Purdue University Press

Chapter Title: Interlinking Institutional Repository Content and Enhancing User Experiences
Chapter Author(s): David Scherer, Lisa Zilinski and Kelley Kimm

Book Title: Making Institutional Repositories Work
Book Editor(s): Burton B. Callicott, David Scherer, Andrew Wesolek
Published by: Purdue University Press. (2016)
Stable URL: http://www.jstor.org/stable/j.ctt1wf4drg.27

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

This book is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/. Funding is provided by Knowledge Unlatched.

*Purdue University Press* is collaborating with JSTOR to digitize, preserve and extend access to *Making Institutional Repositories Work*

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

# 19 Interlinking Institutional Repository Content and Enhancing User Experiences

*David Scherer, Lisa Zilinski, and Kelley Kimm*

In February 2013 the White House Office of Science and Technology Policy announced new requirements for government agencies that fund over $100 million worth of research: The results of funded projects (both the published research and underlying data) must be made publicly and openly available (Holdren, 2013). At Purdue University, the Libraries and the Joint Transportation Research Program (JTRP) are collaborating to produce and disseminate technical report publications and their underlying datasets. In 2014 these two campus partners developed a comprehensive workflow that intersects two separate workflows for gathering and producing these outputs. This new comprehensive workflow allows these interlinked research outputs to be deposited and made publicly available in two unique yet complementary institutional repositories: the Purdue e-Pubs repository and the Purdue University Research Repository (PURR). Although these outputs are deposited in separate repositories, this workflow allows these materials to be interlinked so that users are aware of the other's existence. This case study highlights the development of these two repositories and workflow models and the changes adopted to enhance the content presentation and user experience.

## BACKGROUND

Formed in 1936, the Joint Transportation Research Program (JTRP; https://engineering.purdue.edu/JTRP) is a collaboration between the Indiana Department of Transportation and Purdue University Civil Engineering.



This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

In a typical year JTRP produces 20 to 30 technical reports on a variety of transportation-related issues. These reports are published and made available as free PDF downloads from the JTRP collection (http://docs.lib.purdue.edu/jtrp/) on Purdue e-Pubs, the Purdue Libraries Publishing Division's online publishing platform. The process of publishing these technical reports has evolved through the years so that, beginning in 2006, JTRP began a partnership with the Purdue Libraries to produce and disseminate its technical report publications. In 2010 the partnership expanded further to the Purdue e-Pubs institutional repository, which became both the publishing platform and mode of dissemination.

Beginning in 2013–2014 the Libraries/JTRP partnership extended to include the use of the Purdue University Research Repository (PURR). PURR enables JTRP researchers to publish their datasets online and then link these data to their technical reports via digital object identifier (DOI) (Purdue University Research Repository; https://purr.purdue.edu/). After an initial implementation the workflow model was utilized in the publishing of the first interlinked technical report publication and datasets. It became apparent, however, that something was missing from this process. A means for previewing the datasets was needed to allow for users coming from a variety of platforms or devices (e.g., mobile- or tablet-based platforms) and to ensure a complementary user experience. By interlinking the unique workflows of both of these repositories and providing a common user experience, the repository administrators, research administrators, and editorial manager can coordinate the deposit process of the materials, develop the points of interlinkage, and further ensure that users' needs and experience expectations are being met by repository capabilities and metadata practices.

## INSTITUTIONAL REPOSITORIES AT PURDUE UNIVERSITY

### The Purdue e-Pubs Institutional Document Repository

In 2005, the Purdue University Libraries established the Purdue e-Pubs repository, a traditional institutional repository and online publishing platform for the Libraries Publishing Division. The repository, built upon the Digital Commons platform from bepress, provides free global online open access to scholarship and research authored by Purdue faculty, staff, and

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

students. Since 2010, Purdue e-Pubs has been both the hosting repository and publishing platform for JTRP technical reports. This platform has provided for a holistic production process and standard processes for journal article production and publication. A production editor who manages the review process and production of the technical reports is supported by JTRP funds (Zilinski, Scherer, Bullock, Horton, & Matthews, 2014).

### The Purdue University Research Repository (PURR)

The Purdue University Research Repository (PURR), in collaboration with the Office of the Vice President of Research (OVPR) and Information Technology at Purdue (ITaP), is the Libraries' data repository and was designed to assist Purdue researchers in meeting the data management plan (DMP) requirements of granting agencies. The PURR hub was built using Purdue's own HUBzero open source platform, which "support(s) collaborative development and dissemination of scientific models running in an infrastructure that leverages a 'cloud' of computing resources" (McLennan & Kennell, 2010). PURR was made operational in fall 2011, went live for Purdue users in January 2013, and extends the HUBzero capabilities by allowing users to publish data as scholarship with a DataCite DOI. Some examples of research data are spreadsheets, models, instrument or sensor readings, software source code, surveys, interview transcripts, images, and audiovisual files. In addition to housing and publishing research datasets, PURR allows researchers and graduate students to collaborate on research and create working project spaces.

## COHESIVE MULTIREPOSITORY WORKFLOW MODEL

In 2012, Newton and colleagues reported that publishing and repository services and expertise could be leveraged to provide an enhanced publication with increased discoverability and accessibility. These efforts were further enhanced with the adoption of a second data-focused institutional repository and workflow, which could be used to provide access and disseminate the affiliated datasets. This part of the case study discusses the two workflows used to accomplish the linking of technical reports and datasets—the technical report publication workflow (including the peer-review process) and the PURR dataset publication workflow—and where they intersect.

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

### Joint Transportation Research Program (JTRP) Technical Report Publication Workflow

Purdue e-Pubs is both the hosting repository and the publishing platform for JTRP technical reports, as well as the vehicle for managing the peer-review process. The following is the path from initial report submission to publication:

1. The principal investigator (PI) submits the draft final report with metadata to Purdue e-Pubs.
2. Via e-Pubs, the production editor invites the Study Advisory Committee (SAC) members to review the report.
3. SAC members submit their reviews to e-Pubs.
4. The production editor sends reviews to the PI via e-Pubs.
5. The PI provides a revised report to the project administrator and business owner prior to the closeout SAC meeting.
6. Once the report is approved by the SAC, the PI submits the final report to e-Pubs.
7. The production editor sends the final report to the JTRP managing director, who obtains approval for publication from the Indiana Department of Transportation.
8. Upon approval, the production editor does the following to prepare the report for publication.
    a. Assigns report number and DOI.
    b. Performs light copyediting for consistency.
    c. *Ensures that the PURR DataCite DOI(s) are referenced in the report.*
    d. Manages the typesetting and proof revision process.
    e. Uploads the final typeset report to e-Pubs and completes metadata entry, *including PURR citation(s) with live DOI link(s) to one or more datasets.*
    f. Publishes the report on Purdue e-Pubs and registers the DOI with CrossRef.
    g. Provides the DOI link to the publication to the authors and other interested parties.
    h. Prepares the report to be made available via print on demand and in a free downloadable e-book format.

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

### Purdue University Research Repository (PURR) Dataset Publication Workflow

Most JTRP datasets published to date on PURR are videos linked to technical reports. While the ideal scenario is that the PI creates a DMP and publishes his or her data in PURR, then simply provides the minted DataCite DOI(s) to the production editor before the technical report is sent for typesetting (or includes them in the final report before submission), we are still in the early stages of implementing this workflow. At the time of this writing, what commonly occurs is that the PI provides the production editor with the dataset(s) and metadata, and the production editor publishes them in PURR and ensures that they are referenced properly in the technical report.

The remainder of this section discusses the PURR publication workflow when the PI provides the production editor with the dataset and metadata and requests that the production editor handle the submission and publication.

The production editor performs the following steps to publish a dataset in PURR:

1. Initiates a project in PURR.
   a. Enters a project title and description.
   b. Uploads one or more datasets to the project.
2. Starts a publication. Each dataset is its own publication, and each receives its own DataCite DOI. A project may contain several publications.
   a. Chooses the dataset to publish and makes it available as a downloadable file.
   b. Enters a synopsis.
   c. Enters the abstract text and, if the dataset is a video, a video streaming link. (With video datasets, because we want the video to stream easily on the PURR Web landing page, we upload the video to our YouTube channel and embed the YouTube link in the abstract field. Visitors can view the video immediately on the PURR site, and they can also download the MP4 file.)
   d. Adds authors and tags (key words).
   e. Chooses a publication license.

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

   f. Enters the citation for the related technical report to be published in Purdue e-Pubs.
3. Publishes the dataset.
   a. Reviews all metadata carefully; if the dataset is a video, ensures that it streams; submits the request to publish.
   b. Once PURR has published the dataset and the DataCite DOI is live (generally within 48 hours of request), *adds the citation to the technical report Web landing page on Purdue e-Pubs*.

## COMBINING THE TWO WORKFLOWS: REPOSITORIES IN ACTION

### Linking Publications to Datasets

In short, the production editor performs the following tasks to link a PI's data to his or her technical report:

1. Publishes the dataset in PURR (with a cross-reference to the technical report citation) and obtains the DataCite DOI.
2. Ensures that the PURR DataCite DOI is referenced in the technical report before it goes to typesetting.
3. Adds the PURR citation with DataCite DOI as metadata to the Purdue e-Pubs record for the technical report.
4. Publishes the technical report in Purdue e-Pubs; the landing page includes a cross-reference to the dataset citation.

Sometimes a PI will request linking a dataset to a report after the report has been published in Purdue e-Pubs, and the dataset has not been referenced in the report. In these cases the dataset is published in PURR with a cross-reference citation to the report; then the PURR citation with live DOI link to the dataset is added to the e-Pubs metadata record page.

Figure 19.1 shows the metadata record page for a technical report published on Purdue e-Pubs that is affiliated with a published dataset (in this case an MP4 video). The metadata record contains two citations: (1) a recommended citation for the technical report itself, and (2) a citation that cross-references the dataset. If this report referenced in more than one dataset, the citation for each would be included.

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

**Figure 19.1.** Purdue e-Pubs technical report record page.

Likewise, the PURR Web landing page for this dataset contains two citations: (1) a citation for the dataset itself (Figure 19.2), and (2) a cross-reference citation to the technical report (Figure 19.3). The DOI links are live in all citations.

### Points of Intersection and Linkage

As illustrated in the two repository workflows (Zilinski et al., 2014), there are three primary points of intersection. The initial point of intersection occurs at the point when the PI develops the DMP (Figure 19.4). This ensures that the PI, repository administrators, and production editor are aware in the earliest stages of the research life cycle that a technical report publication will also have datasets. The second point is at the time the DataCite

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

**Figure 19.2.** PURR dataset record page: citation for the dataset.

**Figure 19.3.** PURR dataset record page: citation for the cross-referenced technical report.

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

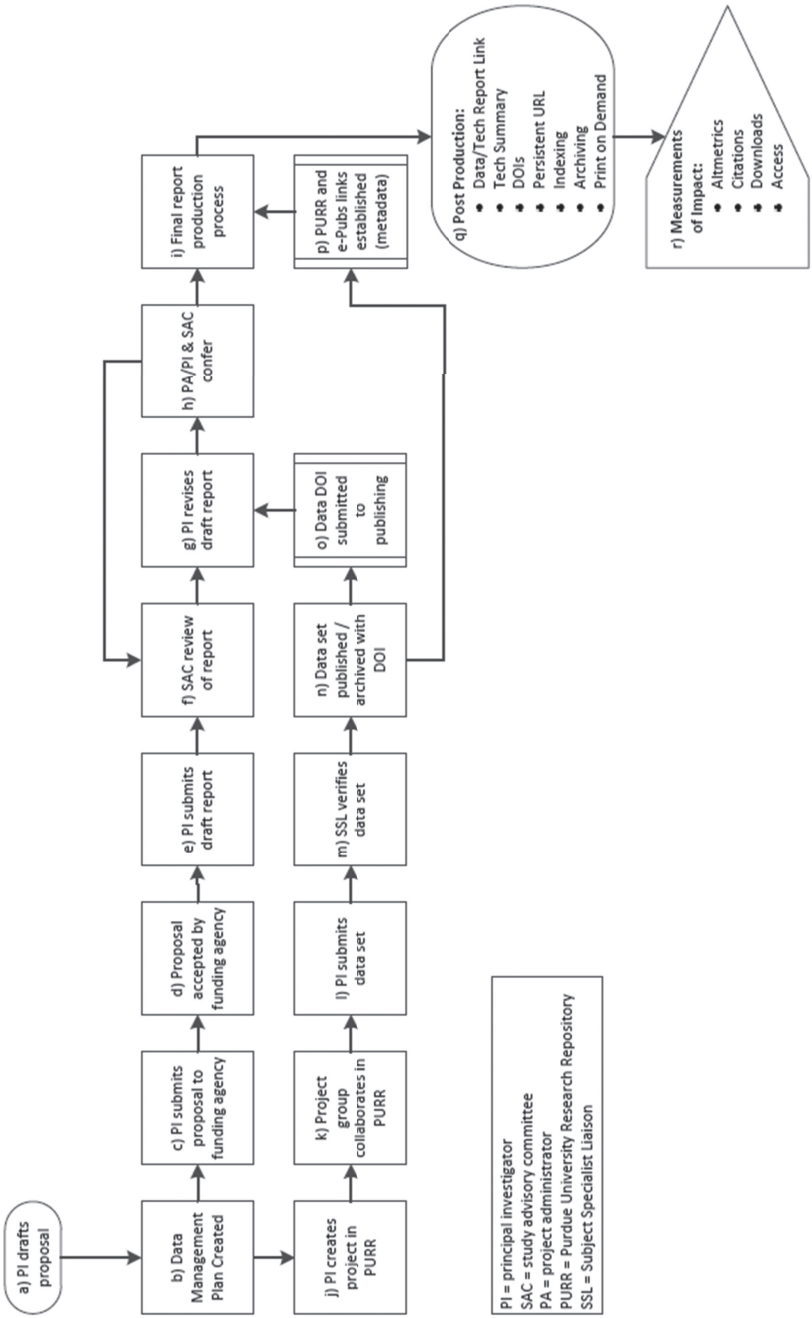


**Figure 19.4.** Purdue e-Pubs and PURR interlinked repository workflow for JTRP.

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

**Figure 19.5.** PURR embedded access-only dataset preview.

DOI is minted. This way the data citation with DOI can be added within the publication. And lastly, the citations and DOIs are embedded into the other objects' corresponding metadata record as a corresponding related object.

## Enhancing the Users' Experiences

Once the integrated workflows were implemented and the technical report publication and datasets were being published and interlinked, it became apparent that another step was necessary. The publications in Purdue e-Pubs were being made available as a downloadable PDF to ensure the widest array of users could access the publications. The datasets for the given reports were in the form of MP4 videos. PURR provided the option for videos to be downloaded too, but the datasets remained in their native file format. This led to a potential issue of not providing users a means to quickly preview the datasets, or a means to play a dataset once it had been downloaded, and required users to have an appropriate video player program. A solution was then developed allowing users to play the dataset

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

in line with their current browser window as a primary option instead of downloading the dataset. This also caused concerns for user experience. This solution required users to have the latest version of their browsers to ensure that the plug-ins created for the tool would play properly.

A final solution was created that would allow an access-only copy of the dataset to be added to an unlisted YouTube channel that could be played from within a wiki-enabled metadata field. Once the files are received the production editor is able to add the video to YouTube and the affiliated streaming linked macro to the wiki-enabled metadata field as described in step 2.C of the PURR workflow. With this new solution the access-only copy is available from the dataset's metadata record while the version of record can still be downloaded from the provided download button (Figure 19.5).

## CONCLUSION

Scherer, Zilinski, and Matthews (2013) discussed several initial lessons learned from interlinking the data publication process with the traditional publication workflow:

1. Linking publishing and data workflows allows collaborators to coordinate resources and anticipate needs at each step of the process.
2. Early interaction with the data repository increases the likelihood that good data management principles and practices would be utilized.
3. Incorporating standard publication attributes increases the visibility and discoverability of the data and traditional publications.
4. Research usage and access metrics can be monitored and evaluated through the use of recognized publication attributes.

As the development of this integrated multirepository workflow model continues, additional lessons have been learned. First, with these new workflows being developed it's crucial to continue to evaluate them for possible revisions and additional steps to enhance the interlinkage of the publications. Second, it's important to involve all vested parties with workflow updates and additional enhancements. Without insight coming from research center administrators, authors, and users, there is no review of services or user experiences to evaluate what must be added or revised. And lastly, it's important that all of these changes are still in line

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms

with the overall goal of increasing access and visibility of the published technical report publications and published datasets. The new video dataset preview capability allows users to interact more fully with the dataset in a way that still allows the version of record to be downloaded for full data manipulation.

## REFERENCES

Holdren, J. P. (2013, February 22). *Memorandum for the heads of executive departments and agencies* [Memorandum from the Executive Office of the President, Office of Science and Technology Policy]. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

McLennan, M., & Kennell R. (2010). HUBzero: A platform for dissemination and collaboration in computational science and engineering. *Computing in Science & Engineering, 12*(2), 48–53. http://dx.doi.org/10.1109/MCSE.2010.41

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

Scherer, D. A., Zilinksi, L. D., & Matthews, C. E. (2013). *Opportunities and challenges of data publication: A case from Purdue* (Proceedings of the Charleston Library Conference). http://dx.doi.org/10.5703/1288284315319

Zilinski, L. D., Scherer, D. A., Bullock, D. M, Horton, D. K., & Matthews, C. E. (2014). Evolution of data creation, management, publication, and curation in the research process. *Transportation Research Record: Journal of the Transportation Research Board*, *2414*, 9–19. http://dx.doi.org/10.3141/2414-02

This content downloaded from 203.129.241.87 on Wed, 29 Nov 2017 11:46:12 UTC
All use subject to http://about.jstor.org/terms