Purdue University Research Repository

Michael Witt
Purdue University

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

Part of the Scholarly Communication Commons
Creative Commons Attribution-Noncommercial 4.0 License
This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.

Repository Citation

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Purdue University Research Repository

Michael Witt
Head, Distributed Data Curation Center
Associate Professor of Library Science
http://www.lib.purdue.edu/research/witt
E-mail: mwitt@purdue.edu
DATA = EVIDENCE

http://epicgraphic.com/data-cake
SOME EXAMPLES OF DATA

Tables and spreadsheets
Sensor and instrument output
Software source code and simulations
Images, video, and audio
Observation logs
Workflows
Interview transcripts
Survey instruments and results
Etc.
Purdue University Research Repository (PURR)

The PURR service is a collaborative effort of the Purdue University Libraries, Executive Vice President for Research and Partnerships, and Information Technology at Purdue. PURR is a designated university core research facility.

Designated community:
Purdue University faculty, staff, and graduate student researchers; their collaborators; and the current and future consumers of their data.
Data is written into the three pillars of our strategic plan:

- **Learning:** “…information literacy defined broadly to include digital information literacy, science literacy, data literacy, health literacy, etc...”
- **Scholarly Communication:** “Lead in data-related scholarship and initiatives”
- **Global Challenges:** “We will lead in international initiatives in information literacy and e-science and ... contribute to international information literacy, learning spaces, data management, and scholarly communication initiatives.”

Publish Datasets with DOIs

Use PURR to publish datasets with Digital Object Identifiers (DOI) that make it easier for people to cite your data and give you credit. Purdue is a founding member of DataCite, the international agency that registers DOIs for data.

Learn More

http://purr.purdue.edu

Start Your Research Project

Create a Data Management Plan
Learn about the detailed requirements for your data management plan (DMP). Funding agency requirements are very specific and our DMP resources can help you to clear up any confusion. Get Started

Upload Research Data to Your Project
Create a project to upload and share your data with collaborators using our step-by-step form to guide you through the process. Invite collaborators from other institutions to join your project. Create a Project

Publish your Dataset
Package, describe, and publish your dataset with a Datacite

Featured Dataset

Maize grain yield record for the WQFS (1998-2012)
By S.M. Brouder, N. De Armond, R.F. Turco, J.J. Volenc
Purdue University

Have a Question?

Username:
Name:
E-mail:
Question:

Digital Curation Centre’s Curation Lifecycle Model: [http://www.dcc.ac.uk/resources/curation-lifecycle-model](http://www.dcc.ac.uk/resources/curation-lifecycle-model)
FACT:

Many funding agencies require data management plans with grant proposals.

Purr.purdue.edu

Purr is your solution for:

Data Management Plans
Ready-made boilerplate text to put in your proposals, tutorials and best practices, support for developing a good data management plan

Collaboration
Purr provides a free HubZero™ project space for Purdue researchers and their collaborators to work together on research and share data and tools online

Publishing Your Data
Publish your research data with digital object identifiers that other scholars can use to easily find and cite your datasets

Archiving Your Data
Your data will be preserved and made accessible for long-term access in a secure, trustworthy digital repository

Over 500 grant proposals from Purdue have included Purr in their data management plans since 2011

Purr is a Purdue Research Core Facility developed by the Office of the Vice President for Research, Purdue Libraries, and Information Technology at Purdue

To learn more, visit Purr at purr.purdue.edu or contact Courtney Earl Matthews purr@purdue.edu 765-496-2770
DATA MANAGEMENT PLANS

- Boilerplate text
- Example DMPs
- DMP Self-Assessment
- DMPTool
- Workshops
- Tutorials
- Reference and consultation with subject-specialist librarian and/or data services specialist

https://purr.purdue.edu/dmp
CREATE A PROJECT

Social Network Graph Analysis (socnet)

Private project by Michael Witt

Files

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Modified</th>
<th>By</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>README_socnet-flickr-2006.pdf</td>
<td>19KB</td>
<td>Jan 30</td>
<td>ma</td>
<td>Social Network...</td>
</tr>
<tr>
<td>socnet-flickr.smrt</td>
<td>3KB</td>
<td>Jan 30</td>
<td>ma</td>
<td>Social Network...</td>
</tr>
</tbody>
</table>

PURR project tutorial video:

http://www.youtube.com/watch?v=q5xGO_oF9uQ
USE PROJECT TO COLLABORATE

Create:
• any Purdue faculty, staff, or graduate student researcher can create projects
• describe the project
• disclaim use of sensitive or restricted data
• receive a default allocation of storage
• register a grant award to increase allocation
• invite collaborators to join project

Collaborate:
• git repository to share and version files (sftp & Google Drive integration)
• virtual machine/s
• wiki
• blog
• to-do list management and project notes
• newsfeed
• stage data publications
### STORAGE ALLOCATION

<table>
<thead>
<tr>
<th>Option:</th>
<th>For what:</th>
<th>Space Available:</th>
<th>For how long:</th>
<th>For who:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default/trial projects</td>
<td>Just trying things out, or don’t need much space</td>
<td>10 GB</td>
<td>3 years</td>
<td>Any Purdue faculty, staff, and graduate student</td>
</tr>
<tr>
<td>Default data publications</td>
<td>Great for small publications</td>
<td>1 GB</td>
<td>10 years</td>
<td>Any Purdue faculty, staff, and graduate student</td>
</tr>
<tr>
<td>Supported projects</td>
<td>Funded projects with PIs from Purdue</td>
<td>100 GB</td>
<td>Life of grant</td>
<td>Purdue faculty with a verifiable grant or account number</td>
</tr>
<tr>
<td>Supported data publications</td>
<td>Publishing work done on a funded project</td>
<td>10 GB</td>
<td>10 years</td>
<td>Purdue faculty with a verifiable grant or account number</td>
</tr>
</tbody>
</table>

**Estimated cost of additional space**

<table>
<thead>
<tr>
<th>Option:</th>
<th>Space Available:</th>
<th>For how long:</th>
<th>For who:</th>
<th>Cost:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra project space</td>
<td>As needed</td>
<td>Per year</td>
<td>Any Purdue faculty, staff, and graduate student</td>
<td>$2.10 per GB per year **</td>
</tr>
<tr>
<td>Extra publication space</td>
<td>As needed</td>
<td>10 years</td>
<td>Any Purdue faculty, staff, and graduate student</td>
<td>$14.30 per GB</td>
</tr>
</tbody>
</table>

[https://purr.purdue.edu/about/pricing](https://purr.purdue.edu/about/pricing)
Graph of Flickr Photo-Sharing Social Network Crawled in May 2006

By David F. Gleich
Purdue University

Crawl of the Flickr photo-sharing social network from May 2006 returning a graph with 820,878 nodes and 9,307,214 edges. Dataset is distributed as a SNAP file with README file with code to read file in Python and MATLAB.

PURR publication tutorial video:
http://www.youtube.com/watch?v=jYBcsfiRhio
• **Executive Committee**: Dean of Libraries, Vice President for Research, Chief Information Officer

• **Steering Committee**: 2 from libraries, 2 from IT, 2 from research office and sponsored programs, 3 domain faculty researchers

• **Personnel**: Project Director (.50), Technologists (3.85), HUBzero Liaison (.35), Metadata Specialist (.20), Digital Archivist (.25), Repository Outreach Specialist (1.0), Data Curator (1.0)

• **Key players**: Subject-specialist librarians & data services specialists
Librarians consult on data management plans in their subject areas.

Creating opportunities for librarians to interact with researchers about data
Librarian is notified by e-mail when a new project is created or a grant is awarded, based on department affiliation of Purdue project owner.

Creating opportunities for librarians to interact with researchers about data
Librarian may consult or collaborate on project if needed.

Creating opportunities for librarians to interact with researchers about data
Librarians review and post submitted datasets.

Creating opportunities for librarians to interact with researchers about data
At the end of initial commitment (10 years), archived and published datasets are remanded to the Libraries’ collection. A librarian working with the digital archivist selects (or not) the dataset for the collection.

Creating opportunities for librarians to interact with researchers about data
2013 was first full year of PURR in operation, to date:

- 1,405 data management plans
- 163 grant awards
- 1,295 registered researchers
- 477 research projects (455 active in last 30 days)
- 120 citations

DMP analysis (n=111 NSF proposals from Purdue, Jan-Jun 2013)

- 49% PURR
- 29% Local computer or server
- 14% Disciplinary repository (e.g., ICPSR, Protein Data Bank, nanoHUB, NEES)
- 8% No data or not applicable