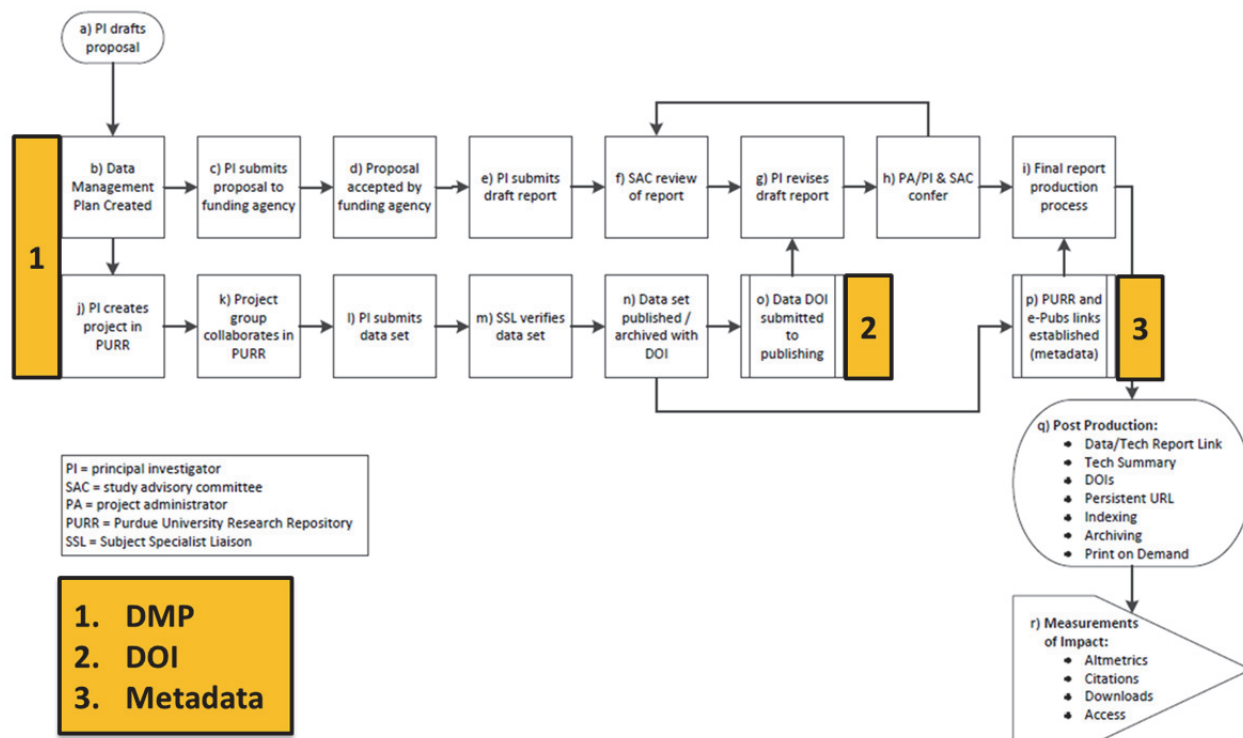
Figure 1. E-Pubs/PURR Comprehensive Workflow

Integrated Comprehensive Publications and Workflows

The comprehensive publication workflow presented aligns the Purdue e-Pubs and PURR workflows (Zilinski et al.). The workflows operate in parallel at three points (Figure 1):

1. At the point the data management plan is created: As early as possible in the research cycle, JTRP, PUP, and PURR identify PIs who are submitting technical reports that contain data and provide education, exposure, and support in the use of PURR. They are made aware that utilizing the integrated workflow will create a richer linked data scholarship package.

2. At the point when then the DOI is minted with PURR: The PI packages, describes, and publishes a dataset with a DOI using PURR. The approved PURR dataset DOI(s) are then submitted to the PUP-JTRP production editor for inclusion in the

technical report. Simultaneously the JTRP production editor registers a DOI for the technical report and provides it to the PURR DDRS for inclusion in the dataset's PURR records.

3. And finally, at the point when the final metadata records are created: The final direct point of communication occurs in the metadata creation stage. In this stage, links are established between the e-Pubs publication and the PURR dataset. PURR and Purdue e-Pubs each issue DOIs for the discrete resources contained in their respective repositories. These DOIs are then used to provide links to and from the metadata record of the respective resource(s) to the related resource(s) in the other repository to create a larger linked scholarly publication package.

In this final step, both metadata linking the technical report to datasets and vice-versa are added to the PURR and e-pubs resource records.
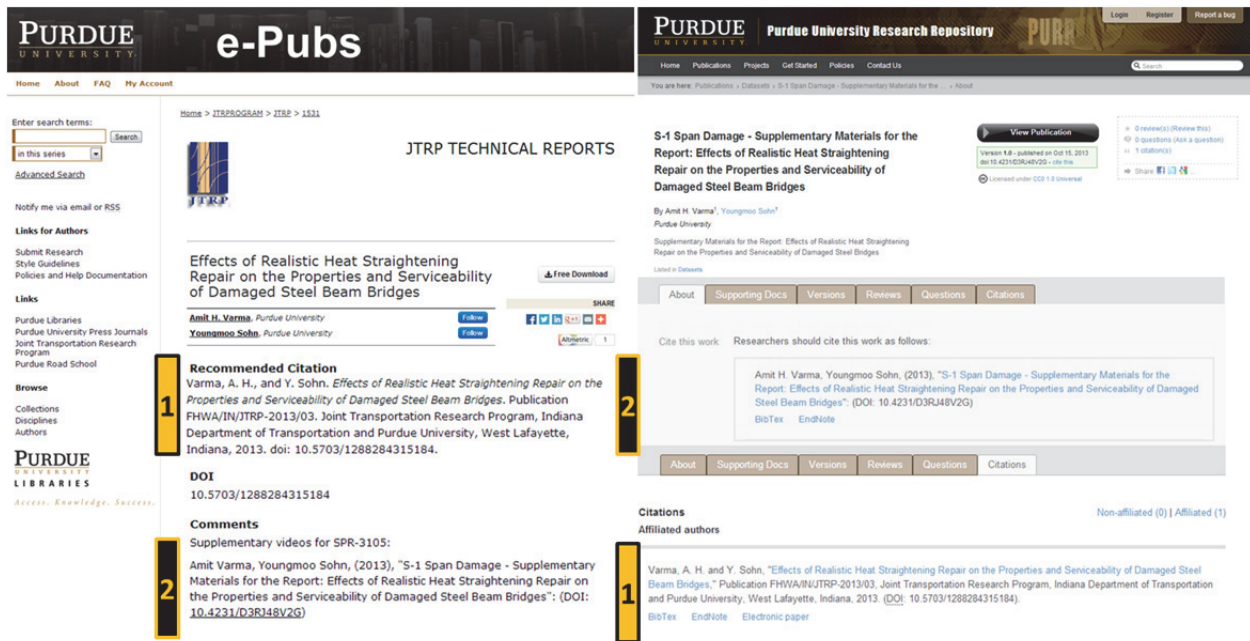
Figure 2. Linking Citations in E-Pubs and PURR

Figure 2 demonstrates these linkages. The Purdue e-Pubs "recommended citation" is provided in the "Citations" tab of the PURR record and the PURR "Cite this work" recommendation is included in the "Comments" field of the Purdue e-Pubs record.

Conclusions

There have been many lessons learned in our early work to incorporate data as scholarship and traditional scholarship into the technical report publication process. These lessons can be translated into best practices. First, linking publishing and data workflows allows for the researcher(s), repository manager(s), editor(s), and disciplinary librarians to coordinate resources and anticipate needs at each step of the process.

Second, early interaction with the data repository was crucial to our success. The sooner an investigator begins working with the PURR services and support group, the higher the likelihood that good data management principles and practices will be employed. In the case of JTRP, standardizing the integration of the PURR service and platform into the comprehensive publication workflow increases the ease of citation of the data and increases the impact of both the published data and the traditional scholarship.

Third, increasing the visibility and discoverability of both the data publications and the traditional technical report publication is accomplished by incorporating standard publication attributes, allowing each publication to appear as a traditional publication. These attributes include an identifiable bibliographic record of metadata, a standard and recognizable citation secure URL, a DOI, and any funding grant information required by the granting agency. Open access to these publications can increase the visibility and access to taxpayer-funded research allowing sponsoring organizations to be compliant with funder requirements.

And finally, through the use of the recognized publication attributes, the research usage and access metrics can be monitored and statistically evaluated through both quantitative and qualitative measurements. The research's impact can be widely monitored and can provide a richer story by having a balance of quantitative and qualitative data. These metrics may include repository download and access data, Google Analytics metrics, Google Scholar data, and altmetric data. A mixture of impact metrics allows both researchers and stakeholders to monitor how research is being recognized and the effect of making these materials openly available.

References

Holdren, J. (2013, February 22). Increasing access to the results of federally funded scientific research. *Office of Science and Technology Policy.* Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Joint Transportation Research Program (JTRP). (n.d.). *Joint Transportation Research Program Collection.* Retrieved from http://docs.lib.purdue.edu/jtrprogram/

National Institutes of Health. (2003, February 26). *Final NIH statement on sharing research data.* Bethesda, MD: National Institutes of Health. Retrieved from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

National Institutes of Health. (2008, January 11). *Revised policy on enhancing public access to archived publications resulting from NIH-funded research*. Bethesda, MD: National Institutes of Health. Retrieved from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html

National Science Foundation. (2013, January). *Chapter II—Proposal preparation instructions*. Arlington, VA: National Science Foundation. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp#dmp

Newton, M. P., Bullock, D. M., Watkinson, C., Bracke, P. J., & Horton, D. (2012).Engaging new partners in transportation research: Integrating publishing, archiving, indexing of technical literature into the research process. *Transportation Research Record: Journal of the Transportation Research Board*, *2291*, 111–123. http://dx.doi.org/10.3141/2291-13

Zilinski, L. D., Scherer, D. A., Bullock, D. M, Horton, D. K., & Matthews, C. E. (Accepted). Evolution of data creation, management, publication, and curation in the research process. Transportation Research Board Annual Meeting, January 13, 2014, Paper No. 14-0664.