Curation Service Models: Purdue University Research Repository

Michael Witt
Purdue University, mwitt@purdue.edu

Follow this and additional works at: http://docs.lib.purdue.edu/lib_fspres
Part of the Library and Information Science Commons

Recommended Citation
http://docs.lib.purdue.edu/lib_fspres/3

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Purdue University Research Repository (PURR, http://research.hub.purdue.edu)

Curation Service Models

Michael Witt
Assistant Professor of Library Science
Purdue University
mwitt@purdue.edu

ASIS&T Research Data Access and Preservation Summit, New Orleans: 2012
Background

• 2006 Libraries’ Strategic Plan; creation of D2C2
• 2007 Data Curation Profiles
• 2008-2009 Purdue e-Data Task Force; NSF DataNet calls; HUBzero collaborations
• 2010 Faculty data committee
• 2011 new Strategic Plan; PURR Working Group
• 2012 PURR project funded
Campus Collaboration

PURR Steering/Executive Committee:
• Dean of Libraries
• Vice President for Information Technology (CIO)
• Vice President for Research

ITaP = Information Technology at Purdue
OVPR = Office of the Vice President for Research
Libraries = Purdue University Libraries
SPS = Sponsored Programs Services

Funded 3-year HUBzero hosting plan (OVPR, Libraries, ITaP)
PURR Working Group

- Associate Vice President for Research, OVPR
- Associate Dean for Digital Programs and Information Access, Libraries
- Associate Dean for Research, Libraries
- Data Services Specialist, Libraries
- Chemical Information Specialist, Libraries
- Assistant Director of Pre-Award Services, SPS
- Managing Director, Launching Centers and Institutes, OVPR
- Visiting Assistant Professor, Libraries
- HUBzero Project Director, ITaP
- University Archivist, Libraries
- HUB Community Liaison, ITaP
- Interdisciplinary Research Librarian, Libraries (chair)
HUBzero

OVPR, Libraries, and ITaP funded 3-year HUBzero hosting plan for PURR.

- HUBero, open source software: http://hubzero.org
- Maintained by HUBzero Consortium
- Over 40 hubs online now, supporting different virtual scientific communities, hundreds of thousands of users
- http://nanoHUB.org - grandfather of the hubs, exemplar
- Built to facilitate virtual communities and online, scientific collaboration, research/teaching
- Collaborate, develop, publish, and manage content as hub resources
- Software tools, documents, multimedia, learning objects, datasets, etc.
- Social network functionality and collaboration features
- LAMP stack, Joomla framework, OpenVZ and Rappture, git, EZID, etc.
Early conceptual model

Service Definition

Roles

Infrastructure

Cost Model

Purdue Libraries e-Data Task Force (internal report), April 16 2008
Draft 7/14/2011
PURR Working Group

PURR Workflow mapped to OAIS Reference Model
(required for TRAC certification)
PURR Demo

- Create a project
- Submit a dataset for preview and then publication and archiving
# PURR Project Space Allocation and Pricing

Storage space for PURR projects currently covered by central resources

<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>500 MB</td>
<td>3 years</td>
<td>Anyone with a Purdue Career Account</td>
</tr>
<tr>
<td>Default data publications</td>
<td>Great for small publications</td>
<td>50 MB</td>
<td>10 years</td>
<td>Anyone with a Purdue Career Account</td>
</tr>
<tr>
<td>Supported projects</td>
<td>Funded projects with PIs from Purdue</td>
<td>100 GB</td>
<td>Life of project</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>Life of project</td>
<td>Anyone with a Purdue Career Account</td>
<td>$2,102.57 per TB per year **</td>
</tr>
<tr>
<td>Extra publication space</td>
<td>As needed</td>
<td>10 years</td>
<td>Anyone with a Purdue Career Account</td>
<td>$14,297.48 per TB</td>
</tr>
</tbody>
</table>
User Support for PURR

- Knowledge Base (FAQ)
- Boilerplate text for data management plan
- Self-assessment tool for data management planning
- Tutorials – e.g., funder-specific guides
- Workshops and video
- Contact Us
  - Bugs = ITaP
  - Grant preparation = SPS
  - Data management or help with PURR = Libraries
- Ask-A-Librarian chat/email “data reference” service
Data Reference: Brainstorming

Data Reference: Implementation

- QuestionPoint software
- Ask-A-Librarian chat/e-mail reference widget on PURR webpages
- Training for front line reference workers with instructions on how to identify and route questions about data or PURR
- If chat, execute script to let patron know question has been referred to a data librarian
- Question is assigned to a group called “data librarians” with a corresponding mailing list
- First data librarian to respond claims question, works with appropriate subject librarian to respond (via email, telephone, or in person)
- Tags question for future reporting
- PURR bookmarks sent to all physical reference desks
Early Adoption of PURR

- 62 projects created (caveat: many are just testing the water)
- 167 registered users

Proposals to date that include PURR as a component of their data management plans since January 2011

- 312 proposals submitted
- 180 pending review now
- 11 grants awarded

34% of NSF proposals from Purdue include PURR (others include USDA, EPA, NASA, NEH, US Forest Service, and various industrial sponsors)
When a Grant is Awarded...

- SPS monitors grants and emails notification of awards that include PURR
- Librarian from Data Services and Research Department contacts appropriate subject librarian to contact PI (based on department affiliation)
- Follow-ups are tracked on wiki, instant feedback from researchers on PURR to project team
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, 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 approved submitted datasets.

Creating opportunities for librarians to interact with researchers about data…
At the end of initial commitment (10 years or the life of the project sponsorship), 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…
PURR Budget January 2012 - June 2013

Libraries
• Digital Library Software Developers
• Digital Data Repository Specialist
• Metadata Specialist (20%)
• Digital Archivist (25%)
• Graduate Assistant (50%)
• ISO 16363 certification

ITaP
• HUB Liaison (25%)
• Technical Project Manager (10%)

Web Developer
• Software Developer (50%)
• Middleware Developer (25%)
• Security Expert (10%)
• Hardware, storage, S&E to scale

Funding from University through the OVPR; PURR has since been identified as a university research core.
We still have a lot of work to do, such as:

• Campus outreach campaign
• Automate workflows
• Extend metadata support and management
• Curation microservices integration
• ISO 16363 certification
• Tool publication
• Assessment - January 2013
Acknowledgements

- Paul Bracke
- D. Scott Brandt
- Jeffrey Bolin
- Richard Buckius
- Jacob Carlson
- Jeremy Garritano
- Amanda Hamaker
- Ned Howell
- Eugenia Kim
- Matthew Kroll
- Gerry McCartney
- Michael McLennan
- Sammie Morris
- James L. Mullins
- Alissa Nedossekina
- Sam Wilson
- Michael Witt

Information Technology at Purdue – Purdue Libraries – Office of the Vice President for Research